

Overview of Gene Regulation in Stem Cell Network to Identify Therapeutic Targets Utilizing Genome Databases

Shihori Tanabe

Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, Tokyo, 158-8501, Japan

Abstract

Title: Overview of gene regulation in stem cell network to identify therapeutic targets utilizing genome databases.

Background: Recent major progress in bioinformatics has enabled the collection and accumulation of so-called big data in medical fields and cell biology. It is important to analyze and interpret these abundant data for appropriate application in therapeutics and the treatment of diseases.

Methods and Findings: Several databases have been introduced worldwide, and the utility of these databases is discussed with literatures. The databases are useful for analyzing genome mutations, gene expression, epigenetic regulation, gene ontology, stem cell phenotype alteration, risk prediction, species differences, and so on. Gene network analysis using these databases may identify targets for therapeutics and treatment relating stem cells.

Conclusions: Molecular network regulation is critical for understanding disease and treatment mechanisms.

Keywords: Bioinformatics, Database, Gene regulation, Stem cell

Abbreviations: AHR: Aryl Hydrocarbon Receptor; BRCA1: Breast Cancer 1 Early Onset; BRCA2: Breast Cancer 2 Early Onset; BSP: Bone Sialoprotein; CCLE: Cancer Cell Line Encyclopedia; CDH1: Cadherin 1 Type 1; CDH2: Cadherin 2 Type 1 N-Cadherin (Neuronal); CEBP: CCAAT-Enhancer-Binding Protein; CGC: Cancer Gene Census; Chibe: Chisio Biopax Editor; COSMIC: Catalogue Of Somatic Mutations In Cancer; DAVID: Database For Annotation; Visualization And Integrated Discovery; Dbgap: Database Of Genotypes And Phenotypes; Dbhimo: Database For Histone-Modifying Enzymes; DDBJ: DNA Data Bank Of Japan; DGV: Database Of Genomic Variants; DLX5: Distal-Less Homeobox5; EGA: European Genome-Phenome Archive; EMT: Epithelial-Mesenchymal Transition; ERG: Epigenetic Regulator Genes; ETS: E26 Transformation-Specific; FGFR2: Fibroblast Growth Factor Receptor 2; GABP: GA-Binding Protein; Gemdbj: Genome Medicine Database Of Japan; GEO: Gene Expression Omnibus; Gtex: Genotype-Tissue Expression; GWAS: Genome-Wide Association Study; HME: Histone-Modifying Enzymes; Hopx: Homeodomain-Only Protein Gene; HSPC: Hematopoietic Stem/Progenitor Cell; IBD: Inflammatory Bowel Disease; ICGC: International Cancer Genome Consortium; ILC3: Group 3 Innate Lymphoid Cell; JSNP: Japanese Single Nucleotide Polymorphisms; KEGG: Kyoto Encyclopedia Of Genes And Genomes; Lncrna: Long Noncoding RNA; MDCK: Madin-Darby Canine Kidney; MHC: Major Histocompatibility Complex; MHCII: MHC Class II; MSC: Mesenchymal Stem Cell; OB-BMST: Osteoblastic Bone Metastasis-Associated Stroma Transcriptome; PGC: Primordial Germ Cell; 15-PGDH: 15-Hydroxyprostaglandin Dehydrogenase; PP2A: Protein Serine/Threonine Phosphatase Type 2A; PPI: Protein-Protein Interaction; PSCA: Prostate Stem Cell Antigen; SNP: Single Nucleotide Polymorphism; TCGA: The Cancer Genome Atlas; TERT: Telomerase Reverse Transcriptase; Th17: T Helper 17; Uniprotkb: Uniprot Knowledgebase

Corresponding author:
Shihori Tanabe

✉ stanabe@nihs.go.jp

Senior Researcher, Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1, Kami-yoga, Setagaya-ku, Tokyo, 158-8501, Japan.

Tel: +81-3-3700-1141 (ext. 304)

Fax: +81-3-3707-6950

Citation: Tanabe S. Overview of Gene Regulation in Stem Cell Network to Identify Therapeutic Targets Utilizing Genome Databases. *Insights Stem Cells*. 2015, 1:1.

Received: October 31, 2015; Accepted: November 17, 2015; Published: November 19, 2015

Introduction

Recent advances in bioinformatics have enabled to predict cancer risk by gene alterations. In breast and ovarian cancer, risk is estimated based on the *breast cancer 1, early onset (BRCA1)* and *breast cancer 2, early onset (BRCA2)* mutations [1]. Increasing amounts of data have recently revealed many candidate genes to be examined for application in target therapies. Variations in the human transcriptome have been revealed using RNA sequence data generated by the Genotype-Tissue Expression (GTEx) project [2]. Comprehensive analyses have shown interindividual differences in gene expression and splicing variability [2]. The existence of tissue-specific transcriptional regulation has also been demonstrated using data from the GTEx project [2]. The differences in gene alterations between individuals highlight the importance of prognostic biomarkers for treatment sensitivity in diseases. A novel method based on OncoFinder pathway activation strength revealed that the JNK pathway (insulin signaling) and the mitochondrial apoptosis pathway are significantly correlated with the response to cetuximab, a monoclonal antibody against epidermal growth factor receptor (EGFR) used in colorectal cancer patients with wild-type K-ras [3].

Genetic variation in *prostate stem cell antigen (PSCA)* has been found to be associated with susceptibility to diffuse-type gastric cancer [4]. Single nucleotide polymorphisms (SNPs) in PSCA have been compared in diffuse-type gastric cancer cases and control subjects, which has revealed statistically significant SNPs [4]. Furthermore, SNPs in PSCA have been found to have a greater effect in diffuse-type than intestinal-type gastric cancer [4]. The allele and genotype frequencies of the rs2976392 SNP are different between the two types of the gastric cancer [4]. The levels of polymorphism may be different among various cell types.

The gene expression in diffuse-type gastric cancer and mesenchymal stem cells (MSCs) have been profiled, which has revealed that combination of *cadherin 1, type 1 (CDH1)* and *cadherin 2, type 1, N-cadherin (neuronal) (CDH2)* distinguishes the cancer cell phenotypes and provided insights of gene regulation in epithelial-mesenchymal transition (EMT) [5-8]. In this review, genome and gene regulation analyses performed with several databases, which are briefly introduced in the text, are overviewed mainly focused in stem cell and cancer networks.

Databases for Gene and Molecular Regulation

Recently, it has become possible to analyze gene regulation using abundant databases (Table 1). One such useful tool is the GTEx database [9]. In this database, a comprehensive overview and assessment of gene regulation in human tissues are provided based on genome-wide association studies (GWASs) [9]. According to the project, all data releases are also available to the public through the Database of Genotypes and Phenotypes (dbGaP), which is a repository that is charged to archive, curate

and distribute information related to genotype-phenotype interactions [10]. Epigenetics, including histone-modifying enzymes (HMEs), can be analyzed using the database for histone-modifying enzymes (dbHiMo) [11]. According to dbHiMo, HMEs are identified using a hidden Markov model-based pipeline, which will be useful tools for revealing epigenetics/epigenomics [11].

De novo transcriptome assembly databases for the central nervous system have been developed, and functional analyses as well as analyses of evolutionary relationships can be performed using BLASTX [12]. Trinity, for de-novo assembly, and Illumina HiSeq2000, for RNA-Seq, were used to conduct analyses that generated an annotation set of 22,604 transcripts that were aligned against the Swiss-Prot database using BLASTX [12]. Genomic signatures undergo evolutionary transitions during environmental alterations, which may be a target for analyses using these databases [13]. The promoter regions of 5865 single-copy orthologs among 10 species were analyzed to calculate a motif score for 188 *Drosophila melanogaster* transcription factors with at least one ortholog in each of the 10 bees, and the motif score was found to correlate with social complexity using phylogenetically independent contrasts [13]. The results of this research showed that the complexity of the gene network has been involved in the evolution of eusociality, indicating that changes in gene regulation were critical for the evolutionary transition in biological organization [13].

The Kyoto Encyclopedia of Genes and Genomes (KEGG), which aims to compile knowledge from information, is a useful tool for analyzing the regulation of genes and genomes [14,15]. KEGG pathway maps are mainly focused on the biological interpretation of genome sequences and other high-throughput data [14,15]. A functional analysis and visualization tool for omics data referred to as FuncTree has been developed using the KEGG database [16]. The role of FuncTree is mapping omics data onto the Functional Tree map, which is a circular dendrogram showing the relationship of biological functions in the KEGG database [16].

The DNA Data Bank of Japan (DDBJ) databases include analytical services for biological information [17]. The DDBJ sequence databases release datasets including genome, genome survey sequence (GSS), transcriptome shotgun assembly (TSA), high-throughput cDNA sequence (HTC), expressed sequence tag (EST), and transcriptome data for various species [17]. The Japanese Single Nucleotide Polymorphisms (JSNP) database includes data on 150,000 SNPs from the Japanese population [18]. The ERGO genome analysis and discovery suite contains biological data from genomics, biochemistry, and high-throughput expression profiling analyses [19]. Metabolic networks and antimicrobial drug targets for Category A-designated bioterrorism agents, including *Bacillus anthracis* (anthrax), *Francisella tularensis* (tularemia) and *Yersinia pestis* (bubonic plague), have been analyzed with ERGO and KEGG [20]. Human genomic variation has been identified and collated into the Genome Variation Database, which is currently

Table 1 Databases for analyzing gene regulation.

Name	Full Name or Source	Reference	URL
GTEx	Genotype-Tissue Expression	9	http://gtexportal.org
dbGaP	Database of Genotypes and Phenotypes	10	http://www.ncbi.nlm.nih.gov/gap
dbHiMo	Database for Histone-modifying Enzymes	11	http://hme.riceblast.snu.ac.kr/
KEGG	Kyoto Encyclopedia of Genes and Genomes	14-16	www.genome.jp/kegg
DDBJ	DNA Data Bank of Japan	17	www.ddbj.nig.ac.jp
JSNP	Japanese Single Nucleotide Polymorphisms	18	http://snp.ims.u-tokyo.ac.jp/
ERGO	Integrated Genomics; Inc.	19;20	www.igenbio.com/ergo
DGV	Database of Genomic Variants	21-23	http://projects.tcag.ca/variation
COSMIC	Catalogue of Somatic Mutations in Cancer	24;25	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
GeMDBJ	Genome Medicine Database of Japan	26	http://gemdbj.nibio.go.jp/
ICGC	International Cancer Genome Consortium	27	http://dcc.icgc.org
TCGA	The Cancer Genome Atlas	29;30	http://cancergenome.nih.gov/dataportal
EGA	European Genome-phenome Archive	31	https://www.ebi.ac.uk/ega/
GEO	Gene Expression Omnibus	32;33	http://www.ncbi.nlm.nih.gov/geo
DAVID	Database for Annotation; Visualization and Integrated Discovery	34	http://david.abcc.ncifcrf.gov
cBioPortal	cBioPortal for Cancer Genomics	35	http://cbioportal.org
UniProtKB	UniProt Knowledgebase	36;37	www.uniprot.org http://www.expasy.org/sprot/ http://www.ebi.ac.uk/swissprot/
CCL	Cancer Cell Line Encyclopedia	38	http://www.broadinstitute.org/ccle/home
NextBio	Illumina	39	http://www.nextbio.com/

referred to as the Database of Genomic Variants (DGV) [21,22]. DGV includes information on copy number variations, SNPs, and genomic variations [21,22]. Using DGV, *par-3 family cell polarity regulator (PARD3)* was found to be microdeleted in squamous carcinomas and glioblastoma [23].

For analyzing somatic mutations in human cancer, the Catalogue Of Somatic Mutations In Cancer (COSMIC) is the comprehensive database [24]. Tumor antigens can be predicted *in silico* using COSMIC and the Cancer Gene Census (CGC) [25]. A database of missense mutation-derived peptides was assembled from the COSMIC database, followed by the identification of candidate peptides that can be related to tumor rejection antigens [25]. Information on mutant genes was collected from CGC, and major histocompatibility complex (MHC) class I binding peptides that are mutated in cancer were analyzed, with their binding being predicted [25]. Sharing of GWAS data related to cancer has been made possible via the Genome Medicine Database of Japan (GeMDBJ) [26]. GeMDBJ contains genome-wide SNP typing data related to Alzheimer's disease, gastric cancer, type 2 diabetes, hypertension, and asthma [26]. The International Cancer Genome Consortium (ICGC) is a database for characterizing genomic abnormalities in different cancer types [27]. The ICGC Data Portal includes ICGC, the Cancer Genome Atlas (TCGA), the Johns Hopkins University data, and the Tumor Sequencing Project [27,28]. TCGA is based on comprehensive sequencing of all protein-coding genes and transcripts in tumors [29]. The involvement of epigenetic regulator genes (ERGs) in human cancer was analyzed using mutation, copy number and expression data from 5943 tumors across 13 TCGA cancer types, which revealed that multiple ERGs, including *enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2)*, are co-regulated in the cell cycle network [30]. The European Genome-phenome Archive

(EGA) is a database for sharing the biomolecular, genetic and phenotypic data collected from human subjects [31]. The policy of distributed access-granting distinguishes EGA from dbGaP [31]. Authorization decisions for dbGaP's datasets are made by the US National Institutes of Health (NIH), whereas the data submitted to EGA are required to be consistent with the original consent agreements, national laws and applicable regulations [31].

Gene expression data collected worldwide are publicly available in the Gene Expression Omnibus (GEO) database [32,33]. GEO provides genomic data generated using microarrays and next-generation sequencing, which enables us to share and analyze the data to explore disease treatment targets [32,33]. The Database for Annotation, Visualization and Integrated Discovery (DAVID) includes a Gene Functional Classification Tool with a novel agglomeration algorithm to categorize genes in function [34]. The DAVID Gene Functional Classification Tool has shifted functional annotation analysis from being term- or gene-centric to biological module-centric [34].

Integrative analysis of cancer genomics and clinical profiles can be performed with the cBioPortal for Cancer Genomics database [35]. The cBioPortal for Cancer Genomics contains genomic data on somatic mutations, DNA copy-number alterations, mRNA and microRNA expression, DNA methylation, protein abundance, and phosphoprotein abundance in cancer, and networks can be analyzed in association with genes [35].

The UniProt consortium has completed the first draft of the complete human proteome in the UniProt Knowledgebase (UniProtKB) Swiss-Prot [36,37]. UniProtKB Swiss-Prot contains manually annotated data on protein functions, enzyme-specific information, biologically relevant domains and sites, post-translational modifications, subcellular locations, tissue

specificity, developmental-specific expression, structures, interactions, splicing isoforms and associated diseases [36,37]. The Broad Institute provides a collection of data from various species for bioinformatic analysis. One of the collaborative studies of the Broad Institute is the Cancer Cell Line Encyclopedia (CCLE) project, which compiles gene expression, chromosomal copy number and parallel sequencing data from human cancer cell lines [38]. The comprehensive database called NextBio is also useful to analyze the therapeutic target of the disease such as cancer [39].

Network Analysis

Gene expression profiles in hepatocellular carcinoma have been analyzed through gene ontology analysis, KEGG and Biocarta pathway enrichment analysis, followed by a protein-protein interaction (PPI) network analysis with Cytoscape software [40]. Genes that are differentially regulated in alcoholic hepatitis were analyzed using Cytoscape in a PPI network [41]. A network analysis was performed based on genes that are regulated in bladder transitional cell carcinoma, which led to the identification of the central nodes in the network and the selection of potential cancer markers [42].

The database cBioPortal for Cancer Genomics was used to analyze *protein serine/threonine phosphatase type 2A (PP2A)* regulation, which revealed that the PP2A complex is deregulated in 59.6% of basal breast tumors [43]. A tool for analyzing human biological pathways referred to as Chisio BioPAX Editor (ChiBE) has been developed using profiling data from the cBioPortal for Cancer Genomics and expression data from GEO [44]. ChiBE links DAVID for further annotation and gene-set-related analysis [44]. A candidate prognostic gene signature in advanced prostate cancer was analyzed with cBioPortal, which interrogates the Memorial Sloan Kettering Cancer Center Prostate Oncogenome Project dataset for changes in the expression of the genes enriched under the KEGG term "Cell Cycle" in clinical prostate cancer [45]. A biomarker panel with 7 genes has been identified in high-risk prostate cancer as a prognostic gene signature [45].

Gene Regulation in the Human Body

Many studies utilize the databases and bioinformatics approach to identify gene regulation in the human body. Intestinal stem cells exhibit variation in their ground states, which have been analyzed using genome bioinformatics techniques, such as microarray analysis, gene set enrichment analysis and exome capture sequencing analysis [46]. Differentiated intestinal stem cells can be marked as goblet (*Muc2⁺*), endocrine (chromogranin A⁺), and Paneth cells and based on polarized villin expression, whereas intestinal stem cells can be distinguished using *SRY (sex determining region Y)-box 9 (SOX9)*, *olfactomedin 4 (OLFM4)*, *prominin 1 (PROM1)*, also known as *CD133* and *leucin-rich repeat-containing G protein-coupled receptor 5 (LGR5)* from tracheobronchial stem cells [46].

In normal human skin, the mutations in 74 cancer genes were analyzed through ultradeep sequencing [47]. *NOTCH1*, whose receptors are key regulators of stem cell biology and targets of inactivating mutations in epithelial cancers, has been found to

be mutated in approximately 20% of normal human skin and to carry driver mutations in 60% of cutaneous squamous cell carcinomas [47]. The main differences between normal cells and cancer cells were found to be the number of driver mutations per cell [47]. The methylation patterns of *runx-related transcription factor 2 (RUNX2)*, *Sp7 transcription factor (SP7)* [also known as *osterix (OSX)*], *distal-less homeobox 5 (DLX5)* and *integrin-binding sialoprotein (IBSP)* [also known as *bone sialoprotein (BSP)*] differ during the osteoblastic differentiation of MSCs [48]. The epigenetic regulation of genes may affect the direction of cell differentiation. The differentiation direction in MSCs might be regulated with gene alterations and paracrine effect of the cells [49,50]. Based on GTEx data, protein-truncating variants were analyzed to reveal clinical interpretations of the genome [51]. It has been revealed that tissue-specific protein truncations exist, although further data collection and analysis will be needed to predict the molecular consequences of these variants [51].

It has been revealed that commensal bacteria-specific CD4⁺ T cells in the intestine are regulated by group 3 innate lymphoid cells (ILC3s) [52]. ILC3 expression of MHC class II (MHCII) is controlled by transcriptional pathways [52]. It has been found that MHCII expression in colonic ILC3s is reduced in pediatric inflammatory bowel disease (IBD), which highlights the significance of analyzing the mechanism of MHCII regulation in ILC3s [52]. The selection of commensal bacteria-specific CD4⁺ T cells in the intestine may be critical for the IBD mechanism [52].

The aryl hydrocarbon receptor (AHR), a type I nuclear receptor and cytosolic transcription factor, regulates T helper 17 (Th17)/T regulatory (Treg) pathways [53]. Th17 cell differentiation is regulated by AHR ligands, which gives rise to the possibility of targeting autoimmune and chronic inflammatory disease correlated with environmental toxins [53]. The pluripotency of primordial germ cells (PGCs) is regulated by *Blimp-1* and *Akt* [54]. *Blimp-1* has been found to suppress the downstream targets of pluripotency network genes such as *Myc*, *Klf4*, *Nanog*, *Oct3/4* and *Sox2* and lead to suppression of the pluripotency of PGCs [54]. In MSCs, the differentiation status can be modulated through epigenetic regulation [55]. Reverted cells from osteogenic-differentiated cells exhibit increased expression of *Nanog*, *Oct4* and *Sox2* [55]. It has been reported that mechanical forces regulating the development, organization, and function of multicellular tissues induce E-cadherin (or Cdh1)-dependent Yap1, a Hippo pathway transcription factor, and β -catenin activation to drive cell cycle entry [56]. When mechanical strain is applied to Madin-Darby canine kidney (MDCK) cells, transient Yap1 activation and subsequent cell cycle reentry occur, and nuclear localization and transcriptional activation of β -catenin are induced, leading to progression into S phase [56]. The Yap1 nuclear exclusion and β -catenin activation induced by mechanical strain require interactions with the extracellular domain of E-cadherin, suggesting a role of the cell adhesive molecule E-cadherin as a mechanical signal transducer [56].

In cellular programming and reprogramming, long noncoding RNAs (lncRNAs) play important roles [57]. In adult tissue stem cells, lncRNAs regulate differentiation and self-renewal, together with other molecules that are dependent on cellular fates [57]. In the

case of the regulation in the brain, the differentiation of embryonic stem cells into neuronal tissues is regulated by TUNA/megamind, a lncRNA required for the maintenance of pluripotency, and PTBP1, Nucleolin (NCL) and hnRNP-K [57]. The regulators inducing pluripotency are different in each dedifferentiated cell type [58]. Reprogramming processes are controlled by Oct4, Sox2, Klf4 and c-Myc in fibroblasts, whereas *mps1*, *plk1* and *cdc2* dedifferentiate zebrafish cardiomyocytes into a pluripotent state [58]. Immune B cells transdifferentiate into macrophages via CCAAT-enhancer-binding protein- α (CEBP α) and CEBP β regulation, and fibroblasts transdifferentiate into cardiomyocytes via regulation of GATA binding protein 4 (GATA4), T-box 5 (Tbx5) and myocyte enhancer factor 2C (Mef2C) [58-60]. EMT is regulated by Snail, and the transcription of Snail is suppressed by Oct4 in the reprogramming process [58]. In terms of tissue generation, the prostaglandin PGE2 supports stem cell expansion, which is supposed to be suppressed by 15-hydroxyprostaglandin dehydrogenase (15-PGDH) [61]. Inhibiting 15-PGDH with the molecule designated SW033291 has been reported to promote tissue repair [61].

Prediction of Risk from Gene Alterations

The gene alterations such as the regulation of Snail gene expression affect the cellular phenotype: Snail causes EMT and tumor progression [62]. The risk of diseases such as cancer may be predicted based on gene alterations. An important point is that gene regulation occurs in molecular networks, which reminds us to consider combinations of genes. Gene alterations are analyzed through network analysis in prostate cancer-induced osteoblastic bone metastasis [63]. Using interspecies difference in bioinformatics, osteoblastic bone metastasis-associated stroma transcriptome (OB-BMST) was generated through comparison with mouse transcripts and human transcripts in implanted cancer and the stroma, which is an elegant method for analyzing stroma genes [63]. According to the generated OB-BMST, *pleiotrophin* (PTN), *Eph receptor 3* (EPHA3) and *fascin actin-bundling protein 1* (FSCN1) were extracted as components of the bone-specific response to prostate cancer-induced osteoblastic bone metastasis [63]. Interspecies differences are interesting and important topics for investigation. A pluripotent state (i.e., the ability to differentiate into any cell type) can allow interspecies chimeric-competent cells to be obtained from region-selective pluripotent stem cells [64]. Interspecies cell type transitions may be a future direction to be explored.

Insights Regarding Genomic Data and Diseases

Genomic alterations such as *BRCA1* and *BRCA2* mutations occur in breast cancer patients and have been revealed as a risk factor for breast cancer [1]. Even in wild-type *BRCA1/2* breast cancer patients in the Sardinian population, it has been revealed that *TOX high mobility group box family member 3* (*TOX3*) and *fibroblast growth factor receptor 2* (*FGFR2*) play roles as susceptibility genes, which includes SNPs in *FGFR2* [65].

It has been revealed that optogenetic stimulation interacts with skeletal muscle via the light-sensitive channel Channelrhodopsin-2 [66]. The force generated from optogenetic-stimulated skeletal muscle leads to cell-type-specific activation, which may be a target of therapy for laryngeal paralysis or other locomotive syndromes [66].

The genetic landscape and genetic determinants of the hematopoietic stem/progenitor cell (HSPC) frequency were analyzed in a GWAS [67]. It was revealed that a *homeodomain-only protein gene* (*Hopx*) locus is associated with the frequency of HSPCs, such as Lineage[Lin]-Sca-1⁺ c-Kit⁺ (LSK) HSPCs, LSKCD150⁺CD48⁻ multipotent progenitors (MPPs), and LSKCD150⁺CD48⁺ cells, which are the most primitive long-term HSCs in mice bone marrow mononuclear cells [67]. In terms of the LSKCD150⁺CD48⁺ population, *Hopx*^{+/+} cells are dominant to *Hopx*^{-/-} cells, which means that the percentage of LSKCD150⁺CD48⁺ cells is low in the mononuclear cells of *Hopx*^{-/-} mice [67]. This finding indicates that *Hopx* plays some role in determining the HSPC frequency [67]. The multimeric GA-binding protein (GABP) transcription factor has been found to selectively bind and activate the mutant *telomerase reverse transcriptase* (*TERT*) encoding the catalytic subunit of telomerase in cancer [68]. GABP was identified as the critical E26 transformation-specific (ETS) transcription factor activating *TERT* expression in the context of highly recurrent promoter mutations [68]. These findings may indicate that the tandem flanking native ETS motifs interact with mutations in cancer, leading to the activation of *TERT*, which enables cells to overcome replicative senescence [68].

Genomic data have provided the abundant information about mutations in cancer, revealing interesting insights in stem cells and cancer [69]. The investigations of cellular networks in terms of gene, genome and proteins would be needed to reveal whole pictures of stem cells [70,71].

Conclusion

Our knowledge has increased due to recent advances in bioinformatics and computational capacity. How to efficiently utilize these data and knowledge is an important issue for future development of the data era. One of the useful directions for utilizing gene information may be the identification of targets for the treatment of diseases, including appropriate predictions.

Since stem cell phenotype alters in disease conditions, it is very important to analyze and understand gene regulation in stem cells. One of the example models is that stem cell clonal expansion exponentially occurred in cancer, which emphasizes the significance of analyzing stem cell alterations relating the combinations of genes and molecules.

Acknowledgments

The author would like to thank the members of the National Institute of Health Sciences and researchers for helpful discussions.

References

- 1 King MC, Marks JH, Mandell JB (2003) New York Breast Cancer Study Group Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* 302: 643-646.
- 2 Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J et al. (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* 348: 660-665.
- 3 Zhu Q, Izumchenko E, Aliper AM, Makarev E, Paz K, et al. (2015) Pathway activation strength is a novel independent prognostic biomarker for cetuximab sensitivity in colorectal cancer patients. *Human Genome Variation* 2: 15009.
- 4 Sakamoto H, Yoshimura K, Saeki N, Katai H, et al. (2008) The Study Group of Millennium Genome Project for Cancer Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 40: 730-740.
- 5 Tanabe S, Aoyagi K, Yokozaki H, Sasaki H (2015) Regulated genes in mesenchymal stem cells and gastric cancer. *World J Stem Cells* 7: 208-222.
- 6 Tanabe S, Komatsu M, Aoyagi K, Yokozaki H, Sasaki H (2015) Implications of epithelial-mesenchymal transition in gastric cancer. *Transl Gastrointest Cancer* 4: 258-264.
- 7 Tanabe S, Aoyagi K, Yokozaki H, Sasaki H (2014) Gene expression signatures for identifying diffuse-type gastric cancer associated with epithelial-mesenchymal transition. *Int J Oncol* 44: 1955-1970.
- 8 Tanabe S, Aoyagi K, Yokozaki H, Sasaki H (2014) Gastric-related markers and their significance in cancer. *J Med Genomics Biomark* 1:002.
- 9 GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648-660.
- 10 Tryka KA1, Hao L, Sturcke A, Jin Y, Wang ZY, et al. (2014) NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res* 42: D975-979.
- 11 Choi J, Kim KT, Huh A, Kwon S, Hong C, et al. (2015) dbHiMo: a web-based epigenomics platform for histone-modifying enzymes. *Database (Oxford)* 2015: bav052.
- 12 Hibsh D, Schori H, Efroni S, Shefi O (2015) De novo transcriptome assembly databases for the central nervous system of the medicinal leech. *Sci Data* 2: 150015.
- 13 Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, et al. (2015) Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348: 1139-1143.
- 14 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M et al. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42: D199-205.
- 15 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109-114.
- 16 Uchiyama T, Irie M, Mori H, Kurokawa K, Yamada T (2015) FuncTree: Functional Analysis and Visualization for Large-Scale Omics Data. *PLoS One* 10: e0126967.
- 17 Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, et al. (2015) The DDBJ Japanese Genotype-phenotype Archive for genetic and phenotypic human data. *Nucleic Acids Res* 43: D18-22.
- 18 Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, et al. (2002) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 30: 158-162.
- 19 Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, et al. (2003) The ERGO genome analysis and discovery system. *Nucleic Acids Res* 31: 164-171.
- 20 Ahn YY, Lee DS, Burd H, Blank W, Kapatral V (2014) Metabolic network analysis-based identification of antimicrobial drug targets in category A bioterrorism agents. *PLoS One* 9: e85195.
- 21 Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36: 949-951.
- 22 MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. (2014) The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42: D986-D992.
- 23 Rothenberg SM, Mohapatra G, Rivera MN, Winokur D, Greninger P, et al. (2010) A genome-wide screen for microdeletions reveals disruption of polarity complex genes in diverse human cancers. *Cancer Research* 70: 2158-2164.
- 24 Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, et al. (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 43: D805-811.
- 25 Khalili JS, Hanson RW, Szallasi Z (2012) In silico prediction of tumor antigens derived from functional missense mutations of the cancer gene census. *Oncoimmunology* 1: 1281-1289.
- 26 Yoshida T, Ono H, Kuchiba A, Saeki N, Sakamoto H (2010) Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. *Cancer Sci* 101: 1582-1589.
- 27 Zhang J, Baran J, Cros A, Guberman JM, Haider S, et al. (2011) International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford)* 2011: bar026.
- 28 Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801-1806.
- 29 Chin L, Hahn WC, Getz G, Meyerson M (2011) Making sense of cancer genomic data. *Genes Dev* 25: 534-555.
- 30 Gnad F, Doll S, Manning G, Arnott D, Zhang Z (2015) Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer. *BMC Genomics* 16 Suppl 8: S5.
- 31 Lappalainen I, Almeida-King J, Kumanduri V, Senf A, Spalding JD, et al. (2015) The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet* 47: 692-695.
- 32 Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210.
- 33 Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41: D991-995.
- 34 Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, et al. (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8: R183.
- 35 Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: p11.

- 36 Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, et al. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31: 365-370.
- 37 UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* 37: D169-174.
- 38 Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603-607.
- 39 Kupersmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, et al. (2010) Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS One* 5.
- 40 Shangguan H, Tan SY, Zhang JR (2015) Bioinformatics analysis of gene expression profiles in hepatocellular carcinoma. *Eur Rev Med Pharmacol Sci* 19: 2054-2061.
- 41 Liu M, Dou Y, Sun R, Zhang Y, Liu Y (2015) Molecular mechanisms for alcoholic hepatitis based on analysis of gene expression profile. *Hepat Mon* 15: e27336.
- 42 Guo B, Che T, Shi B, Guo L, Zhang Z, et al. (2015) Interaction network analysis of differentially expressed genes and screening of cancer marker in the urine of patients with invasive bladder cancer. *Int J Clin Exp Med* 8: 3619-3628.
- 43 Baldacchino S, Saliba C, Petroni V, Fenech AG, Borg N, et al. (2014) Deregulation of the phosphatase, PP2A is a common event in breast cancer, predicting sensitivity to FTY720. *EPMA J* 5: 3.
- 44 Babur Ā, Dogrusoz U, Cakir M, Aksoy BA, Schultz N, et al. (2014) Integrating biological pathways and genomic profiles with ChiBE 2. *BMC Genomics* 15: 642.
- 45 Rajan P, Stockley J, Sudbery IM, Fleming JT, Hedley A, et al. (2014) Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer. *BMC Cancer* 14: 977.
- 46 Wang X, Yamamoto Y, Wilson LH, Zhang T, Howitt BE, et al. (2015) Cloning and variation of ground state intestinal stem cells. *Nature* 522: 173-178.
- 47 Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, et al. (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348: 880-886.
- 48 Farshdousti Hagh M, Noruzinia M, Mortazavi Y, Soleimani M, Kaviani S, et al. (2015) Different Methylation Patterns of RUNX2, OSX, DLX5 and BSP in Osteoblastic Differentiation of Mesenchymal Stem Cells. *Cell J* 17: 71-82.
- 49 Tanabe S (2015) Signaling involved in stem cell reprogramming and differentiation. *World J Stem Cells* 7: 992-998.
- 50 Tanabe S (2014) Role of mesenchymal stem cells in cell life and their signaling. *World J Stem Cells* 6: 24-32.
- 51 Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, et al. (2015) Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348: 666-669.
- 52 Hepworth MR, Fung TC, Masur SH, Kelsen JR, McConnell FM, et al. (2015) Immune tolerance. Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4+ T cells. *Science* 348: 1031-1035.
- 53 Mohinta S, Kannan AK, Gowda K, Amin SG, Perdew GH, et al. (2015) Differential regulation of Th17 and T regulatory cell differentiation by aryl hydrocarbon receptor dependent xenobiotic response element dependent and independent pathways. *Toxicological Sciences* 145: 233-243.
- 54 Nagamatsu G, Saito S, Takubo K, Suda T (2015) Integrative Analysis of the Acquisition of Pluripotency in PGCs Reveals the Mutually Exclusive Roles of Blimp-1 and AKT Signaling. *Stem Cell Reports* 5: 111-124.
- 55 Rui Y, Xu L, Chen R, Zhang T, Lin S, et al. (2015) Epigenetic memory gained by priming with osteogenic induction medium improves osteogenesis and other properties of mesenchymal stem cells. *Sci Rep* 5: 11056.
- 56 Benham-Pyle BW, Pruitt BL, Nelson WJ (2015) Cell adhesion. Mechanical strain induces E-cadherin-dependent Yap1 and β -catenin activation to drive cell cycle entry. *Science* 348: 1024-1027.
- 57 Flynn RA, Chang HY (2014) Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* 14: 752-761.
- 58 El-Badawy A, El-Badri N (2015) Regulators of pluripotency and their implications in regenerative medicine. *Stem Cells Cloning* 8: 67-80.
- 59 Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. *Cell* 117: 663-676.
- 60 Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, et al. (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142: 375-386.
- 61 Zhang Y, Desai A, Yang SY, Bae KB, Antczak MI, et al. (2015) TISSUE REGENERATION. Inhibition of the prostaglandin-degrading enzyme 15-PGDH potentiates tissue regeneration. *Science* 348: aaa2340.
- 62 Wang Y, Shi J, Chai K, Ying X, Zhou BP (2013) The Role of Snail in EMT and Tumorigenesis. *Curr Cancer Drug Targets* 13: 963-972.
- 63 Ozdemir BC, Hensel J, Secondini C, Wetterwald A, Schwaninger R, et al. (2014) The molecular signature of the stroma response in prostate cancer-induced osteoblastic bone metastasis highlights expansion of hematopoietic and prostate epithelial stem cell niches. *PLoS One* 9: e114530. [PMID: 25485970 PMCID: PMC4259356 DOI: 10.1371/journal.pone.0114530]
- 64 Wu J, Okamura D, Li M, Suzuki K, Luo C, et al. (2015) An alternative pluripotent state confers interspecies chimaeric competency. *Nature* 521: 316-321.
- 65 Palomba G, Loi A, Porcu E, Cossu A, Zara I, et al. (2015) Genome-wide association study of susceptibility loci for breast cancer in Sardinian population. *BMC Cancer* 15: 383.
- 66 Bruegmann T, van Bremen T, Vogt CC, Send T, Fleischmann BK, et al. (2015) Optogenetic control of contractile function in skeletal muscle. *Nat Commun* 6: 7153.
- 67 Zhou X, Crow AL, Hartiala J, Spindler TJ, Ghazalpour A, et al. (2015) The Genetic Landscape of Hematopoietic Stem Cell Frequency in Mice. *Stem Cell Reports* 5: 125-138.
- 68 Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SD, et al. (2015) Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* 348: 1036-1039.
- 69 Martincorena I, Campbell PJ (2015) Somatic mutation in cancer and normal cells. *Science* 349: 1483-1489.
- 70 Tanabe S (2015) Origin of cells and network information. *World J Stem Cells* 7: 535-540.
- 71 Tanabe S (2013) Perspectives of gene combinations in phenotype presentation. *World J Stem Cells* 5: 61-67.