Develop a medical diagnostic approach based on the altered RNA-RBP binding dynamics

PI: Prof. Jilin Zhang Co-I: Prof. Jian Yan, Prof. Xin Deng, Prof. Bing Fu (CityUHK BME)

Background

The elderly population in Hong Kong will become doubled by 2046, which means an increase from 1.45 million in 2021 to 2.7million accounting for more than 35% of the inhabitants (https://www.censtatd.gov.hk/en/press_release_detail.html?id=5368). This population structure change poses an emergent and unavoidable burden to the society introduced by natural ageing, particularly to the healthcare and medical systems, as ageing is associated with many types of human diseases[1]. Although the skin is the first barrier of the human body for protection and immune response, ageing-associated human skin diseases have not received adequate attention as neurodegenerative diseases or cancers[2].

Like the ageing-associated Alzheimer's disease, Parkinson's disease and order types of neurodegenerative diseases, a higher prevalence of ageing-associated skin diseases has been observed along the natural ageing process in several populations[3]. Among the surveyed patients, dermatitis, neoplasm and precancerous skin demonstrate a higher incidence besides the infectious conditions[4,5]. Natural ageing could be seen as the result of systematic failure after long-term stress stimuli that overloaded the repairing mechanisms, therefore resulting in compromised cell functions and increased vulnerability towards endogenous toxicity[6,7]. This eventually led to a complete loss of renewal capability while simultaneously being arrested by the irreversible senescent state.

Cells utilize mechanisms that harbor membrane-less organelle (MLO), such as phase-separated stress granules, to respond to stress upon external stimuli [8]. MLO is one type of essential sub-cellular structure that widely and dynamically exists in cells. The condensates observed in MLO are formed by aggregated molecules, or liquid-liquid phase separation (LLPS), which compartmentalize the environment to constrain biochemical reactions. Many studies have indicated that molecule crowding is not a univariate process but can involve many distinct components while demonstrating great dynamics and heterogeneity depending on the cellular context [9]. Apart from maintaining homeostasis after exposure to various types of stress, cells can transform the deviated homeostasis state into an irreversible state by establishing a persistent response to the stress, leading to dysregulated MLO, causing subsequent endogenous toxicity or the onset of parthenogenesis. Indeed, constituent proteins of ribonucleoproteins involved in the stress granules are coupled with atopic dermatitis (AD) and various neurodegenerative diseases [8].

Meanwhile, growing evidence has shown that RNA and RNA binding proteins (RBPs) are indispensable components for condensates in many cellular contexts, including but not limited to the stress granules in neurodegenerative diseases and condensates formed during X chromosome inactivation. Dysregulated RNAs or RBPs are tightly coupled with human diseases, e.g. neurodegenerative and skin diseases [10,11]. The regulatory roles of MLO in human diseases and underlying mechanisms remain under extensive investigation, although either RBP or RNA G-quadruplex (rG4), a non-canonical secondary RNA structure, can trigger the LLPS alone. As RNAs often form complexes with RBPs to exert their functional roles, their dynamic interactions are integral to both transcriptional and post-transcriptional regulation [12]. However, the detailed molecular mechanisms have not been completely elucidated, and using their features to develop and realize diagnostic methodologies remains extremely challenging.

The primary obstacle hindering mechanistic studies is the incomplete understanding of the binding between RBPs and RNAs. This is due to the complex and multi-facet binding activity of RBPs with RNA molecules, which involve linear RNA, RNA structures and dynamic binding change between different binding states [13-15]. Although several *in vivo* and *in vitro* methods, including the HTR-SELEX we established [14], RNAcompete and RNA Bind-n-seq, have been developed to disclose the binding specificities of RBPs, the technical or computational limitations prohibit access to the holistic RNA structure map. The missing structural information of RNAs recognized by RBPs has thus considerably slowed down the decoding regulation of the RNA-RBP

complex in the regulation of MLO.

Another bottleneck is the poorly resolved binding dynamics between RBPs and RNA *in vivo*. Previous studies that revealed binding specificities of RBPs, including the HTR-SELEX we established, have confirmed that RBPs binding specificities are multifaceted, allowing RBPs to bind various linear and/or structural RNAs. More interestingly, besides the hairpin-shaped canonical RNA stem-loop secondary structure, RNAs can also form rG4s that widely exist in the eukaryotic transcriptomes [16]. Even though many RBPs can or are predicted to interact with rG4s, the exact rG4 sequences and whether RBPs have preferred forms of rG4 are not clearly resolved. How the rG4 binding RBPs interact with their linear binding sites has not been thoroughly investigated. Not even mention the dynamics of RBP binding among distinct RNAs recognized by the same RBPs. Nevertheless, without a comprehensive understanding of the RBP binding specificities and structural RNA information, it is impossible to integrate the linear and structural RNA binding dynamics into solving biological questions, reversing the consequence of persistent stress stimuli, and translating the harvested knowledge into diagnostic and therapeutic applications.

Reference

1 Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol* **22**, 75-95 (2021).

2 Sreedhar, A., Aguilera-Aguirre, L. & Singh, K. K. Mitochondria in skin health, aging, and disease. *Cell Death Dis* **11**, 444 (2020).

3 Gu, Y., Han, J., Jiang, C. & Zhang, Y. Biomarkers, oxidative stress and autophagy in skin aging. *Ageing Res Rev* **59**, 101036 (2020).

4 Sinikumpu, S. P., Jokelainen, J., Haarala, A. K., Keranen, M. H., Keinanen-Kiukaanniemi, S. & Huilaja, L. The High Prevalence of Skin Diseases in Adults Aged 70 and Older. *J Am Geriatr Soc* **68**, 2565-2571 (2020).

5 Darjani, A., Mohtasham-Amiri, Z., Mohammad Amini, K., Golchai, J., Sadre-Eshkevari, S. & Alizade, N. Skin Disorders among Elder Patients in a Referral Center in Northern Iran (2011). *Dermatol Res Pract* **2013**, 193205 (2013).

6 Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695-703 (2021).

7 Yousefzadeh, M., Henpita, C., Vyas, R., Soto-Palma, C., Robbins, P. & Niedernhofer, L. DNA damage-how and why we age? *Elife* **10** (2021).

8 Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nat Rev Neurosci* **20**, 649-666 (2019).

9 Rhine, K., Vidaurre, V. & Myong, S. RNA Droplets. *Annu Rev Biophys* **49**, 247-265 (2020).

10 Hanson, K. A., Kim, S. H. & Tibbetts, R. S. RNA-binding proteins in neurodegenerative disease: TDP-43 and beyond. *Wiley Interdisciplinary Reviews: RNA* **3**, 265-285 (2012).

11 Wang, E. T., Taliaferro, J. M., Lee, J. A., Sudhakaran, I. P., Rossoll, W., Gross, C., Moss, K. R. & Bassell, G. J. Dysregulation of mRNA Localization and Translation in Genetic Disease. *J Neurosci* **36**, 11418-11426 (2016).

12 Kharel, P., Fay, M., Manasova, E. V., Anderson, P. J., Kurkin, A. V., Guo, J. U. & Ivanov, P. Stress promotes RNA G-quadruplex folding in human cells. *Nat Commun* **14**, 205 (2023).

13 Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J. Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., Hochman, M., Lambert, N. J., Li, H., McGurk, M. P., Nguyen, T. B., Palden, T., Rabano, I., Sathe, S., Stanton, R., Su, A., Wang, R., Yee, B. A., Zhou, B., Louie, A. L., Aigner, S., Fu, X. D., Lecuyer, E., Burge, C. B., Graveley, B. R. & Yeo, G. W. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719 (2020).

14 Jolma, A., Zhang, J., Mondragón, E., Morgunova, E., Kivioja , T., Yin, Y., Zhu, F., Bourenkov, G., Morris, Q., Hughes, T. R., Maher III, L. J. & Taipale, J. Binding specificities of human RNA binding proteins towards structured and linear RNA sequences. *Genome Research* **30**, 962-973 (2020).

15 Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R. & Burge, C. B. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**, 854-867 e859 (2018).

16 Dumas, L., Herviou, P., Dassi, E., Cammas, A. & Millevoi, S. G-Quadruplexes in RNA Biology: Recent Advances and Future Directions. *Trends Biochem Sci* **46**, 270-283 (2021).

17 Yang, L., Lin, Z., Gao, Y., Zhang, J., Peng, H., Li, Y., Che, J., Zhao, L. & Zhang, J. Populational pan-ethnic screening panel enabled by deep whole genome sequencing. *NPJ Genom Med* **8**, 38 (2023).

18 Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J., Chen, F., Taipale, J., Chan, K. M., Zhang, L. & Yan, J. CRISPR-assisted detection of RNA-protein interactions in living cells. *Nat Methods* **17**, 685-688 (2020).

19 Bourdon, S., Herviou, P., Dumas, L., Destefanis, E., Zen, A., Cammas, A., Millevoi, S. & Dassi, E. QUADRatlas: the RNA G-quadruplex and RG4-binding proteins database. *Nucleic Acids Res* (2022).

20 Shen, Y., Wu, S. Y., Rancic, V., Aggarwal, A., Qian, Y., Miyashita, S. I., Ballanyi, K., Campbell, R. E. & Dong, M. Genetically encoded fluorescent indicators for imaging intracellular potassium ion concentration. *Commun Biol* **2**, 18 (2019).

21 Bischof, H., Rehberg, M., Stryeck, S., Artinger, K., Eroglu, E., Waldeck-Weiermair, M., Gottschalk, B., Rost, R., Deak, A. T., Niedrist, T., Vujic, N., Lindermuth, H., Prassl, R., Pelzmann, B., Groschner, K., Kratky, D., Eller, K., Rosenkranz, A. R., Madl, T., Plesnila, N., Graier, W. F. & Malli, R. Novel genetically encoded fluorescent probes enable real-time detection of potassium in vitro and in vivo. *Nat Commun* **8**, 1422 (2017).

22 Fu, B., Sengupta, K., Genova, L. A., Santiago, A. G., Jung, W., Krzeminski, L., Chakraborty, U. K., Zhang, W. & Chen, P. Metal-induced sensor mobilization turns on affinity to activate regulator for metal detoxification in live bacteria. *Proc Natl Acad Sci U S A* **117**, 13248-13255 (2020).

Work done by us

Our latest research on autosomal recessive disorders revealed that condensates are heavily involved in human autosomal recessive diseases[17]. The mutations of protein filaggrin (FLG) abolish its condensation within keratohyalin granules to compromise skin defence in AD patients. However, the roles of RNAs in this process are yet to be investigated.

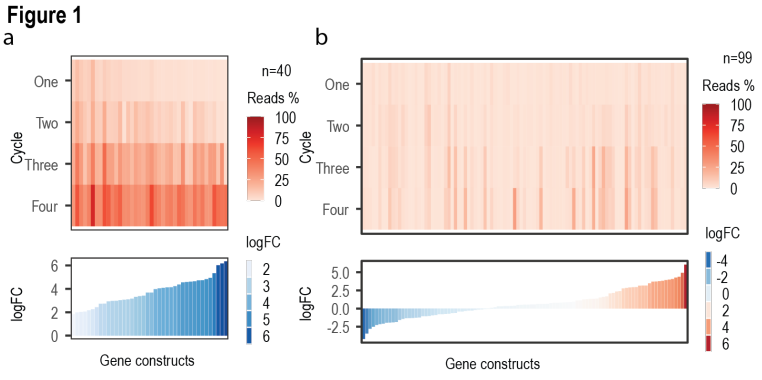
We previously established an *in vitro* assay, HTR-SELEX[14], to capture both linear and structural binding specificities of human RBPs. We also successfully applied HTR-SELEX in various scenarios and proved its robustness [16,18]. We recently observed that under rG4 stabilising condition (Na^+), many RBPs demonstrated strong binding affinity towards rG4 forming oligos (**Figure 1a**). Besides binding to linear RNA sequences and canonical RNA structures (**Figure 1b**), the RNA libraries of some RBPs also showed unexpected rG4 enrichment. We have adapted the HTR-SELEX protocol to stabilise and destabilise rG4, aiming to identify the rG4 forms recognised by RBPs to disclose the principles of rG4 binding specificity and uncover linear or structural RNA motifs masked by rG4 (**Figure 1c**).

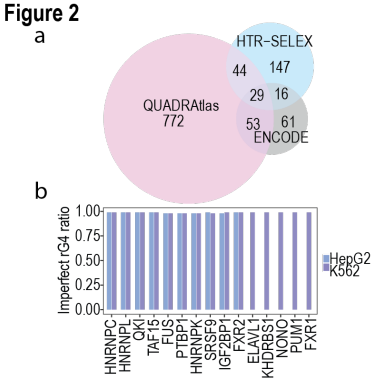
In parallel to the experimental evidence that many RBPs can bind rG4 with strong affinity observed in the HTR-SELEX data, many known rG4 binding RBPs are independently documented in the database Quadratlas [19] (**Figure 2a**). We further re-analysed the ENCODE eCLIP data [13] and confirmed that the rG4 forming sequences were partially recovered by HTR-SELEX (**Figure 2b**). These rG4s were highly enriched in the non-transposable element regions (**Figure 3**), particularly these non-canonical rG4s (**Figure 4**). Moreover, we observed no obvious spatial proximity at the sequence level for most rG4 binding RBPs with known binding specificities toward linear RNAs, suggesting binding with RBPs itself is potentially independent (**Figure 5a**). The explanation of such sparse embedding of linear and structural motifs remains enigmatic, as the full RNA binding spectrum of most RBPs is still poorly investigated. However, we found that some linear RNA motifs recognised by HNRNPC can be nested in the loop region of rG4 sequences (**Figure 5b**). We hypothesise that if linear motifs were not masked by rG4, the binding between rG4 and RBP requires sequence specificity to mediate the binding or recognition of RNA structure.

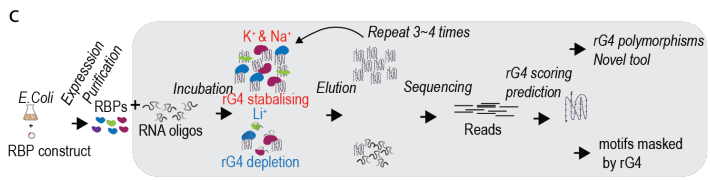
Objectives

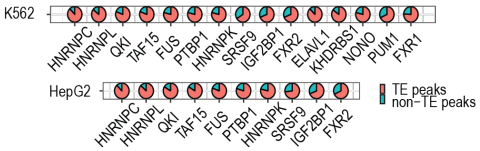
To exploit the potential of using RBP-RNA dynamics for diagnostic purposes, we aim to bridge the knowledge gap between the dynamics of RBP-RNA and the regulation of MLO. Three specific aims will be pursued to lay a solid foundation to complete the overall objective:

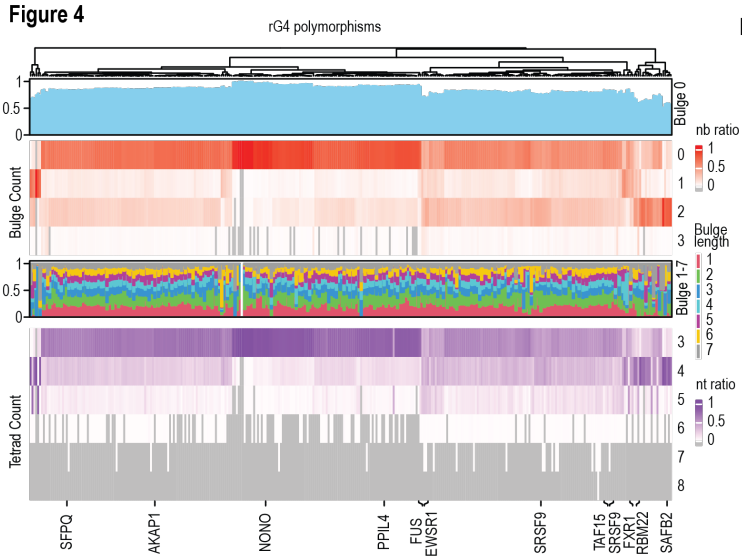
(1) To elucidate the linear and structural basis of RNA molecules bound by the same RBP
(2) To trace altered RNA-RBP dynamics in live cells
(3) To devise a deep-learning algorithm to diagnose skin disease based on RNA-RBP dynamics

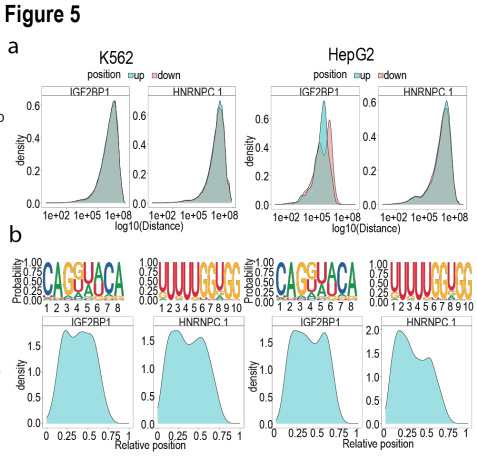












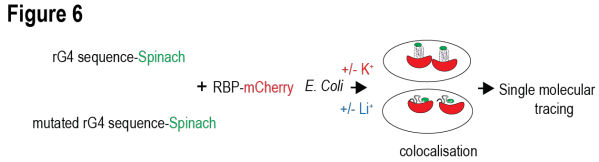


Figure 1. a, b) The rG4 enrichment in the libraries lacking linear sequences or stem-loop structures and libraries with linear and stem-loop RNAs. logFC (log2 fold change) below heatmaps presents the fold change of rG4 percentage between the 4th and 1st cycle. The X-axis includes 99 constructs of 86 RBPs. **c**) Illustration of the HTR-SELEX workflow in rG4 stabilising and depleting conditions.

Figure 2. a) Common rG4 binding RBPs detected by HTR-SELEX, QUADRAtlas and RBPs with available eCLIP. **b**) The ratio of distinct form of non-canonical rG4 in both K562 and HepG2 cell lines.

Figure 3. Most eCLIPs peaks are TE-related.

Figure 4. rG4 polymorphisms recognized by RBPs. The rG4 in the 4th cycle of libraries are predicted. For a given RBP, bulge count, bulge length, and G-tetrad count in rG4 forming ligands are extracted. The symbols of representative RBPs recognizing distinct types of rG4 are labelled.

Figure 5. a) linear RNA motifs embedded in rG4. The top 300 thousand predicted linear motifs are selected to investigate the relative position of nested linear motifs in rG4. **b**) The genomic distance between predicted rG4 forming sequences and linear motif in the eCLIP peak data.

Figure 6. The schematic illustration of single-molecule tracking experiments.