Discovery of novel biomolecular condensate drug targets in oncology using *in silico* predictive tools

Kadi L. Saar[1], Marius Rebmann[1], Mohammed Kanchwala[1], Seema Qamar[1], Prathima Radhakrishnan[1], Jasmine Cornish[1], Julia Doh[1], William Arter[1], Assaf Rotem[1], Mahmoud Ghandi[1], Andrew Seeber[1], Richard C. Centore[1], Martin Kulander[1], Tuomas P.J. Knowles[1,2,3], Sarah A. Teichmann[3,4], Shilpi Arora[1]

[1] Transition Bio, Cambridge, UK & Watertown, MA, US
[2] Yusuf Hamied Department of Chemistry, University of Cambridge, UK
[3] Cavendish Laboratory, Department of Physics, University of Cambridge, UK
[4] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

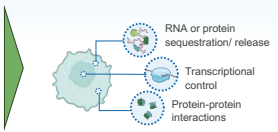
TransitionBio

Abstract # 4891

Intrinsically disordered proteins (IDPs) frequently form biomolecular condensates

Opportunity
~70% of cancer-associated proteins are IDPs

Challenge
IDPs are difficult to drug with traditional screening methodologies

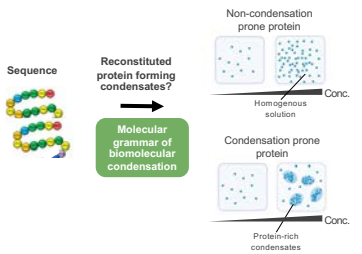


Condensates
IDPs frequently form liquid-like biomolecular assemblies that control macromolecular interactions

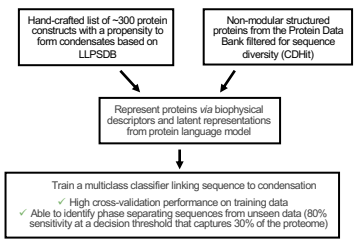
Disease causality
Condensates can drive aberrant transcription, translation & protein interactions

Use of protein language models and biophysical descriptors to develop predictive machine learning models that link protein's sequence to its condensation behavior

Task: Develop a model that links a protein's sequence to its propensity to form condensates

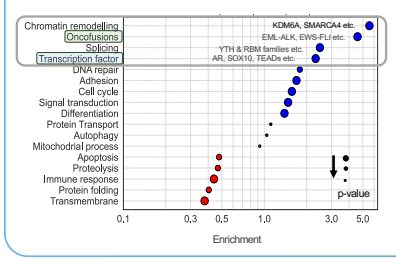


Training: Use experimental data, biophysical descriptors and latent representations from pre-trained protein language model to train the model

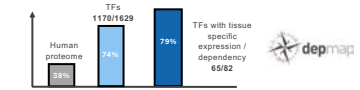


Proteome-wide *in silico* screen using the developed model predicts oncofusions (OFs) and transcription factors (TFs) as protein classes that frequently form condensates

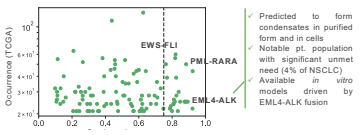
Oncofusions (OFs), transcription factors (TFs), and proteins associated with chromatin remodeling and splicing have high predicted condensation propensity relative to proteome (enrichment calculated with Fisher test)



72% human TFs and 79% of TFs with selective dependency and/or tissue specific expression in *in vitro* models are predicted to form condensates

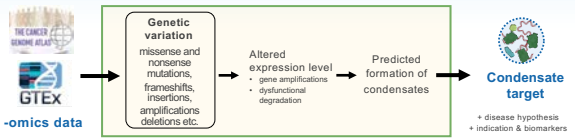


Several frequently occurring OFs are predicted to form condensates

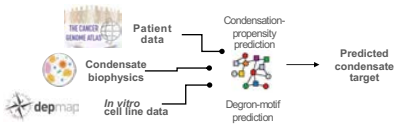


✓ Predicted to form condensates in purified form and in cells
✓ Notable pt. population with significant unmet need (4% of NSCLC)
✓ Available *in vitro* models driven by EML4-ALK fusion

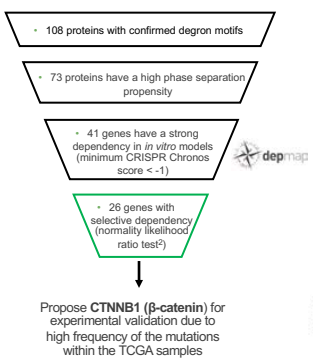
Hypothesis: Genetic variations can lead to aberrant condensation *via* increased intracellular abundance of the target protein



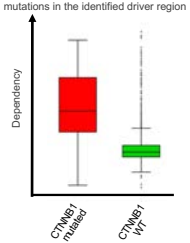
Identification of gain-of-function mutations in degron domains that can lead to target overexpression and aberrant condensation uncovers β-catenin as a target



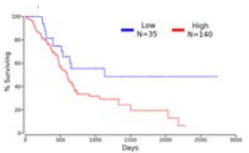
- Identify degron domain mutations within TCGA proteomics samples[1]
- Use developed ML models to evaluate if protein forms condensates
- Evaluate if *in vitro* cell models show selective dependency on the target[2]
- Analyze if mutations elevate dependency in *in vitro* models or lead to reduced survival in patient samples



In vitro CRISPR knock-out dependency data (DepMap) confirms elevated CTNNB1 dependency in the models that carry mutations in the identified driver region



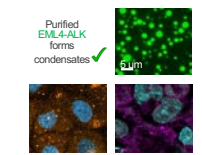
TCGA survival data suggests patients with higher level of beta-catenin have a worse survival prognosis (pancreatic adenocarcinoma cohort)

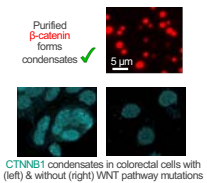


[1] Martinez-Jimenez et al. *Nat Cancer*, 2019
[2] McDonald III et al *Cell*, 2017

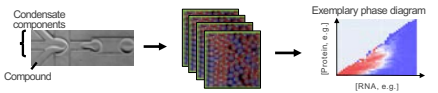
Predicted targets form condensates and are amenable to screening with Transition Bio's platform

1. *In silico* target ID pipeline predicted targets form condensates in a purified form (top) and in the disease models (bottom)





Transition Bio has established a proprietary microfluidics and machine learning enabled molecular screen for identifying compounds that alter condensates

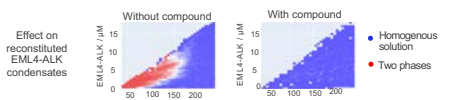


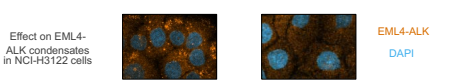
Thousands of microdroplets are generated, each with a different concentration of protein, trigger and test compound

Neural network model detects presence (red) or absence (blue) of condensates in each droplet

Phase diagrams are constructed to identify compounds that shift phase boundary

2. Screen identifies compounds that modulate reconstituted and cellular EML4-ALK condensates





Summary

- We developed machine learning models that link protein sequence to its propensity to form condensates
- **Onco-fusions** and proteins involved in **splicing** and **transcription** are predicted to frequently function *via* condensate-based mechanisms
- We combined –omics data with our predictive condensation models to identify **genetic variations that lead to aberrant condensation**
- Predicted targets from *in silico* target identification pipeline (EML4-ALK, β-catenin) form condensates, phase separate *in vitro* and are amenable to screening with our platform