Trans-omics: how to reconstruct biochemical networks across multiple "omic" layers

Katsuyuki Yugi[a,b], Hiroyuki Kubota[c,d], Atsushi Hatano[a], and Shinya Kuroda[a,e,f]

a. Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
b. PRESTO, Japan Science and Technology Agency, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
c. Division of Integrated Omics, Research Center for Transomics Medicine, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan
d. PRESTO, Japan Science and Technology Agency, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan
e. Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
f. CREST, Japan Science and Technology Agency, Bunkyo-ku, Tokyo 113-0033, Japan

Corresponding author: Yugi, K. (yugi@bs.s.u-tokyo.ac.jp) or Kuroda, S. (skuroda@bs.s.u-tokyo.ac.jp)

Key words
Trans-omics, Metabolome, Proteome, Omic data, Trans-ome–wide association study (trans-OWAS)

ABSTRACT

We propose "trans-omic" analysis for reconstructing global biochemical networks across multiple omic layers by use of both multi-omic measurements and computational data integration. We introduce technologies for connecting multi-omic data based on prior knowledge of biochemical interactions and characterize a biochemical trans-omic network by concepts of a static and dynamic nature. We introduce case studies of

metabolism-centric trans-omic studies to show how to reconstruct a biochemical trans-omic network by connecting multi-omic data and how to analyze it in terms of the static and dynamic nature. We propose a trans-ome–wide association study (trans-OWAS) connecting phenotypes with trans-omic networks that reflect both genetic and environmental factors, which can characterize several complex lifestyle diseases as breakdowns in the trans-omic system.

Trans-omic network across multiple omic layers

Specific "omic" layers can be defined and categorized according to the different basic building blocks of the cell, e.g. DNA, RNA, protein, or metabolite [1, 2] (Figure 1). Many cellular functions are orchestrated by global networks that cut across multiple omic layers, and we define the collection of these networks here as the "trans-omic" network (Figure 1). Most biological studies have been conducted by focusing on a few specific molecules, and the trans-omic network has been built by accumulating literature based on such small-scale analyses. This is a powerful strategy, but the comprehensiveness of each layer is limited. Comprehensive measurement technologies for each omic layer are now becoming available, such as polynucleotide sequencing by next-generation sequencers (genome sequencing [3], RNA sequencing [4, 5], chromatin

immunoprecipitation sequencing [ChIP-seq] [6-8], etc.), mass spectrometry–based phosphoproteomics [9-16], expression proteomics [17, 18] and metabolomics (gas chromatography–mass spectrometry [GC-MS] [19], liquid chromatography–mass spectrometry [LC-MS] [20, 21], capillary electrophoresis–mass spectrometry [CE-MS] [22-24], supercritical fluid chromatography–mass spectrometry [SFC-MS] [25], and nuclear magnetic resonance [NMR] [26, 27]). However, a single omic layer analysis alone does not directly elucidate interaction across multiple omic layers. To overcome the lack of comprehensiveness and the information gap regarding interaction across multiple omic layers, an approach for reconstructing molecular networks by connecting multiple omic data has been proposed [28-42] (Figure 1). Here, we call such an approach "trans-omics." Trans-omics connects multiple omic data. There are two major approaches in reconstructing a trans-omic network: one using prior knowledge of a molecular network and another based only on the data-driven approach without use of prior knowledge [43-46]. The former approach is reconstruction of biochemical networks by connecting multiple omic layers with the support of prior knowledge of molecular networks such as publicly available databases. A reconstructed biochemical trans-omic network inherently provides causality and an input–output relationship at a molecular level, allowing interpretation of the biochemical networks. The biochemical

interactions in a trans-omic network enable us to develop a kinetic model directly from a reconstructed biochemical trans-omic network and to analyze the static and dynamic nature of a trans-omic network defined as static and dynamic signal flow. The latter approach is a data-driven approach that statistically infers associations and correlations between molecules based on multi-omic data. This approach does not require prior knowledge of biochemical interactions and can be applied to a wide range of biological processes. However, a statistically reconstructed trans-omic network does not directly reflect biochemical networks. Therefore, such a network does not provide causality and a biochemical input–output relationship at the molecular level, and it cannot be directly used for analysis of static and dynamic signal flow in a trans-omic network.

In this review, we present an overview of the recent emergence of trans-omic studies using the former approach: reconstruction of a biochemical trans-omic network by using prior knowledge of biochemical interactions. We first summarize five technologies for connecting multi-omic data based on prior knowledge and propose three concepts of the static and dynamic nature of biochemical trans-omic networks. Then, we introduce case studies of biochemical trans-omic networks around metabolic enzymes and metabolites based on prior knowledge of metabolic pathways [31, 34, 37], because prior knowledge in this field is some of the most reliable currently available. Furthermore, we propose a

trans-ome–wide association study (trans-OWAS) that covers both genetic and environmental factor. Because many lifestyle diseases, such as type 2 diabetes mellitus (T2DM), can be regarded as complex multifactorial diseases caused by breakdowns in a trans-omic network, a trans-OWAS can potentially be one approach used in future personalized and systems medicine efforts.

Five technologies for connecting multiple omic data

We first summarize the technologies that connect multiple omic data at a molecular level in a biochemical trans-omic network. Currently available methods that connect omic layers are classified into five categories: (i) metabolic regulation; (ii) transcriptional regulation; (iii) kinase–substrate relationship (KSR); (iv) protein–protein interaction (PPI); and (v) allosteric regulation of enzymes by small compounds (Figure 2A).

Metabolic Regulation

The metabolic regulation class of methods has been used in trans-omic studies that connect the metabolome and other omic layers related to the flow of genetic or environmental information. There are many studies regarding connecting the metabolome layer and other omic layers, such as the transcriptome, proteome,

phosphoproteome, and fluxome. Pioneering works were performed to reveal interactions between the transcriptome and metabolome in *Nicotiana tabacum* [28] and *Arabidopsis thaliana* [29]. In microbiological studies, trans-omic analyses including transcriptome, proteome, metabolome, and metabolic flux in *Escherichia coli* [31] and *Bacillus subtilis* [34] were demonstrated. The signal flow of insulin-signaling–dependent control of metabolites in rat hepatoma FAO cells was reconstructed by connecting phosphoproteome and metabolome [37]. The regulation of transcription in response to perturbations in the nitrogen source was inferred by connecting transcriptome, proteome, and metabolome of *Saccharomyces cerevisiae* [42]. With respect to the connection of protein phosphorylation and metabolism, a link between phosphorylation of metabolic enzymes and metabolic fluxes of *S. cerevisiae* was demonstrated by connecting phosphoproteome, metabolome, and fluxome [35, 41]. Associations of phosphorylated metabolic enzymes and changes in their neighboring metabolites were exhibited by integrating phosphoproteome and metabolome [47]. These authors connected metabolome and other omic layers by projecting them together on metabolic pathway maps. Practical details of omic connection studies with the support of the metabolic pathway map are introduced in the following section ("Three case studies on metabolism-centric trans-omics"). One of the technical bottlenecks of

connecting a metabolome with other omic layers through metabolic enzymes and allosteric regulation is correlating the identities of the same objects in different layers, known as ID conversion. We extensively used the KEGG PATHWAY database for comprehensive ID conversion to connect metabolites (metabolome) and metabolic enzymes (phosphoproteome) in a whole metabolism scale (Figure 2B) [37]. The metabolome and the phosphoproteome data are annotated with the KEGG Compound ID and International Protein Index (IPI) ID [48], respectively. The KEGG entries for metabolites, enzymes, and genes are annotated with KEGG Compound ID, EC number [49], and NCBI geneID, respectively (Figure 2B). KEGG provides cross-reference tables that associate metabolic enzymes and metabolites, in which each EC number of a metabolic enzyme is associated with the KEGG Compound ID of substrate and product metabolites. Likewise, metabolic enzymes and their genes are associated in another cross-reference table provided by KEGG in which the EC number is associated with the NCBI geneID. Therefore, the metabolites were easily associated with the metabolic enzymes using the cross-reference table. Then, we converted the IPI ID that is assigned to phosphorylated metabolic enzymes to the EC number so that we could project the phosphoproteome data on the metabolic pathway map. The IPI ID was initially converted to the NCBI geneID, and then to the EC number. Cross-reference tables

between the IPI ID and the NCBI geneID and between the NCBI geneID and the EC number are provided by EMBL-EBI and KEGG, respectively. Generally, ID conversion within the same omic layer, particularly the transcriptome and the proteome, is easily realized by use of cross-reference tables provided by databases or web services such as BioMart [50], DAVID [51, 52], and bioDBnet [53].

Transcriptional Regulation

The transcriptional regulation class of methods includes those that connect the phosphoproteome or proteome of transcription factors (TFs) and the transcriptome of their target genes. Phosphorylated TFs and their target genes in lipopolysaccharide-stimulated macrophages were connected based on phosphoproteomic data of TFs and microarray data of their target genes [54]. In another work, the binding sites of 119 TFs were determined, and the human transcriptional regulatory network was reconstructed based on ChIP-seq measurements in the ENCODE project [55]. A transcriptional regulatory network within mouse dendritic cells that consists of 1728 activations and 594 repressions by 125 TFs was identified on the basis of transcriptomic data obtained after comprehensive inhibition of the 125 TFs by use of a short hairpin RNA library [56]. The transcriptional regulatory network of human myeloid leukemia cells was reconstructed based on transcriptomic

measurements in combination with promoter analysis [57]. Moreover, for reconstruction of the transcriptional regulatory network, computational methods such as network component analysis (NCA) [58, 59] and limitless arity multiple-testing procedure (LAMP) [60] have been proposed. This class of trans-omic studies also includes many other attempts to relate *cis* and *trans* factors, mainly by using transcriptomic data [61, 62]. It is also likely in the near future to incorporate the metabolome as another key factor in transcriptional regulation, for example, as a donor of chemical groups used for chromatin modification [63, 64].

Kinase–Substrate Relationship

The kinase–substrate relationship (KSR) class of methods has its basis in establishing connections (e.g between a phosphorylated metabolic enzyme and the kinase responsible for its phosphorylation) that are inferred from phosphoproteomic data alone. Although these methods do not directly connect distinct omic layers, they represent an essential step for connecting the phosphoproteome with other omic layers: phosphorylation changes the state of proteins, some of which are functionally associated with other omic layers. KSR inference software includes packages such as Scansite [65], NetPhosK [66], GPS [67], NetPhorest [68, 69], PHOSIDA [70], iGPS [71], NetworKIN [69, 72], and RegPhos [73, 74]. Essentially, these softwares infer KSRs based on

experimentally confirmed consensus amino acid sequence motifs recognized by particular kinases that are provided in public databases such as Phospho.ELM [75], PhosphoSitePlus [76], and PhosphoNetworks [77]. In the case of NetPhorest, the software outputs the probability that a kinase phosphorylates a certain amino acid residue of an input amino acid sequence. The probability is estimated by sigmoid functions whose independent variable is a sequence similarity score between the input sequence and a consensus motif of a particular kinase, and whose dependent variable is the probability calculated in reference to experimentally confirmed KSR data. Recent improvements of KSR inference methods (e.g., PHOSIDA [70], iGPS [71], NetworKIN [69, 72], and RegPhos [73, 74]) emphasize incorporating additional information such as protein localization, kinase accessibility to the phosphorylation sites, and protein–protein interaction (PPI) together with a consensus motif analysis. In particular, incorporating PPI information has been shown to decrease sensitivity moderately but to increase specificity greatly in comparison to the decrease in sensitivity [71]. Thus, using KSR estimation methods that include PPI information is recommended if decreasing false positives is more important than decreasing false negatives.

Protein–Protein Interaction

The protein-protein interaction (PPI) class of methods itself also does not directly

connect distinct omic layers. However, it is an essential step for connecting proteome data to other omic layers. For example, if the interacting proteins are a protein kinase, a TF, and a metabolic enzyme, then the PPI class helps to connect signal transduction (phosphoproteome), transcription (transcriptome), and metabolism (metabolome) [63, 78], respectively. Experimental PPI data accumulated in public databases such as STRING [79] are incorporated in NetworKIN to filter out inferred pairs of kinases and substrates that do not interact with each other. Other reviews provide more detailed overviews of PPI detection technologies and software resources [80-85].

Allosteric Regulation

The allosteric regulation class connects the proteome of metabolic enzymes and metabolites that work as activators or inhibitors of the metabolic enzymes. A sample database for this purpose is BRENDA [86, 87], which provides information on enzymatic assays *in vitro*, including activators and inhibitors of particular enzymes. Recently, another database, ASD [88, 89], has also become available. Other than databases, systematic measurement methods to identify allosteric regulation have been developed by various groups [36, 90-92].

Three distinct concepts in the trans-omic network

A network structure of a biochemical trans-omic network directly involves causality and

the input–output relationship at a molecular level. These features enable us to analyze the static and dynamic nature of a biochemical trans-omic network. A trans-omic network inherently includes three specific concepts related to a network: a map; static signal flow; and dynamic signal flow (Figure 3). We explain three concepts in comparison with a road network as follows. A map of a road network contains all possible routes that one can take. Similarly, a map of a trans-omic network describes all possible interactions between intracellular molecules. A map of a trans-omic network can be composed as a patchwork of individual studies on molecular interactions under the different conditions, such as different tissue and cell types. Since all the molecules are not necessarily co-expressed in a certain tissue and cell type, only part of a map of a trans-omic network exists in a certain biological phenomenon of interest. This part of a map is regarded as a route. For example, a route of a road network is a subset of a map, which is a path leading from a departure point to a destination. Similarly, static signal flow of a trans-omic network corresponds to a route in the map of a road network: it indicates the interactions of only co-expressed molecules in a certain biological phenomenon of interest. Static signal flow can be reconstructed by connecting simultaneously measured multi-omic data. Thus, static signal flow is a qualitative expression and does not involve an amount of flow. A subset of a map that includes an

amount of flow can be defined as dynamic signal flow, which is a static signal flow with quantitative amounts of molecules. Dynamic signal flow corresponds to the traffic in a road network. The traffic of a road network is the quantitative expression of a route, in other words, a subset of a map with an amount of flow. Dynamic signal flow should also be reconstructed by the simultaneously measured multi-omic data under the same conditions. Thus, static signal flow indicates a qualitative molecular interaction, and dynamic signal flow indicates a quantitative molecular interaction. Measurements of time series data using multiple doses of stimulation are useful for precise determination of the dynamic signal flow. The term "network" is likely to be used for a map, static signal flow and dynamic signal flow in different contexts. For example, protein–protein interaction networks obtained by yeast two-hybrid systems [93-96] correspond to maps. Signaling and gene networks underlying specific biological phenomena illustrated with directional arrows correspond to static signal flow. Metabolic flux with quantitatively weighted pathways and molecular activities described by kinetic modeling correspond to dynamic signal flow. Metabolic flux can be regarded as dynamic signal flows because, even at steady state, metabolic flux involves a quantitative amount of flux, although the amount of metabolites remained constant. Pioneering trans-omic works have presented reconstruction of static signal flow by projecting transcriptome, proteome, and

metabolome data on pathway maps of the central carbon metabolism, and they also exhibited dynamic signal flow by measuring or predicting metabolic fluxes, respectively [31] and [34]. Moreover, static and dynamic signal flow related to transcriptional regulation were exhibited by a transcriptional regulatory network and temporal profile of promoter activities that are inferred based on ChIP-chip measurements and NCA [34]. In a third study, static signal flow of insulin action was reconstructed by coordinating metabolome and phosphoproteome data with the support of public databases and web services and dynamic signal flow is also explored using a kinetic model of a local network around liver-type phosphofructokinase 1 (PFKL) [37]. Thus, the concepts of a map and static and dynamic signal flow provide a systematic view of characteristics underlying a trans-omic network.

Three case studies on biochemical trans-omic networks: metabolism-centric trans-omics

Because metabolic pathway maps have been supported by accumulated biochemical studies to date, the omic integration on metabolic pathway maps provides trans-omic networks with more credibility than other molecular networks such as signaling and gene expression alone. Therefore, we introduce three previous studies of

metabolism-centric trans-omic networks as case studies [31, 34, 37] in terms of the five technologies for connecting multi-omic data and the three concepts for the static and dynamic nature of a trans-omic network. In addition, it should be noted that multi-omic measurements of biological samples in these studies were obtained under identical conditions. This is important for reconstructing static and dynamic signal flow. Multi-omic measurements under non-identical conditions might lead to false positives of inferred interactions.

Case study 1: global responses of *E. coli* against genetic and environmental perturbations

In the first studyeffects of genetic and environmental perturbations on multiple omic layers in E. coli were assessed by using 24 single gene disruptants and a wild strain grown at five different growth rates [31]. In this study, the metabolome, expression proteome, transcriptome, and metabolic flux data based on “metabolic regulation” were connected (Figure 2A). The data of metabolome, expression proteome, and transcriptome were projected on the “map” of the central carbon metabolism in *E. coli* that provided “static signal flow” from genetic/environmental perturbations to each omic layer associated with the central carbon metabolism (Figure 4A). They also

exhibited “dynamic signal flow” by projecting the metabolic flux data on the pentose phosphate pathway that constitutes a part of the central carbon metabolism. By these trans-omic reconstruction processes, they found that the *E. coli* cells maintain metabolite levels by two distinct modes of global regulation, flux rerouting and gene expression, in response to single gene disruptions and changes in growth conditions, respectively. Connecting the multiple omic data on the metabolic pathway map enabled identification of the static and dynamic signal flow and revealed these modes of global regulation. Thus, *E. coli* chooses two distinct strategies, flux rerouting and gene expression, to realize robust metabolite level control against genetic and environmental perturbations, respectively.

Case study 2: global dynamic adaptations of *B. subtilis* in response to carbon diauxic shift

In another study, the global response of Bacillus subtilis against the shift of the major carbon source from glucose to malate, and from malate to glucose was characterized [34]. The global responses of *B. subtilis* were assessed from five viewpoints: transcriptome; expression proteome; metabolome; ChIP-chip analysis; and metabolic flux (Figure 4B). The multiple omic data were connected altogether by projection on

maps of central carbon metabolism, thereby identifying static signal flow of the carbon diauxic shifts based on the methods presented in metabolic regulation and transcriptional regulation (Figure 2A). Moreover, they projected computationally estimated metabolic flux and promoter activity on the pathway map to identify dynamic signal flow. The dynamic signal flow described in this study covers the whole of central carbon metabolism [34]. They used metabolic regulation and transcriptional regulation to connect the multiple omic layers and examined time scales of cellular responses based on time-series measurements. They revealed that *B. subtilis* responds to the carbon diauxic shift through two distinct modes of adaptation: faster adaptation by posttranscriptional regulation and slower adaptation by changes in gene expression. When the major carbon source is shifted from glucose to malate, the metabolic fluxes of *B. subtilis* are altered mainly by faster regulation (posttranscriptional regulation), whereas they are changed mainly by slower regulation (gene expression) when the carbon source is shifted from malate to glucose. By connecting multiple omic layers, these two distinct modes of global regulation were found, as was interplay between omic layers in those modes of global regulation. Furthermore, identification of the dynamic signal flow facilitates characterization of time scales of the two distinct modes of global regulation.

Case study 3: reconstruction of the trans-omic network of insulin action in rat hepatoma FAO cells

Regulatory networks surrounding metabolic networks were reconstructed [31] and [34]; however, the network directly from extracellular environments to metabolism has not been reconstructed. We reconstructed a trans-omic network directly from extracellular stimulation (insulin) to metabolism in rat hepatoma FAO cells by connecting metabolome and phosphoproteome layers (Figure 4C) [37]. The phosphoproteome layer was separated into two groups: protein kinases that constitute the insulin-signaling pathway and metabolic enzymes that are substrates of the protein kinases. We used a map of all metabolism including the central carbon metabolism, and the insulin-signaling pathway of the KEGG PATHWAY database to project multiple omic data. We identified static signal flow of insulin according to metabolic regulation, KSR, and allosteric regulation (Figure 2A). According to metabolic regulation, insulin-responsive metabolites were associated with phosphorylated metabolic enzymes whose responsible protein kinases were inferred by use of NetPhorest, a KSR software, and assigned to the insulin-signaling pathway. Overall, the combination of metabolic regulation and KSR allowed us to retrace the signal flow from quantitatively changed

metabolites to the insulin receptor. Subsequently, allosteric regulation of the quantitatively changed metabolites on the metabolic enzymes was incorporated in reference to BRENDA, a database of allosteric regulation. We identified dynamic signal flow around PFKL by using kinetic models. Using the model analysis, functionally non-essential allosteric regulations were trimmed from the original trans-omic network. Our reconstruction study provides a biochemical trans-omic network that includes all reaction steps from input (insulin stimulus) to outputs (the metabolites). In this trans-omic network, we found that 48 phosphorylations of metabolic enzymes out of 71 are novel regulatory pathways. Connecting multiple omic layers allowed identification of insulin signal–dependent regulatory pathways of global metabolism.

Systems medicine and trans-OWAS

It may be possible for trans-omic analysis to be applied to medicine. Advances in measurement technologies and mathematical/computational methods have been promoting systems medicine, which tackles complex diseases [97, 98]. Systems medicine aims to correct the behavior of a group of molecules by using pathway information [99, 100] and is expected to change current reactive medicine, which is enacted after people contract disease, to predictive and personalized medicine based on

genomic data [101-103].

Single Omics-Wide Association Study

Genome-wide linkage analysis between genetic traits and phenotype, also called a genome-wide association study (GWAS), is a promising approach for revealing linkages between an individual's genetic background and potential susceptibility to particular diseases [104]. This approach associates genetic variations with infectious diseases [105] and Mendelian disorders, such as Huntington disease and cystic fibrosis [106]. In addition to GWAS, a single omic layer other than genome (e.g., epigenome [40, 107], transcriptome [108], proteome [108], metabolome [109, 110], and others [15, 111]) and environmental factors (e.g., diet [39, 112] and exposure to chemicals [113]) have also been used for association studies with phenotypes. A phenome-wide association study (PheWAS) assesses whether a genomic region affects multiple phenotypes based on human clinical data and SNP data [114]. Quantitative trait locus (QTL) analysis, an alternative method for disease-related gene discovery, enables us to identify the genomic regions that affect quantitative phenotypes, such as the amount of transcripts, proteins, and metabolites [115-120]. However, QTL has several limitations, such as low mapping resolution and genotypic variation [121]. To resolve these limitations, molecule-based GWAS, in which genomic information is connected with molecules

such as metabolites, has been recently proposed. Metabolite-based GWAS of maize, which can be used against a genetic complex population, identifies associations between genomic region and metabolites at a higher resolution [122, 123]. A pathway-wide association study (PWAS), in which pathway information is used to identify gene sets that are enriched for variants associated with diseases, has also been proposed [124].

Trans-ome–Wide Association Study

Lifestyle diseases, such as hypercholesterolemia and type 2 diabetes mellitus (T2DM), are largely elicited by multiple factors belonging to multiple omic layers that are influenced not only by genetic factors but also by environmental factors linked to lifestyle. GWAS can associate phenotypes only with genetic factors, not with environmental factors. Therefore, only a small proportion of heritability for multifactorial diseases can be explained by GWAS. In T2DM, less than 10% of heritability is explained by genomic variants identified by GWAS, despite the efforts of several GWAS trials [113, 125]. GWAS identifies only phenomenological connections between genotype and phenotype but does not indicate direct biochemical interactions. Therefore, a GWAS approach alone does not provide any substantial information to select an appropriate personalized treatment strategy that may rely on molecular mechanisms [126]. Thus, more globally integrated association studies that reflect both

genomic and environmental information, including RNA, proteins, and metabolites, and that indicate molecular networks are expected for analyzing multifactorial diseases linked with lifestyle and for identifying the molecular pathological mechanisms underlying such diseases.

Here, we propose a trans-OWAS that includes the genome, epigenome, metabolome, proteome, transcriptome, and phenome to identify the global molecular mechanism of multifactorial diseases. In trans-OWAS, the individual network is reconstructed from the multiple omic data, as shown in the case studies. Phenotypes are characterized by using these reconstructed networks. Trans-OWAS has two advantages compared to GWAS: trans-OWAS can associate phenotypes not only with genetic factors but also with environmental factors, and it can elucidate direct molecular networks in trans-omic layers instead of phenomenological relationships (Figure 5A).

Disease states are understood as disorders in a trans-omic network. For example, T2DM, a typical multifactorial disease, can be regarded as a systems breakdown caused by genetic and environmental factors in a trans-omic network. Trans-OWAS can be one of the ideal approaches for T2DM (Figure 5B). Homeostatic feedback between insulin sensitivity and insulin secretion from β cells is a central core for blood glucose regulation, and impairment of the feedback system leads to T2DM. Trans-OWAS can

characterize pathogenesis of T2DM as multiple breakdowns in insulin sensitivity and secretion pathways in a trans-omic network. Consequently, trans-OWAS will reveal the molecular mechanism of pathogenesis of T2DM for each individual patient, because trans-OWAS directly implements both genetic and environmental factor as particular states of a trans-omic network. Thus, trans-OWAS will be an essential tool for personalized diagnosis, prediction of prognosis, and treatment, and may become one of the major approaches in personalized systems medicine.

An integrative network-based association study (INAS), in which single omic data such as transcriptome or interactome are integrated with genomic information to identify the gene regulatory network that elicited the phenotypes, is one example of a trans-OWAS [127, 128]. One of the bottlenecks when performing trans-OWAS is acquisition of a large amount of multi-omic data. Recently, an attempt [38] was presented in which they measure genome, transcriptome, and proteome data from BXD recombinant inbred mice [129] fed a normal diet or a high-fat diet; the data were ideal for trans-OWAS analysis. Furthermore, multi-omic data were also obtained from humans [130]. These studies demonstrate that trans-OWAS will be available in the near future. Trans-OWAS enables us to characterize the pathogenesis of complex multifactorial diseases with both genomic and environmental factors, and to elucidate

their molecular mechanisms in a trans-omic network.

Concluding remarks

We have introduced five technologies, three concepts, and three case studies for biochemical trans-omic networks. However, there still are technological and analytical improvements needed for reconstructing a reliable biochemical trans-omic network. Throughput and comprehensiveness in omic measurements should be improved (Outstanding Questions Box). For data analysis, reliability of pathway information and technologies for connecting different omic layers should be improved and developed (Outstanding Questions Box). A validation method for a reconstructed trans-omic network should be further developed (Outstanding Questions Box). Such improvements will make trans-omic analysis essential and standard in molecular biological studies and medicine in the future.

Acknowledgments

We thank Yibo Wu (Department of Biology, Institute of Molecular Systems Biology, ETH Zurich) for critically reading this manuscript. This work was supported by the Creation of Fundamental Technologies for Understanding and Control of Biosystem Dynamics, CREST,

from the Japan Science and Technology Agency (JST), by a Human Frontier Science Project (HFSP) grant (RGP0061/2011), and by a grant from Japan Diabetes Foundation. K.Y. (Yugi) receives a Grant-in-Aid for Young Scientists (A) (#15H05582) from the Japan Society for the Promotion of Science and Creation of Innovative Technology for Medical Applications Based on the Global Analyses and Regulation of Disease-Related Metabolites, PRESTO, from JST. H.K. (Kubota) receives funding from a Grant-in-Aid for Scientific Research on Innovative Areas (#25117712) from MEXT and Elucidation and regulation in the dynamic maintenance and transfiguration of homeostasis in living body, PRESTO, from JST.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Brazhnik, P., *et al.* (2002) Gene networks: how to put the function in genomics. *Trends Biotechnol* 20, 467-472.
2. Nielsen, J. and Oliver, S. (2005) The next wave in metabolome analysis. *Trends in biotechnology* 23, 544-546.
3. Bentley, D.R., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53-59.
4. Nagalakshmi, U., *et al.* (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344-1349.
5. Cloonan, N., *et al.* (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods* 5, 613-619.
6. Kharchenko, P.V., *et al.* (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26, 1351-1359.
7. Adli, M. and Bernstein, B.E. (2011) Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature protocols* 6, 1656-1668.
8. Yasuoka, Y., *et al.* (2014) Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nature communications* 5, 4322.
9. Ross, P.L., *et al.* (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics : MCP* 3, 1154-1169.
10. Ishihama, Y., *et al.* (2005) Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat Biotechnol* 23, 617-621.
11. Olsen, J.V., *et al.* (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* 127, 635-648.
12. Matsumoto, M., *et al.* (2009) Large-scale proteomic analysis of tyrosine-phosphorylation induced by T-cell receptor or B-cell receptor activation reveals new signaling pathways. *Proteomics* 9, 3549-3563.
13. Humphrey, S.J., *et al.* (2013) Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell metabolism* 17, 1009-1020.
14. Yi, T., *et al.* (2014) Quantitative phosphoproteomic analysis reveals system-wide signaling pathways downstream of SDF-1/CXCR4 in breast cancer stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 111, E2182-2190.
15. Shehzad, Z., *et al.* (2014) A multivariate distance-based analytic framework for connectome-wide association studies. *Neuroimage* 93 Pt 1, 74-94.
16. Emdal, K.B., *et al.* (2015) Temporal proteomics of NGF-TrkA signaling identifies an inhibitory role for the E3 ligase Cbl-b in neuroblastoma cell differentiation. *Science signaling* 8, ra40.
17. Ong, S.E., *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a

simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* 1, 376-386.

18. Stahl-Zeng, J., *et al.* (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Molecular & cellular proteomics : MCP* 6, 1809-1817.

19. Fiehn, O., *et al.* (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18, 1157-1161.

20. Bennett, B.D., *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* 5, 593-599.

21. Tajima, Y., *et al.* (2013) Lipidomic analysis of brain tissues and plasma in a mouse model expressing mutated human amyloid precursor protein/tau for Alzheimer's disease. *Lipids in health and disease* 12, 68.

22. Soga, T., *et al.* (2002) Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Analytical chemistry* 74, 2233-2239.

23. Soga, T. (2006) Differential Metabolomics Reveals Ophthalmic Acid as an Oxidative Stress Biomarker Indicating Hepatic Glutathione Consumption. *The Journal of biological chemistry* 281, 16768-16776.

24. Soga, T., *et al.* (2009) Metabolomic Profiling of Anionic Metabolites by Capillary Electrophoresis Mass Spectrometry. *Anal Chem.* 81, 6165-6174.

25. Bamba, T., *et al.* (2008) High throughput and exhaustive analysis of diverse lipids by using supercritical fluid chromatography-mass spectrometry for metabolomics. *Journal of bioscience and bioengineering* 105, 460-469.

26. Li, M., *et al.* (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 105, 2117-2122.

27. Dona, A.C., *et al.* (2014) Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Analytical chemistry* 86, 9887-9894.

28. Goossens, A., *et al.* (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proceedings of the National Academy of Sciences* 100, 8595-8600.

29. Hirai, M.Y., *et al.* (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* 101, 10205-10210.

30. Joyce, A.R. and Palsson, B.O. (2006) The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7, 198-210.

31. Ishii, N., *et al.* (2007) Multiple High-Throughput Analyses Monitor the Response of E. coli to

Perturbations. *Science* 316, 593-597.
32. Palsson, B. and Zengler, K. (2010) The challenges of integrating multi-omic data sets. *Nat Chem Biol* 6, 787-789.
33. Gerosa, L. and Sauer, U. (2011) Regulation and control of metabolic fluxes in microbes. *Curr Opin Biotechnol* 22, 566-575.
34. Buescher, J.M., *et al.* (2012) Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science* 335, 1099-1103.
35. Oliveira, A.P., *et al.* (2012) Regulation of yeast central metabolism by enzyme phosphorylation. *Molecular systems biology* 8, 623.
36. Link, H., *et al.* (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat Biotechnol*.
37. Yugi, K., *et al.* (2014) Reconstruction of insulin signal flow from phosphoproteome and metabolome data. *Cell reports* 8, 1171-1183.
38. Wu, Y., *et al.* (2014) Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* 158, 1415-1430.
39. Davis, M.A., *et al.* (2014) A dietary-wide association study (DWAS) of environmental metal exposure in US children and adults. *PLoS One* 9, e104768.
40. Rakyan, V.K., *et al.* (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12, 529-541.
41. Oliveira, A.P., *et al.* (2015) Dynamic phosphoproteomics reveals TORC1-dependent regulation of yeast nucleotide and amino acid biosynthesis. *Science signaling* 8, rs4.
42. Oliveira, A.P., *et al.* (2015) Inferring causal metabolic signals that regulate the dynamic TORC1-dependent transcriptome. *Molecular systems biology* 11, 802.
43. Ideker, T. and Lauffenburger, D. (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends in biotechnology* 21, 255-262.
44. Janes, K.A. and Yaffe, M.B. (2006) Data-driven modelling of signal-transduction networks. *Nature reviews. Molecular cell biology* 7, 820-828.
45. Janes, K.A. and Lauffenburger, D.A. (2013) Models of signalling networks - what cell biologists can gain from them and give to them. *Journal of cell science* 126, 1913-1921.
46. Imam, S., *et al.* (2015) Data-driven integration of genome-scale regulatory and metabolic network models. *Frontiers in microbiology* 6, 409.
47. Schulz, J.C., *et al.* (2014) Large-scale functional analysis of the roles of phosphorylation in yeast metabolic pathways. *Science signaling* 7, rs6.
48. Kersey, P.J., *et al.* (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics* 4, 1985-1988.
49. IUBMB (1999) IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and

Nomenclature Committee of IUBMB (NC-IUBMB), newsletter 1999. *European journal of biochemistry / FEBS* 264, 607-609.

50. Kasprzyk, A. (2011) BioMart: driving a paradigm change in biological data management. *Database : the journal of biological databases and curation* 2011, bar049.

51. Huang da, W., *et al.* (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57.

52. Huang da, W., *et al.* (2009) Extracting biological meaning from large gene lists with DAVID. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Chapter 13, Unit 13 11.

53. Mudunuri, U., *et al.* (2009) bioDBnet: the biological database network. *Bioinformatics* 25, 555-556.

54. Weintz, G., *et al.* (2010) The phosphoproteome of toll-like receptor-activated macrophages. *Molecular systems biology* 6, 371.

55. Gerstein, M.B., *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91-100.

56. Amit, I., *et al.* (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326, 257-263.

57. Suzuki, H., *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature genetics* 41, 553-562.

58. Liao, J.C., *et al.* (2003) Network component analysis: reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences of the United States of America* 100, 15522-15527.

59. Kao, K.C., *et al.* (2004) Transcriptome-based determination of multiple transcription regulator activities in Escherichia coli by using network component analysis. *Proceedings of the National Academy of Sciences of the United States of America* 101, 641-646.

60. Terada, A., *et al.* (2013) Statistical significance of combinatorial regulations. *Proceedings of the National Academy of Sciences of the United States of America* 110, 12996-13001.

61. Kim, H.D., *et al.* (2009) Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* 325, 429-432.

62. Iyer, M.K., *et al.* (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics*.

63. Katoh, Y., *et al.* (2011) Methionine adenosyltransferase II serves as a transcriptional corepressor of Maf oncoprotein. *Molecular cell* 41, 554-566.

64. Shyh-Chang, N., *et al.* (2013) Influence of threonine metabolism on S-adenosylmethionine and histone methylation. *Science* 339, 222-226.

65. Obenauer, J.C., *et al.* (2003) Scansite 2.0: Proteome-wide prediction of cell signaling

interactions using short sequence motifs. *Nucleic acids research* 31, 3635-3641.

66. Blom, N., *et al.* (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633-1649.

67. Xue, Y., *et al.* (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular & cellular proteomics : MCP* 7, 1598-1608.

68. Miller, M.L., *et al.* (2008) Linear motif atlas for phosphorylation-dependent signaling. *Science signaling* 1, ra2.

69. Horn, H., *et al.* (2014) KinomeXplorer: an integrated platform for kinome biology studies. *Nature methods* 11, 603-604.

70. Gnad, F., *et al.* (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic acids research* 39, D253-260.

71. Song, C., *et al.* (2012) Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Molecular & cellular proteomics : MCP* 11, 1070-1083.

72. Linding, R., *et al.* (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* 129, 1415-1426.

73. Lee, T.Y., *et al.* (2011) RegPhos: a system to explore the protein kinase-substrate phosphorylation network in humans. *Nucleic acids research* 39, D777-787.

74. Huang, K.Y., *et al.* (2014) RegPhos 2.0: an updated resource to explore protein kinase-substrate phosphorylation networks in mammals. *Database : the journal of biological databases and curation* 2014, bau034.

75. Dinkel, H., *et al.* (2011) Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic acids research* 39, D261-267.

76. Hornbeck, P.V., *et al.* (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic acids research* 40, D261-270.

77. Hu, J., *et al.* (2014) PhosphoNetworks: a database for human phosphorylation networks. *Bioinformatics* 30, 141-142.

78. Bandyopadhyay, S., *et al.* (2010) A human MAP kinase interactome. *Nature methods* 7, 801-805.

79. Franceschini, A., *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 41, D808-815.

80. Cagney, G. (2009) Interaction networks: lessons from large-scale studies in yeast. *Proteomics* 9, 4799-4811.

81. Vidal, M., *et al.* (2011) Interactome networks and human disease. *Cell* 144, 986-998.

82. Rao, V.S., *et al.* (2014) Protein-protein interaction detection: methods and analysis. *International journal of proteomics* 2014, 147648.

83. Westermarck, J., *et al.* (2013) Identification of protein interactions involved in cellular signaling.

Molecular & cellular proteomics : MCP 12, 1752-1763.

84. Shannon, P., *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498-2504.

85. Gehlenborg, N., *et al.* (2010) Visualization of omics data for systems biology. *Nature methods* 7, S56-68.

86. Scheer, M., *et al.* (2011) BRENDA, the enzyme information system in 2011. *Nucleic acids research* 39, D670-676.

87. Chang, A., *et al.* (2014) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic acids research*.

88. Huang, Z., *et al.* (2011) ASD: a comprehensive database of allosteric proteins and modulators. *Nucleic acids research* 39, D663-669.

89. Huang, Z., *et al.* (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. *Nucleic acids research* 42, D510-516.

90. Li, X., *et al.* (2010) Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* 143, 639-650.

91. Gallego, O., *et al.* (2010) A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. *Molecular systems biology* 6, 430.

92. Feng, Y., *et al.* (2014) Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* 32, 1036-1044.

93. Uetz, P., *et al.* (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627.

94. Walhout, A.J., *et al.* (2000) Protein interaction mapping in C. elegans using proteins involved in vulval development. *Science* 287, 116-122.

95. Ito, T., *et al.* (2000) Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proceedings of the National Academy of Sciences of the United States of America* 97, 1143-1147.

96. Ito, T., *et al.* (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4569-4574.

97. Liu, E.T. (2009) Integrative biology--a strategy for systems biomedicine. *Nat Rev Genet* 10, 64-68.

98. Antony, P.M., *et al.* (2012) From systems biology to systems biomedicine. *Current opinion in biotechnology* 23, 604-608.

99. Fitzgerald, J.B., *et al.* (2006) Systems biology and combination therapy in the quest for clinical efficacy. *Nature chemical biology* 2, 458-466.

100. Kitano, H. (2007) A robustness-based approach to systems-oriented drug design. *Nature reviews. Drug discovery* 6, 202-210.
101. Galas, D.J. and Hood, L. (2009) Systems Biology and Emerging Technologies Will Catalyze the Transition from Reactive Medicine to Predictive, Personalized, Preventive and Participatory (P4) Medicine. *Interdisciplinary Bio Cenntral* 1, 6.
102. Auffray, C., *et al.* (2010) Predictive, preventive, personalized and participatory medicine: back to the future. *Genome medicine* 2, 57.
103. Bloss, C.S., *et al.* (2010) Consumer perceptions of direct-to-consumer personalized genomic risk assessments. *Genetics in medicine : official journal of the American College of Medical Genetics* 12, 556-566.
104. Whitcomb, D.C. (2012) What is personalized medicine and what should it replace? *Nature reviews. Gastroenterology & hepatology* 9, 418-424.
105. Chapman, S.J. and Hill, A.V.S. (2012) Human genetic susceptibility to infectious disease. *Nat Rev Genet* 13, 175-188.
106. Stranger, B.E., *et al.* (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187, 367-383.
107. Aberg, K.A., *et al.* (2014) Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* 71, 255-264.
108. Wayengera, M. (2010) On the general theory of the origins of retroviruses. *Theor Biol Med Model* 7, 5.
109. Holmes, E., *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453, 396-400.
110. Cantoria, M.J., *et al.* (2014) Contextual inhibition of fatty acid synthesis by metformin involves glucose-derived acetyl-CoA and cholesterol in pancreatic tumor cells. *Metabolomics* 10, 91-104.
111. Qin, J., *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55-60.
112. Tzoulaki, I., *et al.* (2012) A nutrient-wide association study on blood pressure. *Circulation* 126, 2456-2464.
113. Patel, C.J., *et al.* (2010) An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 5, e10746.
114. Pendergrass, S.A., *et al.* (2013) Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet* 9, e1003087.
115. Brem, R.B., *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752-755.
116. Goring, H.H., *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional

profiling in human lymphocytes. *Nature genetics* 39, 1208-1216.

117. Foss, E.J., *et al.* (2007) Genetic basis of proteome variation in yeast. *Nature genetics* 39, 1369-1375.

118. Picotti, P., *et al.* (2013) A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* 494, 266-270.

119. Dumas, M.E., *et al.* (2007) Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. *Nature genetics* 39, 666-672.

120. Illig, T., *et al.* (2010) A genome-wide perspective of genetic variation in human metabolism. *Nature genetics* 42, 137-141.

121. Kebede, M.A. and Attie, A.D. (2014) Insights into obesity and diabetes at the intersection of mouse and human genetics. *Trends in endocrinology and metabolism: TEM* 25, 493-501.

122. Wen, W., *et al.* (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature communications* 5, 3438.

123. Riedelsheimer, C., *et al.* (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences of the United States of America* 109, 8872-8877.

124. Wang, K., *et al.* (2010) Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11, 843-854.

125. Hara, K., *et al.* (2014) Genetic architecture of type 2 diabetes. *Biochem Biophys Res Commun* 452, 213-220.

126. Nuzhdin, S.V., *et al.* (2012) Genotype-phenotype mapping in a post-GWAS world. *Trends Genet* 28, 421-426.

127. Califano, A., *et al.* (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature genetics* 44, 841-847.

128. Dobrin, R., *et al.* (2009) Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol* 10, R55.

129. Andreux, P.A., *et al.* (2012) Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* 150, 1287-1299.

130. Chen, R., *et al.* (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 148, 1293-1307.

131. Gillet, L.C., *et al.* (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & cellular proteomics : MCP* 11, O111 016717.

132. Wisniewski, J.R., *et al.* (2009) Universal sample preparation method for proteome analysis. *Nature methods* 6, 359-362.

133. Kochanowski, K., *et al.* (2015) Posttranslational regulation of microbial metabolism. *Current*

opinion in microbiology 27, 10-17.

134. Kusano, M., *et al.* (2011) Metabolomic approaches toward understanding nitrogen metabolism in plants. *Journal of experimental botany* 62, 1439-1453.

135. Manning, G., *et al.* (2002) The protein kinase complement of the human genome. *Science* 298, 1912-1934.

Figure legends

Figure 1. Trans-omic network across multiple omic layers (from left to right). Conventionally, a network has been identified by accumulating literature on specific molecules. Measurement of a single omic layer has now become available. Trans-omics is becoming available by connecting multi-omic measurements. A group of molecules with similar chemical properties such as genome, transcriptome, proteome, and metabolome is called an "omic" layer, which can be measured by next-generation sequencers (NGS), microarray, mass spectrometry, and NMR. (This figure partly includes "Process of transcription" by NHS National Genetics and Genomics Education Centre licensed under CC BY 2.0 / modified from original https://www.flickr.com/photos/119980645@N06/13080846733/in/photostream/.)

Figure 2. Technologies that connect multi-omic layers. (A) The classes of the trans-omic network (i) – (v) are indicated. Horizontal lines represent the indicated omic layers. The arrows indicate directions of regulation. (B) Connecting IDs across multiple omic layers. Circles represent IDs. Lines drawn between circles indicate conversion between IDs. The KEGG database plays a pivotal role in connecting multiple omic data by ID manipulation because it provides IDs for each omic layer, cross-reference tables that allow conversion among the IDs, and pathway maps tied with the IDs. Black lines indicate that an ID association or conversion can be performed by use of cross-reference tables provided by KEGG or elsewhere. Red lines are drawn between IDs that require manual conversion.

Figure 3. Three different concepts involved in a trans-omic network in comparison to a road network. A map, a route, and traffic of a road network (left) correspond to a map, a static signal flow, and a dynamic signal flow of a trans-omic network (right), respectively. A route and static signal flow are drawn in blue. Traffic and a dynamic signal flow are drawn in green, orange, and red. The warmer color represents more traffic.

Figure 4. Examples of metabolism-centric trans-omics. Blue and red arrows represent signal flow from genetic and environmental perturbations, respectively. Solid and dashed arrows represent direct and indirect association of molecules, respectively. (A) Global trans-omic responses of *E. coli* including metabolites, transcriptome, expression proteome, and metabolic fluxes against genetic (24 single gene disruptants) and

environmental (five different growth rates) perturbations [31]. (B) Adaptation of *B. subtilis* in a trans-omic network including metabolome, transcriptome, expression proteome, metabolic fluxes, and promoter activities in response to the shift between two major carbon sources, glucose and malate [34]. (C) A global landscape of the trans-omic network including metabolome and phosphoproteome of acute insulin action in rat hepatoma FAO cells [37]. See also a video of this trans-omic network for details (http://www.cell.com/cms/attachment/2020935146/2041143667/mmc7.mp4; Yugi *et al*. (2014) *Cell Rep*., CC BY 3.0).

Figure 5. From GWAS to trans-OWAS. (A, left) GWAS is a linkage analysis that includes the phenotypic relation to a single omic layer (genome). GWAS reflects only genetic factors and the phenomenological relationship between genome and phenome. (A, right) Trans-OWAS is a linkage analysis that includes the phenotypic relation to multiple omic layers. Trans-OWAS reflects both genetic and environmental factors and indicates the molecular relationship of pathogenesis in a trans-omic network. (B) Multifactorial diseases, such as T2DM, appear as breakdowns of the insulin sensitivity pathway (blue) and insulin secretion pathway (red) in a trans-omic network that reflects both genetic and environmental factors. (This figure partly includes ‘Process of transcription’ by NHS National Genetics and Genomics Education Centre licensed under CC BY 2.0/modified from original https://www.flickr.com/photos/119980645@N06/13080846733/in/photostream/, and Figure 1 of "The chromatin signature of pluripotent cells" by Ky Sha and Laurie Boyer, licensed under CC BY 3.0/modified from original http://www.stembook.org/node/585.)

Figure 1

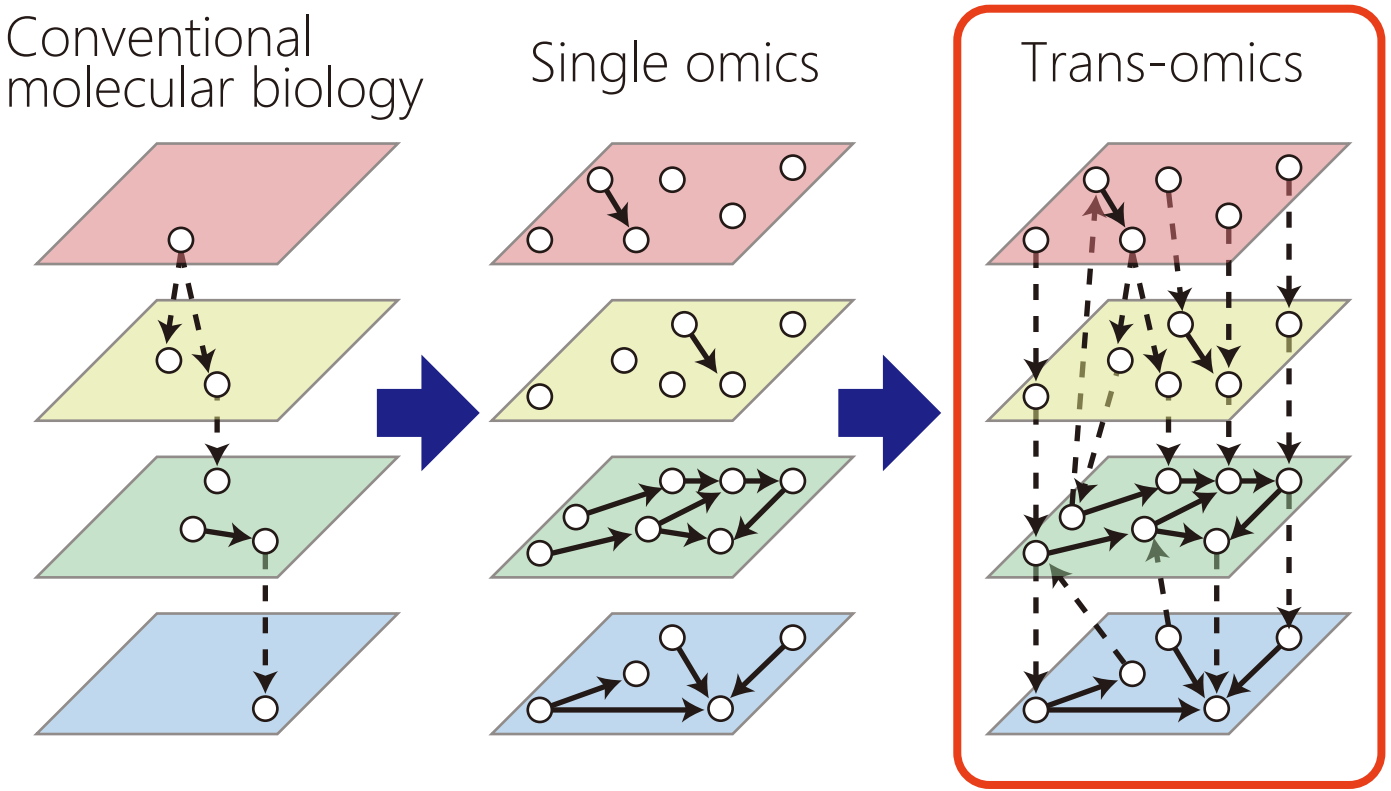

Conventional molecular biology
Single omics
Trans-omics
Genome
Transcriptome
Proteome
Metabolome
Measurement
NGS
RNA-seq (NGS)
Microarray
Mass Spectrometry
Mass Spectrometry
NMR

Figure 2

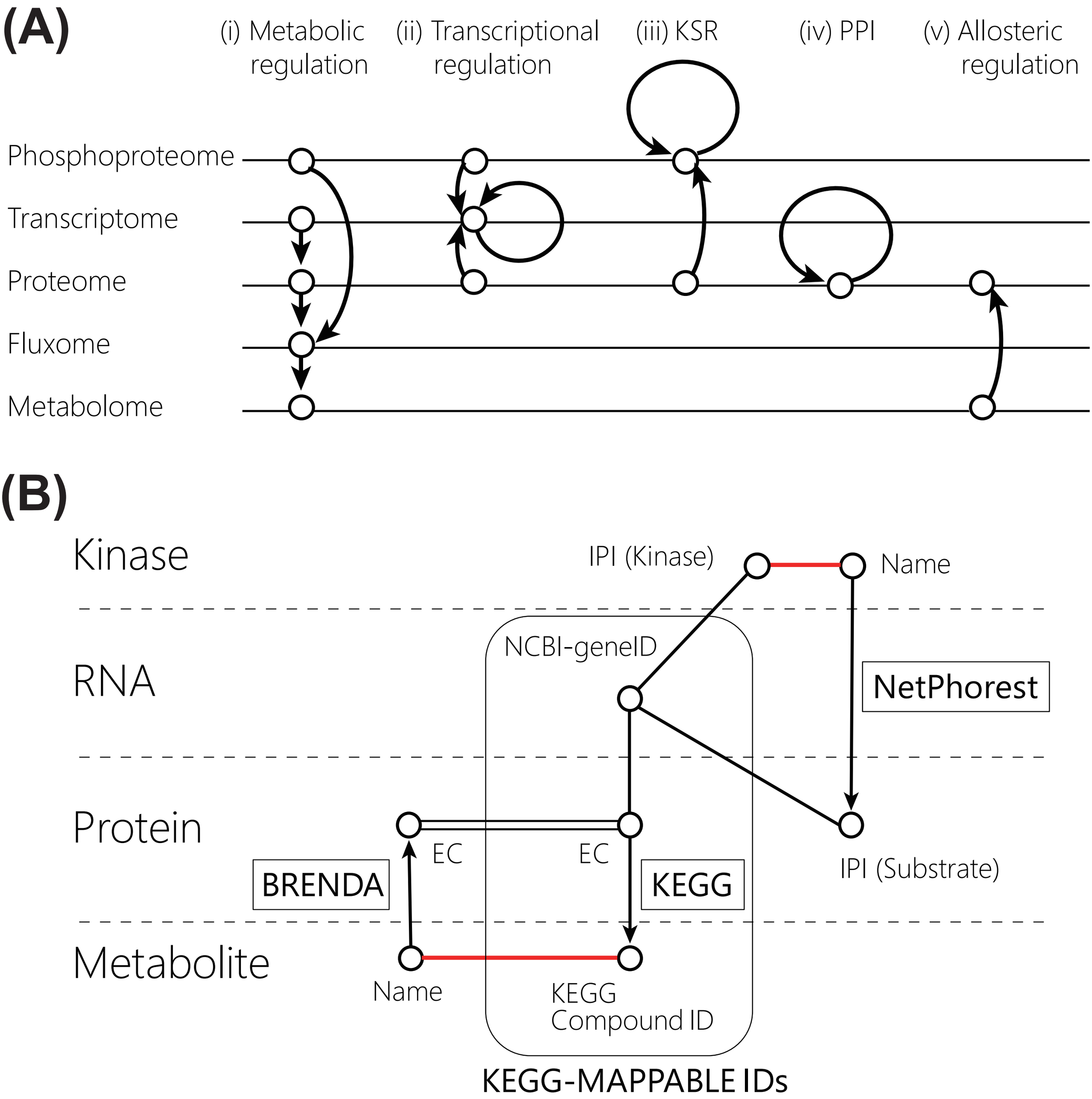
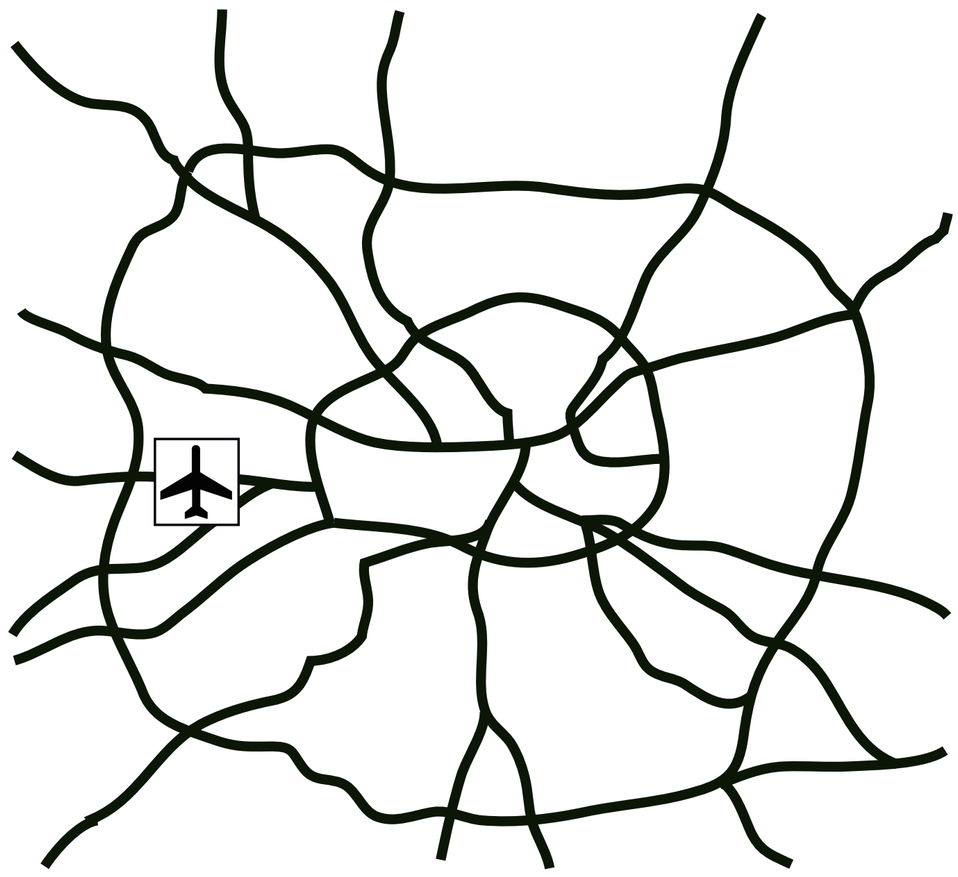

(A)
(i) Metabolic regulation
(ii) Transcriptional regulation
(iii) KSR
(iv) PPI
(v) Allosteric regulation
Phosphoproteome
Transcriptome
Proteome
Fluxome
Metabolome
(B)
Kinase
RNA
Protein
Metabolite
IPI (Kinase)
Name
NCBI-geneID
NetPhorest
EC
EC
BRENDA
KEGG
IPI (Substrate)
Name
KEGG
Compound ID
KEGG-MAPPABLE IDs
ID
Database / web service
Identical
Interconvertible
Manually converted
Signal flow

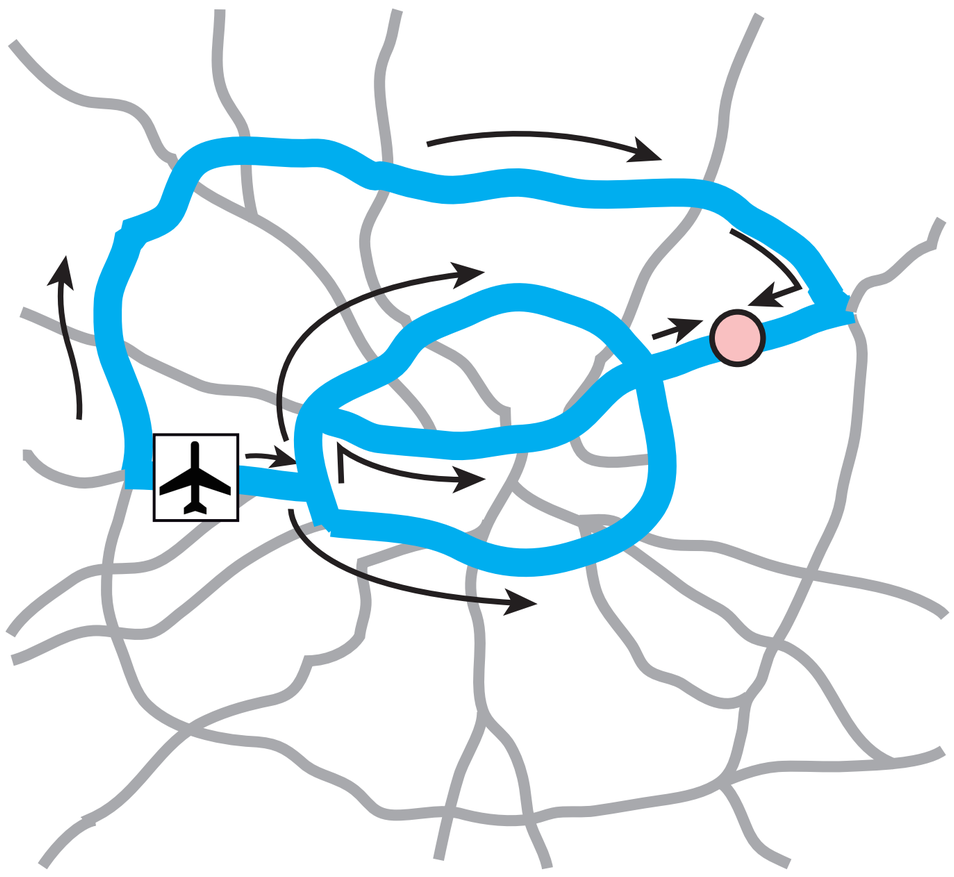
Figure 3

Road Network

Map



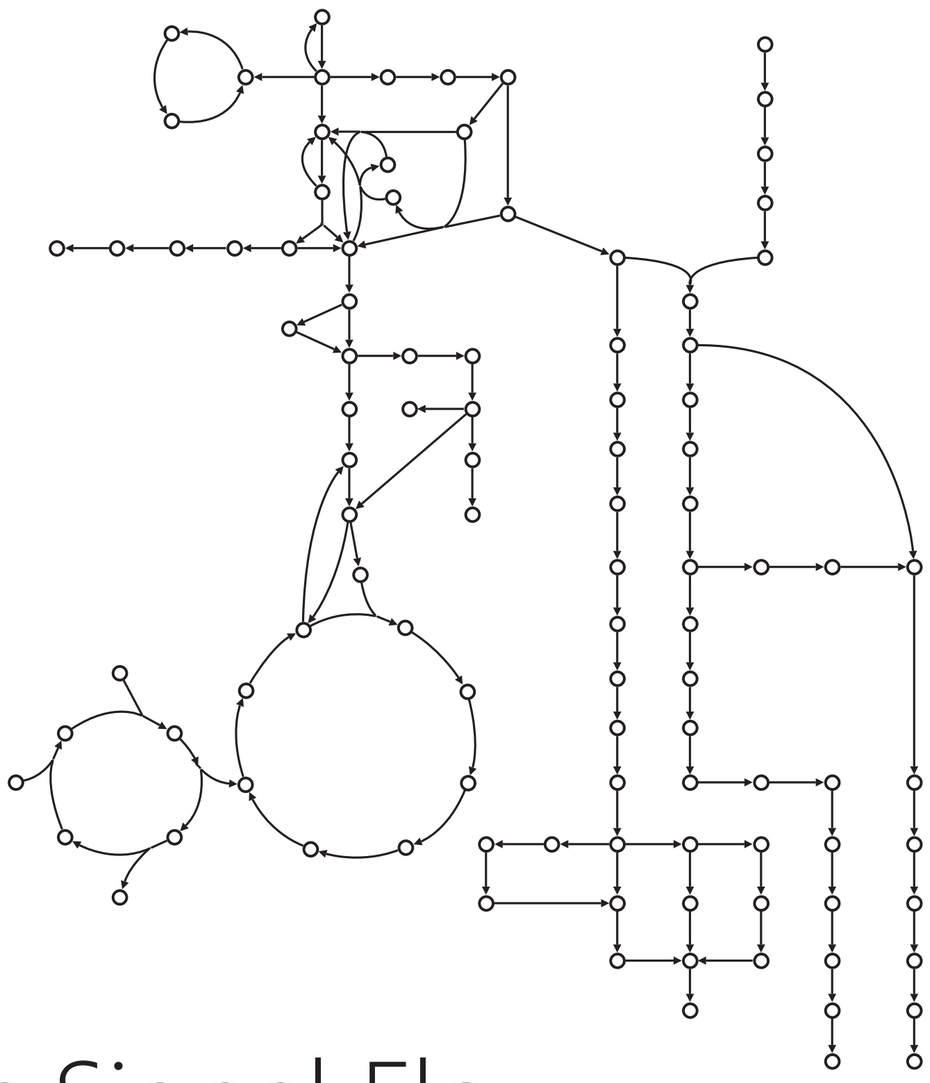
Route



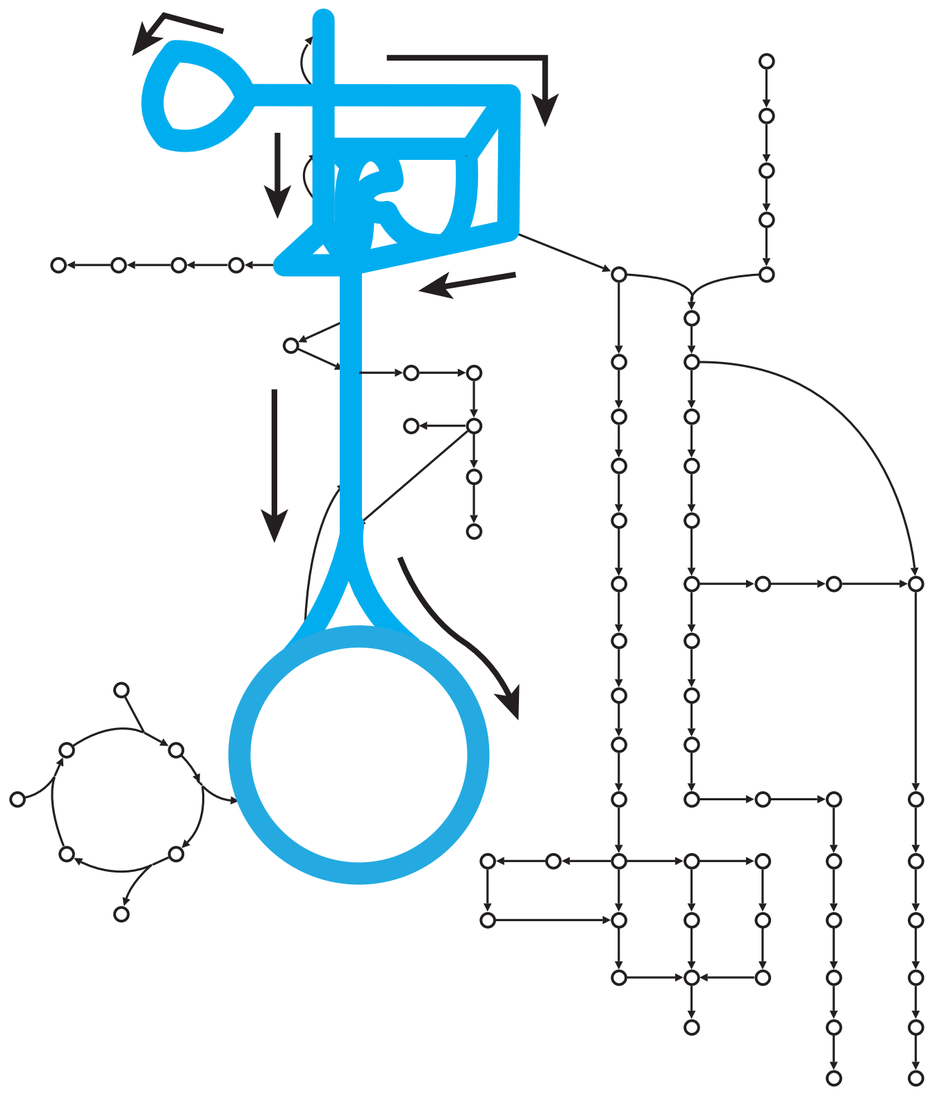
Traffic

Trans-omic Network

Map



Static Signal Flow



Dynamic Signal Flow

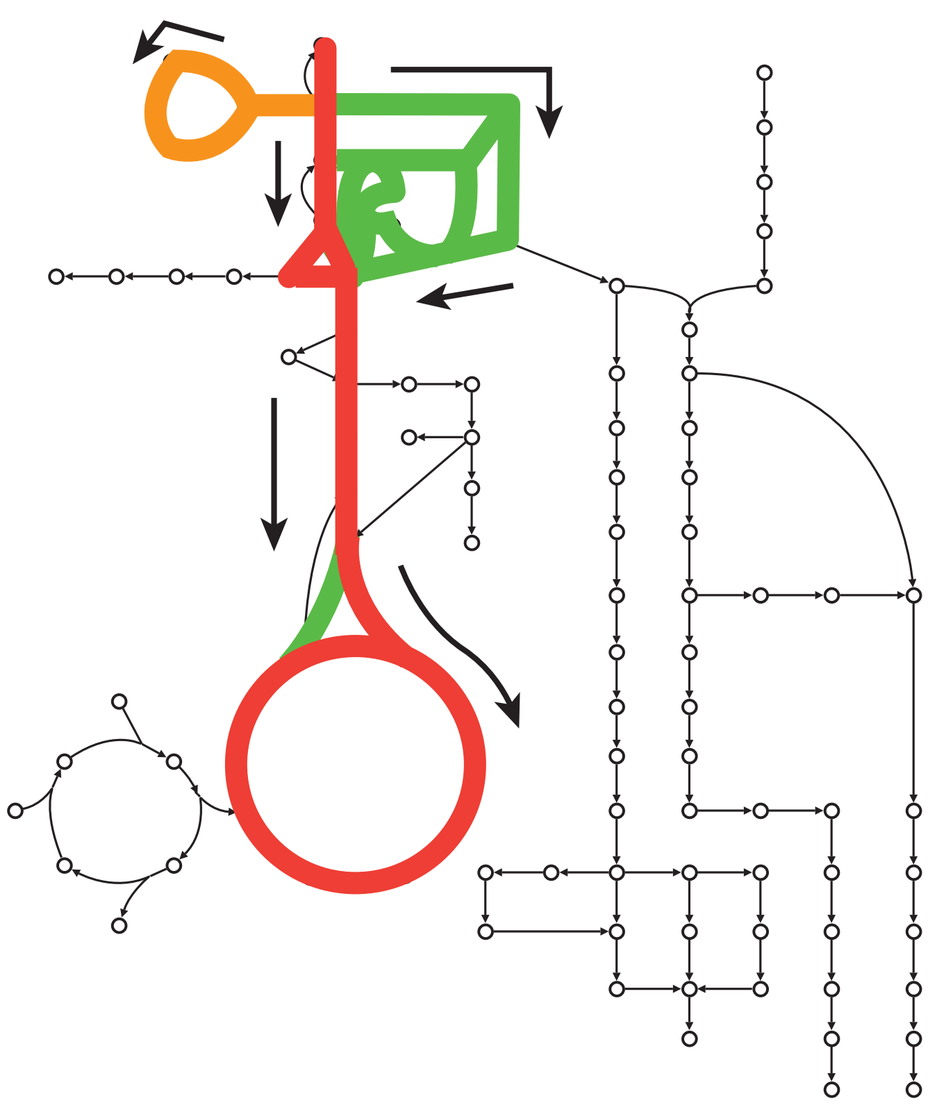
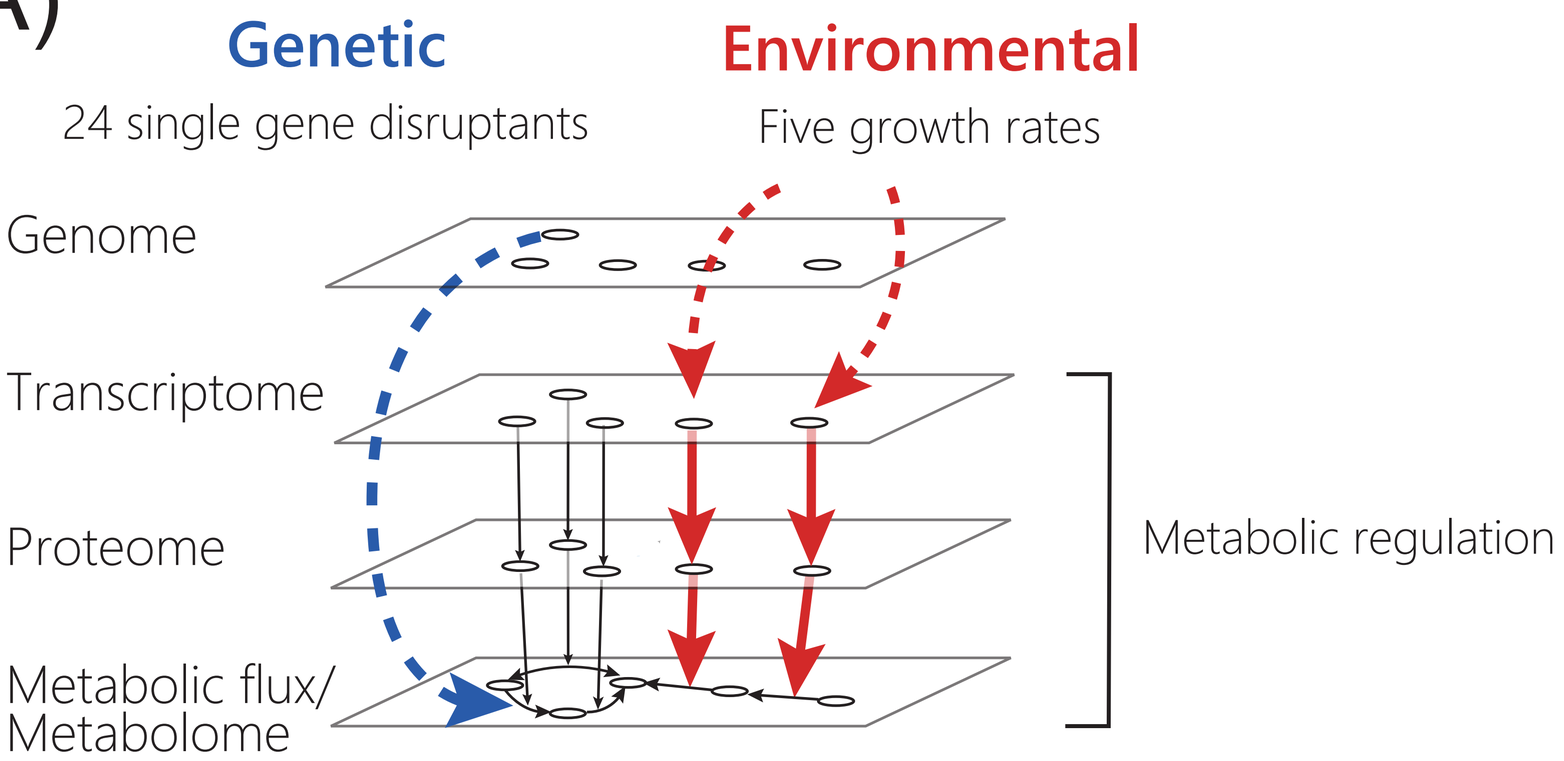
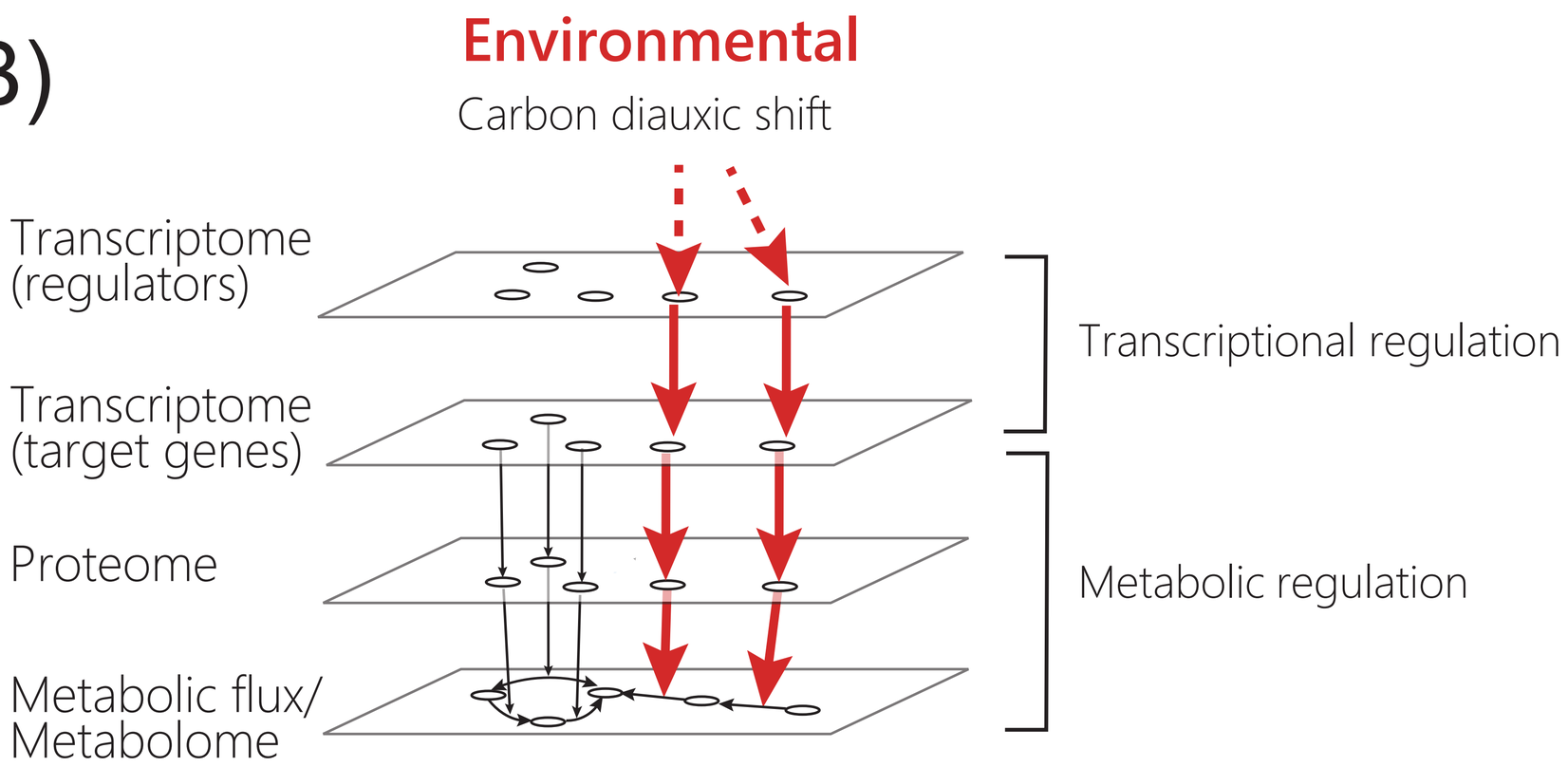


Figure 4

(A)


Genetic
24 single gene disruptants
Environmental
Five growth rates
Genome
Transcriptome
Proteome
Metabolic flux/
Metabolome
Metabolic regulation

(B)


Environmental
Carbon diauxic shift
Transcriptome
(regulators)
Transcriptome
(target genes)
Proteome
Metabolic flux/
Metabolome
Transcriptional regulation
Metabolic regulation

(C)

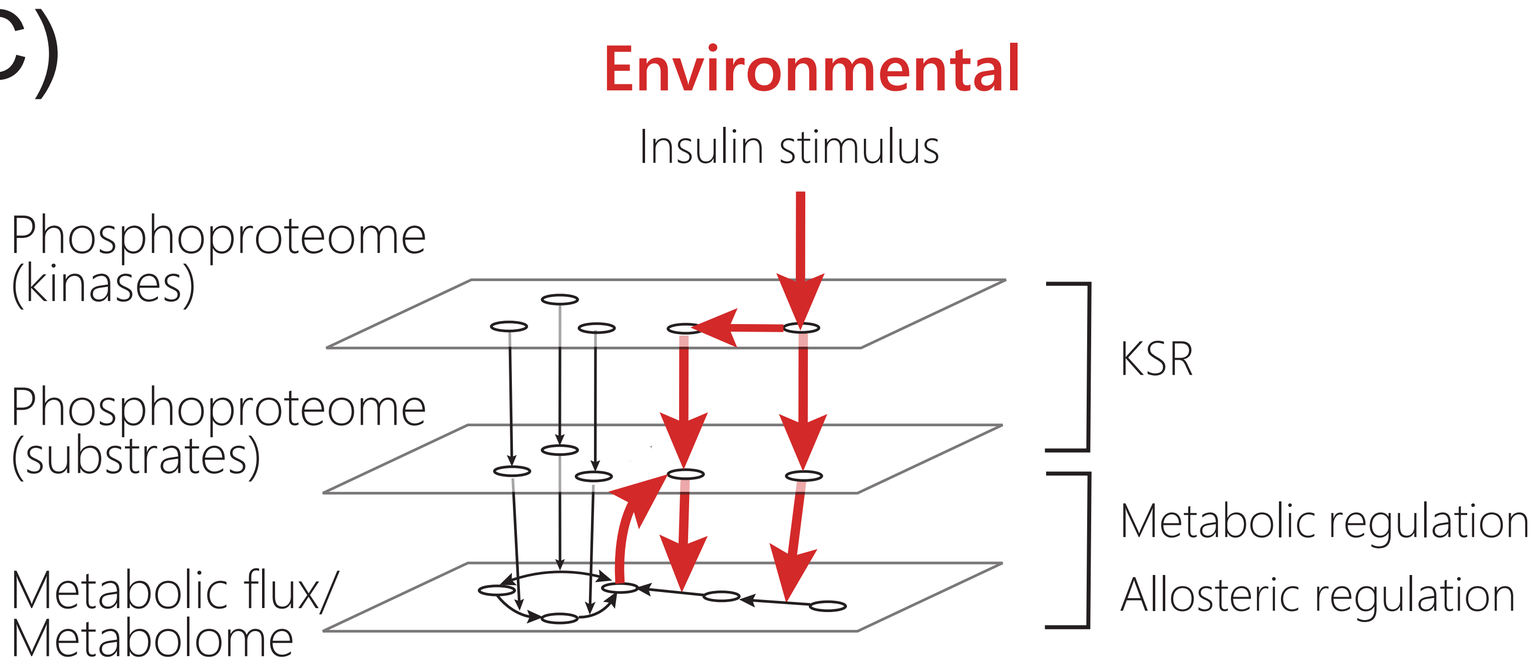

Environmental
Insulin stimulus
Phosphoproteome
(kinases)
Phosphoproteome
(substrates)
Metabolic flux/
Metabolome
KSR
Metabolic regulation
Allosteric regulation

Figure 5

(A)

Present: GWAS

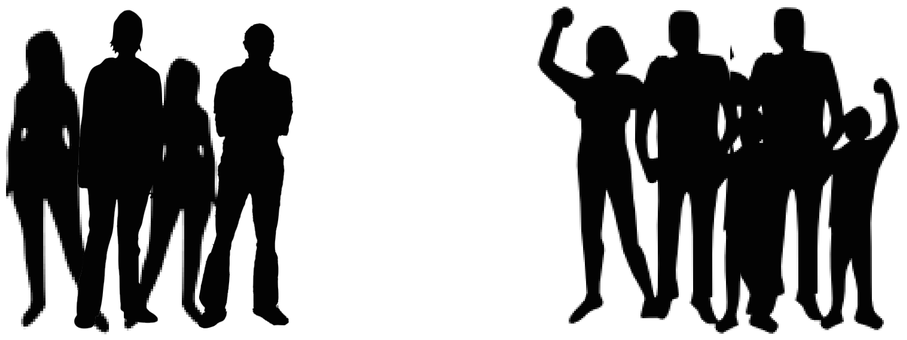
Patients Non-patients

Genetic information

Genome

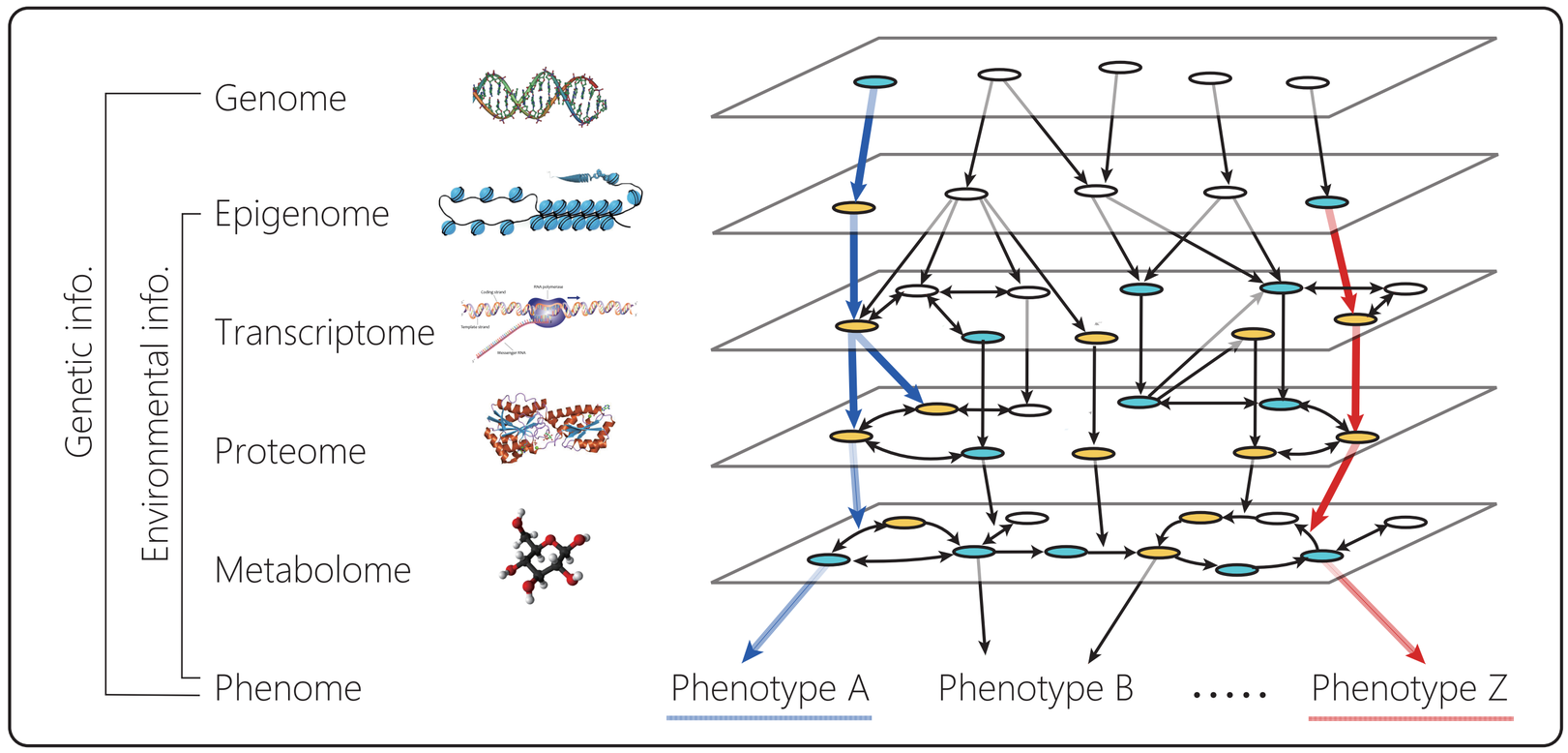
Phenome Phenotype A

Future: Trans-OWAS



Patients Non-patients

Genetic and environmental information



(B)