DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Multi-OMICs and Genome Editing Perspectives on Liver Cancer Signaling Networks

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Lin, Shengda, Yi A. Yin, Xiaoqian Jiang, Nidhi Sahni, and Song Yi. 2016. "Multi-OMICs and Genome Editing Perspectives on Liver Cancer Signaling Networks." BioMed Research International 2016 (1): 6186281. doi:10.1155/2016/6186281. http://dx.doi.org/10.1155/2016/6186281.
Published Version	doi:10.1155/2016/6186281
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:27822072
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Review Article

Multi-OMICs and Genome Editing Perspectives on Liver Cancer Signaling Networks

Shengda Lin,[1] Yi A. Yin,[1,2] Xiaoqian Jiang,[3] Nidhi Sahni,[4] and Song Yi[5]

[1]*Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA*
[2]*Cancer Biology Program, Stanford University School of Medicine, Stanford, CA 94305, USA*
[3]*Division of Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093, USA*
[4]*Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA*
[5]*Department of Genetics, Harvard Medical School, Boston, MA 02115, USA*

Correspondence should be addressed to Nidhi Sahni; nsahni@mdanderson.org and Song Yi; stephen.yi.harvard@gmail.com

The advent of the human genome sequence and the resulting ~20,000 genes provide a crucial framework for a transition from traditional biology to an integrative "OMICs" arena (Lander et al., 2001; Venter et al., 2001; Kitano, 2002). This brings in a revolution for cancer research, which now enters a big data era. In the past decade, with the facilitation by next-generation sequencing, there have been a huge number of large-scale sequencing efforts, such as The Cancer Genome Atlas (TCGA), the HapMap, and the 1000 genomes project. As a result, a deluge of genomic information becomes available from patients stricken by a variety of cancer types. The list of cancer-associated genes is ever expanding. New discoveries are made on how frequent and highly penetrant mutations, such as those in the telomerase reverse transcriptase (*TERT*) and *TP53*, function in cancer initiation, progression, and metastasis. Most genes with relatively frequent but weakly penetrant cancer mutations still remain to be characterized. In addition, genes that harbor rare but highly penetrant cancer-associated mutations continue to emerge. Here, we review recent advances related to cancer genomics, proteomics, and systems biology and suggest new perspectives in targeted therapy and precision medicine.

1. Genetic Alterations in the Cancer Genome: Liver Cancer as an Example

Significant effort has been made to reveal the mutational landscape of cancers. Herein, we use liver cancer as an example to demonstrate recent advances. Primary liver cancer is the sixth most frequent cancer worldwide and a leading cause of death in Asia [1, 2], with hepatocellular carcinoma (HCC) as the most common form, followed by intrahepatic cholangiocarcinoma (IHCC) [2]. Most liver cancers are developed from liver cirrhosis with hepatitis B virus (HBV) and hepatitis C virus (HCV) infections, with alcohol consumption, metabolic diseases, and chemical exposure as major predisposing factors [3–5].

The capacity of next-generation sequencing (NGS) has dramatically increased over the years due to technological advances. Cost per raw megabase of DNA sequence has gone down from over $5000 in 2001 to $0.015 in mid-2015, at a rate faster than Moore's Law [6], allowing more samples to be sequenced in parallel and more powerful statistical analyses to be performed. Whole-genome sequencing (WGS) and exome sequencing have identified various genetic alterations in liver cancer. The first whole-genome sequencing of a HCC genome with HCV revealed more than 11,000 somatic substitutions in the tumor genome and 22 validated chromosomal rearrangements [7]. A later study sequenced 27 HCCs, 25 of which were with HBV or HCV, including 2 sets of multicentric tumors [8]. Multiple chromatin regulator genes, including *ARID1A*, *ARID1B*, *ARID2*, *MLL*, and *MLL3*, were detected in ~50% of the tumors. In addition, HBV integration in the *TERT* locus was frequently observed in a high clonal proportion. Another WGS study

of 88 HCC tumors, among which 81 were associated with HBV, found *CTNNB1* to be the most frequently mutated oncogene (15.9%) and *TP53* to be the most frequently mutated tumor suppressor (35.2%) [9]. Integrated analysis of somatic mutations and focal copy-number changes of 125 HCC tumors and whole exome sequencing on 24 of these samples identified 135 homozygous deletions and 994 somatic mutations [10]. New recurrent alterations in *ARID1A*, *RPS6KA3*, *NFE2L2,* and *IRF2* were found in this study. To date, ~1000 HCCs have been sequenced, which provides a mutational landscape of HCC. Most common mutations, including *TERT* promoter mutations (56%), *TP53* (27%), *CTNNB1* (26%), *ARID2* (7%), *ARID1A* (6%), and *Axin1* (5%), as well as key signaling pathways, such as the canonical WNT signaling pathway and the JAK/STAT pathway, were shown to be altered in liver cancers [9, 11, 12].

Moreover, deep-sequencing technologies have greatly facilitated pathogenic analysis of liver cancers stratified by etiology. A recent exome sequencing study on 243 liver tumors identified 161 putative driver genes associated with 11 recurrently altered pathways [13]. Association of mutations and risk factors defined 3 groups of genes centered on *CTNNB1* (alcohol), *TP53* (hepatitis B virus, HBV), and *AXIN1*. *TERT* promoter mutations and *TP53* alterations were associated with early and advanced stages in the tumors, respectively [13]. *TERT* reactivation is also associated with HBV infection. Although genome integration is not required for HBV replication, fragments of HBV DNA are found in chronic HBV infections and 85–90% of HBV-related HCC [14, 15]. Most of the integration events result in unidirectional upregulation of genes at the integration sites [16]. Consistent with previous analysis of HBV integration sites using PCR [17, 18], high-depth genome sequencing of HBV-positive HCC samples identified frequent HBV integration including the *TERT* locus [8, 19].

As the most frequently mutated target, telomerase plays a central role in liver cancers. Telomerase extends the terminal segment of eukaryotic chromosomes known as the telomeres [1, 2, 20]. Normal cells could only undergo a finite number of divisions in culture before entering a senescence state, a phenomenon discovered by Hayflick and Moorhead in the 1960's [21]. In contrast, cancer cells counter the "end-replication problem" by acquiring the capacity to maintain the telomeres. 80–90% of human cancers sustain their telomeres by reactivating telomerase [22]. The catalytic core of the telomerase consists of the catalytic protein component encoded by *TERT* and the RNA component *TERC* [2, 23]. Additional components, such as dyskerin (DKC) and telomerase Cajal body protein 1 (TCAB1), are required for the holoenzyme to function *in vivo* [3, 4, 24–31]. Telomerase expression is primarily controlled by the transcription of *TERT* [32–34]. Most somatic cells do not express *TERT* and lack telomerase activities [35–37]. The connection between telomere regulation and liver cancers was first studied in Japanese patients in the 1990's. Shortening of telomeres was reported in cirrhosis patients over 45 years with viral hepatitis, and telomerase reactivation was also observed in HCC patients [21, 38–41].

An important insight into the mechanism of *TERT* reactivation was discovered in 2013, when two independent studies identified recurrent somatic mutations in the core promoter of *TERT* genes in different melanoma samples [42, 43]. The most prevalent somatic mutations were two mutually exclusive "C>T" transitions at -124 and -146 from the translational start ATG of the *TERT* gene, respectively. These mutations were subsequently identified in a wide range of other human cancers including HCC, glioma, thyroid, and bladder cancers [44, 45]. Additional less frequent mutations were also detected in the *TERT* promoter, including the tandem mutations "CC>TT" at −124~125 and −138~139 bp from ATG, as well as a A>C transversion at −57 bp from the ATG [43]. These mutations created *de novo* binding motif for Ets/TCF transcription factors. A study of 23 human urothelial cancer cell lines demonstrated that these promoter mutations are correlated with higher levels of *TERT* mRNA, protein, telomerase activity, and telomere length [46]. A member of the Ets family, GABP, was found to be recruited to the mutation site to activate *TERT* [47]. Together, recent findings firmly established that the genetic alterations at the *TERT* promoter play a central role for the cancer-specific telomerase activation. In HCC, the -124C>T accounted for 93% of the total mutations detected, and the frequency of -146C>T was 6%. Promoter mutations were identified in 5 out of 20 macronodules of cirrhotic but not in the 69 cirrhotic tissues, suggesting that the *TERT* promoter mutation is an early genomic alteration that transitions liver cirrhosis to carcinogenesis [48]. Interestingly, *TERT* promoter mutations were not detected in the benign hepatocellular adenoma; in contrast, 7/16 (48%) malignant tumors transformed from HCA and 58/106 (55%) of HCCs in normal liver exhibited the mutations, all of which are significantly associated with mutations activating canonical WNT signaling pathway. Thus, telomerase was activated at a later stage of HCC without cirrhosis.

2. Multi-Omics to Unravel Cancer Mystery: Evolution of Functional Genomics and Proteomics

The advent of human genome sequences has changed our ways to address fundamental questions in human cancer. With information available for thousands of genes, the conventional method of studying one gene (or one protein) at a time could now be complemented by more systematic platforms that study multiple or even all genes at large scale. A potential barrier to this prospect, however, is that most genes have not yet been empirically characterized. For most gene products (or proteins) in the proteome there is a lack of functional information that can be obtained or derived from any biological model. Toward this end, in the recent past, high-throughput functional genomic and proteomic strategies have been invented to facilitate the annotation of large numbers of genes. Such "systems biology" approaches aim to generate quantitative and dynamic models and to interrogate key biological processes with holistic insights (Figure 1). Herein we summarize a few such high-throughput
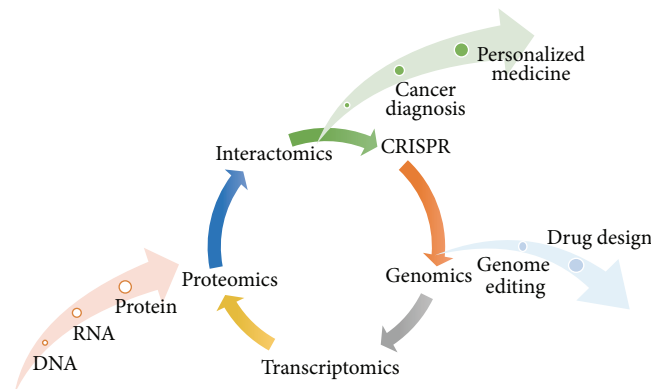
FIGURE 1: Multi-OMIC systems approach to elucidate cancer signaling networks and precision medicine.

genome-wide functional platforms that have been developed [49–53].

2.1. Gene Expression. Expression profiling techniques such as microarray and RNA-seq provide an estimate of mRNAs (transcription levels) present under a given condition in a cell- or tissue-specific fashion. Making transcript level measurements under many different conditions defines a "transcriptome" for a given organism [54]. Microarray is based on the hybridization of a cDNA library to a DNA chip to determine relative abundance of usually fluorophore-labeled targets [55]. RNA-seq takes advantage of next-generation sequencing to quantify the amount of RNAs after reverse transcription [56]. With gene expression data, clustering analyses can be performed to group genes that are similarly expressed. These expression profile clusters often contain functionally related genes that are coregulated and could suggest new functional hypotheses for uncharacterized genes in the same clusters. For correlation measurement, Pearson correlation coefficients are often used with proper titration adjustments [57].

Gene expression is thought to be primarily regulated by transcription factor binding at a given time [56]. Recent studies also revealed important roles of lncRNAs [58] and miRNAs [59] in gene regulation. With the facilitation of modern technologies and next-generation sequencing, RNA-seq gene expression can now be easily performed at the single cell level [60]. However, gene expression at the transcriptional levels may not correlate well with the translational levels [61], so protein-centric studies need to take additional proteomic assays for validation.

2.2. Proteomics. Numerous proteomics approaches have been developed and applied to study large-scale protein functions. Protein localization mapping projects assign functionally related proteins to the same subcellular compartments at similar times, given their possible involvement in similar biological processes [62]. Reverse phase protein array (RPPA) is a proteomics technology that allows for quantitative protein expression measurement at large scale based on high-quality antibodies [63].

On the other hand, large-scale macromolecular interaction screening tools, such as yeast two-hybrid (Y2H) systems or mass spectrometry (TAP/MS), have been widely used to map protein-protein interaction networks in different species, including human. Physically interacting protein partners are believed to share signaling pathways, GO terms, or memberships in protein families [64, 65]. Functionally related gene products often act as macromolecular complexes and form topological modules in the interaction networks, by which hypothesis of function for many unknown proteins could be formulated.

2.3. Data Repositories for Cancer. The Human Gene Mutation Database (HGMD) is a comprehensive repository of germline mutations in genes that are causal for, or are associated with, human disease, including cancer [66]. The ClinVar database [67] from NCBI also contains cancer mutation annotations. Mode of inheritance information for each cancer type can be obtained from two databases: Online Mendelian Inheritance in Man (OMIM) [68] and Universal Protein Resource (UniProt) [69].

The Cancer Genome Atlas (TCGA) is a large repository for genetic mutations in more than 30 cancer types, including ~500 patient samples [70]. TCGA is also an enormous resource for profiling of gene expression, copy-number variation, DNA methylation, and so forth [71, 72]. The International Cancer Genome Consortium (ICGC) is a collaborative organization that aims to coordinate large-scale genomic, transcriptomic, and epigenomic data for over 50 cancer types around the world [73]. The Cancer Cell Line Encyclopedia (CCLE) is another collaborative project with a goal of providing comprehensive genomic data and computational analysis for ~1,000 human cancer cell lines [74]. To facilitate the easy use of multidimensional cancer genomic data, cBioPortal was established to provide a web resource for exploring, visualizing, and analyzing molecular profiling data in cancer tissues and cell lines [75]. Furthermore, large-scale phenotypic analysis can also help characterize genes and suggest potential functional descriptions for many unknown genes. By identifying possible phenotypes attributable to disruptions or alterations in specific genes using technologies

such as knock-outs or RNAi, genes with similar phenotypes can be found that might function together in common functional pathways in a given cellular context [57, 76].

3. An Evolving Systems Biology Toolkit for Better Cancer Precision Medicine

A key leap forward in the development of a cutting-edge cancer research toolkit is to design strategies to flexibly express any genes in the human genome (Figure 1), in order to study them in various cells, under different conditions and in many biological processes of interest. In other words, there is a dire need to develop diverse large-scale functional genomic and proteomic platforms. High-throughput studies [64, 65, 77–82] often require large numbers of protein-encoding genes to be expressed precisely, that is, in-frame without any 5′UTRs, 3′UTRs, or introns, into various expression systems.

3.1. High-Throughput Gateway Technology for Functional Studies. Gateway is a modern molecular technology amenable for high-throughput and automated biomedical experiments. Gateway technology is designed for easy transfer of DNA fragments based on site-specific recombination principle [77, 83, 84]. In this big data era, Gateway has emerged as a cutting-edge tool to facilitate large-scale genomic and functional studies, such as mutagenesis, sequence tagging, protein purification, promoter, and RNA analysis. It has been increasingly appreciated and widely adopted in a variety of cancer research areas.

Gateway technology enables convenient DNA transfer, taking advantage of the recombination machinery between the genomes of bacteria and phage. This process is reversible and involves two enzyme mixes ("BP" and "LR" clonase) and a set of recombination sequences ("*att*" sites). The recombination events are described briefly below. (i) Catalyzed by the BP clonase mix, the *attP* site of the phage DNA recombines with the *attB* site from the bacterial DNA, deriving two new sites, *attL* and *attR*. (ii) Catalyzed by the LR clonase mix, the *attL* and *attR* sites recombine in the excision reaction, reverting back to the *attP* and *attB* sites.

When implementing the Gateway technology in molecular biology, a typical "Gateway Cassette" is designed as a module to insert into a vector. The four recombinational sites (*attB*, *attP*, *attL*, and *attR*) are duplicated and modified. In the BP reaction, we start with a "Donor" vector, containing a Gateway Cassette with *P1* and *P2* sites and usually a chloramphenicol resistance selection marker. The *P1* and *P2* sites on the Donor plasmid recombine with *B1* and *B2* sites, respectively, which flank a DNA sequence of interest. In this way, the DNA of interest can be cloned unidirectionally into the Donor vector. The resulting product is known as an "Entry" clone, containing two *attL* sites, *L1* and *L2*.

Gateway Entry clones can be readily transferred via an "LR" reaction into various expression vectors, known as Destination vectors, for downstream functional studies. In the LR reaction, the *R1* and *R2* sites on Destination plasmids recombine with the *L1* and *L2* sites, respectively, on the Entry clones. Many popular prokaryotic and eukaryotic expression vectors are available, such as yeast two-hybrid AD and DB vectors, fluorescence-based PCA vectors, and LUMIER Myc- and flag-tagged vectors for coimmunoprecipitation. In addition, other existing functional expression vectors can be readily converted to Gateway-compatible Destination Vectors, by inserting a Gateway Cassette. With the fast growing of genomic information and larger-scale research nowadays, the Gateway cloning system apparently emerges as a powerful, high-throughput platform compatible with the current research needs. A collection of genes, as Gateway Entry clones, can be transferred at large scale to one or more Destination Vectors in a simple reaction, manually or robotically.

3.2. The Human ORFeome: A Versatile Tool for Cancer Research. Large libraries of Gateway Entry clones, encompassing all possible open reading frames (ORFs) [84] in the genomes of many species including humans, are necessary for high-throughput functional studies. Ideally, the human "ORFeome" corresponds to all full-length protein encoding genes, including possible variants and isoforms in different tissues, developmental stages, and across the human population. However, identifying such a comprehensive ORFeome collection is apparently challenging, due to limitations in existing experimental strategies [85].

Initial efforts in the construction of a human ORFeome library took advantage of public collections of human cDNAs, such as the Mammalian Gene Collection (MGC) [86]. Using MGC as template for PCR amplification, ~8000 ORFs were Gateway cloned without containing a stop codon; thus, N-terminal and C-terminal protein fusions can both be feasible downstream in Destination vectors. Because there may be multiple splice isoforms and polymorphic variant ORFs for the same gene, the 8000 ORFs represented ~7000 distinct genes. Clones shorter than 100 nucleotides and clones without complete coding sequences (CDS) available in NCBI were eliminated. Successfully cloned ORFs were consolidated as the first version of the human ORFeome collection (hORFeome v1.1) [87]. In 2007, the human ORFeome v3.1, adding ~4,000 new ORFs, brings the total to 12,212 distinct ORFs, representing 10,214 distinct genes [88]. In 2011, the human ORFeome v8.1 was released, containing 16,172 ORFs mapping to 13,833 distinct genes [89]. This extensive ORF library represents an important resource of single-colony, fully sequence-verified human ORFeome Entry clone collection. This set of ORFs ranges in size from 75 to more than 10,000 base pairs. In addition, an Expression Library version of this hORFeome v8.1 was constructed in a lentiviral expression vector that produces consistent titers and gene expression levels and allows delivery to most cell types [89].

4. The Human Interactome: A Scaffold for Functional Proteomics and Evolution in Cancer

Identification of human cancer genes in which mutations are associated with specific clinical manifestations has facilitated our understanding of disease mechanisms. However, like

their normal counterparts, protein products of cancer genes do not function in isolation but are part of highly interconnected cellular signal transduction networks (Figure 1) [90, 91].

4.1. Literature-Derived Interactome (LDI).

Interactome networks could be derived from literature through two different approaches: text mining and manual curation. Text mining is performed computationally by searching for key words in literature databases, such as PUBMED. Manual curation of literature knowledge involves enormous amount of labor and time. However, certain datasets of human molecular interactions have been curated from the literature and stored in public databases, such as BioGRID [92], CORUM [93], BIND [94], DIP [95], STRING [96], HPRD [97], MINT [98], GeneMania [99], and MIPS [100].

4.2. Empirically Derived Interactome (EDI).

Modern molecular biology has brought in many advanced tools for functional studies, but most of them experience limitations when it comes to scale-up to a genome-wide investigation. However, a number of experimental strategies have been employed in large-scale human interactome mapping, such as yeast two-hybrid [79, 101, 102], cofractionation [103], and affinity purification followed by mass spectrometry (AP-MS) [104]. Early efforts using high-throughput systematic yeast two-hybrid platforms have generated preliminary human protein-protein interactome network maps [79, 101]. In 2005, two studies simultaneously reported the first version of human interactome map. 2,754 high-confidence protein-protein interactions among 1,549 proteins were reported in the CCSB-HI1 dataset [79], while 3,186 interactions involving 1,705 proteins were reported in the Stelzl network [101]. A second generation of interactome map was recently published, containing 13,944 interactions among 4,303 distinct proteins [102]. This map covers a vast previously uncharted territory and is 30% larger than the literature of all small-scale studies combined in the past few decades. It is demonstrated to be helpful in predicting novel cancer genes and other disease-associated mechanisms.

Proteome-scale studies of human interactome networks have also been performed using other high-throughput approaches. Based on biochemical fractionation and quantitative mass spectrometry, Havugimana et al. identified a map of 622 protein complexes in human cells. This interactome map profiles 13,993 physical interactions between 3,006 proteins, revealing many interesting biological associations [103]. Lately, another interactome network map systematically charted by affinity purification followed by mass spectrometry (AP-MS) provided another functional view of protein complexes, covering 23,744 interactions among 7,668 proteins with many unexpected hypotheses for previously poorly characterized proteins [104]. Furthermore, another group performed a quantitative network survey to capture human interactome networks with higher resolution in interaction strength and protein abundance [105]. They used quantitative bacterial artificial chromosome with GFP fusion interactomics (QUBIC) and identified 28,504 unique

interactions involving 5,462 proteins. It was demonstrated that weak interactions dominate the network and have topological properties.

To assess the specificity of interactome networks, a random subset of interactions is selected typically for an independent and orthogonal validation to confirm the overall quality of the human interaction networks. It is noteworthy that not all the interactions reported in literature are of high quality or necessarily interpreted as "gold standard." The ones identified by multiple publications or methods tend to be genuine interactions. As expected in any biological assay, the resulting networks exhibit a large fraction of false negatives. To assess the sensitivity of interactome networks, high-confidence subsets of literature-derived interactions can be employed as a comparison for sensitivity measurements.

5. Systems Biology Reveals Functional and Evolutionary Insights into Human Diseases Like Cancer

5.1. Computational Modeling.

Computational modeling has been useful in predicting the functional impact of genes and mutations that are difficult to test experimentally. Polymorphism Phenotyping v2 (PolyPhen-2) [106] was developed to predict the functional significance of a genetic variant based on conservation, protein structure, and other features using naïve Bayes classifier trained by supervised machine-learning. Mapping of genetic variants to Pfam domains (Pfam-A family only) can be performed using the program Hmmer version 3 [107]. The IUPred program [108] can be used to assess the likelihood of residues affected by a genetic variant located in an intrinsically disordered region of the protein. The regular expressions of known eukaryotic linear motifs (ELMs) can be obtained from the ELM database (http://elm.eu.org/). DSSP program [109] can be used to compute solvent accessible area for each residue mutated by a genetic variant. FoldX force-field algorithm [110, 111] can be used to calculate the change in free energy of unfolding ($\Delta\Delta G$) for all mutations that could be mapped to a published crystal structure from Protein Data Bank (PDB) [112]. For interaction interface analysis, the mutated residues can be mapped onto the available structures by using Mechismo (http://mechismo.russelllab.org/), ProtInDB (PROTein-protein INterface residues Data Base), and PDBePISA (Proteins, Interfaces, Surfaces, and Assemblies) [113] servers. The database of three-dimensional interacting domains (3did) documents and predicts high-resolution structures for domain-domain interactions [114].

Recently, a structure-based prediction of a proteome-wide human protein-protein interaction network was released [115]. Through experimental validation of a subset of interactions, this computationally predicted interactome (CPI) network was considered to be of high quality. HINT (High-quality INTeractomes) is a database that extracts high-quality protein-protein interactions [116]. Clusters of cancer mutations in the human proteome can be identified by mutation3D algorithm [117]. Looking into the future, the union of all LDI, EDI, and CPI interactions reveals more

and more comprehensive human interactome networks, and the interaction pairs argue for their potential biological and functional relevance. However, future efforts are still required to interpret condition-specific interactions and to characterize the effects of genomic variation on interaction networks [118], which will in turn generate insights into genotype-phenotype relationships in human.

5.2. Computational Modeling and Network Analysis. Given the highly connected nature of molecular signaling network organization in the cell [119, 120], a conceptual framework was developed to illustrate a global picture (known as "diseasome") of all the known genes involved in human disease. To construct such a "diseasome" network, a compendium of 1,777 human disease genes and 1,286 associated diseases [121] was obtained from the Online Mendelian Inheritance in Man (OMIM) database. In the diseasome network, the human "disease genome" (a long list of known disease genes) was linked to the "disease phenome" (a list of known genetic disorders), deriving a comprehensive set of almost all known gene-disease network associations. This network-based "genome-phenome" profile [122] is a bipartite graph, in which a gene and a disease are linked together if mutations in that gene have been implicated in that disease.

A Human Disease Network (HDN) was derived from the original bipartite "diseasome" landscape [123]. In the HDN network, nodes represent diseases, and edges represent the association between diseases when they share at least one gene in which mutations are associated with both diseases. Overall, 867 of 1,284 diseases have at least one link to other diseases, and 516 diseases form a single connected cluster, the giant component, suggesting that most human diseases share, to some extent, genetic origins. The HDN network is clearly clustered by major disease classes, reflecting visible differences between classes of disorders but commonality in genetic origin within each disease class. Among the most connected diseases is cancer, which is in part due to the many common regulators (such as p53, PTEN, KRAS, ERBB2, and NF1) associated with distinct subtypes of cancer.

Another type of biologically relevant networks concerns disease gene network (DGN) [123]. In the DGN, nodes represent disease genes, and edges represent their association with the same disease. In this network, 1,377 of 1,777 disease genes are connected to at least one other disease gene, and 903 genes form a giant component. The DGN provides a complementary, gene-centered view of the diseasome than the HDN.

Given that interactome networks cover a myriad of genes implicated in human diseases, including cancers, they provide useful insights into possible disease signaling mechanisms. Although existing empirically derived interaction (EDI) networks are far from being complete, the overlap with literature has been shown to be significant [102], demonstrating the high quality of the EDI networks. On the other hand, these networks offer novel biological hypotheses and guide further studies of disease signal transduction in relevant functional contexts. Functional consequences of molecular interactions can be followed up to understand the logic of complex biological networks. Therefore, emerging human interactome networks will eventually facilitate our understanding of human health and disease.

6. Novel Therapeutic Strategies and Precision Medicine

A major problem in cancer treatment is to achieve specific killing of cancer cells while preserving normal cells. Cancer genomes vary from individual to individual.

6.1. New Promises of Gene Therapy from CRISPR. The idea of gene therapy was proposed in the 1970's [124]. The 90's witnessed the first successful gene therapy treating patients with severe combined immune deficiency (SCID) by modifying cells with retroviruses carrying a functional copy of the mutated gene [125–128]. However, complications mostly due to integration of viral vector to oncogenes led to suspension of many clinical trials [129]. Nevertheless, quite a few gene therapy strategies made steady strides entering the new century, including Gendicine (first gene therapy product approved for clinical use in humans) [130], oncolytic virus talimogene laherparepvec, and the immunostimulant sipuleucel-T.

The breakthroughs in CRISPR (clustered regularly interspaced short palindromic repeats) mediated genome editing technology provide us with unparalleled opportunity to bring precision medicine to the genome level [131–134]. Compared to targeting malfunctioned molecules at the protein level, it allows for restoration of proper spatiotemporal regulation of the functional molecules without concerns for dosage responses and side effects [131–134]. By correcting disease-causing mutations in embryonic stem cells, disease prevention is made possible even before the onset of symptoms [135–138]. As proof of principle studies, CRISPR-mediated mutation corrections have been successfully performed using multiple mouse disease models including hereditary tyrosinemia and muscular dystrophy [135, 136], resulting in reversion and prevention of diseases, respectively. Given its great potential, CRISPR/Cas9 can revolutionize personalized cancer treatment: to model functional consequences of recurrent mutations identified through high-throughput sequencing efforts, to discover cancer drug targets by screening protein domains [137], and to inhibit cancer by inactivating driver mutations [138].

CRISPR/Cas9 system edits the genome by first creating DNA double-strand breaks (DSBs) [139, 140]. When DSBs occur, the cells activate one of the three mechanisms to repair double-strand breaks: nonhomologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and homology-directed repair (HDR) [139, 140]. End-joining mechanisms are error-prone and often lead to loss of gene function as a result of random insertions or deletions. In contrast, a DNA sequence, which shares homology with the DSB locus, can be used as a donor template for the HDR pathway to precisely modify the DNA sequence [139, 140]. Sequence-specific endonucleases, such as the zinc-finger nuclease (ZFN) and the transcription-activator-like effector nuclease

(TALEN), can introduce double-strand breaks at specific sites of the genome, which dramatically favors the process of HDR instead of NHEJ [139, 140]. ZFN and TALEN have greatly facilitated genome engineering in a variety of model systems [22, 141, 142]. However, the difficulties of designing and building endonucleases tailored to specific genes of interests and the relatively low cutting efficiency have limited their applications. In recent years, Type II CRISPR system emerged as a useful tool for genome editing, with major advantages in cutting efficiency and versatility [131–134]. The most commonly used CRISPR system was modified from the CRISPR-associated endonuclease 9 (Cas9) in *Streptococcus pyogenes* (SpCas9) [139, 140]. The recombinant Cas9 system consists of three components: the Cas9 protein, the CRISPR RNA (crRNA), and the transactivating crRNA (tracrRNA). The crRNA and tracrRNA are often cloned into a single chimeric guide RNA, known as single guide RNA (sgRNA), resulting in an easy-to-use two-component system [139, 140]. The specificity of the endonuclease was determined by the complementation of the sgRNA and its 20-nucleotide target sequence in the genome [143]. The genomic target sequence must be immediately upstream of a $5'$-NGG protospacer adjacent motif (PAM) [143]. $5'$-NAG can also be tolerated as an alternative PAM [144], albeit with reduced cleavage efficiency [145]. Potential limitations of the CRISPR mainly concern the off-target effects. The seed sequence close to the PAM domain carries more weight in target specificity, while the mismatches further away from the PAM domain and towards the $5'$-end of the targeted genome sequence could be tolerated to certain degree [145–147]. Efforts have been made to evaluate and improve the fidelity of the Cas9 system. For example, the Cas9(D10A) mutant, which functions as an ssDNAase, can be used with a pair of sgRNAs complementary to opposite strands of the target DNA, in order to reduce off-targets [148–150]. This is because the DSB generated at the desired site require both nicking events, while sites with a single nicking event are primarily repaired by the more precise excision-repair mechanisms rather than error-prone end-joining mechanisms. Similarly, catalytically inactive cas9 (dCas9) can be fused to the cleavage domain of the FokI restriction endonuclease [151, 152]. The simultaneous binding of two fusion proteins (fCas9) to target sites that are 13~18 bp apart is required for the DSBs to occur [151, 152], as the FokI only cleaves DNA when dimerized.

The efficiency of CRISPR-mediated genome editing is context dependent, with low efficiencies being observed at high "GC" regions or those in close proximity to heterochromatin [139, 140]. In such cases, multistep targeting might be required. For example, *TERT* promoter mutations reside in a genomic region with ~80% GC content. A two-step approach was employed to introduce *TERT* promoter mutations into hESCs. First, two Cas9/sgRNAs were used to delete a 1.5 kb region at *TERT* promoter encompassing the mutation spot. Second, a sgRNA against the newly synthesized NHEJ-derived junction was coelectroporated with Donor plasmids containing the deleted region with cancer-associated *TERT* promoter mutations [153]. In another study, a two-step "pop-in/pop-out" strategy was used to create N-terminal tagged *TERT* fusion protein. First, homologous recombination was achieved by CRISPR/Cas9 targeting the translational start site of *TERT*, with a donor template containing both the tag and an eGFP expressing cassette flanked by LoxP sites. Successfully targeted cells were selected by flow cytometry. Second, eGFP cassette was removed by Cre-mediated recombination [154].

A potential limit of CRISPR-gene therapy concerns the delivery methods. Recombinant AAV (rAAV) is widely considered to be an ideal viral vehicle for gene therapy, because DNA cargo can persist as episomes in both dividing and quiescent cells state with minimal genome integration. Even though exogenous DNA carried by rAAV has been shown to be effective in correcting mutations like the *Fah* mutation in the liver [155], CRISPR technology could result in higher efficiency of gene correction, as proof-of-concept studies demonstrated by hydrodynamic injection of Cas9/sgRNA and a single-stranded DNA to correct the *Fah* mutation in hepatocytes via homology-directed repair [135]. However, the size of the widely used SpCas9 (~4.2 kb) is approaching the cargo limit of rAAV (~4.5 kb), leaving little room for modification. Recently, a smaller Cas9 from *Staphylococcus aureus* (SaCas9) was described [156]. The authors packaged SaCas9/sgRNA into a single rAAV vector and successfully targeted the Pcsk9 gene in the mouse liver.

A broad community of stakeholders have collaborated closely to forge ahead with precision therapy, especially CRISPR-mediated genome editing. Academic researchers continue to provide more accurate insights into human genetics and molecular basis of diseases, as well as develop more powerful bioinformatics tools for analyzing data at the genome scale. Diagnostic companies develop better tests based on newest data to achieve greater precision in interpreting the likelihood of patient response to the therapy. Pharmaceutical companies strive to increase CRISPR targeting efficiency and minimize off-targeting effects, develop reliable quality control process, and build platforms for reducing the cost of CRISPR-mediated gene targeting. Hospitals and other healthcare providers should actively adopt new technologies for individualized prevention, detection, and treatment of diseases. Meanwhile, efforts should be made to provide easy health data access and share mechanisms and protect patient privacy and data security, as well as create platforms to engage different stakeholders in precision medicine as collaborating partners.

6.2. Cell Transplantation. The advances of genome-engineering techniques, as well as deeper understanding of the expression profiles of stem/progenitor cells, provide better prospects of cell therapy. As mentioned earlier, initial clinical successes by transplanting genetically modified cells to treat SCID provided valuable proof-of-concept. Here, we used hepatocyte transplantation as an example to discuss some of the promises and challenges of cell transplantation.

To date, only a few treatments can increase the life expectancy of liver cancer patients including resection, orthotopic or living donor liver transplantation, radio-frequency ablation/percutaneous ethanol injection, transcatheter arterial chemoembolization, and sorafenib [157].

Although liver transplantation remains the primary therapeutic strategy for end stage liver diseases and acute liver failures, donor shortage remains a primary hurdle. In an aging population, the supply of liver allografts is unlikely to meet the ever-increasing demand. Therapeutic cell transplantations have been brought into preclinical and clinical applications. There are potential advantages of hepatocyte transplantation over liver transplantation, because hepatocyte transplantation is generally considered to be less invasive, and the native livers are not surgically removed to allow other strategies like gene therapy to be performed.

The potential of hepatocytes as the source for cellular therapy has been demonstrated by years of animal experiments. Rodent hepatocytes have remarkable proliferative capacity *in vivo* [158, 159]. Hepatocyte transplantation was effective in correcting metabolic diseases in several rodent models, including the Gunn rat of Crigler-Najjar syndrome type I [160], the $Fah^{-/-}$ (fumarylacetoacetate hydrolase) mouse of tyrosinemia type I [161], the mutant human α1-antitrypsin transgenic mouse [162], and the Long-Evans cinnamon rat of Wilson's disease [163], as well as chemically or surgically induced acute liver failures [164, 165].

A major obstacle to overcome for translating animal studies to human patients is how to obtain enough hepatocytes in a safe transplantation route. The most common route of hepatocyte transplantation is through the portal system. Donor hepatocytes that can be safely infused through the portal vein are usually less than 5% of the liver mass (~2e8 cells/kg), in order to avoid portal hypertension, translocation of the cells to systemic circulation, and embolization in the lung [166]. As many as 70~80% of transplanted hepatocytes are entrapped in the portal space or sinusoids and are subsequently cleared by Kupffer cells and granulocytes [167, 168]. The integration of hepatocytes in the recipient liver is inefficient and requires disruption of hepatic sinusoidal endothelia [169]. Moreover, the initial engraftment of transplanted hepatocytes is unlikely to completely reverse the enzyme deficiency. The continuous repopulation of the recipient liver requires substantial selection advantage of the transplanted hepatocytes, which are artificially created in animal models by extensive parenchymal loss or the proliferative deficiency of endogenous hepatocytes. As a result, repeated hepatocyte transplantation may be required to increase the number of engrafted cells. The population of liver stem cells is potentially a good source of cell transplantation, due to its expandability *in vitro* and bipotent differentiation into hepatocytes and cholangiocytes [170–173]. However, recent studies suggest that mature hepatocytes are responsible for most of the liver repopulation during homeostasis and injuries *in vivo* [174–178]. Within the differentiated hepatocytes, there were subpopulations demonstrating higher repopulating capacities than generic hepatocytes [179, 180]. To identify the best population for cell therapy, comprehensive investigation of the heterogeneity of repopulating cells in the liver is required.

6.3. Systems Biology and Therapeutic Strategies. Until recently, a paradigm of drug discovery has been that for each disease there will be one (or a few) molecular target(s)

that can be affected either positively or negatively by a single chemical compound. This philosophy has clearly been successful for many diseases and has led to the development of "blockbuster" drugs such as the various ACE-inhibitors or Gleevec. However, this one-gene-one-drug approach has given rise to only ~500 drug targets [181] which, after all, represent a tiny portion of the predicted proteome, estimated at ~500,000 proteins taking into account all isoforms and posttranslational modifications [182]. Moreover the "one-gene-one-phenotype" approach is overly simplistic, because one gene can have multiple functions whereas one function can be handled by multiple genes. For example, various "regulatory" proteins such as Ras, Myc, and NF-κB each have disparate functions that are dependent upon cellular context [182]. Clearly, reliance on the "one-gene or one-protein leading to one drug" paradigm will continue to produce useful drugs, but this strategy is increasingly more difficult to implement [183].

Systems biology approaches have been recently applied to enable a holistic view of signaling networks in cancer cells and effectively identify molecular changes in cancer patients (Figure 1) [184]. For instance, global transcriptomic data analyses in B-cell lymphoma from The Cancer Genome Atlas (TCGA) revealed that older patients tend to exhibit decreased metabolism and telomere function, while female patients are likely associated with decreased interferon and PD-1 signaling [185]. In addition, a critical leap forward in proteomics is the gene-centric Human Protein Atlas for expression profiles [186], which resolves tissue-specific proteome variation of the human body [187] and provides significant insights into cancer pathology [188–190].

An alternative strategy is to understand the structural features and properties of molecular and physiological networks [182, 191]. Although this approach may not have immediate returns in terms of successful deployment of useful drugs [182], it will, in the long run, lead to better understanding of how to model networks and how to use those models for *in silico* studies [191]. Examples are how genetic polymorphisms affect responses to individual drugs [192] and how network interactions can be manipulated and altered by the actions of oncogenes and tumor suppressor genes on one side or by pharmacological intervention on the other [193]. A major hurdle to be overcome is the identification of cellular networks and all of their constituent units, along with an understanding of the signaling within networks and between/among networks.

7. Big Data Management and Security of Medical Information

The significant improvement of sequencing technologies makes human genomic data increasingly affordable and available in the era of precision medicine [194]. This paper discusses heavily how massive human genomic data open the door to big data science and speed up discoveries. Despite encouraging future, there are also emerging problems with respect to storing, sharing, and analyzing big human genomic data. The recent NIH data sharing policy change

allows users to store and analyze human genomic data using cloud-computing services, which address some of the issues. But on the other hand, the privacy challenge becomes more prominent with cloud computing as owners lose the full control of the data. It becomes more complicated as copies of data can be stored in a distributed file system or automatically backed up by the cloud service provider. Without necessary protection, it is risky to use the cloud for handling human genomic data, of which information leakage can lead to reidentification [195–199] and might negatively impact patients. The NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genome Data Sharing (GDS) Policy also states that researchers and their institutions are accountable for ensuring the confidentiality of human genomic data, instead of the cloud service provider. There is an imperative need to develop practical and rigorous privacy protection methods to alleviate the technical burden from human genomic researchers. Several recent surveys [197, 200] discussed the relevant techniques. But it remains unclear how these techniques will perform when applied to real human genomic data. There is a lack of direct comparison of different methods in real-world scenarios. Some recent efforts between the computer science community and the biomedical informatics community to jointly tackle the computation and privacy challenges seem promising [201] and more collaborations are necessary to push the fronts.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Shengda Lin and Song Yi contributed equally to this work and should be considered co-first authors.

Acknowledgments

References

[1] B. W. Stewart and C. P. Wild, *World Cancer Report 2014*, World Health Organization, Geneva, Switzerland, 2014.

[2] GBD 2013 Mortality and Causes of Death Collaborators, "Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 385, no. 9963, pp. 117–171, 2015.

[3] A. Forner, J. M. Llovet, and J. Bruix, "Hepatocellular carcinoma," *The Lancet*, vol. 379, no. 9822, pp. 1245–1255, 2012.

[4] Y. Shaib and H. B. El-Serag, "The epidemiology of cholangiocarcinoma," *Seminars in Liver Disease*, vol. 24, no. 2, pp. 115–125, 2004.

[5] J. Yu, J. Shen, T. T. Sun, X. Zhang, and N. Wong, "Obesity, insulin resistance, NASH and hepatocellular carcinoma," *Seminars in Cancer Biology*, vol. 23, no. 6, part B, pp. 483–491, 2013.

[6] K. A. Wetterstrand, *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*, 2015.

[7] Y. Totoki, K. Tatsuno, S. Yamamoto et al., "High-resolution characterization of a hepatocellular carcinoma genome," *Nature Genetics*, vol. 43, no. 5, pp. 464–471, 2011.

[8] A. Fujimoto, Y. Totoki, T. Abe et al., "Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators," *Nature Genetics*, vol. 44, no. 7, pp. 760–764, 2012.

[9] Z. Kan, H. Zheng, X. Liu et al., "Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma," *Genome Research*, vol. 23, no. 9, pp. 1422–1433, 2013.

[10] C. Guichard, G. Amaddeo, S. Imbeaud et al., "Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma," *Nature Genetics*, vol. 44, no. 6, pp. 694–698, 2012.

[11] R. Pinyol, V. Tovar, and J. M. Llovet, "TERT promoter mutations: gatekeeper and driver of hepatocellular carcinoma," *Journal of Hepatology*, vol. 61, no. 3, pp. 685–687, 2014.

[12] J. M. Llovet, A. Villanueva, A. Lachenmayer, and R. S. Finn, "Advances in targeted therapies for hepatocellular carcinoma in the genomic era," *Nature Reviews. Clinical Oncology*, vol. 12, no. 7, pp. 408–424, 2015.

[13] K. Schulze, S. Imbeaud, E. Letouzé et al., "Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets," *Nature Genetics*, vol. 47, no. 5, pp. 505–511, 2015.

[14] C. Bréchot, D. Gozuacik, Y. Murakami, and P. Paterlini-Bréchot, "Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC)," *Seminars in Cancer Biology*, vol. 10, no. 3, pp. 211–231, 2000.

[15] M. Minami, Y. Daimon, K. Mori et al., "Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis," *Oncogene*, vol. 24, no. 27, pp. 4340–4348, 2005.

[16] Z. Jiang, S. Jhunjhunwala, J. Liu et al., "The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients," *Genome Research*, vol. 22, no. 4, pp. 593–601, 2012.

[17] P. Paterlini-Bréchot, K. Saigo, Y. Murakami et al., "Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene," *Oncogene*, vol. 22, no. 25, pp. 3911–3916, 2003.

[18] D. Gozuacik, Y. Murakami, K. Saigo et al., "Identification of human cancer-related genes by naturally occurring hepatitis B virus DNA tagging," *Oncogene*, vol. 20, no. 43, pp. 6233–6240, 2001.

[19] W.-K. Sung, H. Zheng, S. Li et al., "Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma," *Nature Genetics*, vol. 44, no. 7, pp. 765–769, 2012.

[20] C. W. Greider and E. H. Blackburn, "Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts," *Cell*, vol. 43, no. 2, part 1, pp. 405–413, 1985.

[21] L. Hayflick and P. S. Moorhead, "The serial cultivation of human diploid cell strains," *Experimental Cell Research*, vol. 25, no. 3, pp. 585–621, 1961.

[22] N. W. Kim, M. A. Piatyszek, K. R. Prowse et al., "Specific association of human telomerase activity with immortal cells and cancer," *Science*, vol. 266, no. 5193, pp. 2011–2015, 1994.

[23] K. Collins, "The biogenesis and regulation of telomerase holoenzymes," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 7, pp. 484–494, 2006.

[24] Y. Mochizuki, J. He, S. Kulkarni, M. Bessler, and P. J. Mason, "Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 29, pp. 10756–10761, 2004.

[25] S. B. Cohen, M. E. Graham, G. O. Lovrecz, N. Bache, P. J. Robinson, and R. R. Reddel, "Protein composition of catalytically active human telomerase from immortal cells," *Science*, vol. 315, no. 5820, pp. 1850–1853, 2007.

[26] A. S. Venteicher, E. B. Abreu, Z. Meng et al., "A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis," *Science*, vol. 323, no. 5914, pp. 644–648, 2009.

[27] D. Fu and K. Collins, "Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation," *Molecular Cell*, vol. 28, no. 5, pp. 773–785, 2007.

[28] V. Pogačić, F. Dragon, and W. Filipowicz, "Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10," *Molecular and Cellular Biology*, vol. 20, no. 23, pp. 9028–9040, 2000.

[29] F. Dragon, V. Pogačić, and W. Filipowicz, "In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs," *Molecular and Cellular Biology*, vol. 20, no. 9, pp. 3037–3048, 2000.

[30] B. E. Snow, N. Erdmann, J. Cruickshank et al., "Functional conservation of the telomerase protein Est1p in humans," *Current Biology*, vol. 13, no. 8, pp. 698–704, 2003.

[31] P. Reichenbach, M. Höss, C. M. Azzalin, M. Nabholz, P. Bucher, and J. Lingner, "A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed," *Current Biology*, vol. 13, no. 7, pp. 568–574, 2003.

[32] H. J. Muller, "The remaking of chromosomes," *Collecting Net*, vol. 8, pp. 182–198, 1938.

[33] M. Meyerson, C. M. Counter, E. N. Eaton et al., "hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization," *Cell*, vol. 90, no. 4, pp. 785–795, 1997.

[34] T. M. Nakamura, G. B. Morin, K. B. Chapman et al., "Telomerase catalytic subunit homologs from fission yeast and human," *Science*, vol. 277, no. 5328, pp. 955–959, 1997.

[35] B. McClintock, "The stability of broken ends of chromosomes in Zea mays," *Genetics*, vol. 26, no. 2, pp. 234–282, 1941.

[36] C. B. Harley, A. B. Futcher, and C. W. Greider, "Telomeres shorten during ageing of human fibroblasts," *Nature*, vol. 345, no. 6274, pp. 458–460, 1990.

[37] R. C. Allsopp, H. Vaziri, C. Patterson et al., "Telomere length predicts replicative capacity of human fibroblasts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 21, pp. 10114–10118, 1992.

[38] T. Kitada, S. Seki, N. Kawakita, T. Kuroki, and T. Monna, "Telomere shortening in chronic liver diseases," *Biochemical and Biophysical Research Communications*, vol. 211, no. 1, pp. 33–39, 1995.

[39] N. Miura, I. Horikawa, A. Nishimoto et al., "Progressive telomere shortening and telomerase reactivation during hepatocellular carcinogenesis," *Cancer Genetics and Cytogenetics*, vol. 93, no. 1, pp. 56–62, 1997.

[40] Y. Urabe, K. Nouso, T. Higashi et al., "Telomere length in human liver diseases," *Liver*, vol. 16, no. 5, pp. 293–297, 1996.

[41] H. Aikata, H. Takaishi, Y. Kawakami et al., "Telomere reduction in human liver tissues with age and chronic inflammation," *Experimental Cell Research*, vol. 256, no. 2, pp. 578–582, 2000.

[42] F. W. Huang, E. Hodis, M. J. Xu, G. V. Kryukov, L. Chin, and L. A. Garraway, "Highly recurrent TERT promoter mutations in human melanoma," *Science*, vol. 339, no. 6122, pp. 957–959, 2013.

[43] S. Horn, A. Figl, P. S. Rachakonda et al., "TERT promoter mutations in familial and sporadic melanoma," *Science*, vol. 339, no. 6122, pp. 959–961, 2013.

[44] P. J. Killela, Z. J. Reitman, Y. Jiao et al., "TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 6021–6026, 2013.

[45] J. Vinagre, A. Almeida, H. Pópulo et al., "Frequency of *TERT* promoter mutations in human cancers," *Nature Communications*, vol. 4, article 2185, 2013.

[46] S. Borah, L. Xi, A. J. Zaug et al., "TERT promoter mutations and telomerase reactivation in urothelial cancer," *Science*, vol. 347, no. 6225, pp. 1006–1010, 2015.

[47] J. L. Stern, D. Theodorescu, B. Vogelstein, N. Papadopoulos, and T. R. Cech, "Mutation of the TERT promoter, switch to active chromatin, and monoallelic TERT expression in multiple cancers," *Genes & Development*, vol. 29, no. 21, pp. 2219–2224, 2015.

[48] J. C. Nault, M. Mallet, C. Pilati et al., "High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions," *Nature Communications*, vol. 4, article 2218, 2013.

[49] A. J. M. Walhout and M. Vidal, "Protein interaction maps for model organisms," *Nature Reviews Molecular Cell Biology*, vol. 2, no. 1, pp. 55–62, 2001.

[50] M. E. Cusick, N. Klitgord, M. Vidal, and D. E. Hill, "Interactome: gateway into systems biology," *Human Molecular Genetics*, vol. 14, supplement 2, pp. R171–R181, 2005.

[51] Y. Qi and H. Ge, "Modularity and dynamics of cellular networks," *PLoS Computational Biology*, vol. 2, no. 12, article e174, 2006.

[52] B. Causier, "Studying the interactome with the yeast two-hybrid system and mass spectrometry," *Mass Spectrometry Reviews*, vol. 23, no. 5, pp. 350–367, 2004.

[53] H. Ge, A. J. M. Walhout, and M. Vidal, "Integrating 'omic' information: a bridge between genomics and systems biology," *Trends in Genetics*, vol. 19, no. 10, pp. 551–560, 2003.

[54] S. K. Kim, J. Lund, M. Kiraly et al., "A gene expression map for *Caenorhabditis elegans*," *Science*, vol. 293, no. 5537, pp. 2087–2092, 2001.

[55] R. Kiyama and Y. Zhu, "DNA microarray-based gene expression profiling of estrogenic chemicals," *Cellular and Molecular Life Sciences*, vol. 71, no. 11, pp. 2065–2082, 2014.

[56] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[57] K. C. Gunsalus, H. Ge, A. J. Schetter et al., "Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis," *Nature*, vol. 436, no. 7052, pp. 861–865, 2005.

[58] Z. Du, T. Sun, E. Hacisuleyman et al., "Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer," *Nature Communications*, vol. 7, p. 10982, 2016.

[59] H. Guo, N. T. Ingolia, J. S. Weissman, and D. P. Bartel, "Mammalian microRNAs predominantly act to decrease target mRNA levels," *Nature*, vol. 466, no. 7308, pp. 835–840, 2010.

[60] A. K. Shalek, R. Satija, J. Shuga et al., "Single-cell RNA-seq reveals dynamic paracrine control of cellular variation," *Nature*, vol. 510, no. 7505, pp. 363–369, 2014.

[61] R. J. Hause, A. L. Stark, N. N. Antao et al., "Identification and validation of genetic variants that influence transcription factor and cell signaling protein levels," *The American Journal of Human Genetics*, vol. 95, no. 2, pp. 194–208, 2014.

[62] J. C. Simpson, R. Wellenreuther, A. Poustka, R. Pepperkok, and S. Wiemann, "Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing," *EMBO Reports*, vol. 1, no. 3, pp. 287–292, 2000.

[63] S. S. Nishizuka and G. B. Mills, "New era of integrated cancer biomarker discovery using reverse-phase protein arrays," *Drug Metabolism and Pharmacokinetics*, vol. 31, no. 1, pp. 35–45, 2016.

[64] A.-C. Gavin, P. Aloy, P. Grandi et al., "Proteome survey reveals modularity of the yeast cell machinery," *Nature*, vol. 440, no. 7084, pp. 631–636, 2006.

[65] N. J. Krogan, G. Cagney, H. Yu et al., "Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 440, no. 7084, pp. 637–643, 2006.

[66] P. D. Stenson, M. Mort, E. V. Ball, K. Shaw, A. D. Phillips, and D. N. Cooper, "The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine," *Human Genetics*, vol. 133, no. 1, pp. 1–9, 2014.

[67] M. J. Landrum, J. M. Lee, G. R. Riley et al., "ClinVar: public archive of relationships among sequence variation and human phenotype," *Nucleic Acids Research*, vol. 42, no. 1, pp. D980–D985, 2014.

[68] J. Amberger, C. Bocchini, and A. Hamosh, "A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®)," *Human Mutation*, vol. 32, no. 5, pp. 564–567, 2011.

[69] C. UniProt, "Activities at the universal protein resource (UniProt)," *Nucleic Acids Research*, vol. 42, pp. D191–D198, 2013.

[70] C. Kandoth, M. D. McLellan, F. Vandin et al., "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013.

[71] The Cancer Genome Atlas Research Network, "Comprehensive molecular characterization of clear cell renal cell carcinoma," *Nature*, vol. 499, no. 7456, pp. 43–49, 2013.

[72] Cancer Genome Atlas Research Network, "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*, vol. 489, no. 7417, pp. 519–525, 2012.

[73] The International Cancer Genome Consortium, "International network of cancer genome projects," *Nature*, vol. 464, no. 7291, pp. 993–998, 2010.

[74] J. Barretina, G. Caponigro, N. Stransky et al., "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012.

[75] J. Gao, B. A. Aksoy, U. Dogrusoz et al., "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal," *Science Signaling*, vol. 6, no. 269, article pl1, 2013.

[76] F. Piano, A. J. Schetter, D. G. Morton et al., "Gene clustering based on RNAi phenotypes of ovary-enriched genes in C. elegans," *Current Biology*, vol. 12, no. 22, pp. 1959–1964, 2002.

[77] A. J. Walhout, R. Sordella, X. Lu et al., "Protein interaction mapping in *C. elegans* using proteins involved in vulval development," *Science*, vol. 287, no. 5450, pp. 116–122, 2000.

[78] S. Li, C. M. Armstrong, N. Bertin et al., "A map of the interactome network of the metazoan *C. elegans*," *Science*, vol. 303, no. 5657, pp. 540–543, 2004.

[79] J.-F. Rual, K. Venkatesan, T. Hao et al., "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, no. 7062, pp. 1173–1178, 2005.

[80] A.-C. Gavin, M. Bösche, R. Krause et al., "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, pp. 141–147, 2002.

[81] Y. Ho, A. Gruhler, A. Heilbut et al., "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry," *Nature*, vol. 415, no. 6868, pp. 180–183, 2002.

[82] J. Ziauddin and D. M. Sabatini, "Microarrays of cells expressing defined cDNAs," *Nature*, vol. 411, no. 6833, pp. 107–110, 2001.

[83] J. L. Hartley, G. F. Temple, and M. A. Brasch, "DNA cloning using in vitro site-specific recombination," *Genome Research*, vol. 10, no. 11, pp. 1788–1795, 2000.

[84] A. J. M. Walhout, G. F. Temple, M. A. Brasch et al., "GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes," *Methods in Enzymology*, vol. 328, pp. 575–592, 2000.

[85] F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer, "A vision for the future of genomics research," *Nature*, vol. 422, no. 6934, pp. 835–847, 2003.

[86] R. L. Strausberg, E. A. Feingold, L. H. Grouse et al., "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16899–16903, 2002.

[87] J.-F. Rual, T. Hirozane-Kishikawa, T. Hao et al., "Human ORFeome version 1.1: a platform for reverse proteomics," *Genome Research*, vol. 14, no. 10 B, pp. 2128–2135, 2004.

[88] P. Lamesch, N. Li, S. Milstein et al., "hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes," *Genomics*, vol. 89, no. 3, pp. 307–315, 2007.

[89] X. Yang, J. S. Boehm, X. Yang et al., "A public genome-scale lentiviral expression library of human ORFs," *Nature Methods*, vol. 8, no. 8, pp. 659–661, 2011.

[90] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, no. 2, pp. 101–113, 2004.

[91] M. Vidal, "Interactome modeling," *FEBS Letters*, vol. 579, no. 8, pp. 1834–1838, 2005.

[92] A. Chatr-Aryamontri, B.-J. Breitkreutz, R. Oughtred et al., "The BioGRID interaction database: 2015 update," *Nucleic Acids Research*, vol. 43, no. 1, pp. D470–D478, 2015.

[93] A. Ruepp, B. Waegele, M. Lechner et al., "CORUM: the comprehensive resource of mammalian protein complexes—2009," *Nucleic Acids Research*, vol. 38, no. 1, pp. D497–D501, 2009.

[94] G. D. Bader, D. Betel, and C. W. V. Hogue, "BIND: the Biomolecular Interaction Network Database," *Nucleic Acids Research*, vol. 31, no. 1, pp. 248–250, 2003.

[95] I. Xenarios, Ł. Salwínski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Research*, vol. 30, no. 1, pp. 303–305, 2002.

[96] D. Szklarczyk, A. Franceschini, S. Wyder et al., "STRING v10: protein-protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. 1, pp. D447–D452, 2015.

[97] S. Peri, J. D. Navarro, T. Z. Kristiansen et al., "Human protein reference database as a discovery resource for proteomics," *Nucleic Acids Research*, vol. 32, pp. D497–D501, 2004.

[98] A. Chatr-aryamontri, A. Ceol, L. M. Palazzi et al., "MINT: the Molecular INTeraction database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D572–D574, 2007.

[99] J. Vlasblom, K. Zuberi, H. Rodriguez et al., "Novel function discovery with GeneMANIA: a new integrated resource for gene function prediction in *Escherichia coli*," *Bioinformatics*, vol. 31, no. 3, pp. 306–310, 2015.

[100] P. Pagel, S. Kovac, M. Oesterheld et al., "The MIPS mammalian protein-protein interaction database," *Bioinformatics*, vol. 21, no. 6, pp. 832–834, 2005.

[101] U. Stelzl, U. Worm, M. Lalowski et al., "A human protein-protein interaction network: a resource for annotating the proteome," *Cell*, vol. 122, no. 6, pp. 957–968, 2005.

[102] T. Rolland, M. Taşan, B. Charloteaux et al., "A proteome-scale map of the human interactome network," *Cell*, vol. 159, no. 5, pp. 1212–1226, 2014.

[103] P. C. Havugimana, G. T. Hart, T. Nepusz et al., "A census of human soluble protein complexes," *Cell*, vol. 150, no. 5, pp. 1068–1081, 2012.

[104] E. L. Huttlin, L. Ting, R. J. Bruckner et al., "The BioPlex Network: a systematic exploration of the human interactome," *Cell*, vol. 162, no. 2, pp. 425–440, 2015.

[105] M. Hein, N. Hubner, I. Poser et al., "A human interactome in three quantitative dimensions organized by stoichiometries and abundances," *Cell*, vol. 163, no. 3, pp. 712–723, 2015.

[106] I. A. Adzhubei, S. Schmidt, L. Peshkin et al., "A method and server for predicting damaging missense mutations," *Nature Methods*, vol. 7, no. 4, pp. 248–249, 2010.

[107] R. D. Finn, J. Clements, and S. R. Eddy, "HMMER web server: interactive sequence similarity searching," *Nucleic Acids Research*, vol. 39, no. 2, pp. W29–W37, 2011.

[108] Z. Dosztányi, V. Csizmok, P. Tompa, and I. Simon, "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content," *Bioinformatics*, vol. 21, no. 16, pp. 3433–3434, 2005.

[109] C. A. F. Andersen, A. G. Palmer, S. Brunak, and B. Rost, "Continuum secondary structure captures protein flexibility," *Structure*, vol. 10, no. 2, pp. 175–184, 2002.

[110] R. Guerois, J. E. Nielsen, and L. Serrano, "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations," *Journal of Molecular Biology*, vol. 320, no. 2, pp. 369–387, 2002.

[111] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The FoldX web server: an online force field," *Nucleic Acids Research*, vol. 33, no. 2, pp. W382–W388, 2005.

[112] P. W. Rose, C. Bi, W. F. Bluhm et al., "The RCSB Protein Data Bank: new resources for research and education," *Nucleic Acids Research*, vol. 41, no. 1, pp. D475–D482, 2013.

[113] E. Krissinel and K. Henrick, "Inference of macromolecular assemblies from crystalline state," *Journal of Molecular Biology*, vol. 372, no. 3, pp. 774–797, 2007.

[114] R. Mosca, A. Céol, A. Stein, R. Olivella, and P. Aloy, "3did: a catalog of domain-based interactions of known three-dimensional structure," *Nucleic Acids Research*, vol. 42, no. 1, pp. D374–D379, 2014.

[115] Q. C. Zhang, D. Petrey, L. Deng et al., "Structure-based prediction of protein-protein interactions on a genome-wide scale," *Nature*, vol. 490, no. 7421, pp. 556–560, 2012.

[116] J. Das and H. Yu, "HINT: high-quality protein interactomes and their applications in understanding human disease," *BMC Systems Biology*, vol. 6, article 92, 2012.

[117] M. J. Meyer, R. Lapcevic, A. E. Romero et al., "mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome," *Human Mutation*, vol. 37, no. 5, pp. 447–456, 2016.

[118] N. Sahni, S. Yi, M. Taipale et al., "Widespread macromolecular interaction perturbations in human genetic disorders," *Cell*, vol. 161, no. 3, pp. 647–660, 2015.

[119] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, 2001.

[120] J.-D. J. Han, N. Berlin, T. Hao et al., "Evidence for dynamically organized modularity in the yeast protein-protein interaction network," *Nature*, vol. 430, no. 6995, pp. 88–93, 2004.

[121] A. Hamosh, A. F. Scott, J. Amberger, C. Bocchini, D. Valle, and V. A. McKusick, "Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders," *Nucleic Acids Research*, vol. 30, no. 1, pp. 52–55, 2002.

[122] A. J. Butte and I. S. Kohane, "Creation and implications of a phenome-genome network," *Nature Biotechnology*, vol. 24, no. 1, pp. 55–62, 2006.

[123] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 21, pp. 8685–8690, 2007.

[124] T. Friedmann and R. Roblin, "Gene therapy for human genetic disease?" *Science*, vol. 175, no. 4025, pp. 949–955, 1972.

[125] R. M. Blaese, K. W. Culver, A. D. Miller et al., "T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years," *Science*, vol. 270, no. 5235, pp. 475–480, 1995.

[126] A. Abbott, "Gene therapy. Italians first to use stem cells," *Nature*, vol. 356, no. 6369, p. 465, 1992.

[127] M. Cavazzana-Calvo, S. Hacein-Bey, G. De Saint Basile et al., "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease," *Science*, vol. 288, no. 5466, pp. 669–672, 2000.

[128] A. Aiuti, S. Slavin, M. Aker et al., "Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning," *Science*, vol. 296, no. 5577, pp. 2410–2413, 2002.

[129] S. Hacein-Bey-Abina, C. Von Kalle, M. Schmidt et al., "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1," *Science*, vol. 302, no. 5644, pp. 415–419, 2003.

[130] S. Pearson, H. Jia, and K. Kandachi, "China approves first gene therapy," *Nature Biotechnology*, vol. 22, no. 1, pp. 3–4, 2004.

[131] L. Cong, F. A. Ran, D. Cox et al., "Multiplex genome engineering using CRISPR/Cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.

[132] P. Mali, L. Yang, K. M. Esvelt et al., "RNA-guided human genome engineering via Cas9," *Science*, vol. 339, no. 6121, pp. 823–826, 2013.

[133] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, "RNA-programmed genome editing in human cells," *eLife*, vol. 2, Article ID e00471, 2013.

[134] S. W. Cho, S. Kim, J. M. Kim, and J.-S. Kim, "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease," *Nature Biotechnology*, vol. 31, no. 3, pp. 230–232, 2013.

[135] H. Yin, W. Xue, S. Chen et al., "Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype," *Nature Biotechnology*, vol. 32, no. 6, pp. 551–553, 2014.

[136] C. Long, J. R. McAnally, J. M. Shelton, A. A. Mireault, R. Bassel-Duby, and E. N. Olson, "Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA," *Science*, no. 6201, pp. 1184–1188, 2014.

[137] J. Shi, E. Wang, J. P. Milazzo, Z. Wang, J. B. Kinney, and C. R. Vakoc, "Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains," *Nature Biotechnology*, vol. 33, no. 6, pp. 661–667, 2015.

[138] S. Zhen, L. Hua, Y. Takahashi, S. Narita, Y.-H. Liu, and Y. Li, "In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by CRISPR/Cas9," *Biochemical and Biophysical Research Communications*, vol. 450, no. 4, pp. 1422–1426, 2014.

[139] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and applications of CRISPR-Cas9 for genome engineering," *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[140] A. V. Wright, J. K. Nuñez, and J. A. Doudna, "Biology and applications of CRISPR Systems: harnessing Nature's Toolbox for Genome Engineering," *Cell*, vol. 164, no. 1-2, pp. 29–44, 2016.

[141] F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, and P. D. Gregory, "Genome editing with engineered zinc finger nucleases," *Nature Reviews Genetics*, vol. 11, no. 9, pp. 636–646, 2010.

[142] J. K. Joung and J. D. Sander, "TALENs: a widely applicable technology for targeted genome editing," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 1, pp. 49–55, 2013.

[143] E. Deltcheva, K. Chylinski, C. M. Sharma et al., "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III," *Nature*, vol. 471, no. 7340, pp. 602–607, 2011.

[144] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," *Nature Biotechnology*, vol. 31, no. 3, pp. 233–239, 2013.

[145] P. D. Hsu, D. A. Scott, J. A. Weinstein et al., "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nature Biotechnology*, vol. 31, no. 9, pp. 827–832, 2013.

[146] Y. Fu, J. A. Foden, C. Khayter et al., "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells," *Nature Biotechnology*, vol. 31, no. 9, pp. 822–826, 2013.

[147] V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, and D. R. Liu, "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity," *Nature Biotechnology*, vol. 31, no. 9, pp. 839–843, 2013.

[148] S. W. Cho, S. Kim, Y. Kim et al., "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases," *Genome Research*, vol. 24, no. 1, pp. 132–141, 2014.

[149] P. Mali, J. Aach, P. B. Stranges et al., "CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering," *Nature Biotechnology*, vol. 31, no. 9, pp. 833–838, 2013.

[150] F. A. Ran, P. D. Hsu, C.-Y. Lin et al., "Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2013.

[151] S. Q. Tsai, N. Wyvekens, C. Khayter et al., "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing," *Nature Biotechnology*, vol. 32, no. 6, pp. 569–576, 2014.

[152] J. P. Guilinger, D. B. Thompson, and D. R. Liu, "Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification," *Nature Biotechnology*, vol. 32, no. 6, pp. 577–582, 2014.

[153] K. Chiba, J. Z. Johnson, J. M. Vogan, T. Wagner, J. M. Boyle, and D. Hockemeyer, "Cancer-associated TERT promoter mutations abrogate telomerase silencing," *eLife*, vol. 4, Article ID e07918, pp. 1–20, 2015.

[154] L. Xi, J. C. Schmidt, A. J. Zaug, D. R. Ascarrunz, and T. R. Cech, "A novel two-step genome editing strategy with CRISPR-Cas9 provides new insights into telomerase action and TERT gene expression," *Genome Biology*, vol. 16, no. 1, p. 231, 2015.

[155] N. K. Paulk, K. Wursthorn, Z. Wang, M. J. Finegold, M. A. Kay, and M. Grompe, "Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo," *Hepatology*, vol. 51, no. 4, pp. 1200–1208, 2010.

[156] F. A. Ran, L. Cong, W. X. Yan et al., "In vivo genome editing using *Staphylococcus aureus* Cas9," *Nature*, vol. 520, no. 7546, pp. 186–191, 2015.

[157] European Association for the Study of the Liver and European Organisation for Research and Treatment of Cancer, "EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma," *Journal of Hepatology. Supplement*, vol. 56, no. 4, pp. 908–943, 2012.

[158] J. A. Rhim, E. P. Sandgren, J. L. Degen, R. D. Palmiter, and R. L. Brinster, "Replacement of diseased mouse liver by hepatic cell transplantation," *Science*, vol. 263, no. 5150, pp. 1149–1152, 1994.

[159] K. Overturf, M. Al-Dhalimy, C.-N. Ou, M. Finegold, and M. Grompe, "Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes," *The American Journal of Pathology*, vol. 151, no. 5, pp. 1273–1280, 1997.

[160] A. J. Matas, D. E. R. Sutherland, M. W. Steffes et al., "Hepatocellular transplantation for metabolic deficiencies: decrease of plasma bilirubin in Gunn rats," *Science*, vol. 192, no. 4242, pp. 892–894, 1976.

[161] K. Overturf, M. Al-Dhalimy, R. Tanguay et al., "Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I," *Nature Genetics*, vol. 12, no. 3, pp. 266–273, 1996.

[162] B.-S. Ding, D. J. Nolan, J. M. Butler et al., "Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration," *Nature*, vol. 468, no. 7321, pp. 310–315, 2010.

[163] H. Malhi, B. Joseph, M. L. Schilsky, and S. Gupta, "Development of cell therapy strategies to overcome copper toxicity in the LEC rat model of Wilson disease," *Regenerative Medicine*, vol. 3, no. 2, pp. 165–173, 2008.

[164] M. Minato, D. Houssin, I. Demma et al., "Transplantation of hepatocytes for treatment of surgically induced acute hepatic failure in the rat," *European Surgical Research*, vol. 16, no. 3, pp. 162–169, 1984.

[165] D. E. R. Sutherland, M. Numata, A. J. Matas, R. L. Simmons, and J. S. Najarian, "Hepatocellular transplantation in acute liver failure," *Surgery*, vol. 82, no. 1, pp. 124–132, 1977.

[166] S. N. Bhatia, G. H. Underhill, K. S. Zaret, and I. J. Fox, "Cell and tissue engineering for liver disease," *Science Translational Medicine*, vol. 6, no. 245, Article ID 245sr2, 2014.

[167] B. Joseph, H. Malhi, K. K. Bhargava, C. J. Palestro, R. S. McCuskey, and S. Gupta, "Kupffer cells participate in early clearance of syngeneic hepatocytes transplanted in the rat liver," *Gastroenterology*, vol. 123, no. 5, pp. 1677–1685, 2002.

[168] N. Krohn, S. Kapoor, Y. Enami et al., "Hepatocyte transplantation-induced liver inflammation is driven by cytokines-chemokines associated with neutrophils and Kupffer cells," *Gastroenterology*, vol. 136, no. 5, pp. 1806–1817, 2009.

[169] S. Gupta, P. Rajvanshi, R. Sokhi et al., "Entry and integration of transplanted hepatocytes in rat liver plates occur by disruption of hepatic sinusoidal endothelium," *Hepatology*, vol. 29, no. 2, pp. 509–519, 1999.

[170] M. Huch, C. Dorrell, S. F. Boj et al., "In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration," *Nature*, vol. 494, no. 7436, pp. 247–250, 2013.

[171] M. Huch, H. Gehart, R. Van Boxtel et al., "Long-term culture of genome-stable bipotent stem cells from adult human liver," *Cell*, vol. 160, no. 1-2, pp. 299–312, 2015.

[172] C. Dorrell, L. Erker, J. Schug et al., "Prospective isolation of a bipotential clonogenic liver progenitor cell in adult mice," *Genes & Development*, vol. 25, no. 11, pp. 1193–1203, 2011.

[173] K. Furuyama, Y. Kawaguchi, H. Akiyama et al., "Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine," *Nature Genetics*, vol. 43, no. 1, pp. 34–41, 2011.

[174] Y. Malato, S. Naqvi, N. Schürmann et al., "Fate tracing of mature hepatocytes in mouse liver homeostasis and regeneration," *Journal of Clinical Investigation*, vol. 121, no. 12, pp. 4850–4860, 2011.

[175] B. D. Tarlow, C. Pelz, W. E. Naugler et al., "Bipotential adult liver progenitors are derived from chronically injured mature hepatocytes," *Cell Stem Cell*, vol. 15, no. 5, pp. 605–618, 2014.

[176] K. Yanger, D. Knigin, Y. Zong et al., "Adult hepatocytes are generated by self-duplication rather than stem cell differentiation," *Cell Stem Cell*, vol. 15, no. 3, pp. 340–349, 2014.

[177] J. R. Schaub, Y. Malato, C. Gormond, and H. Willenbring, "Evidence against a stem cell origin of new hepatocytes in a common mouse model of chronic liver injury," *Cell Reports*, vol. 8, no. 4, pp. 933–939, 2014.

[178] D. Rodrigo-Torres, S. Affò, M. Coll et al., "The biliary epithelium gives rise to liver progenitor cells," *Hepatology*, vol. 60, no. 4, pp. 1367–1377, 2014.

[179] B. Wang, L. Zhao, M. Fish, C. Y. Logan, and R. Nusse, "Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver," *Nature*, vol. 524, no. 7564, pp. 180–185, 2015.

[180] J. Font-Burgada, S. Shalapour, S. Ramaswamy et al., "Hybrid periportal hepatocytes regenerate the injured liver without giving rise to cancer," *Cell*, vol. 162, no. 4, pp. 766–779, 2015.

[181] J. Kuhlmann, "Alternative strategies in drug development: clinical pharmacological aspects," *International Journal of Clinical Pharmacology and Therapeutics*, vol. 37, no. 12, pp. 575–583, 1999.

[182] S. Huang, "Rational drug discovery: what can we learn from regulatory networks?" *Drug Discovery Today*, vol. 7, supplement 20, pp. S163–S169, 2002.

[183] J. Couzin, "Cancer drugs: smart weapons prove tough to design," *Science*, vol. 298, no. 5593, pp. 522–525, 2002.

[184] H. M. J. Werner, G. B. Mills, and P. T. Ram, "Cancer systems biology: a peek into the future of patient care?" *Nature Reviews Clinical Oncology*, vol. 11, no. 3, pp. 167–176, 2014.

[185] A. Beheshti, D. Neuberg, J. T. McDonald, C. R. Vanderburg, and A. M. Evens, "The impact of age and sex in DLBCL: systems biology analyses identify distinct molecular changes and signaling networks," *Cancer Informatics*, vol. 14, pp. 141–148, 2015.

[186] L. Berglund, E. Björling, P. Oksvold et al., "A genecentric human protein atlas for expression profiles based on antibodies," *Molecular and Cellular Proteomics*, vol. 7, no. 10, pp. 2019–2027, 2008.

[187] M. Uhlén, L. Fagerberg, B. M. Hallström et al., "Proteomics. Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, Article ID 1260419, 2015.

[188] M. Uhlen, P. Oksvold, L. Fagerberg et al., "Towards a knowledge-based Human Protein Atlas," *Nature Biotechnology*, vol. 28, no. 12, pp. 1248–1250, 2010.

[189] M. Uhlen, E. Björling, C. Agaton et al., "A human protein atlas for normal and cancer tissues based on antibody proteomics," *Molecular & Cellular Proteomics*, vol. 4, no. 12, pp. 1920–1932, 2005.

[190] F. Pontén, K. Jirström, and M. Uhlen, "The human protein atlas—a tool for pathology," *Journal of Pathology*, vol. 216, no. 4, pp. 387–393, 2008.

[191] H. Kitano, "Computational systems biology," *Nature*, vol. 420, no. 6912, pp. 206–210, 2002.

[192] M. P. Murphy, "Current pharmacogenomic approaches to clinical drug development," *Pharmacogenomics*, vol. 1, no. 2, pp. 115–123, 2000.

[193] H. Kitano, "Systems biology: a brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002.

[194] J. L. Fox, "Obama catapults patient-empowered Precision Medicine," *Nature Biotechnology*, vol. 33, no. 4, pp. 325–325, 2015.

[195] N. Homer, S. Szelinger, M. Redman et al., "Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays," *PLoS Genetics*, vol. 4, no. 8, Article ID e1000167, 2008.

[196] M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich, "Identifying personal genomes by surname inference," *Science*, vol. 339, no. 6117, pp. 321–324, 2013.

[197] Y. Erlich and A. Narayanan, "Routes for breaching and protecting genetic privacy," *Nature Reviews Genetics*, vol. 15, no. 6, pp. 409–421, 2014.

[198] P. Claes and M. D. Shriver, "Establishing a multidisciplinary context for modeling 3D facial shape from DNA," *PLoS Genetics*, vol. 10, no. 11, Article ID e1004224, 2014.

[199] S. S. Shringarpure and C. D. Bustamante, "Privacy risks from genomic data-sharing beacons," *The American Journal of Human Genetics*, vol. 97, no. 5, pp. 631–646, 2015.

[200] M. Naveed, E. Ayday, E. W. Clayton et al., "Privacy in the genomic era," *ACM Computing Surveys*, vol. 48, no. 1, article 6, 2015.

[201] E. Check Hayden, "Cloud cover protects gene data," *Nature*, vol. 519, no. 7544, pp. 400–401, 2015.