

METHODS OF IDENTIFYING A CONDENSATE PHENOTYPE AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority benefit of U.S. Provisional Patent Application No. 63/155,683, filed on March 2, 2021, and U.S. Provisional Patent Application No. 63/298,171, filed on January 10, 2022, the content of each of which is incorporated herein by reference in their entireties.

TECHNICAL FIELD

[0002] The present application relates to the field of biological condensates.

BACKGROUND

[0003] Conventional disease research and application to therapeutic drug discovery focuses on identifying individual biomolecules (molecular targets) that cause or mediate disease-relevant biology, such as via known cellular pathways. Two principles are integral with this therapeutic drug discovery approach: (1) the molecular target needs to be validated (*e.g.*, using genetics or pharmacology), to provide evidence that modulation of intrinsic molecular target activity and/or interactions with other molecules may act to alleviate disease biology; and, (2) the molecular target needs to be “druggable,” with at least one therapeutic modality (*e.g.*, a small molecule or antibody) capable of effecting the desired impact on the molecular target modulation *in vivo*. This “single target” approach to disease research and therapeutic drug discovery has significant limitations as *in vivo* systems are extraordinarily complex and our understanding of cellular pathways and biomolecular actions and interactions are often incomplete; moreover, not all individual biomolecules, when viewed singularly, are suitable drug targets. Such challenges have contributed to the current shortage of new therapeutic drugs to treat unmet medical needs.

BRIEF SUMMARY

[0004] In some examples, provided herein is a method of identifying a condensate of interest associated with a disease, the method comprising: (a) comparing a first condensate phenotype from a first cell model with a second condensate phenotype from a second cell model, wherein a

difference between the first cell model and the second cell model is attributable to one or more disease-associated factors, and wherein at least one of the one or more disease-associated factors is introduced to either the first cell model or the second cell model; and (b) identifying the condensate of interest as associated with the disease based on a difference between the first condensate phenotype and the second condensate phenotype for the condensate of interest.

[0005] In some examples according to any one of the methods described herein, the first condensate phenotype and the second condensate phenotype are each characterized by one or more phenotypic identifiers. In some examples, the one or more phenotypic identifiers comprise an identifier selected from the group consisting of a condensate presence, absence, level, morphological feature, location, behavior, composition, and material property.

[0006] In some examples according to any one of the methods described herein, the one or more disease-associated factors associated with the disease comprise a factor selected from the group consisting of a genetic variant, post-translational modification variant, exogenous genetic material, presence of an endogenous compound, presence of an exogenous compound, a physical process (*e.g.*, stress), and environmental stimulus.

[0007] In some examples according to any one of the methods described herein, the second cell model is treated and/or engineered based on the one or more disease-associated factors associated with the disease.

[0008] In some examples according to any one of the methods described herein, the first cell model is treated and/or engineered based on the one or more disease-associated factors associated with the disease.

[0009] In some examples according to any one of the methods described herein, the method further comprises obtaining the second cell model.

[0010] In some examples according to any one of the methods described herein, the method further comprises producing the second cell model.

[0011] In some examples according to any one of the methods described herein, the method further comprises obtaining the first cell model.

[0012] In some examples according to any one of the methods described herein, the method further comprises producing the first cell model.

[0013] In some examples according to any one of the methods described herein, the method further comprises obtaining the first condensate phenotype. In some examples, obtaining the first condensate phenotype comprises measuring an association of a first marker with the condensate of interest. In some examples, the first marker is a biological marker. In some examples, the association of the first marker with the condensate of interest is determined using an imaging technique. In some examples, the imaging technique comprises labeling the first marker.

[0014] In some examples according to any one of the methods described herein, the method further comprises obtaining the second condensate phenotype. In some examples, obtaining the second condensate phenotype comprises measuring an association of a second marker with the condensate of interest. In some examples, the second marker is a biological marker. In some examples, the association of the second marker with the condensate of interest is determined using an imaging technique. In some examples, the imaging technique comprises labeling the second marker.

[0015] In some examples according to any one of the methods described herein, the method further comprises determining the difference between the first condensate phenotype and the second condensate phenotype. In some examples, determining the difference between the first condensate phenotype and the second condensate phenotype comprises a qualitative assessment. In some examples, determining the difference between the first condensate phenotype and the second condensate phenotype comprises a quantitative assessment. In some examples, determining the difference between the first condensate phenotype and the second condensate phenotype comprises an *in silico* technique.

[0016] In some examples according to any one of the methods described herein, the condensate of interest is present in, or derived from, the first cell model.

[0017] In some examples according to any one of the methods described herein, the condensate of interest is absent in, or not derived from, the first cell model.

[0018] In some examples according to any one of the methods described herein, the condensate of interest is absent in, or not derived from, the second cell model.

[0019] In some examples according to any one of the methods described herein, the condensate of interest is present in, or derived from, the second cell model.

[0020] In some examples according to any one of the methods described herein, the condensate of interest belongs to a condensate type selected from the group consisting of a stress granule, cleavage body, p-granule, histone locus body, multivesicular body, neuronal RNA granule, nuclear gem, nuclear pore, nuclear speckle, nuclear stress body, nucleolus, Oct1/PTF/transcription (OPT) domain, paraspeckle, perinucleolar compartment, PML nuclear body, PML oncogenic domain, polycomb body, processing body, Sam68 nuclear body, and splicing speckle.

[0021] In some examples according to any one of the methods described herein, the disease is a monogenic disease.

[0022] In some examples according to any one of the methods described herein, the disease is a polygenic disease.

[0023] In some examples according to any one of the methods described herein, the disease is a multifactorial disease.

[0024] In some examples according to any one of the methods described herein, the disease is caused, at least in part, by a stimulus and/or an exogenous agent.

[0025] In some examples according to any one of the methods described herein, the disease is caused by an infectious agent.

[0026] In some examples according to any one of the methods described herein, the first marker and the second marker are the same. In some examples, the first marker and the second marker are different.

[0027] In one aspect, provided herein is a method of identifying a marker useful for identifying a condensate of interest associated with a disease, the method comprising: (a) assessing a plurality of candidate markers, or precursors thereof, for a level of association with one or more disease-associated factors of the disease; (b) assessing at least one of the plurality of candidate markers for a condensate affinity factor; and (c) identifying the marker from the plurality of candidate markers based on the marker having a desired level of association with the one or more disease-associated factors of the disease and having a desired condensate affinity factor.

[0028] In some examples according to any one of the methods described herein, the method further comprises identifying the one or more disease-associated factors of the disease. In some examples, each of the one or more disease-associated factors of the disease is selected from the group consisting of a genetic variant, post translational modification variant, exogenous genetic

material, presence of an endogenous compound, presence of an exogenous compound, a physical process, and environmental stimulus.

[0029] In some examples according to any one of the methods described herein, the method further comprises identifying a gene or a non-coding variant associated with the one or more disease-associated factors of the disease.

[0030] In some examples according to any one of the methods described herein, the level of association of each candidate marker with the one or more disease-associated factors of the disease is based on a disease-causal factor score. In some examples, the disease-causal factor score reflects the strength of association of each candidate marker with the one or more disease-associated factors of the disease. In some examples, the method further comprises assigning each candidate marker with the disease-causal factor score.

[0031] In some examples according to any one of the methods described herein, the condensate affinity factor is based on a condensate-association score. In some examples, the condensate-association score reflects the strength of association of the candidate marker, or a portion thereof, with any condensate, a specific condensate, and/or a macromolecule associated with a condensate. In some examples, the method further comprises assigning the candidate marker with the condensate-association score. In some examples, the condensate-association score is a composite score of a condensate function score and a condensate affinity score. In some examples, the condensate function score is determined based on one or more factors of whether a genetic variation of the candidate marker or a portion thereof or the gene or the non-coding variant associated with the one or more disease-associated factors of the disease: i) is within an intrinsically disordered region (IDR); ii) is subject to a post-translational modification; iii) affects splicing of the candidate marker or the gene associated with the one or more disease-associated factors of the disease; iv) affects a chromatin state close to the gene or the non-coding variant associated with the one or more disease-associated factors of the disease; and v) affects expression of the gene associated with the one or more disease-associated factors of the disease. In some examples, the one or more factors for determining the condensate function score each has a weight contributing to the condensate function score. In some examples, the condensate affinity score is determined based on, in the candidate marker or a portion thereof or the gene associated with the one or more disease-associated factors of the disease, one or more factors of: i) the presence, absence, amount,

and/or degree of an IDR; ii) the presence, absence, amount, and/or degree of a condensate-favoring motif; and iii) the presence, absence, amount, and/or valency of an interacting domain. In some examples, the one or more factors for determining the condensate affinity score each has a weight contributing to the condensate affinity score.

[0032] In some examples according to any one of the methods described herein, the method further comprises identifying a gene expression product based on the identified gene, wherein the gene expression product, or a portion thereof, is used to populate the plurality of candidate markers, or precursors thereof.

[0033] In some examples according to any one of the methods described herein, identifying the marker from the plurality of candidate markers comprises a cumulative score based on the desired level of association with the one or more disease-associated factors of the disease and the desired condensate affinity factor.

[0034] In some examples according to any one of the methods described herein, the marker is a biological marker.

[0035] In some examples according to any one of the methods described herein, the marker is identified *in silico*.

[0036] In some examples according to any one of the methods described herein, the method further comprises verifying the marker as useful for identifying a condensate of interest associated with the disease.

[0037] In other examples, provided herein is a method of identifying a condensate of interest associated with a disease, the method comprising: (a) comparing a first condensate phenotype from a first cell model with a second condensate phenotype from a second cell model, wherein the first condensate phenotype and the second condensate phenotype are obtained using a marker identified using a method described herein (*e.g.*, any one of the methods described above), wherein a difference between the first cell model and the second cell model is attributable to one or more disease-associated factors of the disease, and wherein at least one of the one or more disease-associated factors is introduced to either the first cell model or the second cell model; and (b) identifying the condensate of interest as associated with the disease based on a difference between the first condensate phenotype and the second condensate phenotype for the condensate of interest. In one aspect, provided herein is a method of identifying a condensate of interest associated with a disease, the method comprising: (a) comparing a first condensate phenotype from a first cell model with a second condensate phenotype from a second cell model, wherein the first condensate phenotype and the second condensate phenotype are obtained using a marker identified using the method as described herein, wherein a difference between the first cell model and the second cell model is attributable to one or more disease-associated factors of the disease, and wherein at least one of the one or more disease-associated factors is introduced to either the first cell model or the second cell model; and (b) identifying the condensate of interest as associated with the disease based on a difference between the first condensate phenotype and the second condensate phenotype for the condensate of interest.

[0038] In other examples, provided herein is a method of identifying a compound that modulates a condensate phenotype, the method comprising: (a) admixing the compound and a composition comprising a cell model; and (b) obtaining a resulting condensate phenotype of the composition, wherein a difference between the resulting condensate phenotype and a reference condensate phenotype identifies the compound as modulating the condensate phenotype. In one aspect, provided herein is a method of identifying a compound that modulates a condensate phenotype, the method comprising: (a) admixing the compound and a composition comprising a cell model; and (b) obtaining a resulting condensate phenotype of the composition using a marker identified using the method as described herein, wherein a difference between the resulting condensate phenotype and a reference condensate phenotype identifies the compound as modulating the condensate phenotype.

[0039] In other examples, provided herein is a method of identifying a compound useful for treating a disease, the method comprising: (a) admixing the compound and a composition comprising a first cell model; (b) obtaining a resulting condensate phenotype of the composition, wherein the compound is identified as useful for treating the disease when the resulting condensate phenotype has a desired modulation of a phenotypic identifier associated with one or more disease-associated factors of the disease. In one aspect, provided herein is a method of identifying a compound useful for treating a disease, the method comprising: (a) admixing the compound and a composition comprising a first cell model; (b) obtaining a resulting condensate phenotype of the composition using a marker identified using the method as described herein, wherein the compound is identified as useful for treating the disease when the resulting condensate phenotype has a desired modulation of a phenotypic identifier associated with one or more disease-associated factors of the disease.

[0040] It will also be understood by those skilled in the art that changes in the form and details of the implementations described herein may be made without departing from the scope of this disclosure. In addition, although various advantages, aspects, and objects have been described with reference to various implementations, the scope of this disclosure should not be limited by reference to such advantages, aspects, and objects.

[0041] All references cited herein, including patent applications and publications, are incorporated herein by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] **FIGS. 1A-1D** show fluorescent images of H9C2 cells transiently transfected with a wild type RBM20 polypeptide (**FIG. 1A**), a R636S mutant RBM20 polypeptide (**FIG. 1B**), a R636C mutant RBM20 polypeptide (**FIG. 1C**), and a R636H mutant RBM20 polypeptide (**FIG. 1D**).

[0043] **FIG. 2** shows a schematic of an exemplary workflow for evaluating genes within lead single nucleotide polymorphisms (SNPs).

[0044] **FIG. 3** shows enrichment of SNPs across the genomic loci of KCNJ11 and ABCC8.

[0045] **FIGS. 4A-4B** show fluorescent images of iCell® Cardiomyocytes transiently transfected with GFP fused with a wild type Desmoplakin (DSP) polypeptide (**FIG. 4A**), a S299R mutant DSP polypeptide (**FIG. 4B**), and a Q331ter termination mutant DSP polypeptide (**FIG. 4C**). Cell DNA was also stained for DAPI.