Review

Discovery of molecular associations among aging, stem cells, and cancer based on gene expression profiling

Xiaosheng Wang

Abstract

The emergence of a huge volume of "omics" data enables a computational approach to the investigation of the biology of cancer. The cancer informatics approach is a useful supplement to the traditional experimental approach. I reviewed several reports that used a bioinformatics approach to analyze the associations among aging, stem cells, and cancer by microarray gene expression profiling. The high expression of aging- or human embryonic stem cell-related molecules in cancer suggests that certain important mechanisms are commonly underlying aging, stem cells, and cancer. These mechanisms are involved in cell cycle regulation, metabolic process, DNA damage response, apoptosis, p53 signaling pathway, immune/inflammatory response, and other processes, suggesting that cancer is a developmental and evolutional disease that is strongly related to aging. Moreover, these mechanisms demonstrate that the initiation, proliferation, and metastasis of cancer are associated with the deregulation of stem cells. These findings provide insights into the biology of cancer. Certainly, the findings that are obtained by the informatics approach should be justified by experimental validation. This review also noted that next-generation sequencing data provide enriched sources for cancer informatics study.

Key words Cancer, aging, stem cells, gene expression profiling, cancer informatics

Cancer is closely related to aging in that the incidence of cancer increases exponentially with age. The molecular commonality between human embryonic stem cells (hESCs) and cancer cells suggests a strong linkage between hESCs and cancer cells[1,2]. Thus, to a certain extent, cancer is a developmental disease. In fact, the initiation, proliferation, and metastasis of cancer are often associated with abnormalities in various developmental signatures. An in-depth investigation of the molecular mechanisms that connect aging, stem cells, and cancer could provide meaningful insights into the biology of cancer. The emergence of a huge amount of "omics" data provides a platform for the investigation of the related mechanisms through a computational biology approach. In particular, microarray-based gene expression data that are related to aging, stem cells, and cancer are valuable resources whereby molecular associations among aging, stem cells, and cancer could be discovered.

Molecular Associations Between Aging and Cancer

To date, only one study has investigated the association between aging and cancer by an extensive examination of the expression profiling of aging-related genes in various human tumor types[3].

Identification of human aging-related genes

I have identified 3,359 human aging-related genes on the basis of 10 relevant publications; these genes are involved in the aging of six different tissues: the brain, eye, kidney, muscle, skin, and blood[3-13]. From these 3,359 genes, I identified 69 aging-related signature genes (ASGs) that are commonly expressed in multiple tissues. The functional analysis of the 69-gene set revealed two

Author's Affiliation: Biometric Research Branch, National Cancer Institute, National Institutes of Health, Rockville, MD 20852, USA.

Corresponding Author: Xiaosheng Wang, Biometric Research Branch, National Cancer Institute, National Institutes of Health, Rockville, MD 20852, USA. Email: xiaosheng.wang@nih.gov.

important KEGG pathways: pathways in cancer and MAPK signaling pathway, both of which are significantly relevant to cancer. The gene ontology (GO) analysis of the 69 genes identified significant aging-related GO terms, including oxidative stress, mitochondrial function, apoptosis and senescence, and DNA damage repair. These GO terms are also strongly implicated in cancer. In addition, I obtained 261 human aging-related genes (HAGs) from the Human Ageing Genomic Resources website (http://genomics.senescence.info/genes/allgenes. php) [14], which is an important resource that collects HAGs. Within this 261-gene list, certain genes play important roles in both aging and cancer. For example, FOS regulates cell proliferation, differentiation, apoptosis, and transformation, and has been associated with aging and cancer [15]. FOXO1 belongs to the forkhead family of transcription factors, which are strongly associated with human longevity [16] and cancer [17]. TP63 encodes a member of the P53 family of transcription factors and has been associated with aging and cancer-related pathology [18,19].

Expression profiling of human aging-related genes in tumors

I have identified 28 sets of differentially expressed genes by the 28 normal vs. tumor phenotypes class comparisons in 25 human tumor gene expression datasets; subsequently, I analyzed the overlap between each of these 28 gene sets and each of the aforementioned ASGs and HAGs [3]. Table 1 presents several representative genes that are highly overlapping between both classes of datasets, which include not only the aforementioned FOS, FOXO1, and TP63 but also certain other genes, such as APOD [20], IGF1 [21] and FOXM1 [22] that have also been shown to strongly correlate aging with cancer. Gene function enrichment analyses suggested that these highly overlapping genes were primarily involved in metabolic process, cell cycle regulation, DNA damage response, apoptosis, cell proliferation, and transcriptional regulation. Based on these highly overlapping genes, I have identified several convergent pathways between aging and cancer. These pathways include pathways in cancer, ErbB signaling pathway, MAPK signaling pathway, cell cycle, T-cell receptor signaling pathway, mTOR signaling pathway, B-cell receptor signaling pathway, p53 signaling pathway, insulin signaling pathway, VEGF signaling pathway, and chemokine signaling pathway, among others.

Molecular Associations Between Stem Cells and Cancer

Identification of human stem cell-related molecular signatures

Based on a literature mining approach [1], I have identified four types of hESC-related molecular signatures: genes, pathways, transcription factors (TFs), and microRNAs (miRNAs). The gene signatures include 24 hESC-related gene sets involving the targets of three core TFs (OCT4, SOX2, and NANOG) in hESCs, the targets of oncogene c-Myc, the targets of the tumor suppressor p53, and the targets of polycomb group proteins; these gene sets are essential for controlling the development of hESCs. The pathway signatures include 54 hESC-related pathways that essentially belong to developmental signaling pathways such as the Wnt, Notch, Hedgehog, and Bmi-1 pathways that are necessary for the regulation of stem cell self-renewal and differentiation. The TF signatures include 189 hESC-related TFs that are primarily involved in the regulation of hESC self-renewal and differentiation. Among these TFs, OCT4, SOX2, and NANOG have essential roles in the

Table 1. Genes commonly identified in aging-related and tumor-related gene sets

Gene set	Total number of genes	The number of genes with 5 or more overlaps	The number of genes with a 50% or greater overlapping rate	Highly overlapping representative genes
ASG	69	69 (100%)	17	PGK1, FN1, YWHAZ, AHNAK, NEBL, VCAN, ABI2, PRKCB, WNK1, FGF1, GATM, SFPQ, HPGD, PTGER3, COX7C, LAMP1, H2AFV, APOD, FOXO1, TP63, FOS
HAG	261	234 (90%)	33	CLU, JUND, APP, MAPT, NR3C1, PML, YWHAZ, TCF3, TOP2A, VEGFA, APOE, PRKCA, CDKN2A, HOXB7, IGFBP3, PTK2, SHC1, TERF1, ATP5O, CCNA2, FGFR1, FOXM1, IGF1, TP53

AGS, aging-related signature genes; HAG, human aging-related genes.

transcriptional control of the regulatory circuitry underlying pluripotency [23,24]. The miRNA signatures include 114 hESC-related miRNAs that have important roles in regulating stem cell self-renewal and differentiation[25].

Expression profiling of human stem cell-related molecular signatures in tumors

I have analyzed 51 human gene expression datasets involving 23 tumor types and identified differentially expressed genes among either normal vs. tumor or good prognosis vs. poor prognosis phenotype classes [1]. Furthermore, I identified important tumor-related pathways, TFs, and miRNAs by analyzing gene sets for differential expression among these pre-defined classes.

I have identified 72 sets of differentially expressed genes in the 51 cancer-related human gene expression datasets and found considerable overlaps between each of the 72 cancer-related gene sets and each of the 24 hESC-related gene sets [1]. Figure 1 presents the proportion of genes in each of the 24 hESC-related gene sets that have no less than 10 occurrences in the 72 cancer-related gene sets and demonstrates that a substantial portion of the hESC-related genes are also related to cancer. The highly overlapping genes are mainly involved in cell cycle regulation, DNA damage repair and replication, apoptosis, development and differentiation, cell adhesion, and TF activity.

I have identified 68 groups of pathways by 75 class

comparisons and survival analyses of the cancer-related datasets and found that 26 hESC-related pathway signatures appeared in at least 8 groups [1]. Figure 2 presents the occurrence rate of each of the 26 pathways in the 68 groups of cancer-related pathways. Among the highly overlapping pathways between hESCs and cancer, the cell cycle pathway and MAPK pathway are most prominent. The IGF, ERK, SHH, WNT, PRC2, Notch, PTEN, and TGFβ pathways are also important signaling pathways that link hESCs to cancer.

I have identified 73 groups of TFs in the human cancer gene expression datasets[1]. Among the 189 hESC-associated TFs, 42 TFs appeared in at least 3 groups. Figure 3 presents the occurrence rate of each of the 42 TFs in the 73 groups of cancer-related TFs. The most frequently identified hESC-related TF in tumor was MYC. In fact, MYC is an important TF in both hESCs and cancer cells [23,26-35], and its regulatory networks may account for most of the transcriptional similarity between embryonic stem cells (ESCs) and cancer cells[33]. Several families of hESC-associated TFs, such as MYB, E2F, PAX, SMAD, STAT, POU, SP, and GLI, were shown to be related to cancer (Figure 3).

I have identified 67 groups of miRNAs in the cancer datasets [1]. Among the 114 hESC-associated miRNAs, 50 miRNAs appeared in at least 20 different groups. The most frequently identified miRNA was miR-29c, which occurred 34 times (51% occurrence rate), and the next most common miRNA was miR-200b, which occurred 30 times (45% occurrence rate). Figure 4 lists the 50
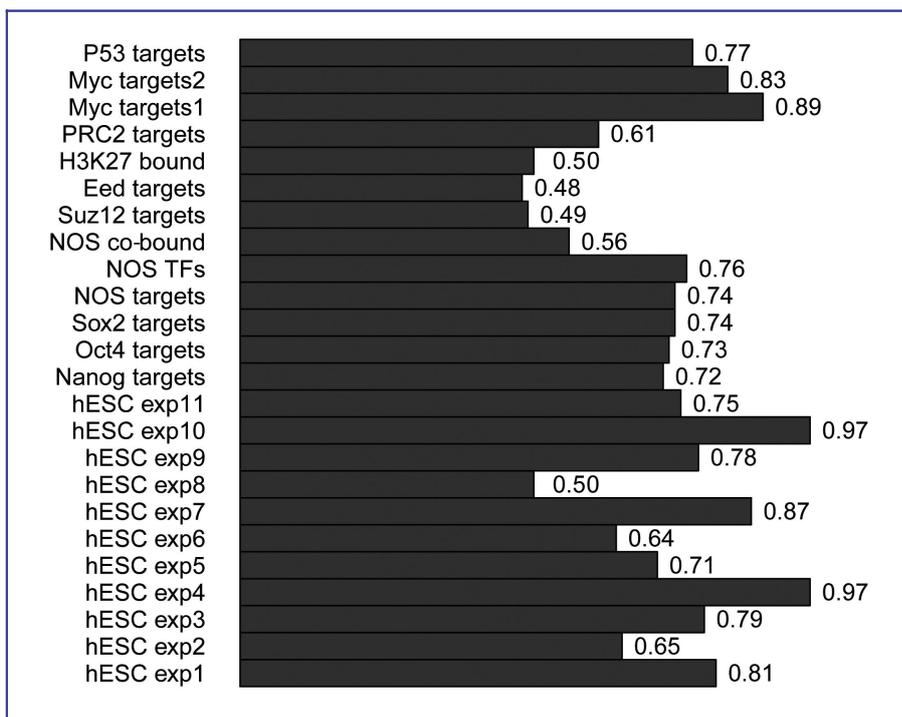
Figure 1. Overlapping rate between human embryonic stem cell (hESC)–related gene sets and differentially expressed gene sets. Each number represents the proportion of genes in the corresponding hESC-related gene set that have no less than 10 occurrences in the 72 differentially expressed gene sets identified in tumors.
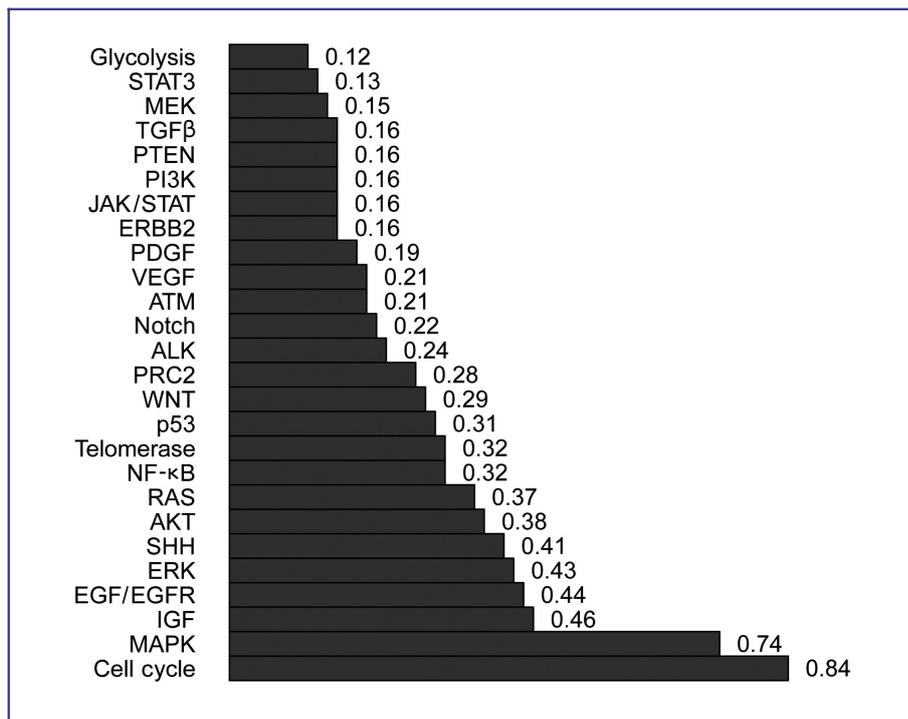
Figure 2. Overlapping rate between hESC-related pathways and differentially expressed pathways. Each number represents the occurrence rate of the corresponding hESC-related pathway in the 68 cancer-related pathway sets.
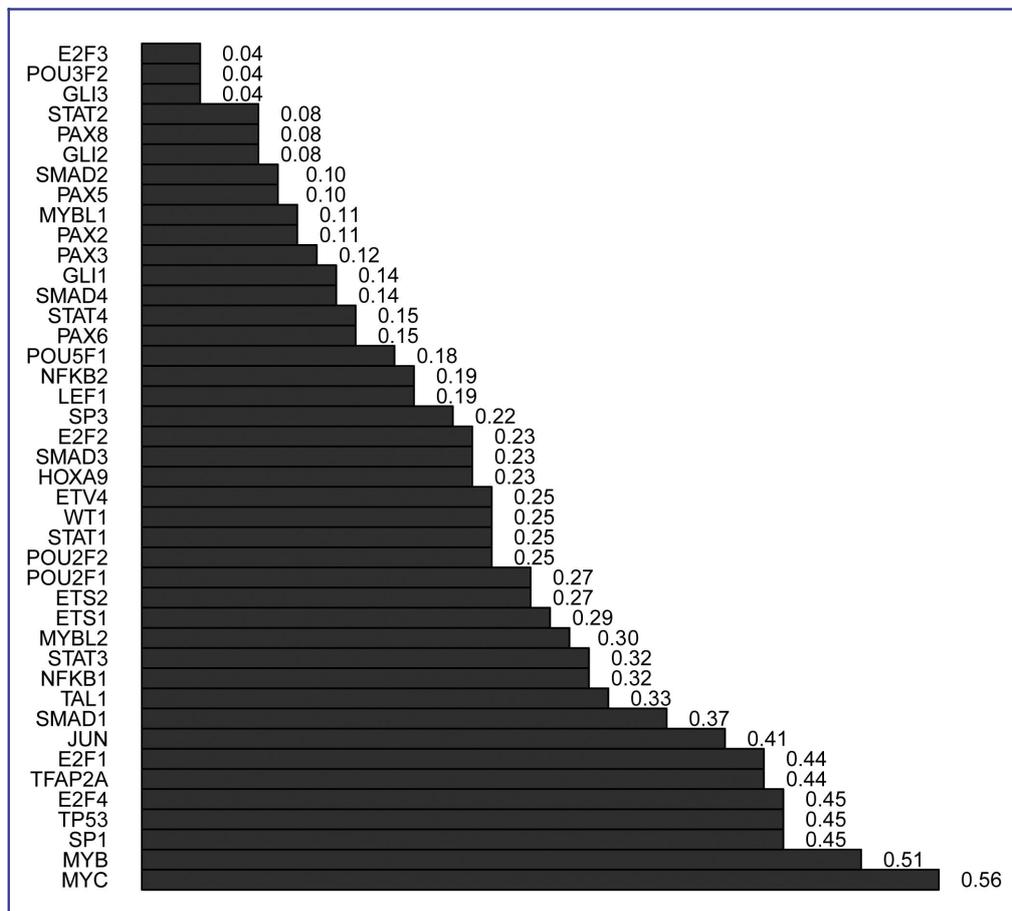
Figure 3. Overlapping rate between hESC-related transcription factors (TFs) and differentially expressed TFs. Each number represents the occurrence rate of the corresponding hESC-related TF in the 68 cancer-related TF sets.
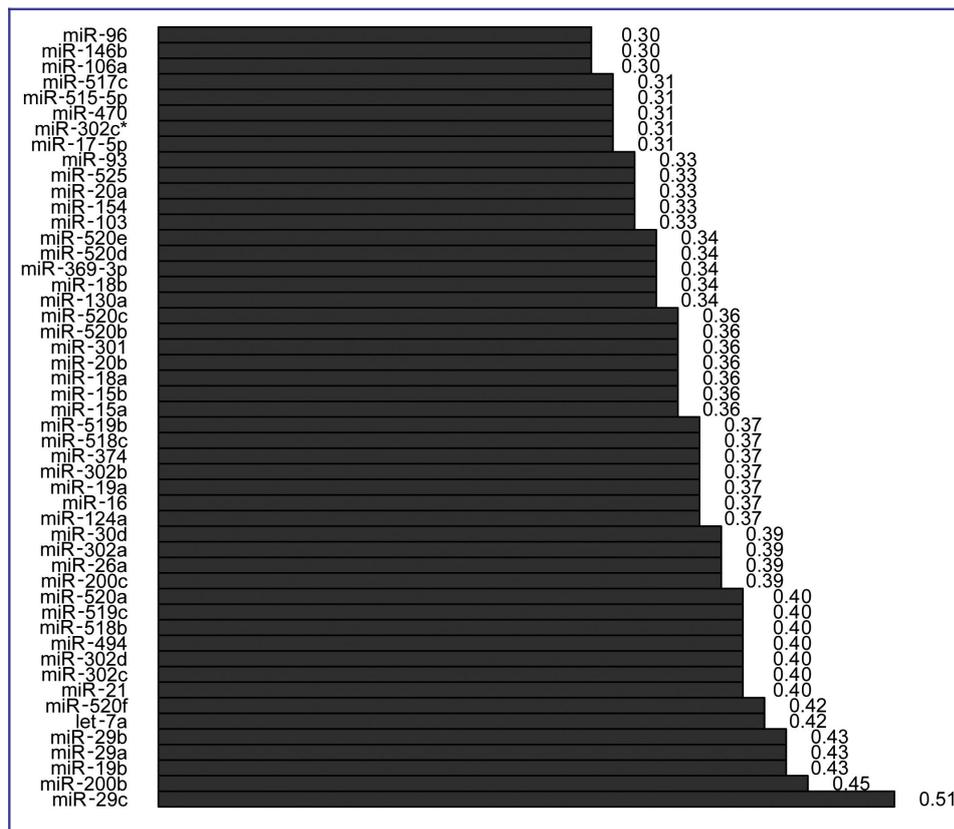
Figure 4. Overlapping rate between hESC-related microRNAs and differentially expressed microRNAs. Each number represents the occurrence rate of the corresponding hESC-related microRNA in the 67 cancer-related microRNA sets.

miRNAs for which the occurrence frequency is no less than 20, indicating that there is a broad range of overlap between "stemness" miRNAs and oncogenic miRNAs. Certain miRNA families, such as the miR-302 family and the miR-200 family, appear to play an important role in the regulation of hESCs and tumorigenesis[36-42].

Other works identifying common molecular signatures between cancer and stem cells

In addition to the previously discussed work [1], several other studies have examined the expression of hESC-associated genes in human cancer [2,29,30,43,44]. For instance, Murat et al.[43] provided the first clinical evidence for the existence of a "glioma stem cell" or "self-renewal" phenotype in the context of the treatment resistance of glioblastoma. Santagata et al. [44] explored the expression levels of the ESC TFs OCT3/4, NANOG, and SOX2 in primary and metastatic germ cell tumors (GCTs). They confirmed NANOG and OCT3/4 as sensitive and specific markers for primary seminoma and embryonal carcinoma and demonstrated that NANOG was a marker for metastatic GCTs. Their findings showed that ESC TFs were useful in the diagnosis of tumors that were metastatic to the retroperitoneum. Ben-Porath et al.[2] identified a subset of hESC-

associated transcriptional regulators, including Nanog, Oct4, Sox2, and c-Myc, that were highly expressed in poorly differentiated tumors. They found a novel link between genes associated with hESC identity and the histopathologic traits of tumors and suggested that these genes might contribute to the stem cell–like phenotypes presented by many tumors. Hassan et al. [29] used microarray gene expression analysis to identify gene set enrichment patterns between human lung adenocarcinoma and squamous cell carcinoma. This analysis revealed that an increased expression of the ESC gene set and decreased expression of the Polycomb target gene set identified poorly differentiated lung adenocarcinoma and that this gene expression signature was associated with markers of poor prognosis and worse overall survival in lung adenocarcinoma. Schoenhals et al.[30] compared the expression levels of the pluripotency factors OCT4, SOX2, KLF4, and MYC in 40 human tumor types with the levels in the normal tissue counterparts using publicly available gene expression data and found significant overexpression of at least one of these factors in 18 out of the 40 investigated cancer types. Furthermore, they found that these genes were associated with tumor progression or poor prognosis. Overall, these studies revealed that "stemness" gene

expression signatures were associated with tumor malignancies and therefore might be informative molecular predictors for cancer therapy outcomes[45].

Discussion

Cancer is related to aging. Many aging-related genes demonstrate differential expression profiling between normal tissues and tumor. Two significant aging-related pathways, pathways in cancer and MAPK signaling pathway, are also implicated in carcinogenesis. Certain critical molecular mechanisms underlie both aging and cancer; these mechanisms include cell cycle regulation, metabolic process, DNA damage response, apoptosis, p53 signaling pathway, immune/inflammatory response, etc. It is worthy of note that the immune/inflammatory response is an important link between aging and cancer, given that the convergent pathways between aging and cancer include multiple pathways that correlate with the immune/inflammatory response, such as T-cell receptor signaling pathway, mTOR signaling pathway, B-cell receptor signaling pathway, and chemokine signaling pathway (see Table 5 of reference [3]).

Cancer is also related to development. Various hESC-specific characteristics, including genes, pathways, TFs, and miRNAs, were often differentially expressed between normal and tumor phenotypes or among cancer subtypes with distinct clinical outcomes. The signaling pathways like the cell cycle, MAPK, SHH, WNT, PRC2, Notch, PTEN, and TGFβ pathways, that played an important role in control of the hESC fate were also strongly implicated in the initiation, proliferation, and metastasis of cancer. The hESC-specific TFs like OCT4 and c-Myc were important in the maintenance of the undifferentiated state of cancer cells. The miRNAs that were overexpressed in undifferentiated hESCs like miRNA-302 and miRNA-200 families were also involved in cancerous pathology. The mechanism of cell cycle regulation provides the primary commonality between hESCs and cancer cells in that the hESC-associated signatures most frequently identified in tumors are primarily involved in the regulation of cell cycle[1].

The huge amount of available "omics" data facilitates the investigation of the associations among aging, stem cells, and cancer using the computational biology approach. This approach constitutes an important supplement to the traditional experimental approach. However, the findings obtained through the computational biology approach are only reliable after they have been experimentally validated.

With the development of next-generation sequencing (NGS) technology, increasing quantities of cancer-related sequencing data are emerging. In contrast to microarray methods, NGS can directly determine the cDNA sequences and quantify the expression of diverse RNA molecules from genomes over a wide range of levels[46]. RNA sequencing (RNA-Seq) is useful for studying complex transcriptomes and revealing sequence variations, such as SNPs, in the transcribed regions[47]. RNA-Seq has a large dynamic range of expression levels over which transcripts can be detected and is therefore highly accurate for quantifying expression levels. Moreover, this approach also has high levels of reproducibility and requires less RNA samples. The bioinformatics investigation of NGS data will provide a promising direction for cancer informatics research.

References

[1] Wang X. Computational analysis of expression of human embryonic stem cell-associated signatures in tumors. BMC Res Notes, 2011,4:471.

[2] Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell–like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet, 2008,40:499–507.

[3] Wang X. Microarray analysis of ageing-related signatures and their expression in tumors based on a computational biology approach. Genomics Proteomics Bioinformatics, 2012,10:136–141.

[4] Lu T, Pan Y, Kao SY, et al. Gene regulation and DNA damage in the ageing human brain. Nature, 2004,429:883–891.

[5] Erraji-Benchekroun L, Underwood MD, Arango V, et al. Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. Biol Psychiatry, 2005,57:549–558.

[6] Hawse JR, Hejtmancik JF, Huang Q, et al. Identification and functional clustering of global gene expression differences between human age-related cataract and clear lenses. Mol Vis, 2003,9:515–537.

[7] Melk A, Mansfield ES, Hsieh SC, et al. Transcriptional analysis of the molecular basis of human kidney aging using cdna microarray profiling. Kidney Int, 2005,68:2667–2679.

[8] Rodwell GEJ, Sonu R, Zahn JM, et al. A transcriptional profile of aging in the human kidney. PLoS Biol, 2004,2:e427.

[9] Welle S, Brooks AI, Delehanty JM, et al. Gene expression profile of aging in human muscle. Physiol Genomics, 2003,14:149–159.

[10] Welle S, Brooks AI, Delehanty JM, et al. Skeletal muscle gene expression profiles in 20–29 year old and 65–71 year old women. Exp Gerontol, 2004,39:369–377.

[11] Zahn JM, Sonu R, Vogel H, et al. Transcriptional profiling of aging in human muscle reveals a common aging signature. PLoS Genet, 2006,2:e115.

[12] Lener T, Moll PR, Rinnerthaler M, et al. Expression profiling of aging in the human skin. Exp Gerontol, 2006,41:387–397.

[13] Visala Rao D, Boyle GM, Parsons PG, et al. Influence of ageing, heat shock treatment and in vivo total antioxidant

status on gene-expression profile and protein synthesis in human peripheral lymphocytes. Mech Ageing Dev, 2003,124: 55–69.

[14] de Magalhaes JP, Budovsky A, Lehmann G, et al. The human ageing genomic resources: online databases and tools for biogerontologists. Aging Cell, 2009,8:65–72.

[15] Choi EK, Uyeno S, Nishida N, et al. Alterations of c-fos gene methylation in the processes of aging and tumorigenesis in human liver. Mutat Res, 1996,354:123–128.

[16] Flachsbart F, Caliebe A, Kleindorp R, et al. Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc Nat Acad Sci U S A, 2009,106:2700–2705.

[17] Arden KC. Multiple roles of FOXO transcription factors in mammalian cells point to multiple roles in cancer. Exp Gerontol, 2006,41:709–717.

[18] Keyes WM, Mills AA. p63: a new link between senescence and aging. Cell Cycle, 2006,5:260–265.

[19] Tomkova K, Tomka M, Zajac V. Contribution of p53, p63, and p73 to the developmental diseases and cancer. Neoplasma, 2008,55:177–181.

[20] Muffat J, Walker DW. Apolipoprotein D: an overview of its role in aging and age-related diseases. Cell Cycle, 2010,9:269 – 273.

[21] Anisimov VN. Insulin/IGF-1 signaling pathway driving aging and cancer as a target for pharmacological intervention. Exp Gerontol, 2003,38:1041–1049.

[22] Laoukili J, Stahl M, Medema RH. FoxM1: at the crossroads of ageing and cancer. Biochim Biophys Acta, 2007,1775:92–102.

[23] Young RA. Control of the embryonic stem cell state. Cell, 2011,144:940–954.

[24] Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 2005,122:947–956.

[25] Gangaraju VK, Lin H. MicroRNAs: key regulators of stem cells. Nat Rev Mol Cell Biol, 2009,10:116–125.

[26] Smith K, Dalton S. Myc transcription factors: key regulators behind establishment and maintenance of pluripotency. Regen Med, 2010,5:947–959.

[27] Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell, 2008,133:1106–1117.

[28] Kim J, Chu J, Shen X, et al. An extended transcriptional network for pluripotency of embryonic stem cells. Cell, 2008,132:1049–1061.

[29] Hassan KA, Chen G, Kalemkerian GP, et al. An embryonic stem cell-like signature identifies poorly differentiated lung adenocarcinoma but not squamous cell carcinoma. Clin Cancer Res, 2009,15:6386–6390.

[30] Schoenhals M, Kassambara A, De Vos J, et al. Embryonic stem cell markers expression in cancers. Biochem Biophys Res Commun, 2009,383:157–162.

[31] Kashyap V, Rezende NC, Scotland KB, et al. Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and

stem cell microRNAs. Stem Cells Dev, 2009,18:1093–1108.

[32] Ben-David U, Benvenisty N. The tumorigenici ty of human embryonic and induced pluripotent stem cells. Nat Rev Cancer, 2011,11:268–277.

[33] Kim J, Woo AJ, Chu J, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell, 2010,143:313–324.

[34] Kidder BL, Yang J, Palmer S. Stat3 and c-myc genome-wide promoter occupancy in embryonic stem cells. PLoS One, 2008,3:e3932.

[35] Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 2007,131:861–872.

[36] Lin SL, Chang DC, Ying SY, et al. MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways. Cancer Res, 2010,70:9473–9482.

[37] Lin SL, Chang DC, Chang-Lin S, et al. Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. RNA, 2008,14:2115–2124.

[38] Lee NS, Kim JS, Cho WJ, et al. miR-302b maintains " stemness" of human embryonal carcinoma cells by post-transcriptional regulation of Cyclin D2 expression. Biochem Biophys Res Commun, 2008,377:434–440.

[39] Murray MJ, Saini HK, van Dongen S, et al. The two most common histological subtypes of malignant germ cell tumour are distinguished by global microRNA profiles, associated with differential transcription factor expression. Mol Cancer, 2010,9: 290.

[40] Gregory PA, Bert AG, Paterson EL, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol, 2008,10:593–601.

[41] Shimono Y, Zabala M, Cho RW, et al. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. Cell, 2009,138:592–603.

[42] Vallejo DM, Caparros E, Dominguez M. Targeting notch signalling by the conserved miR-8/200 microRNA family in development and cancer cells. EMBO J, 2011,30:756–769.

[43] Murat A, Migliavacca E, Gorlia T, et al. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. J Clin Oncol, 2008,26: 3015–3024.

[44] Santagata S, Ligon KL, Hornick JL. Embryonic stem cell transcription factor signatures in the diagnosis of primary and metastatic germ cell tumors. Am J Surg Pathol, 2007,31:836– 845.

[45] Glinsky GV. "Stemness" genomics law governs clinical behavior of human cancer: implications for decision making in disease management. J Clin Oncol, 2008,26:2846–2853.

[46] Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet, 2011,12:87–98.

[47] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 2009,10:57–63.