Article Watch: July 2021

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Avenue, Athens GA 30606. Tel; (706) 713-2216: Fax; (706) 713-2221: Email; cslaught@uga.edu or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

NUCLEIC ACID SEQUENCING

Garg S, Fungtammasan A, Carroll A, Chou M, Schmitt A, Zhou X, Mac S, Peluso P, Hatas E, Ghurye J, Maguire J, Mahmoud M, Cheng H, Heller D, Zook J M, Moemke T, Marschall T, Sedlazeck F J, Aach J, Chin C-S, Church G M, Li H. Chromosome-scale, haplotype-resolved assembly of human genomes. ***Nature Biotechnology*** **39;2021:309-312.**

Kronenberg Z N, Rhie A, Koren S, Concepcion G T, Peluso P, Munson K M, Porubsky D, Kuhn K, Mueller K A, Low W Y, Hiendleder S, Fedrigo O, Liachko I, Hall R J, Phillippy A M, Eichler E E, Williams J L, Smith T P L, Jarvis E D, Sullivan S T, Kingan S B. Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. ***Nature Communications*** **12;2021:1935.**

Porubsky D, Ebert P, Audano P A, Vollger M R, Harvey W T, Marijon P, Ebler J, Munson K M, Sorensen M, Sulovari A, Haukness M, Ghareghani M, Lansdorp P M, Paten B, Devine S E, Sanders A D, Lee C, Chaisson M J P, Korbel J O, Eichler E E, Marschall T, Human Genome Structural Variation C. Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. ***Nature Biotechnology*** **39;2021:302-308.**

Ebert P, Audano P A, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder M J, Sulovari A, Ebler J, Zhou W, Serra Mari R, Yilmaz F, Zhao X, Hsieh P, Lee J, Kumar S, Lin J, Rausch T, Chen Y, Ren J, Santamarina M, Höps W, Ashraf H, Chuang N T, Yang X, Munson K M, Lewis A P, Fairley S, Tallon L J, Clarke W E, Basile A O, Byrska-Bishop M, Corvelo A, Evani U S, Lu T-Y, Chaisson M J P, Chen J, Li C, Brand H, Wenger A M, Ghareghani M, Harvey W T, Raeder B, Hasenfeld P, Regier A A, Abel H J, Hall I M, Flicek P, Stegle O, Gerstein M B, Tubio J M C, Mu Z, Li Y I, Shi X, Hastie A R, Ye K, Chong Z, Sanders A D, Zody M C, Talkowski M E, Mills R E, Devine S E, Lee C, Korbel J O, Marschall T, Eichler E E. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. ***Science*** **372;2021:eabf7117.**

https://doi.org/10.7171/jbt.2021-3202-005

It has long been recognized that genome assembly based upon short-read sequences without resolution (phasing) of haplotypes results in under-representation of structural variation (inversions, deletions, duplications and insertions of >50 bp in length). It may also introduce assembly errors in regions where haplotypes differ in sequence. Genome assembly without haplotype resolution also obscures essential information about the way allelic variation affects epigenetic alterations and gene expression changes, and compromises biologic inference about heritable disease and evolutionary change. Methodologic developments by several groups involving long-read and strand-specific sequencing now accomplish *de novo*, long-range (telomere-to-telomere) assembly of completely phased whole human diploid genomes, and do so without the restrictive necessity for a reference sequence or parental genomic data. Garg *et al.* and Kronenberg *et al.* begin with high-fidelity long-read sequence data generated by the Pacific Biosciences (PacBio) technology. They perform phased assembly with input of information from Hi-C, which records long-range chromosomal interactions preferentially within strands. Porubsky *et al.* also begin with long-read sequence data (either from PacBio or Oxford Nanopore Technologies), but they employ the Strand-seq method for phasing. In Strand-seq a thymidine analog labels the nascent strand during DNA replication and removes that strand from sequencing so that only the template strand of each homologous chromosome pair is sequenced. The resulting data allow contig assemblies to be assigned to chromosomes in order to phase them. Inference of the order of contigs within each chromosome is based on low-frequency switching between parental strands in the parental cell due to sister chromatid exchange. Contigs that are closer together share the same template strand more often than more distant contigs. The Strand-seq methodology differs from the Hi-C methodology in its requirement for the extra step of labeling dividing cells in culture. The Strand-seq technology is then deployed by Ebert *et al.* to produce fully phased genome assemblies for 35 individuals of diverse ethnicity. Their study reveals

a previously unrecognized prevalence of structural variants, including a high fraction of structural variants resulting from homology-directed repair, nonallelic homologous recombination or variable number tandem repeats. The study also examines the evolutionary trajectory of human retrotransposons and further identifies new associations between disease-associated variants and changes in gene expression. These new capabilities for genomic haplotype phasing presage widespread application of the methodology in clinical and population studies.

METABOLOMICS

Hartmann F J, Mrdjen D, Mccaffrey E, Glass D R, Greenwald N F, Bharadwaj A, Khair Z, Verberk S G S, Baranski A, Baskar R, Graf W, Van Valen D, Van Den Bossche J, Angelo M, Bendall S C. Single-cell metabolic profiling of human cytotoxic T cells. *Nature Biotechnology* 39;2021:186-197.

Much interest is presently focused on the linkage between patterns of cellular metabolism in immune cells and immune cell function expressed in terms of migration, proliferation, and effector molecule secretion. Improved understanding of the metabolic control of immune cell function is expected to improve discrimination of functionally specialized immune cell subsets and possibly to lead to medicinal control of immune cells in disease. Hartmann *et al.* here describe methodology for broad-based characterization of cytotoxic T cell metabolic status at the single-cell level. They focus especially on *ex vivo* cells, which better represent function in the physiologic context than *in vitro* models. As a proxy for metabolic flux through pathways of interest the authors use a panel of over 41 antibodies against metabolite transporters, metabolic enzymes, regulatory modifications such as protein phosphorylation, modifiers of mitochondrial dynamics, signaling molecules and transcription factors. The multiplexed quantification of these targets is compared with T cell subsets identified by lineage markers. The authors perform their measurements by multiplexed mass cytometry using antibodies conjugated to isotopes of heavy metals that provide mass-tag barcoding. They demonstrate concordance between their proxy metabolic measurements and metabolic/respiratory activity, although it is acknowledged that circumstances producing lack of concordance may nonetheless be envisaged. The mass-tagged antibodies are also used in multiplexed ion beam imaging studies with time-of-flight mass spectrometry to determine the distribution of cell subsets in tissue sections. Because the methodology is applicable to fixed cells, the authors are able to apply it to clinical samples of colorectal carcinoma. They show that $CD8^+$ T cells associated with tumors are enriched in cells expressing CD39 and PD1, markers of T cell exhaustion, yet the $CD39^+/PD1^+$ cells may be categorized into two populations with high and low metabolic activity respectively. Interestingly, the $CD39^+/PD1^+$ cells with low metabolic activity are not found within the tumor boundaries. These results suggest that study of metabolic activity might further illuminate the interaction between cytotoxic T cells and tumor cells and possibly provide improved clinical markers for stratification of response to immunotherapy.

MACROMOLECULAR SYNTHESIS & SYNTHETIC BIOLOGY

Vorobieva A A, White P, Liang B, Horne J E, Bera A K, Chow C M, Gerben S, Marx S, Kang A, Stiving A Q, Harvey S R, Marx D C, Khan G N, Fleming K G, Wysocki V H, Brockwell D J, Tamm L K, Radford S E, Baker D. De novo design of transmembrane β barrels. *Science* 371;2021:eabc8182.

Integral membrane β-barrel proteins are of particular technological interest because they form nanopores in membranes that have found uses including DNA sequencing, small-molecule sensing and water filtration. The selectivity of the pores is potentially customizable by choice of resides affecting the polypeptide backbone conformation and the amino acids exposed within the pore channel. Vorobieva *et al.* now show that β barrel proteins may be designed *de novo*, thereby increasing latitude for customization. Naturally occurring β barrel proteins are folded and inserted into membranes with the help of chaperones and other accessory proteins, but the authors now show that their designed proteins may be correctly folded and inserted into synthetic bilipid membranes without accessory proteins. In iterative design-build-test cycles the authors discover that to accomplish successful membrane insertion the β-turn sequences that connect successive β-strands in the barrel must be amenable to local destabilization during transition from the aqueous environment to the membrane environment so that they can translocate through the membrane during folding. Premature stabilization of β-turns results in the formation of undesired β-sheet structures. Within the membrane environment the turns are stabilized by the interaction between the strands that they connect. The authors successfully design transmembrane pores and show that NMR and crystal structures match closely to the designed features.

PROTEOMICS

He Y, Rashan E H, Linke V, Shishkova E, Hebert A S, Jochem A, Westphall M S, Pagliarini D J, Overmyer

K A, Coon J J. Multiomic single-shot technology for integrated proteome and lipidome analysis. *Analytical Chemistry* 93;2021:4217-4222.

Both proteomic and lipidomic analyses are conventionally performed by reverse-phase analyte separation coupled with on-line mass spectrometric identification. Remarkably, however, the two kinds of analysis have remained uncombined. He *et al.* take initial steps toward combination of proteomic and lipidomic analysis into a single analytical method. They identify conditions for liquid chromatography and mass spectrometry that best satisfy the joint requirements of proteome and lipidome. Sample preparation is initiated with a phase separation between a methanol/*t*-butyl ether mixture and water to separate organic-soluble lipids from aqueous-soluble proteins. The proteins are precipitated, then redissolved for digestion with lysyl endopeptidase and trypsin. Joint separation is performed on a Waters C18 reverse-phase BEH column (150 mm × 1.0 mm × 2.1 μm particle size). Lipid extract is injected first, then peptides. The combined analyte mixture is eluted in ammonium formate with a non-linear gradient from 0 to 100% of an isopropanol/acetonitrile mixture (90:10 v/v). Peptides elute predominantly earlier than lipids. The analytes are subjected to electrospray ionization. For peptides, the positive ion mode is used, but polarity switching is initiated once lipid elution predominates. The system is robust and yields identification of over 2,800 protein groups and 325 lipids from *Saccharomyces cerevisiae*. He *et al.* analyze a series of yeast strains featuring deletion of genes involved in lipid biosynthesis, showing that their integrated methodology for coanalysis of proteins and lipids reveals relationships between proteome and lipidome. Future improvements in depth/speed of analysis are expected, and deployment for sample-limited applications such as clinical analyses is anticipated.

FUNCTIONAL GENOMICS AND PROTEOMICS

Payne A C, Chiang Z D, Reginato P L, Mangiameli S M, Murray E M, Yao C-C, Markoulaki S, Earl A S, Labade A S, Jaenisch R, Church G M, Boyden E S, Buenrostro J D, Chen F. In situ genome sequencing resolves DNA sequence and structure in intact biologic samples. *Science* 371;2021:eaay3446.

Powerful new techniques have recently become available for investigation of 3-D genomic architecture on a single-cell basis. For example, single-cell Hi-C permits high-resolution analysis of chromatin architecture based on DNA sequencing, and single-cell multiplexed DNA fluorescence *in-situ* hybridization (FISH) permits high-resolution analysis of targeted sequence features based on imaging. Payne *et al.* now contribute a new technique to add to this family of methodologies. They combine *de novo* sequencing with spatial imaging of the same cells. They first use Tn5 transposase to randomly incorporate sequencing adaptors into fixed genomic DNA for creation of a genomic DNA library *in situ*. The DNA fragments are circularized by ligation of a DNA hairpin to each end. Hairpins contain a unique molecular identifier and sequencing primer sites. The resulting circular templates are subjected to rolling circle amplification *in situ* to create localized amplicons that constitute ∼ 400-500 nm features for *in situ* sequencing. The unique molecular identifiers are read by sequential rounds of sequencing by ligation with fluorescence imaging *in situ*. This step may be followed by immunostaining to localize cellular landmarks as required. The *in situ* amplicons are then dissociated, amplified by polymerase chain reaction (PCR) to produce an *in vitro* sequencing library, and subjected to paired-end sequencing on a conventional Illumina sequencer *ex situ*. The reads are tagged with the unique molecular identifier that specifies their *in situ* localization. It finally remains to match the *in situ* (unique molecular identifier) and *ex situ* sequence reads. The authors execute this protocol to analyze 106 cultured human fibroblasts and 113 cells in 57 intact early mouse embryos at various stages, localizing hundreds of thousands of DNA sequences in the individual cells. In early mouse embryos, they identify parent-specific changes in genome structure between embryonic stages, and provide evidence that epigenetic memory of chromosome positioning is transmitted during early cell divisions.

Xu J, Kudron M M, Victorsen A, Gao J, Ammouri H N, Navarro F C P, Gevirtzman L, Waterston R H, White K P, Reinke V, Gerstein M. To mock or not: a comprehensive comparison of mock IP and DNA input for ChIP-seq. *Nucleic Acids Research* 49;2021:e17-e17.

Chromatin immunoprecipitation (IP) followed by DNA sequencing (ChIP-seq) is widely employed for identification of transcription factor (TF) binding sites in the genome. The technique involves cross-linking DNA with adjacent proteins using formaldehyde, then shearing the DNA to small fragments by sonication. An antibody against the TF of interest is then used to precipitate TF-DNA complexes. The DNA is then sequenced to identify the TF binding sites. A bias resulting from the greater susceptibility to shearing of open conformation DNA is controlled by performing sequencing of an aliquot that has not undergone IP. Controlling for non-specific binding by the antibody can be accomplished by comparison with a

mock IP that employs either an antibody of irrelevant specificity or a form of the TF lacking the epitope tag that the antibody recognizes. However, this control is rarely performed because it generally yields much less DNA and the results are felt to be too noisy. Xu *et al.* perform systematic side-by-side comparisons of DNA input and mock IP controls and show that the DNA input control alone is inadequate to exclude spurious site identifications. They observe that the prevalence of spurious sites identified by mock IP is correlated with transcriptional activity and therefore with chromatin accessibility, presumably because larger numbers of different proteins bind to conformationally open regions. Moreover, complex samples such as whole tissues of organisms have more spurious sites than cell lines. The authors conclude that DNA input and mock IP controls provide complementary information and both kinds of control should be deployed in ChIP-seq studies.

MACROMOLECULAR CHARACTERIZATION

Sharma D, Zagore L L, Brister M M, Ye X, Crespo-Hernández C E, Licatalosi D D, Jankowsky E. The kinetic landscape of an RNA-binding protein in cells. *Nature* 591;2021:152-156.

The interaction between cellular RNAs and RNA binding proteins (RNPs) is conventionally investigated using a protocol in which RNA-RNP complexes are crosslinked by ultraviolet (UV) light, the RNP of interest is immunoprecipitated, and the bound RNA's are identified by high-throughput sequencing. The complexes identified in this way represent interactions at steady state. They do not provide information about the kinetics of complex formation within cells. Reaction rates have not been amenable to quantification using this methodology because the rate of crosslinking is slow compared to RNA association and dissociation rates. Sharma *et al.* now overcome this limitation by rapid crosslinking with a pulsed femtosecond UV laser. By altering laser power and protein concentration they control crosslinking time in a way that permits the progress of association and dissociation reactions to be followed. Rate constants can therefore be calculated. The authors validate this methodology for 2 purified RNPs by showing that RNA affinities calculated from the rate constants they measure *in vitro* reproduce affinities previously measured by fluorescence anisotropy. They then apply their kinetic methodology to investigate the interactions of an RNA binding protein in cells. They choose a protein called deleted in azoospermia-like (DAZL), an RNP that regulates mRNA stability, or translation, or both in oocyte maturation and spermatogenesis by binding to 3′-UTRs of mRNAs. The authors investigate DAZL binding kinetics in mouse GC-1 cells, a cell line similar to spermatocytes. Rate constants k_{on} and k_{off} and crosslinking rate constants are calculated for sites on thousands of individual mRNAs, each at different laser powers and DAZL concentrations (under control of a doxycycline-inducible promotor). The data indicate that DAZL occupancy by mRNA is low, turnover (binding/release) is rapid on the physiologic timescale, and DAZL may bind to multiple sites on an mRNA. Such characteristics help explain the physiologic function of this RBP. Kinetic analysis of RNA and RNP interactions is similarly expected to illuminate RNP function in other systems.

SEPARATIONS

Delano M, Walter T H, Lauber M A, Gilar M, Jung M C, Nguyen J M, Boissel C, Patel A V, Bates-Harrison A, Wyndham K D. Using hybrid organic–inorganic surface technology to mitigate analyte interactions with metal surfaces in UHPLC. *Analytical Chemistry* 93;2021:5773-5781.

There are numerous advantages to construction of high-performance liquid chromatography (HPLC) systems using stainless steel. However, these advantages are partially offset by interaction of stainless steel surfaces with some analytes, particularly analytes with multiple phosphate or carboxylate groups, which results in degradation of peak shape and losses of recovery. Stainless steel is also susceptible to corrosion, particularly with acidic mobile phases or mobile phases containing halides. Stainless steel may also release iron ions into the mobile phase that can interact with analytes. Delano *et al.* report development of an ethylene-bridged siloxane polymer modification for stainless steel surfaces to mitigate these effects. The chemical composition is related to that of ethylene-bridged hybrid (BEH) chromatographic particles. It is formed on metal surfaces using a vapor deposition process. The authors test its performance in ultra-HPLC (UHPLC) instruments and columns and show significantly improved results in separations of nucleotides, a phosphopeptide and an oligonucleotide. A protein (enolase) also shows less iron ion adduction by mass spectrometry. Their work serves as a reminder that the surfaces encountered by analytes may affect separation in ways that can be mitigated by appropriate chemical treatment.

Broman A, Lenshof A, Evander M, Happonen L, Ku A, Malmström J, Laurell T. Multinodal acoustic trapping enables high capacity and high throughput enrichment of extracellular vesicles and microparticles in miRNA and MS proteomics studies. *Analytical Chemistry* 93;2021:3929-3937.

Chen Y, Zhu Q, Cheng L, Wang Y, Li M, Yang Q, Hu L, Lou D, Li J, Dong X, Lee L P, Liu F. Exosome detection via the ultrafast-isolation system: EXODUS. ***Nature Methods*** **18;2021:212-218.**

Extracellular vesicles (EVs) are receiving increasing attention as mediators of cell signaling and as biomarkers whose composition indicates the functional state of their cells of origin. EVs have traditionally been purified by differential ultracentrifugation, but this methodology is laborious and often less effective than might be wished. One effective alternative is acoustic trapping. In this technique EVs are trapped between polystyrene seed particles focused within a microfluidic channel by a standing ultrasonic wave. Broman *et al.* here describe improved instrumentation for implementation of ultrasonic trapping. They substitute the single-node resonance of standard equipment with a larger channel that supports multinode resonance, increasing seed particle capacity 40× and throughput 25-40×. Standard microfluidic flow rates of <50 µL/min are increased to help process larger-volume samples. The authors purify EVs from urine and demonstrate processing of 1-3 mL samples in ~8-10 min. Another alternative to ultracentrifugation for purification of EVs is ultrafiltration. Chen *et al.* purify exosomes (a category of EV with diameter 30-150 nm) by passage through a fluid-filled channel bounded by two aluminum oxide membranes. By adjusting the pressure between outlets attached to the two membranes, one membrane is held at positive pressure, the other at negative pressure so that filtration occurs in the direction of the transverse pressure gradient. Exosomes are held within the fluid channel because the membrane pores are small enough to exclude them. Harmonic oscillators provide vibration within the channel, and the pressure gradient is cyclically reversed. These processes help resuspend the particles within the channel, and hence limit fouling and prevent aggregation. In this device a 10 mL urine sample is processed in ~10 min.

IMAGING

Asher W B, Geggier P, Holsey M D, Gilmore G T, Pati A K, Meszaros J, Terry D S, Mathiasen S, Kaliszewski M J, Mccauley M D, Govindaraju A, Zhou Z, Harikumar K G, Jaqaman K, Miller L J, Smith A W, Blanchard S C, Javitch J A. Single-molecule FRET imaging of GPCR dimers in living cells. ***Nature Methods*** **18;2021:397-405.**

The formation of homodimeric or heterodimeric complexes by G protein-coupled receptors (GPCRs) within biologic membranes has stimulated much interest in GPCR oligomerization. This interest is driven by the hope that oligomerization might be amenable to control for pharmaceutical purposes. In principle, single-molecule Förster resonance energy transfer (smFRET) offers a suitable modality for the study of such oligomerization. It provides quantitative information about molecular distances <10 nm between fluorescent probes for study of the dynamics of molecular interactions. However, it has seldom been used for interactions between integral membrane proteins in living cells because of various technical limitations. Asher *et al.* now incorporate several recent developments that overcome limitations in expressing, labeling and tracking transmembrane proteins by smFRET and deploy the methodology for study of GPCRs. Firstly, they regulate receptor density in CHO cells with a tetracycline-inducible promoter. For fluorescent labeling of the interacting species they then employ $SNAP_{fast}$ tags comprised of an enzyme that spontaneously forms covalent linkages with benzylguanine fluorophores at high efficiency. They choose self-healing fluorophores, Lumidyne 555p and 655, which provide photostability without the need for harsh chemical conditions. These fluorophores are also membrane-impermeant, which limits fluorescent background from the cell interior. Finally, to further reduce background they use total internal reflection microscopy to image the proteins diffusing within the plasma membrane. Using this methodology, the authors investigate representatives of class A, B and C GPCRs. They find that the µ-opioid receptor (class A) and the secretin receptor (class B) exist as monomers at low surface densities, but whereas the µ-opioid receptor remains monomeric at higher densities, the secretin receptor forms long-lived complexes. The metabotropic glutamate receptor 2 (class C), however, exists as a constitutive dimer in a density-independent fashion. The methodology is expected to be widely applicable for the study of dynamics of other membrane proteins in living cells.

CELL BIOLOGY

Barennes P, Quiniou V, Shugay M, Egorov E S, Davydov A N, Chudakov D M, Uddin I, Ismail M, Oakes T, Chain B, Eugster A, Kashofer K, Rainer P P, Darko S, Ransier A, Douek D C, Klatzmann D, Mariotti-Ferrandiz E. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. ***Nature Biotechnology*** **39;2021:236-245.**

This benchmarking study systematically compares the numerous methods for characterization of the T cell receptor (TCR) repertoire in circulating human $CD4^+$ T cells. The repertoire of specificities for different peptide antigens is generated by combinatorial pairing between TCR α and β chains; by combinatorial recruitment among the 47 V

segment genes and 61 J segment genes to make TCR α chains, and the 48 V segment genes, 2 D segment genes and 12 J segment genes to make TCR β chains; and by junctional diversity generated by random excision and addition of nucleotides at the V(D)J junctions. The TCR sequencing methods may be categorized as either genomic DNA- or RNA (*i.e.,* cDNA)-based. RNA-based methods use either multiplex PCR with panels of V and J primers or rapid amplification of complementary DNA ends (RACE)-PCR, which may incorporate unique molecular identifiers (UMIs) to limit PCR amplification bias and sequencing errors. The authors encounter substantial differences in accuracy and inter- and intramethod reproducibility among the 9 protocols they test. Some differences are anticipated, and some less so. Not surprisingly, a small input number of T cells limits detected diversity and rare clone detection. Most RACE methods capture less diversity in TCR α than TCR β, perhaps because the number of transcripts is 2-3× lower for TCR α. Methods based on genomic DNA show a low degree of overlap with those based on RNA, perhaps reflecting the larger copy number of RNA molecules than DNA. Multiplex PCR is poorly reproducible, suggesting the choice of multiplexing primers might bias amplification of some clones. And finally, non-UMI methods show greater sensitivity for detection of rare clones than UMI methods, perhaps because of the reads-per-UMI cut-off incorporated into the UMI methodology. TCR repertoire has inherent importance related to the effects of clinical conditions (*e.g.,* autoimmune disease, malignancy, infection) and interventions (*e.g.,* vaccination, immunotherapy). The strengths and weaknesses of the different methods are relevant to the choice of method. The findings of this study will also be of interest to all investigators studying gene expression in complex cellular assemblages.

DRUG DEVELOPMENT

Theodoris C V, Zhou P, Liu L, Zhang Y, Nishino T, Huang Y, Kostina A, Ranade S S, Gifford C A, Uspenskiy V, Malashicheva A, Ding S, Srivastava D. Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease. *Science* 371;2021:eabd0724.

Drug candidates are conventionally screened for cellular effects on just one or a very small number of disease-relevant parameters. The authors of the present study propose methodology by which effects of drug candidates on the entire ensemble of pathways dysregulated in a disease process are included in testing. They pilot their approach in a study of calcific aortic valve disease (CAVD), the most common valvular heart disease in developed countries and in the aging population. CAVD is characterized by calcification of the aortic valve leaflet. The authors follow up previous observations that haploinsufficiency of NOTCH1 can cause CAVD. The mechanism involves de-repression of osteoblast-like gene networks in cardiac valve cells. In the present work the authors derive induced pluripotent stem cells (iPSCs) from NOTCH1-haploinsufficient human endothelial cells and test 1595 small molecules for their effect on the expression of 119 genes representing a NOTCH1-dependent network by RNA-seq. Using a machine-learning approach they identify molecules that correct the network signature similar to $NOTCH^{+/-}$ endothelial cells to a $NOTCH^{+/+}$ signature. The authors identify 8 such molecules. The one with the strongest restorative effect on the network, XCT790, is an inverse agonist of estrogen-related receptor α (ERRα). The authors go on to show that XCT1790 reduces aortic valve thickness, calcification, and echocardiographic signs of aortic stenosis in a mouse model of NOTCH1 haploinsufficiency, and reduces the number of aortic valve cells expressing RUNX2, a transcription factor associated with osteoblast differentiation. Network-based drug screening that utilizes iPSCs and machine learning is expected to provide an effective strategy for design of therapies for models of other human diseases.