



PERSPECTIVE OPEN

A new precision medicine initiative at the dawn of exascale computing

Ruth Nussinov^{1,2}, Hyunbum Jang¹, Guy Nir^{3,4,7}, Chung-Jung Tsai¹ and Feixiong Cheng^{5,6}

Which signaling pathway and protein to select to mitigate the patient's expected drug resistance? The number of possibilities facing the physician is massive, and the drug combination should fit the patient status. Here, we briefly review current approaches and data and map an innovative patient-specific strategy to forecast drug resistance targets that centers on parallel (or redundant) proliferation pathways in specialized cells. It considers the availability of each protein in each pathway in the specific cell, its activating mutations, and the chromatin accessibility of its encoding gene. The construction of the resulting Proliferation Pathway Network Atlas will harness the emerging exascale computing and advanced artificial intelligence (AI) methods for therapeutic development. Merging the resulting set of targets, pathways, and proteins, with current strategies will augment the choice for the attending physicians to thwart resistance.

Signal Transduction and Targeted Therapy (2021)6:3

; <https://doi.org/10.1038/s41392-020-00420-3>

INTRODUCTION

Precision medicine aims to identify patient-specific drug targets.^{1–5} To date, the approaches have largely focused on (i) identification of proteins with driver mutations; the mutations can be strong drivers, weak drivers, rare drivers, or latent drivers;^{2,5–11} (ii) decision on how to target: should the mutant protein be targeted with a combination of drugs, e.g. one orthosteric and the other allosteric, or should a second protein also be targeted, in which case should the second protein be from the same pathway (the more frequent case) or from a different pathway.¹² If from a different (redundant, or parallel) pathway, the protein is generally selected based on the physician's prior knowledge; and finally, once identified, (iii) selection of drugs targeting these. Drug discovery can be via large-scale screening, in silico docking, structure-based drug design, or drug repurposing.^{13–18} The approaches rely on vast quantities of data, high-resolution structural data, highly efficient state-of-the-art algorithms to sift through these and large-scale scientific computation.¹ They are mapped in the diagram in Fig. 1. These approaches empowered significant progress since the launch of the precision medicine initiative, with breakthrough discoveries identifying activating mutations in key oncogenic proteins and their isoforms, their patterns and mechanisms.¹⁹ However, the complexity and challenge of identifying drug resistance targets call for broadening the current strategies and marshalling new ones from a different standpoint.

The framework outlined here maps a conceptually innovative complementary Precision Medicine Initiative that embraces the principles of cell proliferation to counter drug resistance. Different than current strategies it proposes (i) to identify or predict all

potential *proliferation pathways* in the cell. Proliferation pathways constitute the 'stockpile' of drug resistance pathways; (ii) to investigate the chromatin accessibility of genes encoding each protein in each proliferation pathway in the specific cell to confirm pathway availability in drug resistance, as well as their cell-specific expression levels; (iii) to identify their driver mutations, and the mechanisms of pathway activation; and finally, exploiting artificial intelligence (AI) methodologies, to (iv) integrate multi-omics cancer data and networks' perturbations for therapeutic development. Thus, it aims to stop cell proliferation by identifying all possible proliferation pathways and *predicting which gene is likely to become the next driver* of cancer in the specific patient cell. This innovative and comprehensive strategy is computationally intensive and proposes to exploit the emerging exascale computing in the last two steps. We expect that a Proliferation Pathway Atlas incorporating such data would be an invaluable resource to the community.

PROLIFERATION PATHWAYS ARE CRITICAL IN CANCER

Proliferation pathways link the cellular environment to the cell cycle. Genetic (mutations) and epigenetic alterations can lead to pathway hyperactivation, fueling cancer progression, as does inactivation of tumor suppressors. Protein-protein interactions of major cancer drivers are enriched in mutations, hijacking proliferative signaling networks.

Pathways crosstalk. Crosstalk emerges due to shared interactions and elements.^{20,21} It can influence their expression level and

¹Computational Structural Biology Section, Frederick National Laboratory for Cancer Research in the Laboratory of Cancer Immunometabolism, National Cancer Institute, Frederick, MD 21702, USA; ²Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; ³Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; ⁴Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA; ⁵Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106, USA and ⁶Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

Correspondence: Ruth Nussinov (NussinoR@mail.nih.gov)

⁷Present address: Department of Biochemistry & Molecular Biology, Department of Neuroscience, Cell Biology and Anatomy, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555, USA

Received: 14 August 2020 Revised: 27 October 2020 Accepted: 30 October 2020

Published online: 06 January 2021

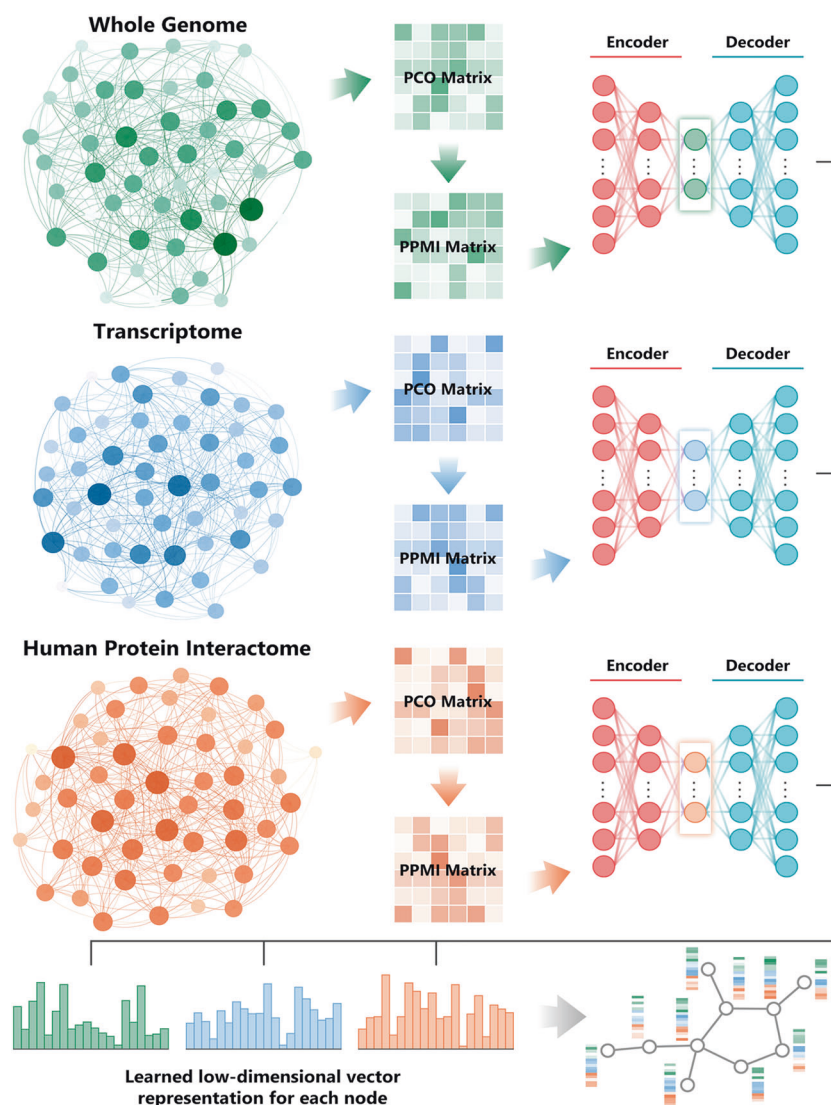


Fig. 1 A workflow illustrating an artificial intelligence-based, exascale computing framework for integrative analysis of multi-omics data for precision cancer medicine. PCO, probabilistic co-occurrence; PPMI, positive pointwise mutual information. The detailed description about a deep neural network-based multi-omics and network integration can be found in a recent study⁸¹

function. Under physiological conditions, crosstalk enables cells to cope with perturbations of homeostasis. In drug resistance, inhibition of a signaling pathway can promote activation of a survival pathway that bypasses the inhibited pathway. Insight into connections between signaling pathways and foresight into their distinct activation can be powerful in the treatment of cancer.²²

What distinguishes proliferation pathways from other signaling pathways?

The Ras/phosphoinositide 3-kinase (PI3K)/Akt and Ras/extracellular signal-regulated kinase (ERK) pathways provide good examples of proliferation pathways.^{23,24} Ras is activated by stimulated receptor tyrosine kinases (RTKs). Ras mutants are involved in roughly a third of the cancers. The identity of the other two-third proliferation pathways is only partially known.

Proliferation involves cell growth and division. Proliferation can take place through many pathways and is particularly active during development. It is also essential in adult homeostasis. Signaling pathways that control cell proliferation²⁵ can act by linking the cellular environment to progression through the G1 (Gap 1) phase of the cell cycle (Fig. 2). Progression through G1 is controlled by retinoblastoma protein (pRb) whose

phosphorylation by the G1 cyclin-dependent kinases (CDKs) promote passage of the cell cycle to the S (Synthesis) phase. The pRb pathway (thus G1) is mainly regulated by cyclins and CDK inhibitors with inputs from major cellular signaling pathways. pRb tumor suppressor binds to the E2F1 transcription factor (TF), repressing the G1/S transition; phosphorylation of pRb proteins by CDKs liberates E2F, promoting the transition to S phase.

Criteria for identifying proliferation pathways

A pathway that promotes cell proliferation (i) can lead to activation of TFs that induce expression of proteins acting in multiple pathways, including oncogenic functions such as proliferation and survival, with some of these (ii) entering the cell cycle. Cyclin-D, whose synthesis is initiated during the cell cycle G1 phase and is involved in regulating cell cycle progression provides an example (Fig. 2). The cyclin-D/CDK4 complex, which consists of cyclin-D and CDK4, or CDK6, a serine-threonine kinase, is essential for the progression of the cell from the G1 to the S phase, for the Start or G1/S checkpoint. Some proteins control operations critical for cell cycle progression. Cyclin-D transcription is activated through the growth factor-stimulated RTK proliferation pathway which expresses Myc, a TF that controls transcription

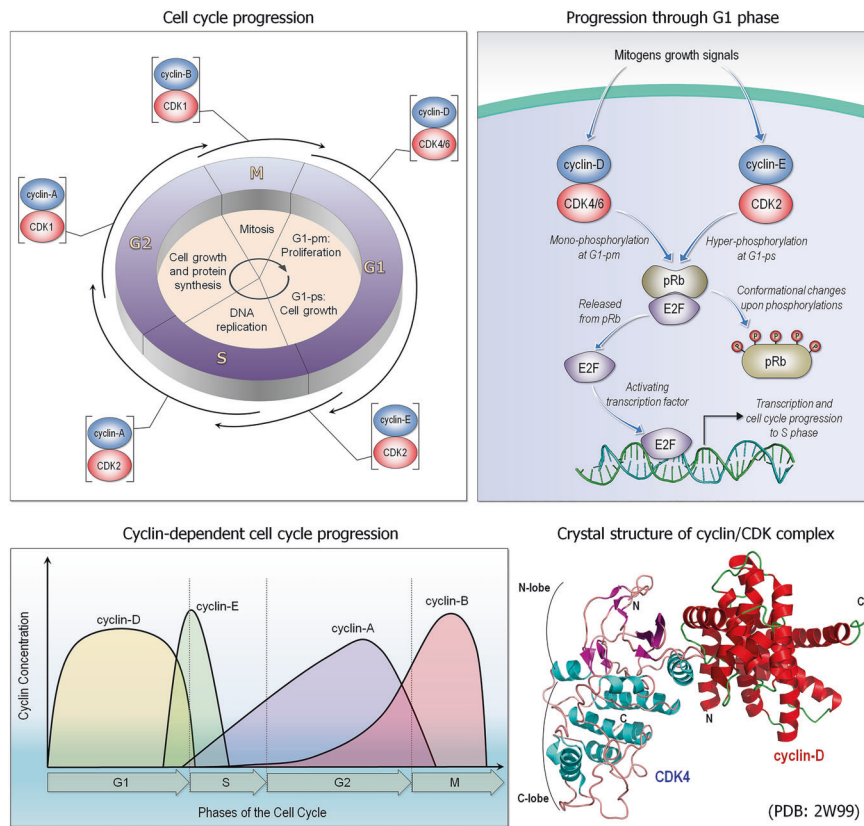


Fig. 2 Cyclin-dependent kinases (CDKs) cell cycle. CDKs and their cyclin partners form complexes, regulating the progression through the cell cycle (upper left panel). In early G1 phase (G1-pm), retinoblastoma protein (pRb) becomes phosphorylated by the complex, a pair of cyclin-D and CDK4/6 (upper right panel). In late G1 phase (G1-ps), pRb is hyper-phosphorylated by cyclin-E/CDK2 complex, undergoing a large conformational change. This conformational change fails to assemble with E2F, a transcription factor, which promotes to progress the G1/S transition. Cyclins bind with the dependent kinases and their concentration varies during the cell cycle (lower left panel). A crystal structure illustrates an example of cyclin/CDK complex (PDB: 2W99) (lower right panel). Apo-CDKs exhibit little kinase activity; CDKs become active kinases when interacting with the regulatory protein called a cyclin

of several cell cycle-regulating genes, including cyclin-D. Myc promotes the cell cycle primarily through its role in cellular growth control. c-Myc target genes include regulators of cell growth; but also, those functioning in cell division pathways. Among c-Myc target genes that regulate cell growth are those associated with ribosomal protein transcription and translation, including translation initiation factors such as eukaryotic translation initiation factor 4E (eIF4E). Active RTK signals through the two major signaling pathways; c-Myc is involved in both. Notably, a proliferation pathway, such as MAPK can also activate gene sets for immune response.²⁶

EXAMPLES OF PROLIFERATION PATHWAYS

Wnt/ β -catenin, Notch, Hedgehog, transforming growth factor β (TGF- β) and Hippo are implicated in developmental processes and proliferation. Janus kinase/signal transducer and activator of transcription (JAK/STAT) is an example of a proliferation pathway through a cytokine receptor (IL7). Here we focus on development-related pathways and discuss the first three. Embryogenesis and tumorigenesis share coordinated mechanisms of proliferation, differentiation, and migration.²⁷

Wnt/ β -catenin signaling

The Wnt signaling cascade is a main regulator of development, controlling the growth of embryonic stem cells and adult cell specialization (Fig. 3). The pathway is also frequently active in cancer.²⁸ Wnt growth factors alter gene expression by stimulating different classes of receptors. They lead to cell proliferation

through their impact on the cell cycle.²⁹ Wnt pathway components, such as β -catenin, Dishevelled (Dsh, Dvl in mammals), Frizzled (Frz, a Wnt receptor), low-density lipoprotein receptor-related protein 6 (LRP6, a Wnt co-receptor), and Axin have been associated with cell cycle regulation, centrosome biology, and cell division. Several Wnt pathway components play essential roles during mitosis, which is proposed to also regulate Wnt signaling via cyclin-Y/CDK14 phosphorylation of LRP6.³⁰ They also control cell morphogenesis, affecting the cytoskeleton and the mitotic spindle. Wnt-stimulated signaling activates β -catenin which interacts with DNA-bound TFs of the T-cell factor (TCF) family. β -catenin switches inactive TCF into a transcriptional activator of its target genes.³¹ Chromatin remodeling complexes can bind β -catenin and promote transcriptional activation of TCF-responsive reporter genes. Transcriptional co-activators, such as p300 and cAMP-response element binding protein (CREB) can alter chromatin structure through histone acetyltransferase to stimulate transcriptional activity. In the absence of a Wnt signal, β -catenin is degraded by a complex which includes the Axin scaffold protein, glycogen synthase kinase 3 β (GSK3 β), and adenomatous polyposis coli (APC). TCF is bound to the Groucho repressor; binding of Wnt to its receptors induces dissociation of the complex. β -catenin binds TCF in the nucleus.

Hedgehog signaling

Hedgehog communicates between cells. It is important for organ development, regeneration and homeostasis; it is frequently modulated in cancer.²⁷ It cross-talks with e.g. transforming growth factor β (TGF β s), Wnt, Notch, and the Sonic hedgehog (Shh). The

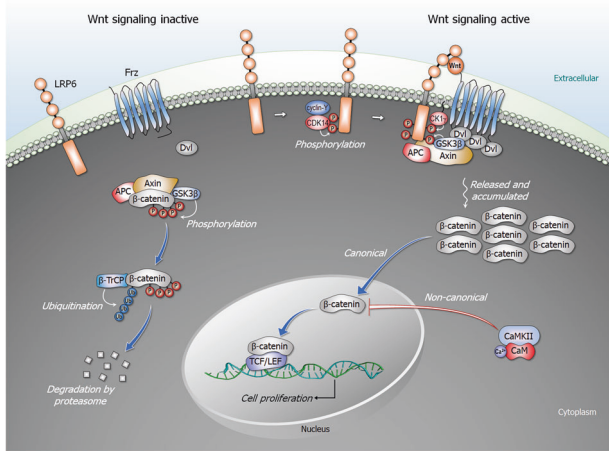


Fig. 3 The Wnt signaling pathways. In the absence of Wnt signal, the destruction complex constituted by Axin, adenomatosis polyposis coli (APC), and glycogen synthase kinase 3 β (GSK3 β) leads to phosphorylation of β -catenin by GSK3 β in the cytoplasm. The phosphorylated β -catenin is subsequently ubiquitinated by β -TrCP and targeted for proteasomal degradation. The canonical Wnt/ β -catenin pathway is activated by upon binding of Wnt to its Frizzled (Frz) receptor and low-density lipoprotein receptor-related protein 6 (LRP6) co-receptor. LRP6 phosphorylated by cyclin-Y/CDK14 complex, GSK3 β , and casein kinase 1 γ (CK1 γ) recruits the destruction complex and Dishevelled (Dsh, Dvl) in mammals) to the plasma membrane. Dvl is activated through polymerization, inhibiting the destruction complex. This results in accumulation of unphosphorylated β -catenin in the cytoplasm and its subsequent translocation into the nucleus, leading to cell proliferation. In the non-canonical Wnt/calcium pathway, calmodulin-bound CaMKII (calcium/calmodulin-dependent protein kinase II) negatively regulates the canonical β -catenin/TCF/LEF signaling through phosphorylation of TCF, which inhibits β -catenin/TCF-mediated transcription

Shh pathway can involve canonical or non-canonical signaling. The first is receptor ligand-dependent when Shh binds to Ptch (a 12-transmembrane protein) at the membrane; the second is through downstream smoothed (Smo).³² Smo regulates Gli transcription factors processing and activation, which activate target genes. Non-canonical activation is Gli-independent. Hedgehog signaling upregulates multiple proteins, including N-Myc (a member of the Myc family), forkhead box M1 (FoxM1), and Cdc25B, which activates the cyclin-dependent kinase CDC2. It also upregulates *CCND1*, *CCND2*, and *CCNE*. Cyclin-D1, cyclin-D2, and cyclin-E which drive cell-cycle progression at the G1/S phase, while FoxM1, cyclin-B1, and Cdc25B act at the G2/M (mitotic) phase. Thus, hedgehog signals drive cell-cycle progression through multiple cell cycle regulators.

Notch signaling

Notch signaling takes place via cell-cell communication, where transmembrane ligands on one cell activate those of the other. The cleaved receptor is translocated to the nucleus.³³ Notch intracellular domain (NICD) forms a trimeric complex with CSL (CBF1, Suppressor of Hairless, Lag-1; a transcription factor that activates genes downstream in the Notch pathway) and Mastermind-like (MAML) transcriptional coactivator, which converts CSL from a repressor to an activator and initiates transcription of Notch downstream target genes. In the absence of Notch signaling, CSL represses transcription; following activation by Notch, it is converted into a transcriptional activator and activates transcription of the same genes. Notch signaling with its CSL cofactor can maintain cells in an undifferentiated state, consequently associated with cancer. It controls cell lineage and tissue development, blocking differentiation thus retaining stem

or progenitor cells, or governing the balance between cell fates. Notch signaling mediates G1/S cell-cycle progression in T-cells via cyclin-D3 and its dependent kinases and activates cell cycle reentry and progression in quiescent cardiomyocytes. Notch signaling acts before cell division to promote asymmetric cleavage and cell fate of neural precursor cells; its activation can inhibit proliferation of endothelial cells by delaying cyclin-D/CDK4-mediated phosphorylation of the retinoblastoma protein. It also regulates variant cell cycles to control cell size³⁴ and more.

EGFR signaling

Epidermal growth factor receptor (EGFR) pathway is a classic example of proliferation pathway that can lead to G1 cell cycle progression, through cyclin-D expression, CDK4/6 activation, and the repression of cyclin-dependent kinase inhibitor proteins (CDKi) by EGFR signaling pathways.

Selection of the proliferation pathway to drug

Halting proliferation by drugging the pathway most likely to become the next driver in the patient cell is a powerful and compelling amplification of current therapeutic approaches. It considers cancer evolution dynamics which to date has been missing. The challenge is however in the knowledge of (i) all possible proliferation pathways, (ii) the accessibility of each gene encoding a protein in each pathway in the specific cancer cell, including (iii) expression data, and (iv) the driver mutations in each gene.

GENES OF TARGETED PATHWAYS SHOULD BE ACCESSIBLE

To be a good drug candidate, the proteins in the proliferation pathway should be available in the specific cell. This requires that the genes encoding the pathway proteins are accessible to the transcription machinery or can become accessible upon a 'modest' change in the chromatin structure. Not all proteins are expressed in all cells. Chromatin availability status is cell type, lineage and state-dependent.^{12,35} Genes active in developmental or embryonic pathways can become densely packed in the chromatin and inaccessible. Further, because signaling in a skin cell differs from that in a kidney cell, proliferation pathways in drug resistance are likely to differ between these cells.³⁶ Oncogenic cells manifest tissue-specific tendencies,^{36,37} with distinct cells having preferred proliferation profiles. Accessibility is controlled by cell-specific chromatin-binding factors,^{38,39} including e.g. pioneer transcription factors that locally unfold the condensed chromatin and nucleosomes. Accessibility can also be regulated by the proliferation pathway itself, as in the case of Notch³⁵ and its epigenetics.⁴⁰

Experimental accessibility data are limited. Predicting the three-dimensional genome organization and chromatin accessibility is also challenging. High-resolution structural data provide structural detail, allow mapping of genomes, insight into effects of mutations and dysregulation that traditional methods that identify the genes with active histone modification markers, such as H3K27ac, H3K4ac3 are unable to provide. Simulations with parameterization based on the free-energy landscape theory,^{41,42} genomics and epigenomics data, reproduced chromosome conformation capture data (Hi-C)^{12,43-47} and super-resolution microscopy.^{42,48,49} They permitted predicting chromatin structures at 5 kilobase resolution starting from genomics and epigenomics data that are available for hundreds of cell types, including cancer cells.⁴² Integration of Hi-C data with conventional microscopy led to more accurate prediction of genome organization.⁵⁰ More recently, Hi-C data and super-resolution imaging were brought together through integrative modeling of genomic regions (IMGR), thus achieving high spatial and genomic resolution, while maintaining the single-cell identity.⁵¹⁻⁵³ IMGR can be broadly divided into three steps. In step one, models are constructed of Hi-

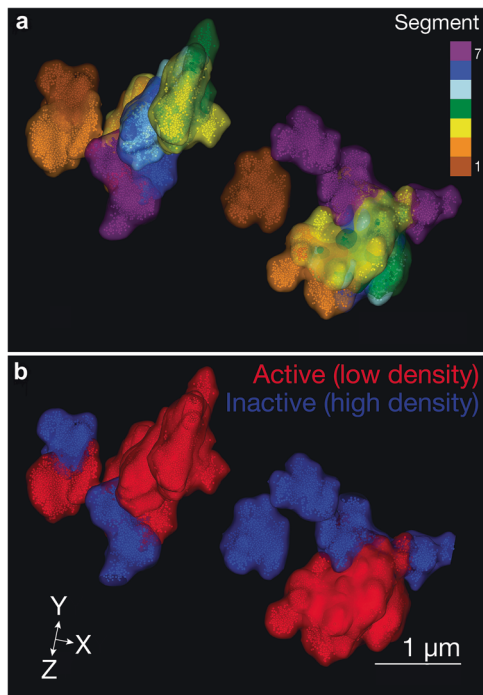


Fig. 4 OligoSTORM image of chromatin density. (a) Seven chromosomal segments (chr19:8.68–15.2; hg19) rendered as isosurface clusters in a PGP1 fibroblast. Each color represents a different chromosomal segment. (b) The seven segments color-coded according to their transcriptional activity and epigenetics marks.⁵¹ (blue—inactive, red—active). Here, the density of the inactive is higher than the active chromatin

C data.^{54,55} In step two, these models are rigidly fitted onto structures resolved by super-resolution microscopy. The top 5% that fit the most qualify to the next step, which is the flexible fitting. In flexible fitting, the polymer chains are allowed to swivel around TAD borders, which are expected to be more flexible. The model that best fits each super-resolved structure is chosen. Such a technology promotes optimism that a precision level that unearths the chromatin status of driver genes is reachable; *genes with sparse chromatin density would suggest that they are drug resistance candidates*. Integrative successes promise increasingly detailed mapping of dynamic chromatin maps of single cells.

IMGR is especially beneficial when integrating with images (Fig. 4). Here, we focus on 7 chromosomal segments out of the 9 imaged using sequential OligoSTORM, and color-code them as either active (red) or inactive (blue).⁵¹ OligoSTORM^{52,56} is the integration of Oligopaints Fluorescence in situ Hybridization (FISH) probes,⁵⁷ with the super-resolution technology called Stochastic Optical Reconstruction Microscopy (STORM).⁵⁸ Sequential OligoSTORM^{51,53,59} allows imaging of multiple genomic loci, going much beyond the limitations of spectral resolution (Fig. 4A). Even though these chromosomal segments were imaged at the ~ Mb scale, with IMGR, their genomic resolution can improve to 10 kb and better, which is two orders of magnitude higher for some of these segments. Interestingly, the density of the inactive chromatin in this PGP1f (Personal Genome Project, participant 1 fibroblasts) nucleus is higher than that of the active chromatin (Fig. 4B). Drugs may find active, cell-type specific, chromatin target more efficiently. OligoSTORM gene-specific visualizing technologies, or IMGR, can learn whether gene accessibility is influential in successful drug therapy. The efficiency of drug therapy might also be dependent upon the structural variation between homologous chromosomes.⁵¹

IDENTIFYING DRIVER MUTATIONS WITH EXASCALE COMPUTING

Proliferation pathways are activated by driver mutations. Their identification involves algorithmic strategies, statistical evaluation and databases.^{2,19,60–64} Since the methods are statistics-based, the mutations are mostly identified based on their frequencies of occurrence. Recently, however, an increasing number of statistically rare mutations were identified in patients, raising the question of how to identify rare, and weak drivers which are often observed only in certain tissues thus overall infrequent.^{5,19,62,65–72} K-Ras4B^{A146T} is one example where the mechanism is understood. Different than K-Ras4B^{G12D}, a strong driver that blocks GTP hydrolysis and is expressed in many cancers, including pancreatic and colon, the weaker K-Ras4B^{A146T} which acts by promoting guanine nucleotide exchange factor (GEF)-mediated GDP by GTP exchange, transforms colon but is not sufficiently powerful in transforming pancreatic cancer cells.⁷³ “Latent” mutations, that need an emerging ‘helper’ mutation with additive effects for observable pathological consequences are especially challenging to identify. Mechanistically, whether frequent or rare, mutations that release autoinhibition are often driver mutations,⁶³ clusters of mutations also tend to contain drivers, including rare, and latent.^{5,19,62} Identification of driver mutations, including weak, rare and latent, in each protein in all proliferation pathways requires immense computational power. These mutations are determined not based on their statistics, but by their ability to shift the protein conformation from an inactive to the active state. Identification of each mutation in each protein necessitates powerful computing to observe whether it executes this shift, expressed by conformational change. Such computing power is forecast to reach to scientific community. Exascale computing systems are capable of a billion (i.e. a quintillion) calculations per second. This scale permits executing such long timescales explicit solvent simulations which are required to capture the redistributions of the ensembles. These indicate the population time of conformations where the mutation switches the protein from the inactive to the active state. Figure 5 illustrates why massive compute time is necessary.

ARTIFICIAL INTELLIGENCE, MULTI-OMICS DATA, AND NETWORK PERTURBATIONS FOR THERAPEUTIC DEVELOPMENT

The human genome project accelerated genetic and genomic studies such as The Cancer Genome Atlas (TCGA) to inform precision medicine drug discovery.¹ The underlying hypothesis of cancer systems biology is that sub-cellular networks gradually rewire throughout disease initiation, progression, and maintenance, leading to progressive shifts of local and global network properties and systems states,⁷⁴ including protein-protein interactions and gene regulatory network, all controlling cancer initiation and drug responses (Fig. 5). Genome alterations, amplification, deletion, translocation, and mutations can only be selected for in cells if they encode changes, or perturbations, in the human interactome and systems properties of the affected cells.^{75,76} Personalized treatment needs to be designed to deal with such perturbations; rather than only with genomic events. Analysis of over 2.5 million nonsynonymous somatic mutations derived from 6,789 tumor exomes across 14 cancer types from TCGA, showed that Individualized Network-based Co-Mutation (INCM)-inferred putative genetic interactions are correlated with patient survival and drug responses in cancer cell lines.⁷⁵ Drug-target network analysis revealed candidate therapeutic pathways that target tumor vulnerabilities and identified several potential pharmacogenomics biomarkers. A Genome-wide Positioning Systems network (GPSnet) algorithm incorporated individual patient’s DNA and RNA profiles into the human protein-protein interactome network to prioritize targets and repurposed drugs

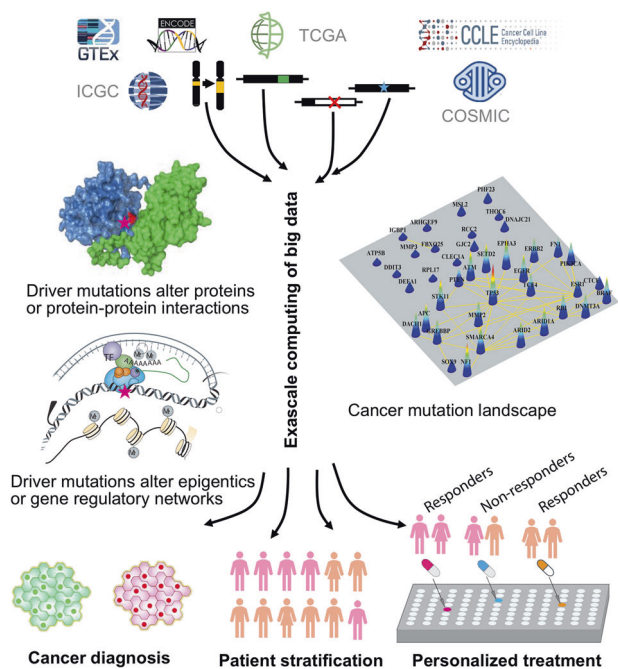


Fig. 5 A proposed atlas and its associated portal for precision cancer medicine using exascale computing of “big” data

for cancer.¹⁴ A GPSnet-predicted and experimentally validated drug, ouabain, revealed potential antitumor activities in lung adenocarcinoma by uniquely targeting a HIF1 α /LEO1-mediated cell metabolism pathway.¹⁴

The human interactome networks already contributed to understanding tumorigenesis and rapid identification of driver genes in human cancer and drug treatment.^{1,15,77,78} Cancer networks, and broadly sub-cellular systems, require information and models at multi-dimensional levels, including cells, tissues, organs, and organisms, which are missing in traditional computational approaches. Cancer therapy is moving from drug-centered to patient-centered approach. This requires paradigm shifts along the entire drug development process and multi-omics data integration. The increase in data (including DNA/RNA sequencing data) and the difficulty of data analysis, will also be aided by exascale computing. Advances in AI have been applied to cancer medicine, particularly in large-scale, integrative analyses of multi-omics and biological networks. Still, development and application of AI methods in precision medicine are still in its infancy.

Cancer data come from high-dimensional sources, electronic health care records (imaging, laboratory results, diagnosis codes), genetic testing, among others. An oncologist has to evaluate vast amounts of information, including the patient’s history, family history, genomic sequences, medications, and more, to guide rapid clinical decision. Among the multiple AI techniques, deep neural networks have gained attention in precision cancer medicine, especially for imaging data analysis^{79,80} and complex biological network integration.^{16,81} Saltz et al. presented convolutional neural network (CNN) models to analyze 5,200 digital images from 13 cancer types.⁷⁹ They demonstrated that tumor-infiltrating lymphocyte maps identified by CNN models were correlated with patient survival, tumor types, and immune profiles.⁷⁹ A one-class logistic regression (OCLR) machine-learning algorithm incorporated transcriptomic and epigenetic profiles from cancer patients for assessing the degree of oncogenic dedifferentiation.⁸² OCLR identified previously undiscovered biological mechanisms associated with the dedifferentiated oncogenic state quantified by stemness indices, a key measurement of cancer progression.⁸² Indices predicted by OCLR

revealed novel targets and possible targeted therapies by specifically targeting tumor differentiation.⁸²

AI approaches excel at automatically recognizing complex patterns in multi-omics data and providing quantitative assessment of genetic regions, omic layers, and pathways associated with tumorigenesis and precision medicine drug discovery (Fig. 1). deepDTnet, a network-based deep learning methodology was developed for novel target identification and drug repurposing via a heterogeneous drug-gene-disease network embedding 15 types of chemical, genomic, phenotypic, and cellular network profiles.⁸¹ DCell, a visible neural network embedded in the hierarchical structure of 2526 subsystems comprising a eukaryotic cell,⁸³ showed consistent results with laboratory observations when evaluated on several million genotypes.⁸³ Its framework may be applied to tumor cells although they are highly complex systems with millions of components and interactions. An AI-based, exascale computing framework that incorporates genome/transcriptome/proteome data, human protein-protein interactome, public drug-target databases (Fig. 1), along with functional validation or patient data validation offers powerful tools for accelerating precision cancer medicine.

Coupled with identification of mutations, enabled by powerful exascale computing at the single protein level, can create a comprehensive and rounded computational framework, whose organization will integrate all components.

CONCLUSIONS: STOP CELL PROLIFERATION

The potential of precision medicine to sustain human health has captivated the imagination of the scientists and the public. The National Cancer Institute described precision medicine as “an approach to patient care that allows doctors to select treatments that are most likely to help patients based on a genetic understanding of their disease”. However, *exactly how to select* has been unclear. The number of possibilities is massive, and the drug combination should fit the patient status. Significant progress has been made since the launch of the precision medicine initiative. However, to date its success has been limited. A major reason is the emergence of drug resistance.

Here we map a new concept: stopping cancer cell proliferation by targeting the proliferation pathway and genes that are likely to be the next drivers in the expected emergence of drug resistance. Current technologies, which can already obtain gene-scale resolution of chromatin increasingly allow forecasting such set of drug resistance targets through identification of proliferation pathways and the accessible genes encoding them. While here we focus on proliferation pathways, for completeness, in the future survival pathways and others critical in drug resistance should also be included.

In the biological sciences, exascale computing in the next decade is expected to be dominated by hybrid modeling, molecular dynamics, free-energy simulations, drug design, and discovery, and modeling the behavior of molecular assemblies and cell actions exploiting imaging at different scales. The concept described here fits well into these capabilities aiming to arrest cell proliferation in drug resistance.

ACKNOWLEDGEMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. This research was supported [in part] by the Intramural Research Program of NIH, National Cancer Institute, Center for Cancer Research.

ADDITIONAL INFORMATION

Competing interests: The authors declare no competing interests.

REFERENCES

- Cheng, F., Liang, H., Butte, A. J., Eng, C. & Nussinov, R. Personal mutanomes meet modern oncology drug discovery and precision health. *Pharm. Rev.* **71**, 1–19 (2019).
- Nussinov, R., Jang, H., Tsai, C. J. & Cheng, F. Review: precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLoS Comput. Biol.* **15**, e1006658 (2019).
- Manem, V. S. K., Salgado, R., Aftimos, P., Sotiriou, C. & Haibe-Kains, B. Network science in clinical trials: a patient-centered approach. *Semin Cancer Biol.* **52**, 135–150 (2018).
- Dugger, S. A., Platt, A. & Goldstein, D. B. Drug development in the era of precision medicine. *Nat. Rev. Drug Discov.* **17**, 183–196 (2018).
- Nussinov, R., Jang, H., Tsai, C. J. & Cheng, F. Precision medicine review: rare driver mutations and their biophysical classification. *Biophys. Rev.* **11**, 5–19 (2019).
- Nussinov, R. & Tsai, C. J. 'Latent drivers' expand the cancer mutational landscape. *Curr. Opin. Struct. Biol.* **32**, 25–32 (2015).
- Carter, H. et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res.* **69**, 6660–6667 (2009).
- Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
- Nussinov, R., Jang, H. & Tsai, C. J. The structural basis for cancer treatment decisions. *Oncotarget* **5**, 7285–7302 (2014).
- Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl Acad. Sci. USA* **113**, 14330–14335 (2016).
- Dimitrakopoulos, C. M. & Beerenwinkel, N. Computational approaches for the identification of cancer genes and pathways. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, e1364 (2017).
- Nussinov, R., Tsai, C. J. & Jang, H. Are parallel proliferation pathways redundant? *Trends Biochem. Sci.* **45**, 554–563 (2020).
- Zeng, X. et al. Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics* **36**, 2805–2812 (2020).
- Cheng, F. et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat. Commun.* **10**, 3476 (2019).
- Huang, Y. et al. A systems pharmacology approach uncovers wogonoside as an angiogenesis inhibitor of triple-negative breast cancer by targeting Hedgehog signaling. *Cell Chem. Biol.* **26**, 1143–1158 (2019). e1146.
- Zeng, X. et al. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* **35**, 5191–5198 (2019).
- Cheng, F. In silico oncology drug repositioning and polypharmacology. *Methods Mol. Biol.* **1878**, 243–261 (2019).
- Cheng, F. et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* **9**, 2691 (2018).
- Gao, J. et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* **9**, 4 (2017).
- Rowland, M. A., Fontana, W. & Deeds, E. J. Crosstalk and competition in signaling networks. *Biophys. J.* **103**, 2389–2398 (2012).
- Adelaja, A. & Hoffmann, A. Signaling crosstalk mechanisms that may fine-tune pathogen-responsive NFκB. *Front Immunol.* **10**, 433 (2019).
- Prahallad, A. & Bernards, R. Opportunities and challenges provided by crosstalk between signalling pathways in cancer. *Oncogene* **35**, 1073–1079 (2016).
- Nussinov, R., Tsai, C. J. & Mattos, C. 'Pathway drug cocktail': targeting Ras signaling based on structural pathways. *Trends Mol. Med.* **19**, 695–704 (2013).
- Fernandes, M. S., Sanches, J. M. & Seruca, R. Targeting the PI3K signalling as a therapeutic strategy in colorectal cancer. *Adv. Exp. Med Biol.* **1110**, 35–53 (2018).
- Duronio, R. J. & Xiong, Y. Signaling pathways that control cell proliferation. *Cold Spring Harb. Perspect. Biol.* **5**, a008904 (2013).
- Wei, X. et al. The evolutionarily conserved MAPK/Erk signaling promotes ancestral T-cell immunity in fish via c-Myc-mediated glycolysis. *J. Biol. Chem.* **295**, 3000–3016 (2020).
- Carballo, G. B., Honorato, J. R., de Lopes, G. P. F. & Spohr, T. A highlight on sonic hedgehog pathway. *Cell Commun. Signal* **16**, 11 (2018).
- Nusse, R. & Clevers, H. Wnt/β-catenin signaling, disease, and emerging therapeutic modalities. *Cell* **169**, 985–999 (2017).
- Bryja, V., Cervenka, I. & Cajanek, L. The connections of Wnt pathway components with cell cycle and centrosome: side effects or a hidden logic? *Crit. Rev. Biochem Mol. Biol.* **52**, 614–637 (2017).
- Davidson, G. The cell cycle and Wnt. *Cell Cycle* **9**, 1667–1668 (2010).
- Franz, A., Shlyueva, D., Brunner, E., Stark, A. & Basler, K. Probing the canonicity of the Wnt/Wingless signaling pathway. *PLoS Genet.* **13**, e1006700 (2017).
- Blotta, S. et al. Canonical and noncanonical Hedgehog pathway in the pathogenesis of multiple myeloma. *Blood* **120**, 5002–5013 (2012).
- Siebel, C. & Lendahl, U. Notch signaling in development, tissue homeostasis, and disease. *Physiol. Rev.* **97**, 1235–1294 (2017).
- Von Stetina, J. R., Frawley, L. E., Unhavaithaya, Y. & Orr-Weaver, T. L. Variant cell cycles regulated by Notch signaling control cell size and ensure a functional blood-brain barrier. *Development* **145**, dev157115 (2018).
- Ludwig, L. S. et al. Transcriptional states and chromatin accessibility underlying human erythropoiesis. *Cell Rep.* **27**, 3228–3240 (2019). e3227.
- Sack, L. M. et al. Profound tissue specificity in proliferation control underlies cancer drivers and aneuploidy patterns. *Cell* **173**, 499–514 (2018). e423.
- Haigis, K. M., Cichowski, K. & Elledge, S. J. Tissue-specificity in cancer: the rule, not the exception. *Science* **363**, 1150–1151 (2019).
- Monroe, T. O. et al. YAP partially reprograms chromatin accessibility to directly induce adult cardiogenesis in vivo. *Dev. Cell* **48**, 765–779 (2019). e767.
- Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
- Wang, H. et al. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc. Natl Acad. Sci. USA* **111**, 705–710 (2014).
- Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc. Natl Acad. Sci. USA* **113**, 12168–12173 (2016).
- Qi, Y. & Zhang, B. Predicting three-dimensional genome organization with chromatin states. *PLoS Comput. Biol.* **15**, e1007024 (2019).
- Zhou, J. et al. Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl Acad. Sci. USA* **116**, 14011–14018 (2019).
- Gursoy, G., Xu, Y., Kenter, A. L. & Liang, J. Computational construction of 3D chromatin ensembles and prediction of functional interactions of alpha-globin locus from 5C data. *Nucleic Acids Res.* **45**, 11547–11558 (2017).
- Baxter, J. S. et al. Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.* **9**, 1028 (2018).
- Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).
- Oluwadare, O., Highsmith, M. & Cheng, J. An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol. Proced. Online* **21**, 7 (2019).
- Cheng, R. R. et al. Exploring chromosomal structural heterogeneity across multiple cell lines. *Elife* **9**, e60312 (2020).
- Contessoto, V. G. et al. The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome. *bioRxiv* <https://doi.org/10.1101/2019.12.20.885145> (2020).
- Abbas, A. et al. Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat. Commun.* **10**, 2049 (2019).
- Nir, G. et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
- Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
- Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
- Bau, D. & Marti-Renom, M. A. Genome structure determination via 3C-based data integration by the Integrative Modeling Platform. *Methods* **58**, 300–306 (2012).
- Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol.* **13**, e1005665 (2017).
- Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
- Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
- Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
- Masica, D. L. et al. CRAVAT 4: cancer-related analysis of variants toolkit. *Cancer Res* **77**, e35–e38 (2017).
- Brown, A. L., Li, M., Goncarencu, A. & Panchenko, A. R. Finding driver mutations in cancer: elucidating the role of background mutational processes. *PLoS Comput Biol.* **15**, e1006981 (2019).
- Nussinov, R., Tsai, C. J. & Jang, H. Why are some driver mutations rare? *Trends Pharm. Sci.* **40**, 919–929 (2019).
- Nussinov, R., Tsai, C. J. & Jang, H. Autoinhibition can identify rare driver mutations and advise pharmacology. *FASEB J.* **34**, 16–29 (2020).

64. Chen, H. et al. Comprehensive assessment of computational algorithms in predicting cancer driver mutations. *Genome Biol.* **21**, 43 (2020).
65. Rogers, M. F., Gaunt, T. R. & Campbell, C. CScape-somatic: distinguishing driver and passenger point mutations in the cancer genome. *Bioinformatics* **36**, 3637 (2020).
66. Loganathan, S. K. et al. Rare driver mutations in head and neck squamous cell carcinomas converge on NOTCH signaling. *Science* **367**, 1264–1269 (2020).
67. Myers, M. B., McKim, K. L., Wang, Y., Banda, M. & Parsons, B. L. ACB-PCR quantification of low-frequency hotspot cancer-driver mutations. *Methods Mol. Biol.* **2102**, 395–417 (2020).
68. Guo, Y. et al. Recent progress in rare oncogenic drivers and targeted therapy for non-small cell lung cancer. *Onco Targets Ther.* **12**, 10343–10360 (2019).
69. Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).
70. Harrison, P. T., Vyse, S. & Huang, P. H. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Semin Cancer Biol.* **61**, 167–179 (2020).
71. Allen, A. et al. Rare BRAF mutations in pancreatic neuroendocrine tumors may predict response to RAF and MEK inhibition. *PLoS ONE* **14**, e0217399 (2019).
72. Song, J., Peng, W. & Wang, F. A random walk-based method to identify driver genes by integrating the subcellular localization and variation frequency into bipartite graph. *BMC Bioinform.* **20**, 238 (2019).
73. Poulin, E. J. et al. Tissue-specific oncogenic activity of KRAS(A146T). *Cancer Discov.* **9**, 738–755 (2019).
74. Cheng, F. et al. Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. *Mol. Biol. Evol.* **31**, 2156–2169 (2014).
75. Liu, C. et al. Individualized genetic network analysis reveals new therapeutic vulnerabilities in 6,700 cancer genomes. *PLoS Comput. Biol.* **16**, e1007701 (2020).
76. Liu, C. et al. Computational network biology: data, models, and applications. *Phys. Rep.* **846**, 1–66 (2020).
77. Cheng, F., Jia, P., Wang, Q. & Zhao, Z. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* **5**, 3697–3710 (2014).
78. Cheng, F., Kovacs, I. A. & Barabasi, A. L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 (2019).
79. Saltz, J. et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**, 181–193 (2018). e187.
80. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510 (2018).
81. Zeng, X. et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* **11**, 1775–1797 (2020).
82. Malta, T. M. et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354 (2018). e315.
83. Ma, J. et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15**, 290–298 (2018).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020