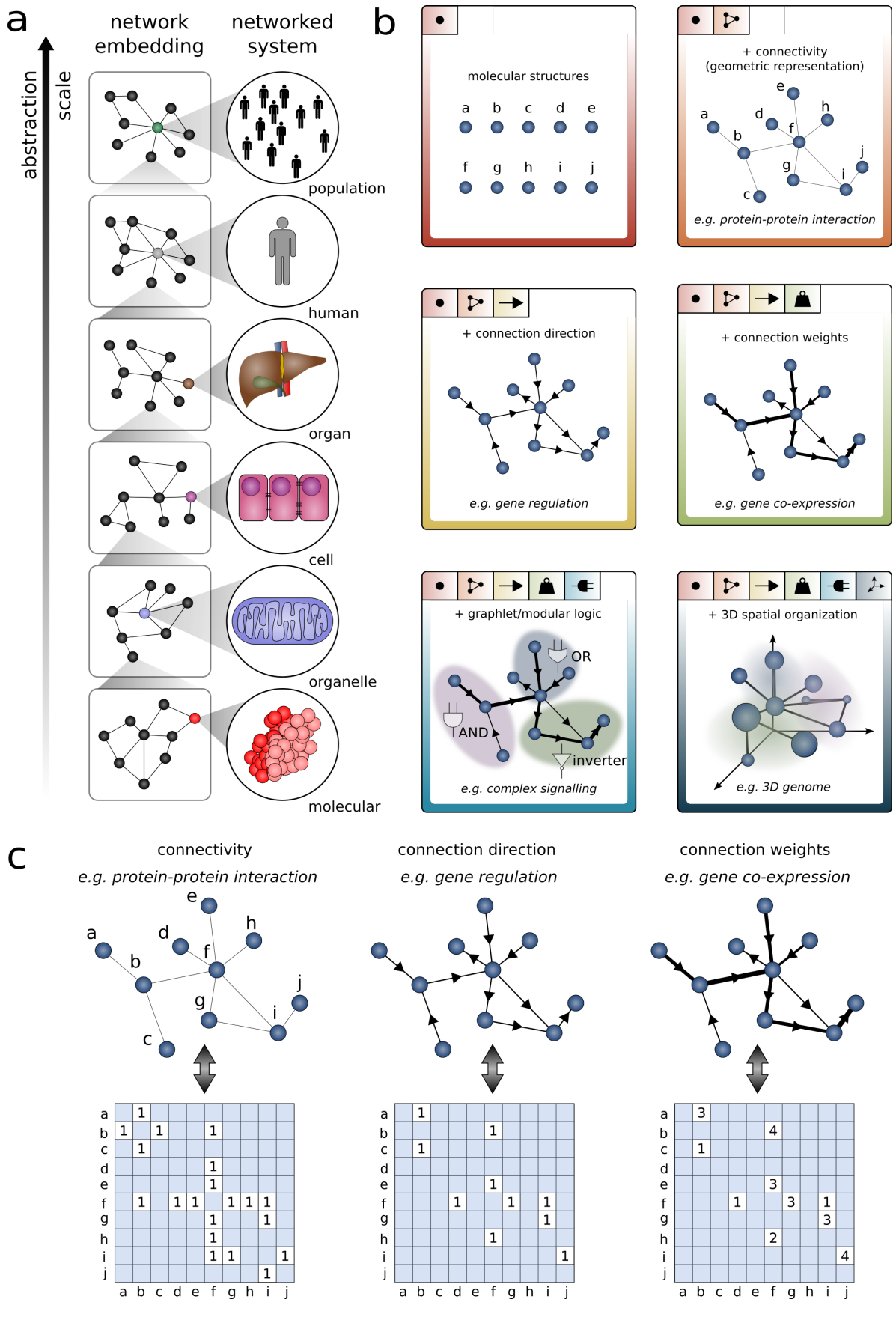
a
network embedding
networked system
abstraction
scale
population
human
organ
cell
organelle
molecular
b
molecular structures
a b c d e
f g h i j
+ connectivity
(geometric representation)
e.g. protein-protein interaction
+ connection direction
e.g. gene regulation
+ connection weights
e.g. gene co-expression
+ graphlet/modular logic
OR
AND
inverter
e.g. complex signalling
+ 3D spatial organization
e.g. 3D genome
c
connectivity
e.g. protein-protein interaction
connection direction
e.g. gene regulation
connection weights
e.g. gene co-expression

Figure 1: Network representations. ***1.a*** Molecular networks form a functional base layer for a number of higher-order biological networks, including networks of organelles (e.g. vesicular transport), cellular networks (e.g. neural), and population scale networks (e.g. disease transmission). ***1.b*** Abstract network representations can be built through a progressive layering of information and logic, according to the network under study. For instance, the addition of directional information to a network may particularly important when representing a gene regulatory network. ***1.c*** Matrices are useful for representing certain network variables, like the pattern of connection and connection weights.

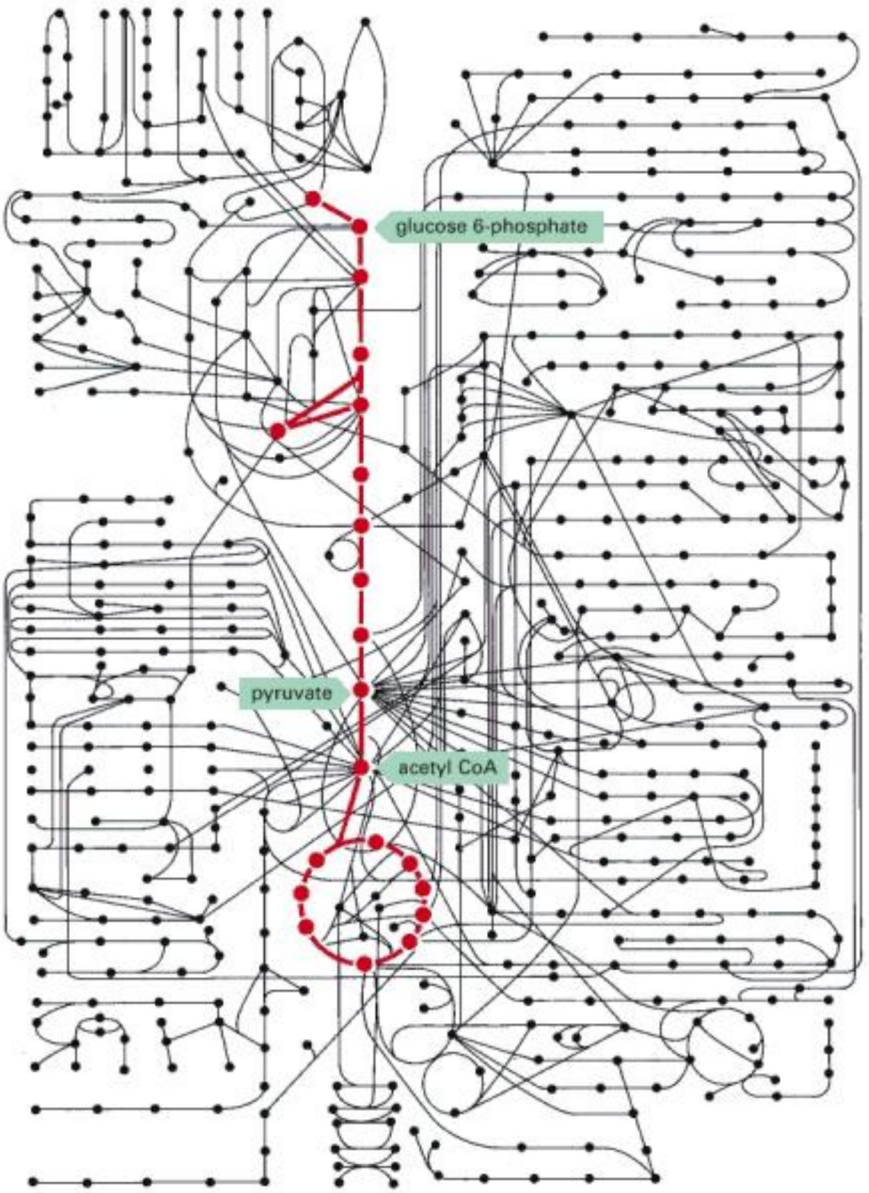

Figure 2: Glycolysis and the citric acid cycle. Despite the complexity of the complete human metabolic network, the core subgraphs of glycolysis and the citric acid cycle can be appreciated in their global context through selective focus. The network structure of glycolysis is linear while

that of the citric acid cycle is cyclical. The two subgraphs are deeply enmeshed within the other processes of metabolism.*

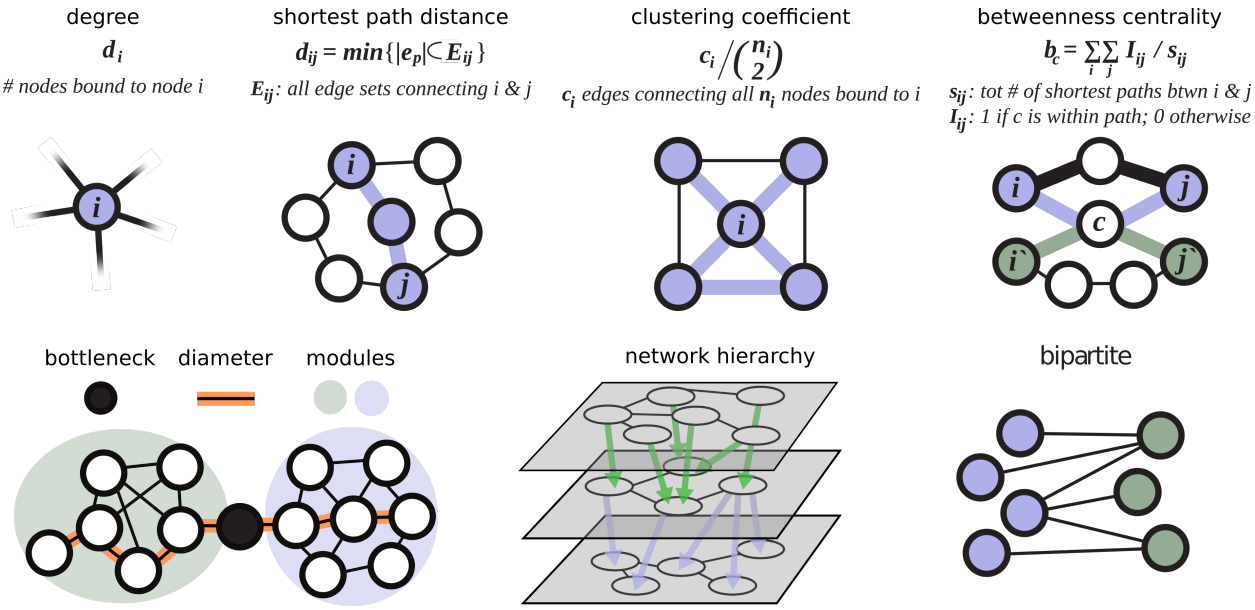

Figure 3: Pictorial glossary. Shown are a series of schematic representations of common network concepts and measures. Many of these metrics (such as degree, clustering coefficient, and betweenness centrality) are used as measures of node importance or influence. Node and edge metrics may be used by algorithms to elucidate higher-order topological features of networks (such as modules and diameter). Hierarchical structures have been used to organize many types of systems, including regulatory networks.

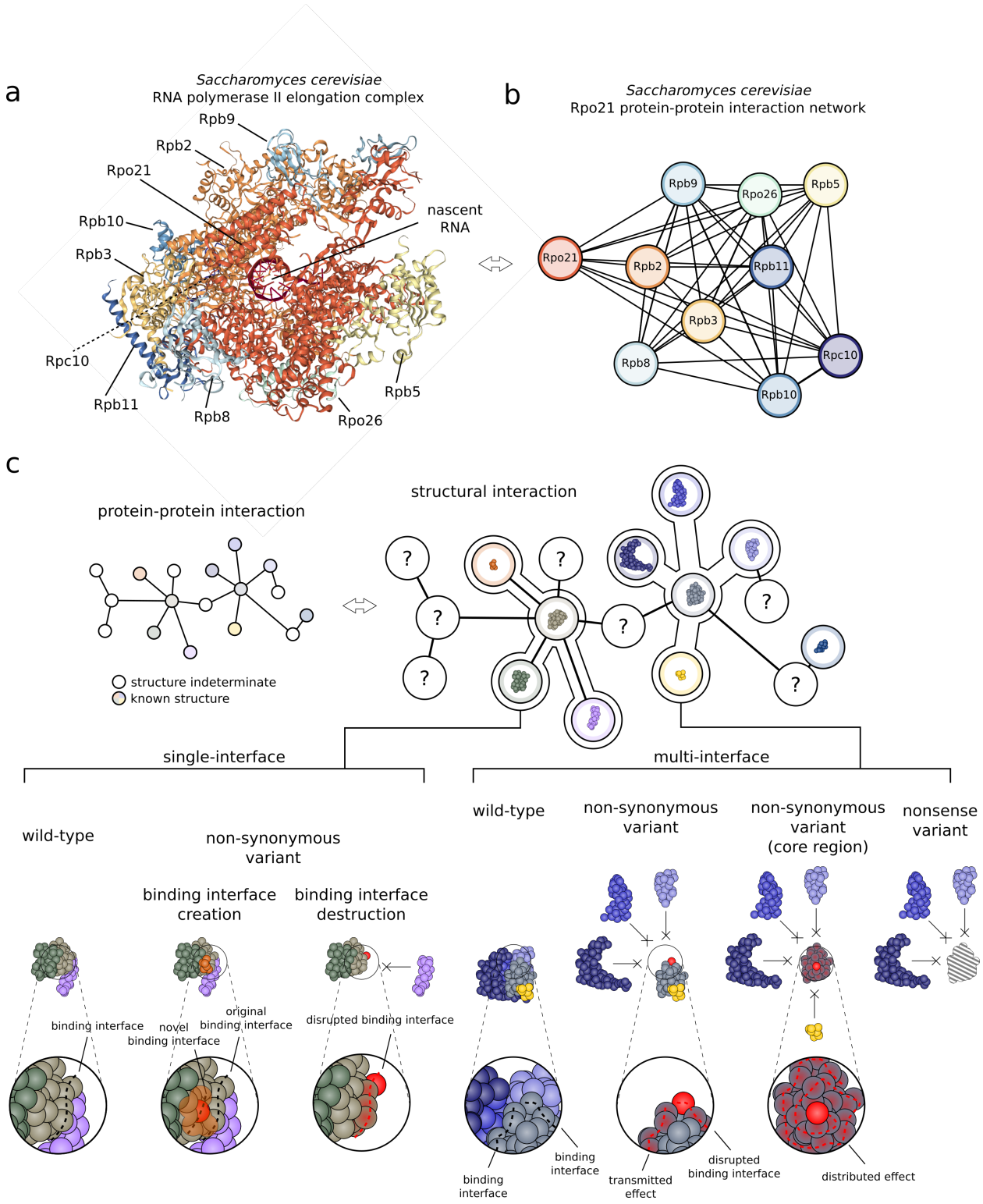

Figure 4: Structural interaction networks. The molecular interaction network of the RNA polymerase II elongation complex in *Saccharomyces cerevisiae* can be represented structurally (**4.a****) or may be represented as an abstract molecular interaction network (***4.b******)

The knowledge of molecular structure that is lost in translation to an abstract network representation, may be important for interpreting certain observed molecular network phenomena. ***4.c*** Three-dimensional protein structure data can be mapped onto protein interaction networks to construct structural interaction networks (SINs). SINs provide physical intuition and nuance for the interactions in a protein interaction network. For instance, a SIN can help distinguish interaction involving single or multiple interfaces. This can be helpful in identifying permanent and transient interactions in the network. High-resolution definitions of various interactions are helpful when prioritizing disease-associated variants, to gain mechanistic insights. For example, disease-associated non-synonymous variants can either create or destroy a binding interface of an individual protein. This, in turn, will influence its interaction with other proteins in the network, which can drive disease progression. Furthermore, variants influencing core and surface of proteins will affect interactions in different ways. For example, for a given protein, mutations on its surface will mostly effect interactions involving a particular interface, whereas those in the core will disrupt all interactions equally.

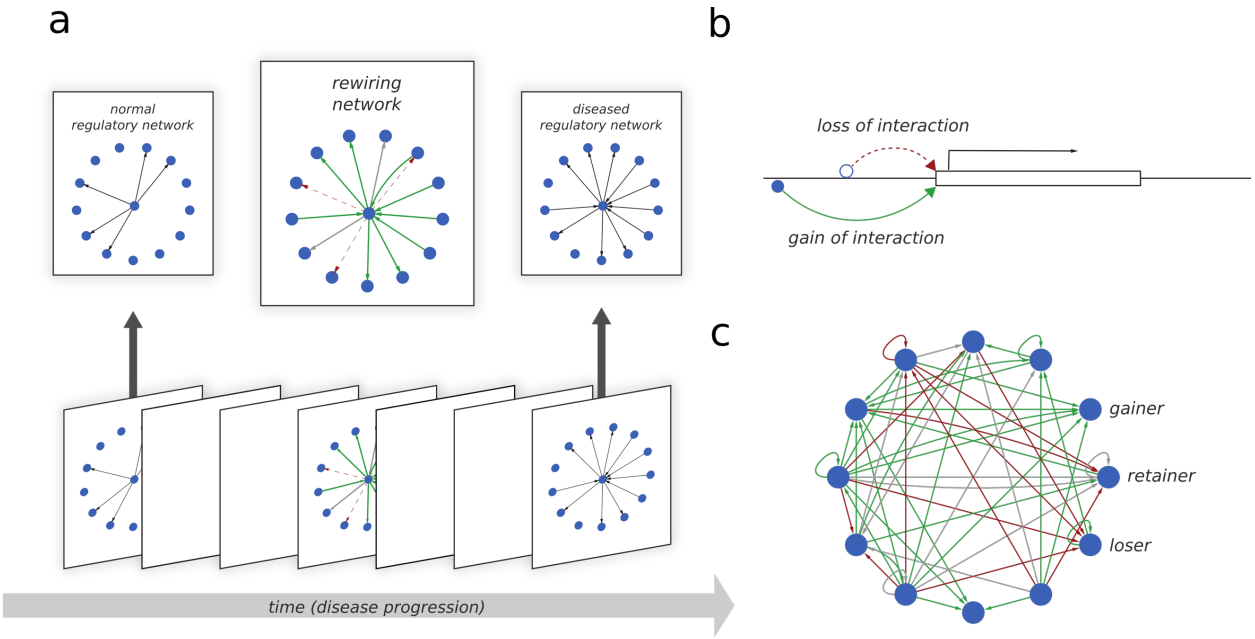

Figure 5. Network rewiring. ***5.a*** A schematic diagram illustrates the progression of a regulatory network from normal to a diseased state. The state of the regulatory network at a specific point in time is depicted as a snapshot. ***5.b*** Binding profiles of regulatory proteins can be used to infer both gain and loss of interaction in different cell states. ***5.c*** By reconstituting the time progression of the regulatory network, the resulting network rewiring can summarize the dynamic changes in regulatory elements.

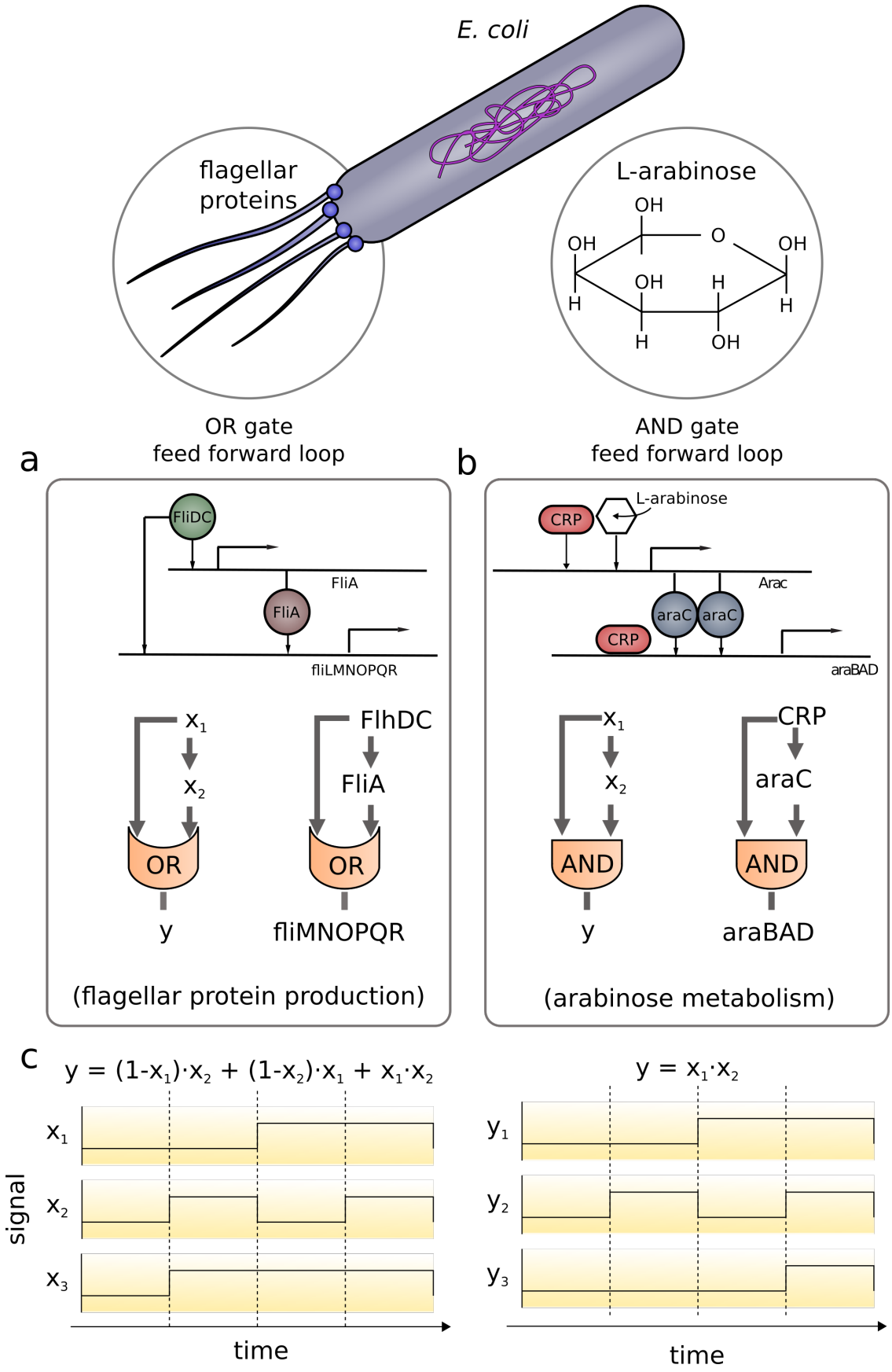
E. coli
flagellar proteins
L-arabinose
OH
O
OH
OH
H
OH
H
H
H
OH
OR gate
feed forward loop
AND gate
feed forward loop
a
b
FliDC
FliA
FliA
fliLMNOPQR
L-arabinose
CRP
Arac
araC
araC
CRP
araBAD
x1
x2
OR
y
FlhDC
FliA
OR
fliMNOPQR
(flagellar protein production)
x1
x2
AND
y
CRP
araC
AND
araBAD
(arabinose metabolism)
c
y = (1-x1)·x2 + (1-x2)·x1 + x1·x2
y = x1·x2
x1
x2
x3
signal
time
y1
y2
y3
time

Figure 6: Feedforward loops (FFLs) - a frequently observed motif in molecular networks. *6.a* An example of a coherent FFL active in the regulation of flagellar protein production in *E. coli*. The *FlhDC* promoter directs the production of *FliA*, which activates Class 2 operon genes *fliMNOPQR*. *FlhDC* also acts additively to activate *fliMNOPQR*. b) Also in *E. coli*, the presence of arabinose induces the formation of the AraC-arabinose complex, which is essential to transcribe the *ara* operon. CRP and cyclic AMP are required in this process. c) A typical AND gate FFL requires the presence of both input elements to produce the output, whereas the presence of either of the two inputs is sufficient to generate the output in a typical OR gate FFL.

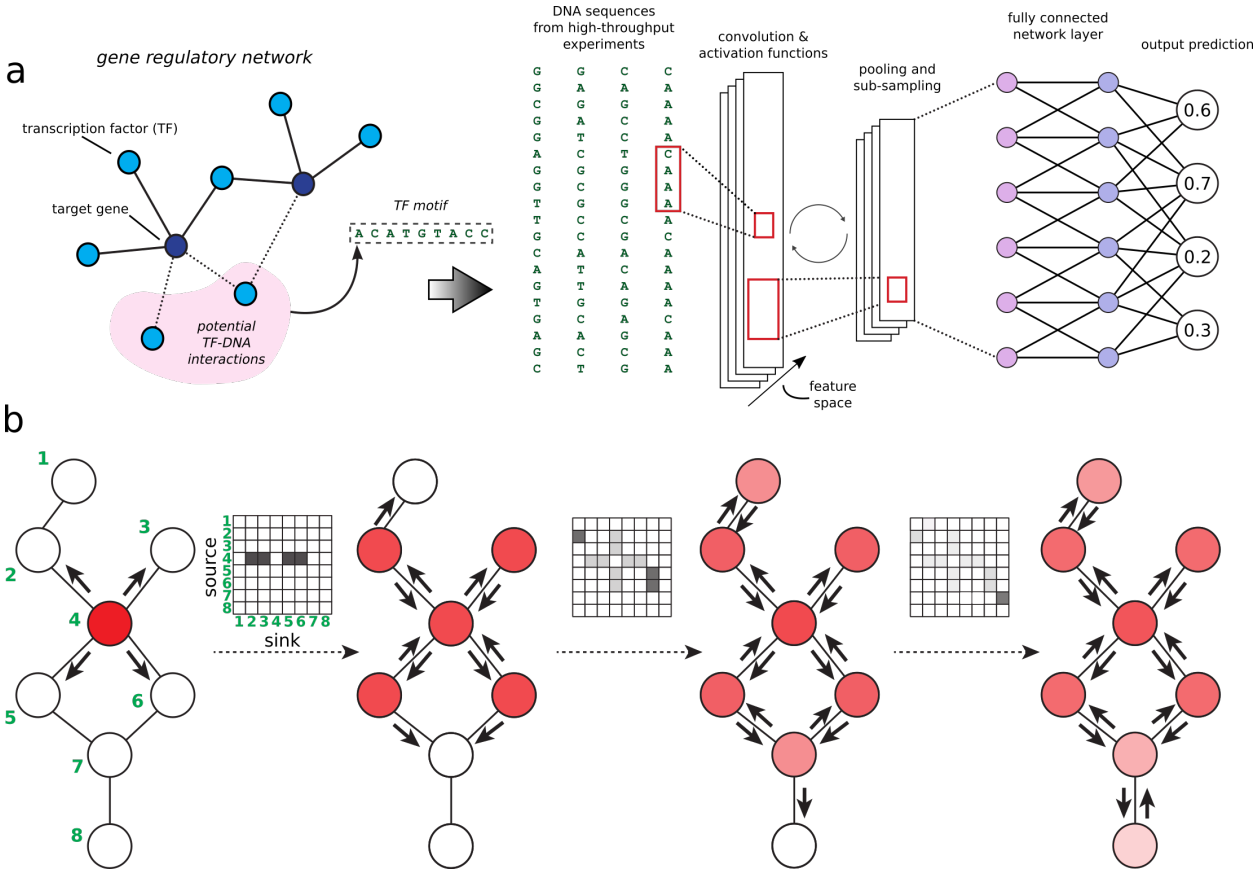

Figure 7: Network algorithms. ***7.a*** The general structure of a convolutional neural network with sample input and output (similar to DeepBind). Here we are trying to detect TF binding sites. If we have high throughput sequencing data containing sequences of potential TF binding sites, we can output the probability a particular sequence is a TF binding site. Training data consists of sequences with experimentally determined binding scores. The convolution layer performs feature extraction by convolving the input matrix with a convolution matrix called a kernel or feature detector. The resulting matrix is the feature map, which in this example would be sequence motifs. An activation function operation (e.g. rectified linear unit - ReLU) introduces nonlinearity into the model. Pooling and sub-sampling provide dimensionality reduction of the feature map; the depth of the feature map corresponds to the number of kernels used in the convolution step. The fully connected layer uses the feature maps to make predictions about the input. Here, the output is how likely a given input sequence is a DNA-binding site. ***7.b*** Network propagation. *Left to right*: Shown are a series of steps by which information (sometimes termed “heat” in networks literature) propagates through a network. This information originates

in node 4 (often a gene believed to be disease-associated with high confidence) and subsequently flows to neighboring nodes 2, 3, 5, and 6. In the next step, this signal may partially flow back into node 4, as well as neighboring nodes 1 and 7, before eventually reaching node 8. Matrices represent the heat being contributed from source to sink nodes. When applied to large networks, the resultant distribution of heat throughout the network may enable one to assign well-defined modules.

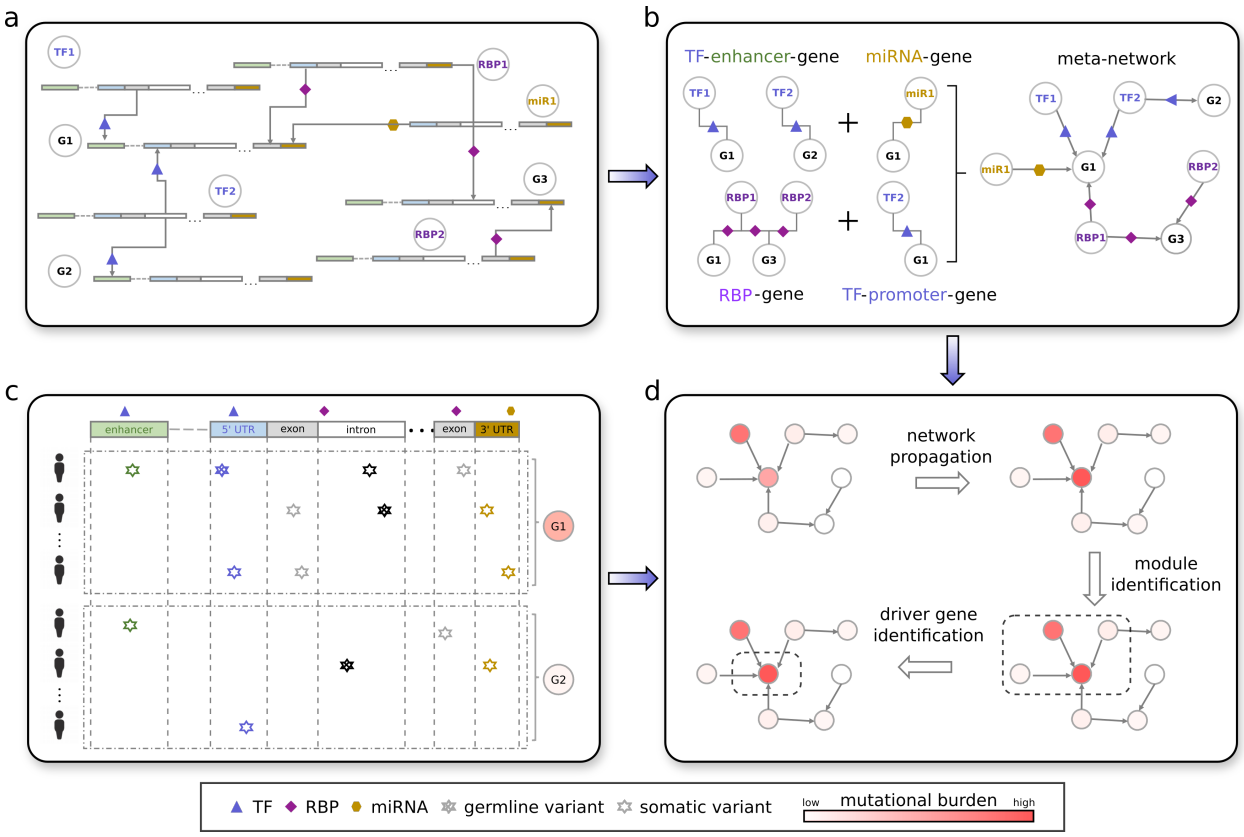

Figure 8: Cancer gene networks: ***8.a*** Gene interactions from multiple regulatory levels may be integrated together to form a meta-network (***8.b***). ***8.c*** By pooling variants from multiple patients, and mapping these mutations to extended gene regulatory regions, an aggregated mutational burden score can be defined; ***8.d*** Through techniques like network propagation, highly mutated subnetworks and key genes can be identified.

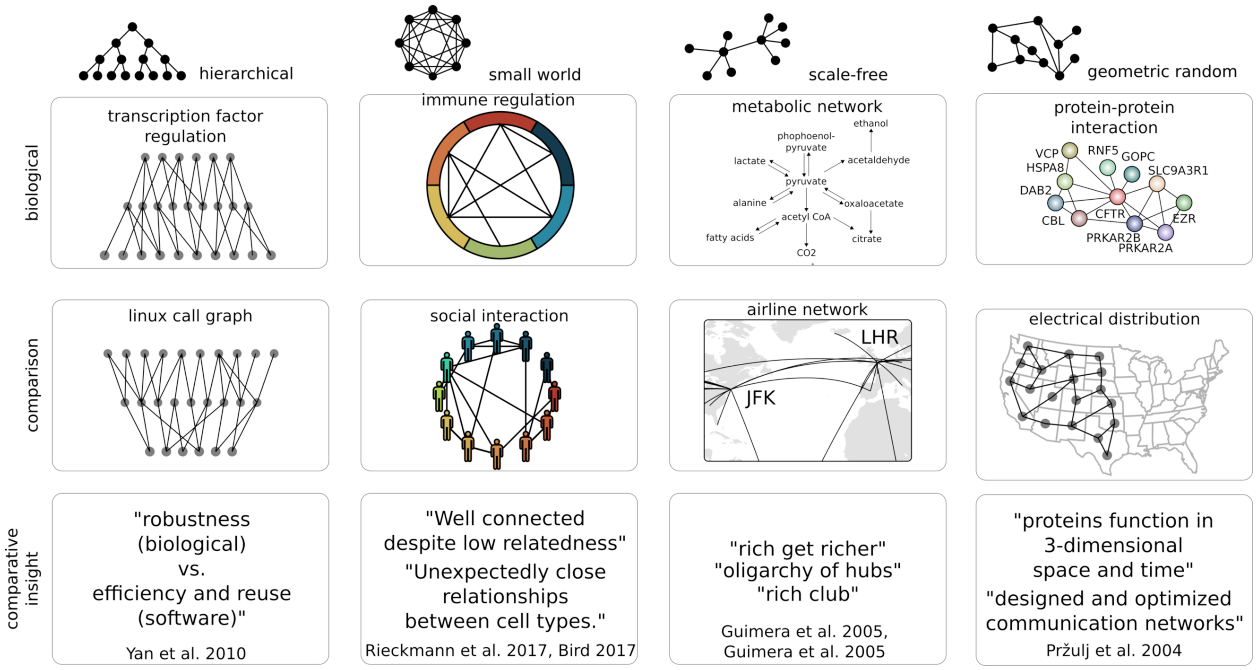

Figure 9: Cross-disciplinary network comparisons: By comparing networks across disciplines, we may learn more about the network about the structure and function of both biological networks and man-made networks. For example, by comparing airline flight routes to the human metabolic network, we learn that both follow a 'scale-free' distribution. A similar 'rich-get-richer' evolutionary process may apply to both networks. Just as flight options are most easily expanded by connecting to an already well-connected airport, pyruvate and acetyl-CoA may function as hub-metabolites, facilitating molecular transitions between biochemical pathways.

** Figure 2 from Alberts B., Johnson A., Lewis J., Raff M., Roberts K., & Walter P. Molecular Biology of The Cell. 5th edition. New York, (NY): Garland Science, Taylor & Francis Group; 2008. (reproduction rights pending publisher approval).*
***Figure 4.a adapted from the RCSB PDB (www.rcsb.org) of PDB ID of 1I6H (1), visualized with NGL Viewer (2).*
****Figure 4.b adapted from the STRING v10 protein-protein interaction database, showing experimentally determined interactions (3).*

1. *Gnatt AL, Cramer P, Fu J, et al. 2001. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. Science (80-.). 292(5523):1876–82*
2. *Rose AS, Hildebrand PW. 2015. NGL Viewer: a web application for molecular visualization. Nucleic Acids Res. 43(W1):W576–79*
3. *Szklarczyk D, Franceschini A, Wyder S, et al. 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43(D1):*