Human Genome Center

Laboratory of Functional Analysis In Silico
機能解析イン・シリコ分野

Professor	Kenta Nakai, Ph.D.	教　授	博士(理学)　中　井　謙　太
Associate Professor	Sung-Joon Park, Ph.D.	准教授	博士(工学)　朴　　聖　俊
Assistant Professor	Martin Loza, Ph.D.	助　教	博士(生命機能学)　ローザ　マーティン

Laboratory of Genome Database
ゲノムデータベース分野

Professor	Kenta Nakai, Ph.D.	教　授	博士(理学)　中　井　謙　太

The mission of our laboratory is to conduct computational ("in silico") studies on the functional aspects of genome information. At present, we mainly focus on the analysis of regulatory information of gene expression in the non-coding region, using a variety of next generation sequencing (NGS) data. In addition, we are actively collaborating with researchers from various fields.

1. Epigenetic characterization of housekeeping core promoters and their importance in tumor suppression

Martin Loza, Alexis Vandenbon[1], and Kenta Nakai
[1]Institute for Life and Medical Sciences, Kyoto University, Japan.

In this research, we elucidate the presence of around 11,000 housekeeping *cis*-regulatory elements (HK-CREs) and describe their main characteristics. Besides the trivial promoters of housekeeping genes, most HK-CREs reside in promoter regions and are involved in a broader role beyond housekeeping gene regulation. HK-CREs are conserved regions rich in unmethylated CpG sites. Their distribution highly correlates with that of protein-coding genes, and they interact with many genes over long distances. We observed reduced activity of a subset of HK-CREs in diverse cancer subtypes due to aberrant methylation, particularly those located in chromosome 19 and associated with zinc finger genes. Further analysis of samples from 17 cancer subtypes showed a significantly increased survival probability of patients with higher expression of these genes, suggesting them as housekeeping tumor suppressor genes. Overall, our work unravels the presence of housekeeping CREs indispensable for the maintenance and stability of cells.

2. A graph-embedding approach dissecting enhancer-cofactor-promoter interactions

Sung-Joon Park and Kenta Nakai

The spatial genome organization mediates the functional impact of distal chromosomal interactions.

Particularly, enhancer-promoter interactions have been intensively studied by cutting-edge computational algorithms. However, we still have a very limited understanding of how enhancer signals transmit to their target promoters through highly intricate regulatory networks. Here, we developed a new computational framework. The method combined a regression modeling that predicts gene expression by inferring important promoter-distal and -proximal gene regulators with a graph-embedding algorithm that detects cell-type specific and conserved regulatory interactions in complex gene regulatory networks. We applied the method to human naïve and germinal center B cells. Consequently, we identified sets of promoter-distal transcription factors and architectural cofactor proteins, which are appropriately co-regulated to prevent malignancy. These results highlight the importance of understanding the cis- and trans-regulatory interactions composed in the transcriptional domain. Our approach offers an alternative method to understanding enhancer biology mediated by protein-protein interactions in the 3D genome organization.

3. HyGAnno: Hybrid graph neural network-based cell type annotation for single-cell ATAC sequencing data

Weihang Zhang, Yang Cui, Bowen Liu, Martin Loza, Sung-Joon Park and Kenta Nakai

Reliable cell type annotations are crucial for investigating cellular heterogeneity in single-cell omics data. Although various computational approaches have been proposed for single-cell RNA sequencing (scRNA-seq) annotation, high-quality cell labels are still lacking in single-cell ATAC sequencing (scATAC-seq) data because of extreme sparsity and inconsistent chromatin accessibility between datasets. Here, we present a novel automated cell annotation method called HyGAnno that transfers cell type information from a well-labeled scRNA-seq reference to an unlabeled scATAC-seq target via a parallel graph neural network in a semi-supervised manner. Unlike existing methods that utilize only gene expression or gene activity features, HyGAnno integrates genome-wide accessibility peak features to facilitate the training process. In addition, HyGAnno reconstructs a reference-target cell graph that can be used to detect cells with low prediction reliability according to their specific graph connectivity patterns. HyGAnno was tested using large datasets and demonstrated the advantages of accurate cell annotation, interpretable cell embedding, robustness to noisy reference data, and adaptability to tumor tissues.

4. Spatial Transcriptomics Analysis via Image-Aided Graph Contrastive Learning for Domain Exploration and Alignment-Free Integration

Yitao Yang, Yang Cui, Xin Zeng, Yubo Zhang, Martin Loza, Sung-Joon Park and Kenta Nakai

Biological tissues, as intricate networks of varied cell types, perform essential life functions through unique spatial configurations. Recent spatial transcriptomics techniques, such as 10x Visium, have greatly enhanced our ability to map genetic data within these configurations, offering deeper insights into the genetic makeup of tissues in health and disease and advancing our grasp of molecular and physiological intricacies. We propose Spatial Transcriptomics and Image-based Graph learning (STAIG), a deep learning framework for advanced spatial region identification in spatial transcriptomics. STAIG integrates gene expression, spatial coordinates, and histological images using graph contrastive learning, excelling in feature extraction and enhancing analysis on datasets with or without histological images. It has outperformed existing methods in recognizing spatial regions within Visium and other platform datasets. STAIG was specifically engineered to counter batch effects and to allow for the integration of tissue slices without pre-alignment. In applications to human breast cancer and zebrafish melanoma, STAIG has identified regions with high precision, uncovering new insights into tumor microenvironments. STAIG thus offers a versatile tool for probing cellular structures and interactions, enriching our understanding of spatial transcriptomics.

5. TF-EPI: An Interpretable Enhancer Promoter Interaction Detection Method Based on Large Language Model

Bowen Liu, Weihang Zhang, Xin Zeng, Martin Loza, Sung-Joon Park and Kenta Nakai

The detection of enhancer-promoter interactions (EPIs) is crucial for understanding gene expression regulation, disease mechanisms, and more. In this study, we developed TF-EPI, a deep learning model based on Transformer, designed to detect these interactions solely from DNA sequences. The performance of TF-EPI surpassed other state-of-the-art methods in multiple benchmark datasets. Importantly, by utilizing the attention mechanism of the Transformer, we identified distinct cell type-specific motifs and sequences in enhancers and promoters, which we validated against databases like Jaspar and UniBind, highlighting the potential of our method in discovering new biological insights. Moreover, our analysis of the transcription factors corresponding to these motif and short sequence pairs revealed the heterogeneity

and commonality of gene regulatory mechanisms and demonstrated the ability to identify TFs relevant to the source information of the cell line. Finally, the introduction of transfer learning can mitigate the challenges posed by cell type-specific gene regulation, yielding enhanced accuracy in cross-cell line EPI detection results. Overall, our work unveils important sequences information for the investigation of enhancer-promoter pairs based on the attention mechanism of Transformer, which provides an important milestone in the investigation of cis-regulatory grammar.

6. Computational Transcriptomic Analysis Identifies a Novel Immune-dysregulated TNBC Subtype Regulated by STAT3

Yang Cui and Kenta Nakai

Triple-negative breast cancer (TNBC) is the most aggressive subtype of breast cancer and frequently appears to be resistant to treatment due to its extensive intratumoral heterogeneity. The tumor microenvironment (TME) plays a critical role in TNBC immunomodulation and progression. In this study, we leveraged TNBC bulk RNA-seq data and TNBC scRNA-seq data to investigate the TNBC TME. Utilizing CIBERSORTx to deconvolve the bulk RNA-seq data with scRNA-seq as a reference, we identified cell compositions within the TME. Non-negative matrix factorization further revealed three unique clusters, with patients in Cluster 2 (C2) exhibiting significantly shorter survival times. Next, we trained a random forest classifier on TNBC bulk RNA-seq data to identify the C2 patients within the scRNA-seq dataset. We found a reduced abundance of immune cells, including CD4+ T cells, CD8+ T cells, DC, Macrophage, and NK cells, and an increase in breast cancer cells in C2, indicating an immunosuppressive environment. We further conducted cell-cell communication (CCC) analysis and found compromised MHC-I and MHC-II signaling in CD8+ T cells and macrophages, respectively, in C2. Followed SCENIC analysis unveiled the gene regulatory network of the CD8+ T cells in C2, revealing an enrichment of regulon STAT3, which is associated with T cell dysfunction. Our study unveils an immunodysregulated subtype within TNBC, and provides valuable insights into its heterogeneity.

7. Development of the splice site prediction model that can account for long-distance effects using Transformer.

Yuna Miyachi and Kenta Nakai

RNA splicing is an essential process in the regulation of gene expression. Recently, deep learning methods have been applied to predict splice sites and elucidate their complex mechanisms. However, there are no existing methods that account for both long-range sequence effects and high interpretability.

To solve this problem, we created a new splice site prediction model consisting of a convolution layer and an attention mechanism, taking as input sequences up to 100k in length. The combination of the convolutional layer and the attention mechanism allowed the model to learn both local and global features. The results show that the proposed model outperforms the state-of-the-art SpliceAI model in terms of precision and F1 score on a dataset of Gencode annotations. We also performed predictions on an aberrant splicing dataset generated from DBASS and compared the performance of the proposed method with SpliceAI. We found that our model is more sensitive to changes in splice sites than SpliceAI if the threshold is set sufficiently low. We also examined which sequences are important for splicing by analyzing the attention weights. We found several exonic splicing enhancer candidates in addition to the GT-AG rule, which is known as a consensus sequence for donor and acceptor sites of intron. Altogether, our proposed model has high interpretability and achieved high prediction performance, which could be useful for disease diagnosis and mechanistic elucidation, such as the detection of pathogenic mutations that cause aberrant splicing.

8. Comparative Single-cell Analyses of Human and Mouse Dendritic Cell Progenitors

Phit Ling Tan, Florent Ginhoux[2], Kenta Nakai
[2]Singapore Immunology Network (SIgN), A*STAR

The dendritic cells (DC) population comprises a heterogeneous family of immune cells, including plasmacytoid DC (pDC) and two subsets of conventional DC (cDC1 and cDC2). Despite the well characterization of mature DC, the origins and differentiation pathways of human DC are still not clearly elucidated as compared to the mouse counterpart. In this study, we compare human and mouse bone marrow cell types in order to find out the progenitors of differentiated DC populations. We integrated multiple public datasets via Mutual Nearest Neighbors (Haghverdi et al., 2018). We also predicted transcription factor (TF) regulons and gene regulatory networks with SCENIC+ (Bravo and De Winter et al., 2023). In the future, we intend to identify the homologous TFs of the same lineage across humans and mouse to understand the genetic mechanisms of DC specification.

9. Discovery of human oncogenes guided by bioinformatics analysis of multi-omics data.

Martin Loza, Alexis Vandenbon[1], Gözde Korkmaz[3], and Kenta Nakai

[3]School of Medicine, Koç University, Turkey

Cis-regulatory elements, as enhancers and promoters, orchestrate the spatial and temporal expression of genes and are indispensable for organismal development and disease protection. In our latest publication (Loza et al., 2023), we uncovered the existence of a subset of housekeeping regulatory elements with potential tumor suppression capabilities.

Functional genomics that is based on functional phenotypic screening has become highly prominent in cancer-focused studies to identify the function of genes and regulatory elements essential for the malignant phenotype. In the last decade, CRISPR-based screening has been stated as a fundamental method for genome-wide loss-of-function (LOF) screening. Pooled library approaches make it possible to uncover novel genes that promote malignancy, and until now, several studies have demonstrated the applicability of CRISPR-based genome-wide LOF screenings to identify novel oncogenic genes.

Altogether, the scope of this study is to validate the functionality of the housekeeping core promoters identified as potential tumor suppressors. Using CRISPR-based screening, we will validate their targets as potential human oncogenes for diverse cancer sub-types.

10. Comparative single-cell transcriptomic analysis reveals key differentiation drivers and potential origin of vertebrate retina

Xin Zeng, Fuki Gyoja[4], Yang Cui, Martin Loza, Takehiro G. Kusakabe[4], Kenta Nakai
[4]Faculty of Science and Engineering, Konan University

The retina plays a vital role in capturing and processing visual information in vertebrates, allowing them to adapt to diverse environments. Single-cell expression profiles of the vertebrate retina have been described; however, a deeper understanding of the expression patterns in the context of development and evolution among homologous cell types is lacking. To identify such shared patterns, we examined and compared approximately 230,000 retinal cells from three species: mouse, chicken, and zebrafish. We found significant similarities among homologous cell types, indicating inherent regulatory patterns. To understand these shared patterns, we constructed gene regulatory networks for each developmental stage for the three species. We identified 690 regulons with 381 regulators across the three species. In addition, we identified ten common cell type-specific regulators and 16 preserved regulons. Using RNA velocity analysis, we pinpointed conserved driver genes key to retinal cell differentiation in both mouse and zebrafish, and by intersecting regulators, we extracted the crucial regulators that facilitate these differentiation processes. Finally, investigation of the potential origins of photoreceptors and retinal ganglion cells by examining conserved expression patterns among the three vertebrate species and the invertebrate Ciona intestinalis revealed functional similarities in light transduction mechanisms between Ciona photoreceptor-related cells and vertebrate retinal cells. Our findings offer insights into the conserved regulatory frameworks intrinsic, evolutionarily preserved differentiation programs, and ancestral origins of vertebrate retinal cells.

11. Exploring the role of hysteresis in macrophage lipid metabolism and robustness of repolarization memory

Yubo Zhang, Kei Ishida[5], Wenbo Yang, Yutaro Kumagai[6], Sung-Joon Park, Jun Kunisawa[6] and Kenta Nakai
[5]National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN)
[6]National Institute of Advanced Industrial Science and Technology (AIST)

Macrophages exhibit polarization, leading to functionally distinct phenotypic changes. This process of polarization and subsequent de-/re-polarization results in hysteresis, where the cells retain genetic and epigenetic markers from previous states, impacting their core functions. To probe the role of hysteresis in lipid metabolism, we utilized computational analysis on an extensive public dataset. Our approach involved reanalyzing 154 RNA-seq datasets pertaining to (re)polarized macrophages, including those with lipid accumulation. We applied Weighted Gene Correlation Network Analysis (WGCNA) and conducted functional enrichment analysis. This led to the identification of 10 significant gene modules, including one uniquely associated with polarized macrophages without de-/re-polarization history, linked to cholesterol biosynthesis. Conversely, a module enriched in inflammatory and LPS response was closely related to cells with a history of specific macrophage de-/re-polarization. To further validate the expression trends of hysteresis genes in de-/repolarized macrophages, we are using lab techniques like PCR and Western Blot, aligning these with prior research. We extended the de-/repolarization duration to bolster the credibility of our results. Our research aims to establish the reproducibility of the hysteresis effect and highlight the importance of its inhibition in regulating lipid metabolism. The complex regulation of cellular memory, however, still presents significant unknowns. Herein, we delve into our ongoing efforts to elucidate the role of macrophage hysteresis. This research was presented at InCOB2023.

12. Identification of transcription factors that contribute to enhancer-promoter communication in mESCs by heterogeneous graph neural network

Yubo Zhang, Xufeng Shu[7], Yitao Yang, Wenbo Yang, Piero Carninci[7] and Kenta Nakai
[7]RIKEN Center for Integrative Medical Sciences

Transcription factors (TFs) play a critical role in the communication between enhancers and promoters, essential for gene expression in mouse embryonic stem cells (mESCs). Moving beyond the traditional focus on pairwise TF interactions, our research delves into a more sophisticated network involving multiple interactions and transcriptional condensates. We employ state-of-the-art sequencing methods and deep learning techniques to unravel complex TF interactions within enhancer-promoter loops, aiming to uncover crucial, yet hidden, TF pairs that influence mESC identity and differentiation. In this endeavor, we use CAGE-seq to identify promoter and enhancer regions in mESCs precisely and complement this with RADICL-seq and PCHi-C for detecting chromatin loops. By integrating computational analysis of TF binding sites with existing protein-protein interaction data, we create an extensive network graph featuring enhancers, promoters, and TFs. The innovative co-contrastive learning method, HeCo, allows us to produce simplified representations of TF nodes. This methodology aids in predicting missing links in the TF network, which we hypothesize to be critical in enhancer-promoter interactions. These predictions, including 20 high-confidence TF-TF interaction pairs, are currently being validated through experimental wet-lab techniques. Our overarching objective is to lead the development of a detailed graph that maps the Enhancer-TFs-Promoter/Genes regulatory network in mESCs. This project is aimed at pinpointing essential TFs and regulatory components involved in enhancer-promoter communications and loop formations. Through this, we hope to facilitate the prediction and discovery of novel TF interactions, thereby advancing the understanding of complex biological networks in stem cell research.

13. Identification of COPS5 as a novel biomarker of diffuse large B cell lymphoma by a machine-learning algorithm

Yutong Dai, Sung-Joon Park, Jingmei Li[8], Keita Yamamoto[8], Susumu Goyama[8] and Kenta Nakai
[8]Graduate School of Frontier Sciences, The University of Tokyo

The transition from generalized medicine to precision medicine has been pivotal in cancer research, emphasizing the need for biomarkers. Despite challenges in identifying meaningful biomarkers, machine-learning approaches for biomarker discovery have been widely adopted. Particularly, Deep learning, while powerful, faces criticism for its interpretability and performance issues due to missing data.

To address these challenges, our study employs the Joint Non-Negative Matrix Factorization (JNMF), an unsupervised algorithm, to unravel cancer biomarkers from multi-omics data. Unlike prior pan-cancer studies, our approach integrates JNMF, gene signature and pathway analysis, and patient data validation, ensuring high interpretability and clinical relevance. The JNMF efficiently handles sparse datasets, demonstrated through simulations and the Cancer Cell Line Encyclopedia (CCLE) datasets. The analysis identifies 40 reproducible and robust feature sets, termed Common Pattern Modules (CPMs), revealing cancer-specific patterns. Notably, the hematopoietic and lymphoid-specific CPM highlights genomic features of diffuse large B-cell lymphoma (DLBCL). Pathway analysis associates DLBCL key pathways, including NF-κB, TP53, JAK-STAT, and PI3K, with COPS5 identified as a key upstream regulator of DLBCL-related pathways in this CPM. Survival analysis using The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) data establishes COPS5 overexpression as a predictor of poor patient survival, corroborating its role in DLBCL. Moreover, Consistent with our observation of overexpression of COPS5, MYC, BCL6, TP53, and STAT3 in DLBCL-specific CPM, a positive correlation between COPS5 expression levels and these genes was also observed in DLBCL patients. More significantly, CRISPR/Cas9 knockout experiments further validate COPS5's significance in promoting malignant growth in mature B-cell neoplasms.

In conclusion, the integrated analysis unveils DLBCL genomic characteristics, spotlighting COPS5 as a potential prognostic biomarker. The methodology holds promise for extracting clinically significant biomarkers from diverse omics data, advancing precision medicine in cancer research.

14. GlycoMSParser: an automated pipeline for comprehensive permethylated glycan analysis

Huan-Chuan Tseng, Wei-Kang Tseng[9], Yuen-Hsien Tseng[10], Kenta Nakai
[9]National Chengchi University, Taiwan
[10]National Taiwan Normal University, Taiwan

We're trying to establish a pipeline that is applicable to automatically annotate glycans (polysaccharides) extracted from proteins obtained from biological samples, which were generally analyzed and annotated manually. Software for annotating spectra based on in silico calculated glycan fragments and database search still needs man to operate it with proper knowledge and conditions based on the experimental

method. In our research, we developed one non-machine learning algorithm that can annotate glycan assisted with glycosyltransferase gene expression (if it exists and the functions are known) and can predict and annotate similar glycan structures under the same experimental method in unknown samples. We're still improving the whole pipeline and testing it with more data and parameters to get more comprehensive and convincing results for publishing this research.

15. Cross-sample genetic correlations reveal a causal relationship between human complex traits

Yi Zheng, Zheng Ning[11], Yudi Pawitan[11], Xia Shen[11], Kenta Nakai
[11]Karolinska Institutet, Sweden

Ascertaining causality in human complex traits is crucial for understanding biology and disease prevention. In this research, we introduce a novel concept termed cross-sample genetic correlation (X-rG). It is calculated from GWAS summary statistics of two phenotypes from different cohorts and serves as an indicator to identify potential causal relationships between complex traits. Through mathematical modeling and numerical simulations, we have found that when the difference in genetic correlation between two cohorts is attributable to causal effects, a significant disparity in X-rG values is observed. Conversely, variation in genetic correlations due to pleiotropy leads to symmetric X-rG values. Leveraging this feature, we screened for asymmetry in X-rG among 211 complex traits between male and female cohorts in the UK Biobank, identifying 207 pairs of traits with potential causal relationships. The findings underscore the pivotal roles of education and occupation type in exposure and outcome, respectively, which are beneficial for improving health indicators and well-being. This method provides an effective complement to causal reasoning for complex traits and diseases and has significant implications for the development of precision medicine strategies.

16. Discovery of antibodies with potential therapeutic applications aided by machine learning and artificial intelligence.

Diana Hinojosa-Trujillo[12], Bowen Liu, Weihang Zhang, Melissa Garcia-Vega[12], Sofia Hernandez-Valenzuela[12], Monica Resendiz-Sandoval[12], Martin Loza, and Jesus Hernandez[12]
[12]Centro de Investigacion en Alimentacion y Desarrollo, Mexico

The use of antibodies to improve human health is an emerging area of biotechnology and immunology. They represent a therapeutic strategy with several applications for managing various types of cancer, infectious and inflammatory diseases, and other noninfectious diseases. Several pharmaceutics and research groups have worked in the last years to find new therapeutic applications for antibodies, including antibody modifications to enhance their efficacy and stability and facilitate their administration. One critical step in this area is antibody discovery, which represents the major challenge of the field. The new computational and experimental methodologies offer important advantages for discovering and testing new antibodies. Single-cell RNAseq is the best example of these new tools since it allows isolating antibody sequences from B-cells of individuals immunized with the therapeutic target. However, analyzing a significant number of sequences takes time and effort. The use of computational approaches like deep learning, machine learning, and artificial intelligence can potentially assist in discovering new antibodies, providing a tool for the prediction of neutralizing profiles, thus virtually reducing the time, effort, and resources invested.

In this project, we aim to build a bioinformatics pipeline based on artificial intelligence to discover new neutralizing antibodies against SARS-CoV-2. The final platform could be extended to assist in the discovery of antibodies against other diseases using public libraries or new ones generated by Prof. Hernández´s Laboratory.

17. Identifying pioneer molecular modulating global chromatin accessibility by genome-wide ATAC-see screening and ATAC-seq analysis

Sung-Joon Park, Yasuyuki Ohkawa[13], Kenta Nakai, Yusuke Miyanari[14]
[13] Medical Institute of Bioregulation, Kyushu University
[14]WPI Nano Life Science Institute, Kanazawa University

The dynamics of chromatin accessibility play a crucial role in a local environment for accomplishing various biological functions. Here, to understand the molecular mechanisms underlying the control of chromatin accessibility, we conducted a genome-wide CRISPR screening combined with an optimized ATAC-see that covers more than 19,000 human genes in eHAP cells, along with developing a computational analytic pipeline. This examination detected genes previously unknown to modulate global chromatin accessibility, including TFDP1, HNRNPU, EIF3D, and THAP11. Particularly, the transcription factor TFDP1 markedly impacted global chromatin accessibility through transcriptional regulation of canonical histones. That is, the TFDP1-depleted cells showed increased global chromatin accessibility and enhanced the efficiency of DNA-associated applications,

including iPSC reprogramming. These results highlight the potential that the manipulation of chromatin accessibility by altering key molecules is a promising tool for enhancing the efficiency of various cell engineering applications.

18. Resting heart rate and risk of dementia: a Mendelian Randomization Study in the International Genomics of Alzheimer9s Project and UK Biobank

Xingxing Chen[15,16], Yi Zheng, Jun Wang[17], Blake Yue[18,19], Xian Zhang[16], Kenta Nakai , Lijing L Yan[15,16,20,21]

[15]School of Public Health, Wuhan University, China
[16]Duke Kunshan University, Global Health Research Center, China
[17]Huazhong University of Science and Technology, China
[18]Edith Cowan University, Australia
[19]Auckland University of Technology, New Zealand
[20]Duke University, USA
[21]Peking University, China

Observational studies have demonstrated that a higher resting heart rate (RHR) is associated with an increased risk of dementia. However, it is not clear whether the association is causal. This study aimed to determine the causal effects of higher genetically predicted RHR on the risk of dementia. We performed a generalized summary Mendelian randomization (GSMR) analysis to analyze the corresponding effects of higher RHR on the following different outcomes: 1) diagnosis of AD (International Genomics of Alzheimer's Project), 2) family history (maternal and paternal) of AD from UK Biobank, 3) combined meta-analysis including these three GWAS results. Further analyses were conducted to determine the possibility of reverse causal association by adjusting for RHR modifying medication. The results of GSMR showed no significant causal effect of higher genetically predicted RHR on the risk of AD (β_{GSMR} = 0.12, P = 0.30). GSMR applied to the maternal family history of AD (β_{GSMR} = -0.18, P = 0.13) and to the paternal family history of AD (β_{GSMR} = -0.14, P = 0.39) showed the same results. Furthermore, the results were robust after adjusting for RHR modifying drugs (β_{GSMR} = -0.03, P = 0.72). In conclusion, our study did not find any evidence that supports a causal effect of higher RHR on dementia. Previous observational associations between RHR and dementia are likely attributed to the correlation between RHR and other cardiovascular diseases.

19. DeepBIO: An automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation, and visualization analysis

Ruheng Wang[22], Yi iang[22], Junru Jin[22], Chenglin Yin[22], Haoqing Yu[22], Fengsheng Wang[22], Jiuxin Feng[22], Ran Su[23], Kenta Nakai, Quan Zou[24], & Leyi Wei[22]

[22]Shandong University, China
[23]Tianjin University, China
[24]University of Electronic Science and Technology of China, China

DeepBIO is the first-of-its-kind automated and interpretable deep-learning platform for high-throughput biological sequence functional analysis. It is a one-stop-shop web service that enables researchers to develop new deep-learning architectures to answer any biological question. Specifically, given any biological sequence data, DeepBIO supports a total of 42 state-of-the-art deep-learning algorithms for model training, comparison, optimization, and evaluation in a fully automated pipeline. DeepBIO provides a comprehensive result visualization analysis for predictive models covering several aspects, such as model interpretability, feature analysis, and functional sequential region discovery. We expect DeepBIO to ensure the reproducibility of deep-learning biological sequence analysis, lessen the programming and hardware burden for biologists, and provide meaningful functional insights at both the sequence level and base level from biological sequences alone. DeepBIO is publicly available at https://inner.wei-group.net/DeepBIO.

20. 19n01, a broadly neutralizing antibody against omicron BA.1, BA.2, BA.4/5, and other SARS-CoV2 variants of concern

Melissa Garcia-Vega[12], Edgar A. Melgoza-Gonzalez[12], Sofia Hernandez-Valenzuela[12], Diana Hinojosa-Trujillo[12], Monica Resendiz-Sandoval[12], Mara Anais Llamas-Covarrubias[25], Martin Loza-Lopez, Olivia Valenzuela[26], Alan Soto-Gaxiola[27], Miguel A. Hernandez-Onate[28], Veronica Mata-Haro[28], Irene Cassaniti[39], Jose Camilla Sammartino[29], Alessandro Ferrari[29], Luca Simonelli[30], Mattia Pedotti[30], Rui Sun[11], Fanglei Zuo[11], Fausto Baldanti[29], Luca Varani[30], Harold Marcotte[11], Qiang Pan-Hammarstrom[11], and Jesus Hernandez[12]

[25]Research Institute for Microbial Diseases, Osaka University
[26]Division de Ciencias de la Salud, Universidad de Sonora, Mexico
[27]Secretaria de Salud del Estado de Sonora, Mexico
[28]Centro de Investigacion en Alimentacion y Desarrollo, Mexico
[29]Fondazione IRCCS Policlinico San Matteo, Italy
[30]Universita` della Svizzera italiana (USI), Italy

This study reports the isolation and characterization of a human monoclonal antibody (mAb) called 19n01. This mAb was isolated by using single-cell RNAseq of B cells from donors infected with the an-

cestral strain. This mAb possesses a potent and broad capacity to bind and neutralize all previously circulating variants of concern (VOCs), including Omicron sublineages BA.1, BA.2, and BA.4/5. The pseudovirus neutralization assay revealed robust neutralization capacity against the G614 strain, BA.1, BA.2, and BA.4/5, with inhibitory concentration (IC50) values ranging from 0.0035 to 0.0164 mg/mL. The microneutralization assay using the G614 strain and VOCs demonstrated IC50 values of 0.013–0.267 mg/mL. Biophysical and structural analysis showed that 19n01 cross-competes with ACE2 binding to the receptor-binding domain (RBD), and the kinetic parameters confirmed the high affinity against the Omicron sublineages (KD of 61 and 30 nM for BA.2 and BA.4/5, respectively). These results suggest that the 19n01 is a remarkably potent and broadly reactive mAb.

Publication list

Ruheng Wang, Yi iang, Junru Jin, Chenglin Yin, Haoqing Yu, Fengsheng Wang, Jiuxin Feng, Ran Su, Kenta Nakai, Quan Zou, & Leyi Wei. DeepBIO: An automated and interpretable deep-learning platform for high-throughput biological sequence prediction, functional annotation, and visualization analysis. *Nucl. Acids Res*., **51**(7), 3017-3029 (2023). (Breakthrough Article)

Shu Liu, Fujio Toriumi, Xin Zeng, Mao Nishiguchi, and Kenta Nakai. SignedS2V: Structural embedding method for signed networks. in (Hocine Cherifi, Rosario Nunzio Mantegna, Luis M. Rocha, Chantal Cherifi, & Salvatore Miccichè eds.) Complex Networks and Their Applications XI. (Proceedings of The Eleventh *International Conference on Complex Networks and Their Applications:* COMPLEX NETWORKS 2022) Volume 1, pp.337-349 (2023). Springer ISBN: 978-3-031-21126-3

Esra Büşra Işık, Michelle D. Brazas, Russell Schwartz, Bruno Gaeta, Patricia M. Palagi, Celia W. G. van Gelder, Prashanth Suravajhala, Harpreet Singh, Sarah L. Morgan, Hilyatuz Zahroh, Maurice Ling, Venkata P. Satagopam, Annette McGrath, Kenta Nakai, Tin Wee Tan, Ge Gao, Nicola Mulder, Christian Schönbach, Yun Zheng, Javier De Las Rivas & Asif M. Khan, Grand challenges in bioinformatics education and training (career feature). *Nature Biotechnology*, **41**, 1171-1174 (2023).

Dan Li, Yoshikazu Johmura, Satoru Morimoto, Miyuki Doi, Keiko Nakanishi, Manabu Ozawa, Yuji Tsunekawa, Akane Inoue-Yamauchi, Hiroya Naruse, Takashi Matsukawa, Yukio Takeshita, Naoki Suzuki, Masashi Aoki, Ayumi Nishiyama, Xin Zeng, Chieko Konishi, Narumi Suzuki, Atsuya Nishiyama, Alexander Stephen Harris, Mariko Morita, Kiyoshi Yamaguchi, Yoichi Furukawa, Kenta Nakai, Shoji Tsuji, Satoshi Yamazaki, Yuji Yamanashi, Shoichi Shimada, Takashi Okada, Hideyuki Okano, Tatsushi Toda, Makoto Nakanishi, LONRF2 is a protein quality control ubiquitin ligase whose deficiency causes late-onset neurological deficits. *Nature Aging,* **3**(8), 1001-1019 (2023).

Zhongshen Li, Junru Jin, Wenjia He, Wentao Long, Haoqing Yu, Xin Gao, Kenta Nakai, Quan Zou, and Leyi Wei. CoraL: interpretable contrastive meta-learning for the prediction of cancer-associated ncRNA-encoded small peptides. *Brief. Bioinform.,* **24**(6), bbad352 (2023).

Martin Loza, Alexis Vandenbon, and Kenta Nakai. Epigenetic characterization of housekeeping core promoters and their importance in tumor suppression. *Nucleic Acids Research,* (2023): gkad1164.

Yubo Zhang, Wenbo Yang,Yutaro Kumagai, Martin Loza,Weihang Zhang, Sung-Joon Park, and Kenta Nakai. Multi-omics computational analysis unveils the involvement of AP-1 and CTCF in hysteresis of chromatin states during macrophage polarization. *Front. Immunol*., **14**, 1304778 (2023).

Satoko Ishii, Taishi Kakizuka, Sung-Joon Park, Ayako Tagawa, Chiaki Sanbo, Hideyuki Tanabe, Yasuyuki Ohkawa, Mahito Nakanishi, Kenta Nakai, and Yusuke Miyanari, Genome-wide ATAC-see screening identifies TFDP1 as a modulator of global chromatin accessibility, *Nature Genetics*, in press

García-Vega Melissa, Edgar A. Melgoza-González, Sofía Hernández-Valenzuela, Diana Hinojosa-Trujillo, Mónica Reséndiz-Sandoval, Mara Anais Llamas-Covarrubias, Loza Martin, et al. "19n01, a broadly neutralizing antibody against omicron BA. 1, BA. 2, BA. 4/5, and other SARS-CoV-2 variants of concern." *iScience,* **26**, no. 4 (2023).

Jiravejchakul Natnicha, Gabriela L. Abe, Loza Martin, Soyoung Park, Ponpan Matangkasombut, Jun-Ichi Sasaki, Satoshi Imazato, Diego Diez, and Daron M. Standley. "Intercellular crosstalk in adult dental pulp is mediated by heparin-binding growth factors Pleiotrophin and Midkine." *BMC Genomics,* **24**, no. 1 (2023): 184.