Genome **Biology**

REVIEW

The next frontier of systems biology: higher-order and interspecies interactions

Michael A Fischbach[1]* and Nevan J Krogan[2]*

Abstract

Systems approaches are not so different in essence from classical genetic and biochemical approaches, and in the future may become adopted so widely that the term 'systems biology' itself will become obsolete.

Systems biology means different things to different people, and one can envisage it more as a strategy for studying biological systems than as a field of biology. Systems approaches have been very successful in the realms of biochemistry and genetics, especially for genetically tractable organisms, and have led to a deluge of mechanistic insights into a variety of biological areas. The 'systematic' nature of the approach involves testing or assaying all components of a biological milieu simultaneously, in an unbiased fashion, with no prior assumptions of what will be found. However, these modern approaches are not so different when compared to more classical genetic and biochemical strategies. Finally, we anticipate that the next frontier of systems biology will involve both higher-order interactions and the study of interspecies relationships in a systematic fashion.

A decade ago, Bruce Alberts, Andrew Murray and Lee Hartwell noted that cellular components are organized into functional groups, or modules, and that the reductionist approach of studying each component in isolation was limiting [1,2]. Recent efforts in systems biology have taken advantage of this observation by using unbiased approaches to define the protein complexes that comprise these modules. For example, two groups have used a systematic affinity tag/purification and mass spectrometry approach to identify hundreds of protein
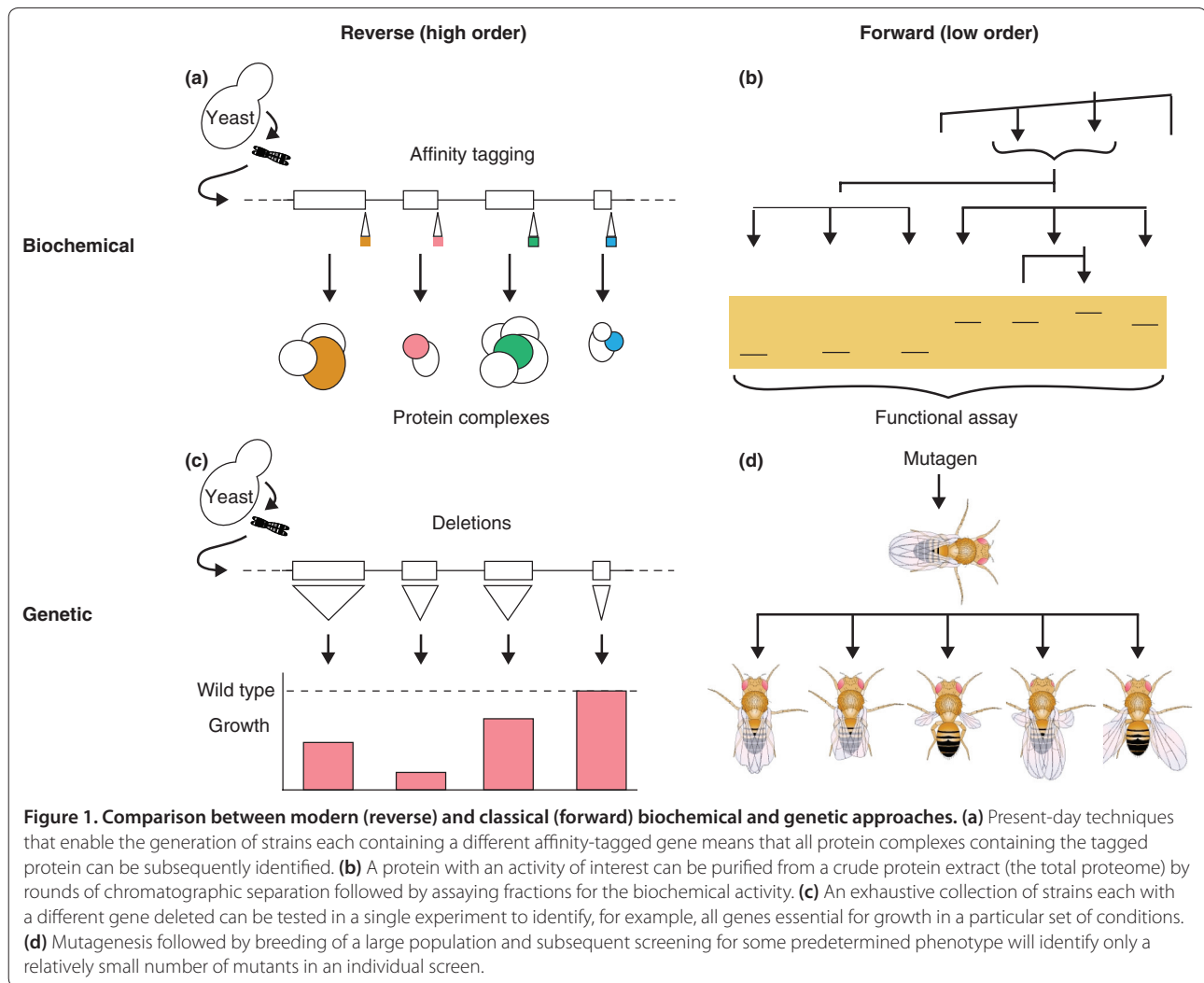
complexes in the budding yeast *Saccharomyces cerevisiae*, many of which were previously unknown [3,4] (Figure 1a). Global efforts to define protein complexes have been extended to the prokaryotes *Escherichia coli* [5,6] and *Mycoplasma pneumoniae* [7] as well as to mammalian cells [8,9]. Highlighting the power of these approaches to rapidly uncover new biology in mapping out the circuit diagram of a cell, Kühner *et al.* [7] characterized 62 homomultimeric and 116 heteromultimeric soluble protein complexes in *M. pneumoniae*, and the majority of these were novel. A similar proportion of novel findings were uncovered when this unbiased proteomic approach was applied to other prokaryotic organisms [5,6] and higher organisms [8,9].

In comparison, consider a classic biochemistry experiment: in 1958, Arthur Kornberg and co-workers purified DNA polymerase from *E. coli* by fractionating a crude protein extract and testing individual fractions for a DNA-replicating activity [10,11]. At first glance, Kornberg's experiments might seem a world apart from the *M. pneumoniae* effort; the former identified a single enzyme while the latter defined nearly all of the protein complexes in the cell. However, these classical and modern approaches are in fact surprisingly similar (Figure 1b), as both Kornberg and Kuhner *et al.* were performing unbiased, systematic screens of bacterial proteomes. Indeed, their major difference is one of scale, not type: Kornberg sought to identify a single molecular machine with a specific function, whereas Kuhner *et al*'s goal was to identify all of the molecular machines. While the latter studies do not address the complexes' functions, one can now leverage other information or strategies to subsequently scan the defined molecular machines to infer their functions. For example, one can use bioinformatics approaches, such as finding homologs in other organisms, and infer the evolutionary conservation of similar functions. Also, comparing this information with other types of data, if they exist, can also be illuminating. For example, a three-pronged interrogation of the poorly studied *M. pneumoniae* used not only proteomic techniques as described above [7], but also global studies of the transcriptome [12] and metabolome [13]. Ultimately, this information can be integrated to

*Correspondence: krogan@cmp.ucsf.edu; fischbach@fischbachgroup.org
[1]Department of Bioengineering and Therapeutic Sciences and California Institute of Quantitative Biosciences, University of California, San Francisco, San Francisco, CA 94158, USA
[2]Department of Cellular and Molecular Pharmacology and California Institute of Quantitative Biosciences, University of California, San Francisco, San Francisco, CA 94158, USA

BioMed Central

Figure 1. Comparison between modern (reverse) and classical (forward) biochemical and genetic approaches. (a) Present-day techniques that enable the generation of strains each containing a different affinity-tagged gene means that all protein complexes containing the tagged protein can be subsequently identified. **(b)** A protein with an activity of interest can be purified from a crude protein extract (the total proteome) by rounds of chromatographic separation followed by assaying fractions for the biochemical activity. **(c)** An exhaustive collection of strains each with a different gene deleted can be tested in a single experiment to identify, for example, all genes essential for growth in a particular set of conditions. **(d)** Mutagenesis followed by breeding of a large population and subsequent screening for some predetermined phenotype will identify only a relatively small number of mutants in an individual screen.

ascertain the functions of individual proteins and complexes, and their proposed biochemical activities can be tested in a more traditional fashion.

Genetic analyses have also greatly benefited from global systems approaches. For example, Ron Davis, Mark Johnston and colleagues [14] generated a genome-wide collection of *S. cerevisiae* gene deletion mutants, which enabled them to identify genes essential for growth under standard laboratory conditions. Unbiased screening of this genome-wide mutant library using reverse genetics (the approach in which the function of a gene is identified starting with the DNA sequence rather than the phenotype) to identify gene function through the response of the mutants to different culture conditions, different drugs, and by gene-expression profiling [15-17] has led to a deluge of functional insights into nearly all the biological process in the yeast cell (Figure 1c). Genome-wide knockout libraries have now been created in other genetically tractable organisms, including *E. coli* [18] and

Schizosaccharomyces pombe [19], and similar functional studies are now being carried out in these.

Forward genetics - the process of screening large numbers of organisms to identify those with a variant phenotype and then identifying the mutant gene responsible - was pioneered by Thomas Hunt Morgan in the early 1900s. Morgan selected phenotypic variants of the fruit fly *Drosophila melanogaster* generated after chemical mutagenesis, such as those with white rather than red eyes [20], or wings shorter than normal [21], and performed cross-breeding experiments to identify single heritable mutant genes (Figure 1d). As more and more *Drosophila* mutant strains were generated, these studies led to the generation of the first genetic map, based on recombination frequencies, by one of Morgan's students, Alfred Sturtevant [22]. Similar mutagenesis approaches have been carried out in other organisms, but tricks have been developed to help make many organisms more genetically tractable. For example, in budding yeast,

the location in the chromosomes of genes mutated by the random insertion of a transposon can be pinpointed by detecting the transposon itself [23]. Again, these experiments collectively represent genome-wide screens, since in chemical or transposon mutagenesis each gene in the organism is, in theory, subjected to the mutagen, although in this case, only the mutations that produce a desired effect will be identified.

Collectively, comparisons between the classical and modern approaches demonstrate their similarity: they involve systematically testing or assaying all components of a biological milieu in an unbiased fashion. The primary difference is their dimensionality; for classical genetics and biochemistry, a single gene or protein was often the answer, whereas a systems biologist seeks many answers at once even if the questions are not defined at the outset. Importantly, combining perturbations yields additional information as it enables the analysis of how the parts interact - the result could be the entire circuit diagram of a cell.

## Higher-order experiments as a future focal point of systems biology

If modern systems biology is only a short leap from classical biochemistry and genetics, how will future experiments in systems biology continue the trend of increased dimensionality? We believe that some of the greatest gains will be made in two areas: multiple perturbations within a species; and interspecies interactions.

Multiple perturbations within a species

While systematic single-mutant analysis has revealed much in terms of gene function, the advent of methodology for creating double mutants *en masse* in a variety of organisms, including *S. cerevisiae* [24], *S. pombe* [25] and *E. coli* [26,27], has greatly accelerated the characterization of biological pathways and their interconnections.

Since single-gene perturbations often provide limited phenotypic consequences, the ability to generate double mutants allows a deeper probing of phenotypic space (Figure 2). Ultimately, this approach creates a powerful phenotypic signature for a given mutant (that is, how a mutant interacts genetically with all other mutants it is queried against), which can be used to group functionally related sets of genes. While initially this strategy is often not considered as 'hypothesis-driven', it is most certainly a 'hypothesis generator', with some of the most interesting connections revealed being completely unanticipated. For example, a direct connection between the nuclear pore and repair of damaged DNA during DNA replication by pore-associated enzymes was uncovered in yeast using these strategies [28].

Of course, triple perturbations within a single organism are also possible (for example, a triple mutant, or a
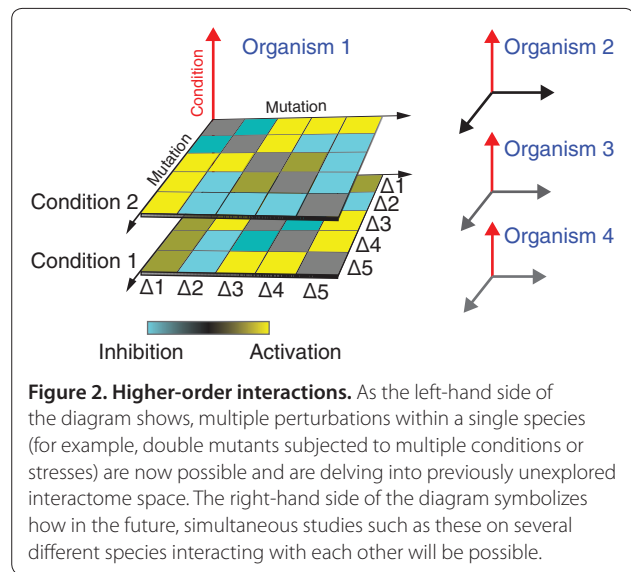
Figure 2. Higher-order interactions. As the left-hand side of the diagram shows, multiple perturbations within a single species (for example, double mutants subjected to multiple conditions or stresses) are now possible and are delving into previously unexplored interactome space. The right-hand side of the diagram symbolizes how in the future, simultaneous studies such as these on several different species interacting with each other will be possible.

double mutant put under a given stress condition), which reveal even more about complex biological phenomena (Figure 2). For example, Trey Ideker and colleagues have generated a quantitative genetic-interaction map in budding yeast using double mutants in the presence of an exogenous DNA-damaging agent, an additional perturbation that delved into previously unexplored interactome space (S Bandyopadhyay *et al.*, personal communication).

Interspecies interactions

Systems biology does not end at the cell membrane; interactions between cells of different species are governed by the same principles as those between functional modules. Genetic and biochemical interspecies interactions can be just as significant as those within a species. For example, a polymorphism in the mammalian tripartite motif family protein TRIM5α modulates the infectivity of HIV in Old World monkeys [29], representing a genetic interaction between a mammalian and a viral gene. Likewise, during bacterial and viral infections of animals, direct interspecies protein-protein interactions can occur when pathogen-encoded proteins hijack cellular processes by binding to and perturbing the activity of host protein complexes. For example, the *Pseudomonas* type III secretion system delivers the bacterial toxin ExoS into host cells where it functions as a GTPase-activating protein for the host's Rho-family GTPases. Their activation results in perturbation of the actin cytoskeleton, a prime target of these GTPases in eukaryotic cells [30]. Interspecies genetic interactions between pathogens such as HIV and *Mycobacterium tuberculosis* and their hosts have already been studied systematically [31-34]. For example, genome-wide RNA interference screens targeting human genes in the

context of infection with HIV and tuberculosis have been carried out. These studies have identified sets of host factors that are required for infection, providing a more global functional view of pathogenesis [31-34].

Future efforts are likely in three areas. First, work such as that on HIV and *M. tuberculosis* is likely to be extended to studying not only other host-pathogen interactions, but also host-symbiont interactions such as those between gut epithelial cells and *Bacteroides* spp. [35], to determine how *Bacteroides* metabolites influence the host and how the host response in turn modulates the cell state of *Bacteroides*. Second, the effects of small molecules are likely to be added as a condition; the importance of this is that the resulting three-way host-pathogen-small molecule system comes close to mimicking an infected human patient being treated with a drug (Figure 2). Third, the development of suitable intraspecies variants will allow the investigation of communication between cells of the same species in the context of an interspecies system such as host-bacterium symbiosis. Such systems will have the power to detect genetic interactions relevant to paracrine signaling in eukaryotic cells, and to quorum sensing and other intraspecies signaling in prokaryotic cells.

Changes over space and time

Most systems-biological experiments study genetic and biochemical interactions at a single time point. But many interesting biological processes involve temporal or spatial dynamics - for example, cell migration down a gradient of chemoattractant or a pulse of signaling in response to an extracellular growth factor - and so another form of higher-dimension systems biology will be the study of how cellular modules change over space and time. Another area in which dimensionality is likely to increase is where the assay is used as a readout. The most common assays are the simplest: cell growth and reporter gene expression. As high throughput mass spectrometry, transcriptional profiling, and DNA sequencing become more common, assays that scan an entire genome, proteome, or metabolome will generate richer data for each set of perturbations.

In conclusion, there are two reasons for systematic approaches gaining so much traction among biologists. First, screening all the genes or proteins in an organism is not that much more difficult than analyzing a small subset, and robotics and high-throughput screening techniques are now within the reach of most labs. Second, the costs of systems biology scale sub-linearly while the payoffs scale super-linearly. Put simply, screening 100 times as many genes yields more than 100 times the information; the additional information consists in learning how groups of genes behave, enabling functional modules to be identified and characterized. As

a result, we believe systems biological approaches will be adopted broadly, perhaps even becoming standard practice in experiments on genetically tractable organisms. Indeed, broad acceptance of systematic approaches could render the term 'systems biology' obsolete, which would surely be a mark of its success.

References
1. Alberts B: **The cell as a collection of protein machines: preparing the next generation of molecular biologists.** *Cell* 1998, **92**:291-294.
2. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402**:C47-C52.
3. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, *et al.*: **Proteome survey reveals modularity of the yeast cell machinery.** *Nature* 2006, **440**:631-636.
4. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, *et al.*: **Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*.** *Nature* 2006, **440**:637-643.
5. Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A: **Interaction network containing conserved and essential protein complexes in *Escherichia coli*.** *Nature* 2005, **433**:531-537.
6. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A, *et al.*: **Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins.** *PLoS Biol* 2009, **7**:e96.
7. Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castaño-Diez D, Chen WH, Devos D, Güell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC: **Proteome organization in a genome-reduced bacterium.** *Science* 2009, **326**:1235-1240.
8. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Hériché JK, Sykora MM, Augsburg M, Hudecz O, Buschhorn BA, Bulkescher J, Conrad C, Comartin D, Schleiffer A, Sarov M, Pozniakovsky A, Slabicki MM, Schloissnig S, Steinmacher I, Leuschner M, Ssykor A, Lawo S, Pelletier L, Stark H, Nasmyth K, Ellenberg J, Durbin R, Buchholz F, Mechtler K, Hyman AA, Peters JM: **Systematic analysis of human protein complexes identifies chromosome segregation proteins.** *Science* 2010, DOI:10.1126/science.1181348.
9. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D: **Large-scale mapping of human protein-protein interactions by mass spectrometry.** *Mol Syst Biol* 2007, **3**:89.
10. Lehman IR, Bessman MJ, Simms ES, Kornberg A: **Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*.** *J Biol Chem* 1958, **233**:163-170.
11. Bessman MJ, Lehman IR, Simms ES, Kornberg A: **Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction.** *J Biol Chem* 1958, **233**:171-177.
12. Güell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L: **Transcriptome complexity in a genome-reduced bacterium.** *Science* 2009, **326**:1268-1271.
13. Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH, Wodke JA, Güell M, Martínez S, Bourgeois R, Kühner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutiérrez-Gallego R, Russell RB, Gavin AC, Bork P,

Serrano L: Impact of genome reduction on bacterial metabolism and its regulation. *Science* 2009, **326**:1263-1268.

14. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, *et al.*: Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999, **285**:901-906.

15. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH: Functional discovery via a compendium of expression profiles. *Cell* 2000, **102**:109-126.

16. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, *et al.*: Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002, **418**:387-391.

17. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G: The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 2008, **320**:362-365.

18. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006, **2**:2006.0008.

19. Bioneer *Schizosaccharomyces pombe* [http://pombe.bioneer.co.kr]

20. Morgan TH: The origin of five mutations in eye color in *Drosophila* and their modes of inheritance. *Science* 1911, **33**:534-537.

21. Morgan TH: The origin of nine wing mutations in *Drosophila*. *Science* 1911, **33**:496-499.

22. Sturtevant AH: The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* 1913, **14**:43-59.

23. Snyder M, Elledge S, Davis RW: Rapid mapping of antigenic coding regions and constructing insertion mutations in yeast genes by mini-Tn10 "transplason" mutagenesis. *Proc Natl Acad Sci USA* 1986, **83**:730-734.

24. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C: Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 2001, **294**:2364-2368.

25. Roguev A, Wiren M, Weissman JS, Krogan NJ: High-throughput genetic interaction mapping in the fission yeast *Schizosaccharomyces pombe*. *Nat Methods* 2007, **4**:861-866.

26. Butland G, Babu M, Díaz-Mejía JJ, Bohdana F, Phanse S, Gold B, Yang W, Li J, Gagarinova AG, Pogoutse O, Mori H, Wanner BL, Lo H, Wasniewski J, Christopolous C, Ali M, Venn P, Safavi-Naini A, Sourour N, Caron S, Choi JY, Laigle L, Nazarians-Armavil A, Deshpande S, Joe S, Datsenko KA, Yamamoto N, Andrews BJ, Boone C, Ding H, *et al.*: eSGA: *E. coli* synthetic genetic array analysis. *Nat Methods* 2008, **5**:789-795.

27. Typas A, Nichols RJ, Siegele DA, Shales M, Collins SR, Lim B, Braberg H, Yamamoto N, Takeuchi R, Wanner BL, Mori H, Weissman JS, Krogan NJ, Gross CA: High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat Methods* 2008, **5**:781-787.

28. Nagai S, Dubrana K, Tsai-Pflugfelder M, Davidson MB, Roberts TM, Brown GW, Varela E, Hediger F, Gasser SM, Krogan NJ: Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* 2008, **322**:597-602.

29. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J: The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 2004, **427**:848-853.

30. Aktories K, Schmidt G, Just I: Rho GTPases as targets of bacterial protein toxins. *Biol Chem* 2000, **381**:421-426.

31. Kumar D, Nath L, Kamal MA, Varshney A, Jain A, Singh S, Rao KV: Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. *Cell* 2010, **140**:731-743.

32. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ: Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008, **319**:921-926.

33. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK: Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 2008, **135**:49-60.

34. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS: Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 2008, **4**:495-504.

35. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI: A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 2003, **299**:2074-2076.