



Expression levels of 63 p53-related genes add up to similar values in 24 different tissues and are unified in cancer

B. Altenberg^a, A. Rapp^{b,1}, E. Schmitt^c, K.O. Greulich^{b,*}

^a *Bioinformatics Group, European Molecular Biology Laboratory, Meyerhofstrasse 1, D 69120 Heidelberg, Germany*

^b *Single Cell and Single Molecule Techniques, Leibniz Institute for Age Research, Beutenbergstrasse 11, D 07745 Jena, Germany*

^c *Biocomputing, Leibniz Institute for Age Research, Beutenbergstrasse 11, D 07745 Jena, Germany*

Received 27 October 2006; accepted 19 June 2007

Abstract

The expression patterns of 62 genes interacting with p53 have been investigated in 24 normal and cancerous tissues using NIH's dbEST library. The expression levels of individual genes, such as the *TTP53* gene itself, but also other genes, vary up to 33-fold among the 24 different tissues and no consistent pattern can be recognized. However, when expression levels for all 63 genes are summed, these "cumulated levels" are surprisingly constant over the 24 investigated normal tissues. In cancers, the variation is further reduced. Essentially, the cumulated expression levels in cancer are independent of those in normal tissue. We furthermore constructed a linear statistical classifier, i.e., a *weighted* sum of gene expression levels, which robustly distinguishes normal from cancer tissue independent of the particular kind of tissue. Thus, despite very large differences for individual genes and considerable changes during carcinogenesis, the cumulated expressions have narrowly defined levels.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Data mining; dbEST database; Cancer genes; Statistical classifier

A defect in only one gene as a cause of disease is rather an exception than the rule. On the other hand, though it may be envisioned for the future, at the present state of knowledge it appears impossible to understand disease as a complete interaction network of some 10,000 genes in a cell. What appears more reasonable is the attempt to find ensembles of genes that cooperate in the genesis of disease and to understand them as functional units. If such an ensemble remains unaffected during the progress of disease it may be ruled out as a driving force of the latter. Particularly interesting is carcinogenesis. When an ensemble of genes behaves similarly (for example is over- or underexpressed as a whole) in a wide variety of cancer tissues, compared to their normal counterparts, it may be seen as a general feature of cancer and does not contribute to the genesis of just one or a few selected cancers. Alternatively, when such an ensemble of genes behaves differently in a wide variety of

normal tissues but more similarly in the corresponding cancer tissues, this suggests functional relevance,

Indeed a few such ensembles of genes, namely a few classical biochemical pathways, are accepted as acting as a whole. For example, glycolysis is elevated in most cancers (the Warburg effect [1]) and has recently gained new interest [2]. Almost all the genes in the glycolysis pathway are overexpressed in a wide variety of phenotypically very different cancer tissues [3]. In addition to these more classical functional ensembles of genes others may be envisioned, for example, nodes in interaction networks [4]. Other ensembles may be those that are differentially sensitive (noisy) toward external stimuli. The expression of housekeeping genes does not reveal such sensitivity, while genes involved in environmental responses do so [5]. Also, ensembles of genes have been discovered that usually are expressed only in one given healthy tissue but also are group-wise switched on in cancers of another tissue, in whose normal counterpart they are not expressed [6]. As already mentioned above, particularly interesting are markers that indicate cancer in general, i.e., which are over- or underexpressed in a wide variety of cancers

* Corresponding author. Fax: +49 3641 656410.

E-mail address: kog@fli-leibniz.de (K.O. Greulich).

¹ Present address: Sir William Dunn School of Pathology, University of Oxford, South Park Road, Oxford OX1 3RE, UK.

(common cancer biomarkers [7]) or are ubiquitous cancer genes [8].

In human cells, one of the best known candidates for such a functional ensemble of cooperating genes is the tumor suppressor *TP53* and 62 of its interacting proteins. The product of this gene is located at the intersection between several pathways involved in carcinogenesis. Together with 62 other proteins of four pathways, *TP53* has a central function in the regulation of the cell cycle and apoptosis [9,10]. Vast amounts of data on its expression and mutations in cancer are available (for example [10–12]), but many effects are ascribed solely to this gene and its product. It is, however, also known that molecules up- and downstream in the signaling cascade of *TP53* also contribute to transforming a cell into a neoplastic phenotype. Studies on *TP53* as an ensemble with its binding proteins are rare [10,13] and, particularly, direct comparisons of a wide variety of phenotypically very different tissues and their cancers are still missing.

We have performed an in silico study on the expression of 62 *TP53*-binding genes and *TP53* itself in 24 different healthy tissues that represent approximately 70% of all clinical cases [3] and their corresponding cancer tissues. There is some discussion on the genes that should be ascribed to this pathway. We have used genes listed in *TP53*-dependent pathways by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg) (see also Materials and methods). Omitting one or a few genes will probably not seriously affect the statements made in the following.

In silico studies have become possible since a vast amount of gene expression data is now available. The NIH provides a

database (dbEST) that collects and normalizes gene expression data, published in the literature, and offers such collected data for a large number of annotated genes [14]. We used this database [15–18] to get an overview on the expression behavior of genes of *TP53*-related genes [19].

Currently it is not possible to allow discrimination of cancer subclasses or tumor stages or ethnic variability. The dbEST database simply does not provide such data for all 24 investigated cancers. However, a major result of the present study will be that cancers with very different phenotypes are surprisingly uniform and thus the lack of fine distinction probably does not invalidate the general message of the present work. In other words, the aim of the present study is to provide a crude grid of the roles of 63 genes in 24 cancers, i.e., 3024 data points. This grid, as soon as more detailed studies are available in literature, may then become a sort of guideline to characterize subclasses of different cancers and specific aspects of individual patients.

Results

Table 1 gives the expression data, i.e., the a_{ji} , for the 63 genes in 24 different healthy and cancerous tissues as they can be derived from data such as in Table 3 under Materials and methods, multiplied by 100 as will be mentioned there. Below each of the two blocks for normal tissue and cancer in Table 1 the sums of the expression levels of the all genes in a specific cancer are given.

Fig. 1 shows the expression levels for the *TP53* gene in 24 selected tissues and the corresponding cancers. For tissues for

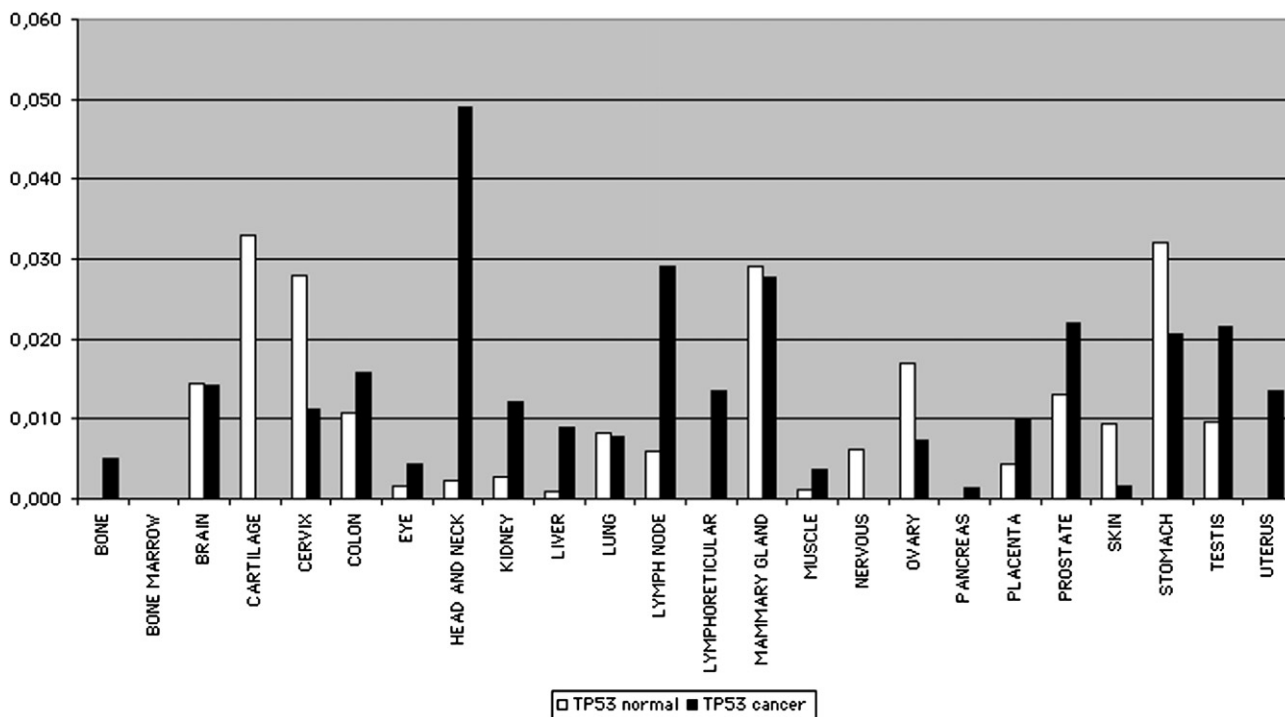


Fig. 1. Expression of the *TP53* gene in 24 tissues and the corresponding cancers. For normal nervous tissue, the level is 0.033, for liver it is 0.001, i.e., the variation is a factor of 33. Also, in cancer some expression levels are increased, some of them are reduced, i.e., no consistent pattern emerges when only this single gene is regarded. A similar observation is made for other genes such as *STAT1*, *CDKN1A*, or *SHC1*.

Table 1
Expression data for the 63 p53-signaling genes in 24 different cancers (raw data obtained via tables of the type of Table 3 multiplied by 100 for easier reading)

Gene name	Description	Cytogenetic location	Bone	Bone marrow	Brain	Cartilage	Cervix	Colon	Eye	Head and neck	Kidney	Liver	Lung	Lymph node	Lympho-reticular	Mammary gland	Muscle	Nervous tissue	Ovary	Pancreas	Placenta	Prostate	Skin	Stomach	Testis	Uterus
<i>(A) Normal</i>																										
<i>AATF</i>	Apoptosis antagonizing transcription factor	17q11.2-q12	0.000	0.000	0.008	0.007	0.000	0.000	0.011	0.004	0.006	0.006	0.006	0.009	0.006	0.005	0.015	0.001	0.000	0.000	0.006	0.008	0.002	0.004	0.023	0.003
<i>APAF1</i>	Apoptotic peptidase activating factor	12q23	0.000	0.006	0.001	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.006	0.000	0.002	0.005	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
<i>ATM</i>	Ataxia telangiectasia mutated	11q22-q23	0.007	0.000	0.004	0.007	0.000	0.000	0.007	0.016	0.005	0.006	0.003	0.037	0.000	0.015	0.013	0.004	0.000	0.000	0.004	0.004	0.002	0.004	0.007	0.008
<i>ATR</i>	Ataxia telangiectasia and Rad3 related	3q22-q24	0.000	0.000	0.003	0.000	0.000	0.000	0.002	0.002	0.003	0.006	0.003	0.015	0.000	0.004	0.006	0.001	0.017	0.027	0.002	0.001	0.004	0.004	0.006	0.005
<i>Bak1</i>	BCL2-antagonist/killer 1	6p21.3	0.000	0.000	0.001	0.000	0.000	0.004	0.002	0.000	0.002	0.001	0.005	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.004	0.004	0.001	0.000
<i>BAX</i>	BCL2-associated X protein	19q13.3-q13.4	0.000	0.006	0.000	0.000	0.000	0.004	0.003	0.000	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.007	0.001	0.000
<i>BCL2</i>	B-cell CLL/lymphoma 2	18q21.33 18q21.3	0.000	0.013	0.003	0.000	0.000	0.011	0.001	0.004	0.008	0.002	0.005	0.001	0.000	0.020	0.006	0.001	0.000	0.000	0.018	0.005	0.013	0.032	0.005	0.003
<i>BID</i>	BH3 interacting domain death agonist	22q11.1	0.000	0.000	0.004	0.000	0.000	0.000	0.005	0.000	0.002	0.000	0.003	0.004	0.000	0.002	0.000	0.005	0.000	0.000	0.001	0.008	0.006	0.000	0.001	0.010
<i>BRCA1</i>	Breast cancer 1, early onset	17q21	0.007	0.013	0.001	0.000	0.000	0.004	0.001	0.000	0.001	0.002	0.001	0.021	0.000	0.004	0.002	0.002	0.009	0.000	0.000	0.001	0.000	0.000	0.002	0.000
<i>BRCA2</i>	Breast cancer 2, early onset	13q12.3	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.002	0.000	0.003	0.002	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.003	0.000
<i>CASP2</i>	Caspase 2	7q34-q35	0.000	0.000	0.001	0.000	0.000	0.004	0.006	0.000	0.002	0.003	0.001	0.030	0.000	0.004	0.002	0.006	0.000	0.000	0.006	0.005	0.006	0.004	0.002	0.000
<i>CASP9</i>	Caspase 9	1p36.3-p36.1	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.005	0.003	0.001	0.004	0.000	0.002	0.000	0.001	0.009	0.000	0.005	0.002	0.002	0.000	0.005	0.018
<i>CCNE2</i>	Cyclin E2	8q22.1	0.000	0.006	0.003	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.000	0.013	0.000	0.002	0.005	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.004	0.000
<i>CCNG1</i>	Cyclin G1	5q32-q34	0.000	0.019	0.013	0.020	0.000	0.007	0.028	0.013	0.019	0.007	0.009	0.009	0.000	0.005	0.010	0.013	0.000	0.000	0.004	0.008	0.034	0.014	0.004	0.018
<i>CCNG2</i>	Cyclin G2	4q21.1	0.000	0.006	0.006	0.013	0.000	0.004	0.006	0.002	0.005	0.000	0.006	0.012	0.000	0.013	0.005	0.013	0.000	0.000	0.005	0.007	0.004	0.000	0.001	0.005
<i>CDC14A</i>	CDC14 cell divis. cycle 14 homolog A	1p21	0.007	0.000	0.001	0.000	0.000	0.004	0.002	0.000	0.001	0.002	0.003	0.004	0.000	0.002	0.002	0.003	0.000	0.000	0.001	0.004	0.000	0.004	0.010	0.000
<i>CDC2</i>	Cell division cycle 2, G1 to S and G2 to M	10q21.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.006	0.005	0.052	0.006	0.002	0.006	0.001	0.000	0.000	0.000	0.000	0.004	0.000	0.006	0.000
<i>CDC25A</i>	Cell division cycle 25A	3p21	0.000	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.002	0.000	0.001	0.001	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>CDC25C</i>	Cell division cycle 25C	5q31	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.006	0.006	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.007	0.000
<i>CDK4</i>	Cyclin-dependent kinase 4	12q14	0.015	0.000	0.004	0.000	0.000	0.004	0.015	0.000	0.012	0.003	0.007	0.016	0.000	0.002	0.002	0.007	0.000	0.000	0.011	0.008	0.011	0.004	0.001	0.008
<i>CDK7</i>	Cyclin-dependent kinase 7	5q12.1	0.000	0.006	0.003	0.000	0.000	0.004	0.002	0.004	0.005	0.006	0.005	0.009	0.000	0.004	0.002	0.005	0.000	0.000	0.005	0.005	0.000	0.000	0.006	0.008
<i>CDKN1A/P21</i>	Cyclin-dep. kinase inhibitor 1A (p21, Cip1)	6p21.2	0.058	0.019	0.011	0.079	0.000	0.032	0.015	0.027	0.004	0.003	0.037	0.000	0.000	0.020	0.010	0.004	0.009	0.013	0.017	0.014	0.076	0.021	0.004	0.013
<i>CDKN2A</i>	interleukin 6 (interferon, beta 2)	9p21	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.003	0.001	0.003	0.011	0.002	0.001	0.001	0.000	0.000	0.003	0.000	0.000	0.000	0.003	0.000
<i>CHEK1</i>	CHK1 checkpoint homolog (<i>S. pombe</i>)	11q24-q24	0.022	0.000	0.001	0.000	0.000	0.000	0.001	0.002	0.002	0.003	0.001	0.006	0.000	0.000	0.002	0.001	0.009	0.000	0.002	0.001	0.007	0.000	0.007	0.000
<i>CHEK2/RAD53</i>	CHK2 checkpoint homolog (<i>S. pombe</i>)	22q11 22q12.1	0.000	0.000	0.000	0.000	0.000	0.004	0.005	0.000	0.001	0.003	0.002	0.006	0.000	0.000	0.001	0.004	0.009	0.000	0.003	0.001	0.000	0.000	0.003	0.000
<i>CSPG2/VCAN</i>	Versican	5q14.3	0.022	0.026	0.013	0.020	0.000	0.018	0.000	0.000	0.006	0.006	0.014	0.000	0.000	0.007	0.020	0.013	0.034	0.000	0.016	0.008	0.013	0.000	0.003	0.031
<i>CX3CL1</i>	Chemokine (C-X3-C motif) ligand 1	16q13	0.000	0.000	0.007	0.000	0.000	0.000	0.007	0.002	0.002	0.005	0.011	0.000	0.000	0.015	0.001	0.019	0.000	0.000	0.002	0.006	0.002	0.000	0.001	0.000

(continued on next page)

Table 1 (continued)

Gene name	Description	Cytogenetic location	Bone	Bone marrow	Brain	Cartilage	Cervix	Colon	Eye	Head and neck	Kidney	Liver	Lung	Lymph node	Lympho-reticular	Mammary gland	Muscle	Nervous tissue	Ovary	Pancreas	Placenta	Prostate	Skin	Stomach	Testis	Uterus
<i>DAPK1</i>	Death-associated protein kinase 1	9q34.1	0.000	0.000	0.002	0.000	0.000	0.007	0.006	0.002	0.001	0.004	0.007	0.000	0.000	0.011	0.012	0.008	0.009	0.013	0.044	0.001	0.009	0.000	0.001	0.003
<i>DAXX</i>	Death-associated protein 6	6p21.3	0.015	0.013	0.001	0.013	0.000	0.004	0.013	0.020	0.002	0.003	0.005	0.006	0.011	0.020	0.005	0.003	0.000	0.000	0.002	0.001	0.007	0.004	0.010	0.005
<i>DDR1</i>	Discoidin domain rec. family, member 1	6p21.3	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.013	0.001	0.000	0.000	0.000	0.000	0.000
<i>E2F1</i>	APBA2BP, amyloid beta (A4) precursor	20q11.22	0.022	0.019	0.007	0.007	0.000	0.004	0.008	0.002	0.005	0.002	0.004	0.006	0.006	0.000	0.006	0.015	0.009	0.000	0.004	0.007	0.007	0.000	0.005	0.005
<i>E2F3</i>	E2F transcription factor 3	6p22	0.000	0.006	0.002	0.007	0.000	0.000	0.002	0.002	0.002	0.006	0.003	0.004	0.000	0.002	0.004	0.001	0.000	0.000	0.001	0.002	0.002	0.000	0.003	0.000
<i>ERK/EPHB2</i>	EPHB2, EPH receptor B2	1p36.1-p35	0.000	0.000	0.001	0.000	0.000	0.021	0.013	0.000	0.001	0.000	0.004	0.000	0.000	0.000	0.004	0.003	0.000	0.000	0.004	0.000	0.000	0.000	0.003	0.000
<i>FADD</i>	Fas (TNFRSF6)-assoc. via death domain	11q13.3	0.000	0.000	0.003	0.000	0.000	0.004	0.009	0.000	0.004	0.002	0.004	0.000	0.000	0.000	0.001	0.003	0.009	0.000	0.008	0.001	0.011	0.000	0.002	0.000
<i>FANCA</i>	Fanconi anemia, complementation group A	16q24.3	0.000	0.000	0.001	0.000	0.000	0.011	0.002	0.000	0.001	0.010	0.001	0.027	0.011	0.000	0.002	0.003	0.017	0.000	0.004	0.000	0.000	0.007	0.005	0.005
<i>GADD45A</i>	Growth arrest/ DNA-damage-inducible	1p31.2-p31.1	0.000	0.000	0.001	0.013	0.028	0.000	0.006	0.000	0.015	0.004	0.003	0.001	0.000	0.002	0.000	0.004	0.009	0.000	0.003	0.002	0.007	0.000	0.000	0.016
<i>HDAC1</i>	Histone deacetylase 1	1p34	0.015	0.006	0.005	0.007	0.000	0.018	0.013	0.024	0.008	0.006	0.012	0.021	0.061	0.004	0.004	0.002	0.034	0.000	0.008	0.008	0.006	0.007	0.004	0.005
<i>IL6</i>	Interleukin 6 (interferon, beta 2)	7p21	0.022	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.002	0.000	0.004	0.001	0.006	0.000	0.001	0.000	0.017	0.000	0.002	0.000	0.011	0.000	0.000	0.000
<i>JUN</i>	V-jun sarc. virus 17 oncogene homolog	1p32-p31	0.029	0.019	0.007	0.007	0.000	0.021	0.020	0.009	0.011	0.004	0.010	0.006	0.000	0.035	0.004	0.016	0.017	0.000	0.013	0.019	0.011	0.004	0.007	0.044
<i>KRAS</i>	V-Ki-ras2 rat sarc. oncogene homolog	12p12.1	0.000	0.006	0.003	0.000	0.000	0.000	0.001	0.002	0.004	0.002	0.004	0.038	0.000	0.002	0.005	0.009	0.009	0.000	0.008	0.001	0.013	0.000	0.003	0.000
<i>LIG4</i>	Ligase IV, DNA, ATP-dependent	13q33-q34	0.000	0.000	0.004	0.000	0.000	0.000	0.001	0.000	0.000	0.004	0.000	0.003	0.000	0.000	0.007	0.003	0.000	0.000	0.001	0.001	0.004	0.018	0.000	0.000
<i>MDM2</i>	Mdm2/p53 binding protein	12q14.3-q15	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.001	0.000	0.004	0.009	0.000	0.002	0.005	0.000	0.000	0.000	0.003	0.002	0.000	0.000	0.001	0.000
<i>MDM4</i>	Mdm4/p53 binding protein	1q32	0.007	0.000	0.004	0.000	0.000	0.007	0.006	0.007	0.004	0.003	0.007	0.016	0.006	0.015	0.009	0.005	0.009	0.000	0.005	0.006	0.000	0.014	0.008	0.013
<i>MSH2</i>	MutS homolog 2 nonpolyposis type 1	2p22-p21	0.000	0.000	0.007	0.000	0.000	0.000	0.001	0.004	0.006	0.001	0.003	0.010	0.017	0.000	0.002	0.013	0.000	0.000	0.001	0.000	0.006	0.004	0.007	0.000
<i>MYC</i>	V-myc myelocytomatosis homolog	8q24.12-q24.13	0.000	0.000	0.007	0.026	0.000	0.004	0.001	0.002	0.005	0.004	0.002	0.000	0.000	0.009	0.004	0.008	0.000	0.000	0.004	0.005	0.011	0.004	0.001	0.008
<i>P300/EP300</i>	EP300, E1A binding protein p300	22q13.2	0.000	0.000	0.002	0.000	0.000	0.004	0.002	0.004	0.002	0.002	0.008	0.025	0.000	0.007	0.000	0.005	0.000	0.000	0.011	0.004	0.006	0.007	0.006	0.005
<i>P38/MAPK1</i>	MAPK1, Mitogen-activated prot. Kinase 1	22q11.2 22q11.21	0.007	0.006	0.041	0.013	0.000	0.004	0.016	0.009	0.014	0.007	0.011	0.034	0.000	0.025	0.010	0.035	0.000	0.000	0.007	0.007	0.035	0.011	0.013	0.031
<i>PCAF</i>	P300/CBP-associated factor	3p24	0.000	0.006	0.005	0.000	0.000	0.000	0.005	0.004	0.002	0.005	0.005	0.006	0.000	0.002	0.005	0.001	0.009	0.000	0.021	0.005	0.004	0.004	0.002	0.003
<i>RB1</i>	Retinoblastoma 1	13q14.2	0.000	0.000	0.004	0.000	0.000	0.000	0.002	0.004	0.004	0.007	0.002	0.030	0.000	0.004	0.007	0.003	0.009	0.000	0.007	0.000	0.007	0.018	0.008	0.008
<i>REF3L</i>	REV3-like, catal. subun. of DNA pol zeta	6q21	0.000	0.000	0.003	0.007	0.000	0.004	0.009	0.000	0.003	0.006	0.005	0.010	0.006	0.005	0.016	0.006	0.009	0.000	0.006	0.007	0.002	0.004	0.004	0.018
<i>SHC1</i>	SHC (Src homolog 2 transforming prot. 1	1q21	0.007	0.013	0.020	0.053	0.000	0.007	0.008	0.011	0.039	0.006	0.014	0.007	0.039	0.004	0.013	0.007	0.017	0.000	0.014	0.025	0.061	0.004	0.006	0.008
<i>SIRT1</i>	Sirtuin 1	10q21.3	0.015	0.000	0.002	0.000	0.000	0.004	0.002	0.000	0.001	0.004	0.002	0.007	0.006	0.004	0.002	0.000	0.000	0.000	0.002	0.001	0.004	0.000	0.004	0.010

<i>STAT1</i>	Sign. Transduc.+activ. of transcript. 1	2q32.2	0.029	0.013	0.017	0.073	0.000	0.007	0.014	0.004	0.008	0.014	0.010	0.013	0.000	0.015	0.006	0.009	0.000	0.000	0.011	0.012	0.013	0.018	0.010	0.016
<i>STAT5A</i>	Sign. Transduc.+activ. of transcript.5A	17q11.2	0.015	0.000	0.001	0.007	0.000	0.004	0.007	0.002	0.003	0.004	0.005	0.004	0.022	0.005	0.002	0.002	0.017	0.000	0.004	0.005	0.006	0.004	0.001	0.005
<i>TNF</i>	TNF (superfamily, member 2)	6p21.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.009	0.000	0.002	0.001	0.000	0.000	0.001	0.000
<i>TP53</i>	Tumor protein p53 (Li-Fraumeni)	17p13.1	0.000	0.000	0.014	0.033	0.028	0.011	0.002	0.002	0.003	0.001	0.008	0.006	0.000	0.029	0.001	0.006	0.017	0.000	0.004	0.013	0.009	0.032	0.010	0.000
<i>TP53BP2</i>	Tumor protein p53 binding protein, 2	1q42.1	0.000	0.000	0.019	0.007	0.000	0.004	0.006	0.007	0.006	0.001	0.002	0.003	0.000	0.009	0.002	0.005	0.000	0.000	0.004	0.007	0.006	0.000	0.001	0.003
<i>TP73</i>	Tumor protein p73	1p36.3	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.006	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.000
<i>TP73L</i>	TP73L, Tumor protein p73-like	3q28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.011	0.000	0.001	0.000
<i>TRADD</i>	TNFRSF1A-associated via death domain	16q22	0.000	0.000	0.001	0.000	0.000	0.004	0.002	0.000	0.009	0.003	0.002	0.015	0.006	0.002	0.000	0.000	0.000	0.000	0.001	0.008	0.004	0.000	0.001	0.003
<i>WRN</i>	Werner syndrome	8p12-p11.2	0.000	0.000	0.001	0.000	0.000	0.000	0.004	0.000	0.000	0.006	0.002	0.006	0.000	0.000	0.004	0.000	0.000	0.000	0.001	0.001	0.000	0.004	0.001	0.003
<i>WT1</i>	Wilms tumor 1	11p13	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.015	0.023
<i>XRCC4</i>	XRCC4 DNA repair protein	5q13-q14	0.000	0.000	0.001	0.007	0.000	0.000	0.002	0.002	0.003	0.001	0.000	0.003	0.000	0.004	0.002	0.002	0.000	0.000	0.002	0.000	0.000	0.000	0.007	0.000
	Sum up		0.319	0.231	0.283	0.423	0.056	0.257	0.316	0.220	0.262	0.198	0.283	0.614	0.221	0.349	0.274	0.290	0.316	0.067	0.339	0.266	0.462	0.264	0.269	0.368
<i>(B) Cancer</i>																										
<i>AATF</i>	Apoptosis antagonizing transcript. factor	17q11.2-q12	0.003	0.007	0.012	0.008	0.022	0.011	0.004	0.003	0.007	0.013	0.008	0.007	0.016	0.011	0.004	0.000	0.004	0.005	0.018	0.001	0.016	0.006	0.009	0.008
<i>APAF1</i>	Apoptotic peptidase activating factor	12q23	0.000	0.000	0.001	0.000	0.000	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.007	0.001
<i>ATM</i>	Ataxia telangiectasia mutated	11q22-q23	0.000	0.003	0.002	0.000	0.000	0.005	0.004	0.003	0.000	0.002	0.007	0.015	0.000	0.002	0.000	0.000	0.000	0.001	0.005	0.006	0.002	0.002	0.000	0.006
<i>ATR</i>	Ataxia telangiectasia and Rad3 related	3q22-q24	0.007	0.000	0.001	0.000	0.002	0.007	0.004	0.011	0.000	0.003	0.002	0.007	0.000	0.007	0.000	0.000	0.000	0.001	0.000	0.007	0.005	0.006	0.005	0.004
<i>BAK1</i>	BCL2-antagonist/killer 1	6p21.3	0.002	0.007	0.004	0.003	0.007	0.003	0.000	0.000	0.004	0.002	0.002	0.004	0.009	0.001	0.000	0.000	0.003	0.001	0.000	0.004	0.004	0.008	0.005	0.010
<i>BAX</i>	BCL2-associated X protein	19q13.3-q13.4	0.002	0.000	0.001	0.000	0.000	0.001	0.002	0.040	0.000	0.002	0.003	0.000	0.002	0.001	0.000	0.000	0.000	0.003	0.003	0.000	0.001	0.000	0.001	0.002
<i>BCL2</i>	B-cell CLL/lymphoma 2	18q21.33 18q21.3	0.002	0.003	0.003	0.005	0.000	0.001	0.004	0.004	0.004	0.000	0.008	0.018	0.022	0.013	0.004	0.000	0.016	0.001	0.003	0.003	0.002	0.005	0.016	0.001
<i>BID</i>	BH3 interacting domain death agonist	22q11.1	0.015	0.007	0.008	0.003	0.004	0.005	0.004	0.001	0.003	0.004	0.001	0.007	0.000	0.001	0.000	0.000	0.001	0.011	0.000	0.003	0.006	0.000	0.002	0.004
<i>BRCA1</i>	Breast cancer 1, early onset	17q21	0.000	0.000	0.003	0.000	0.011	0.003	0.004	0.001	0.002	0.007	0.003	0.004	0.004	0.002	0.000	0.000	0.003	0.000	0.003	0.003	0.014	0.003	0.007	0.004
<i>BRCA2</i>	Breast cancer 2, early onset	13q12.3	0.002	0.003	0.000	0.000	0.000	0.003	0.004	0.001	0.003	0.000	0.001	0.018	0.000	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.004	0.000
<i>CASP2</i>	Caspase 2	7q34-q35	0.005	0.003	0.002	0.008	0.004	0.004	0.000	0.010	0.002	0.001	0.002	0.007	0.000	0.006	0.000	0.000	0.006	0.001	0.003	0.000	0.006	0.000	0.011	0.004
<i>CASP9</i>	Caspase 9	1p36.3-p36.1	0.002	0.000	0.007	0.000	0.020	0.003	0.002	0.003	0.004	0.003	0.005	0.000	0.013	0.004	0.014	0.000	0.002	0.003	0.003	0.001	0.000	0.002	0.003	0.005
<i>CCNE2</i>	Cyclin E2	8q22.1	0.002	0.010	0.002	0.000	0.004	0.001	0.000	0.001	0.000	0.000	0.002	0.004	0.000	0.003	0.000	0.000	0.002	0.001	0.000	0.003	0.002	0.000	0.007	0.000
<i>CCNG1</i>	Cyclin G1	5q32-q34	0.017	0.045	0.011	0.005	0.020	0.014	0.036	0.002	0.012	0.031	0.011	0.025	0.013	0.019	0.011	0.000	0.007	0.003	0.060	0.007	0.006	0.003	0.025	0.023
<i>CCNG2</i>	Cyclin G2	4q21.1	0.007	0.000	0.003	0.013	0.004	0.003	0.004	0.002	0.003	0.004	0.002	0.007	0.002	0.003	0.000	0.000	0.001	0.000	0.005	0.009	0.002	0.002	0.002	0.008

(continued on next page)

Table 1 (continued)

Gene name	Description	Cytogenetic location	Bone	Bone marrow	Brain	Cartilage	Cervix	Colon	Eye	Head and neck	Kidney	Liver	Lung	Lymph node	Lympho-reticular	Mammary gland	Muscle	Nervous tissue	Ovary	Pancreas	Placenta	Prostate	Skin	Stomach	Testis	Uterus	
<i>CDC14A</i>	CDC14 cell divis. cycle 14 homolog A	1p21	0.003	0.000	0.002	0.003	0.000	0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.002	0.000	0.002	
<i>CDC2</i>	Cell division cycle 2, G1 to S and G2 to M	10q21.1	0.007	0.010	0.003	0.008	0.002	0.005	0.006	0.000	0.004	0.003	0.003	0.007	0.002	0.021	0.004	0.000	0.004	0.003	0.008	0.003	0.003	0.003	0.007	0.011	
<i>CDC25A</i>	Cell division cycle 25A	3p21	0.002	0.000	0.001	0.000	0.004	0.002	0.013	0.000	0.001	0.002	0.001	0.000	0.025	0.001	0.021	0.000	0.002	0.004	0.005	0.003	0.003	0.005	0.005	0.004	
<i>CDC25C</i>	Cell division cycle 25C	5q31	0.009	0.000	0.003	0.000	0.002	0.005	0.000	0.001	0.000	0.004	0.001	0.004	0.000	0.003	0.000	0.000	0.001	0.001	0.003	0.000	0.002	0.000	0.006	0.014	
<i>CDK4</i>	Cyclin-dependent kinase 4	12q14	0.015	0.003	0.042	0.011	0.065	0.022	0.015	0.005	0.023	0.028	0.025	0.000	0.045	0.031	0.659	0.000	0.045	0.027	0.045	0.016	0.055	0.005	0.018	0.036	
<i>CDK7</i>	Cyclin-dependent kinase 7	5q12.1	0.012	0.000	0.001	0.016	0.016	0.003	0.002	0.004	0.004	0.002	0.004	0.011	0.002	0.013	0.000	0.000	0.000	0.000	0.015	0.000	0.003	0.005	0.002	0.007	
<i>CDKN1A/P21</i>	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	6p21.2	0.009	0.021	0.022	0.048	0.031	0.014	0.073	0.008	0.022	0.019	0.116	0.025	0.007	0.010	0.032	0.000	0.008	0.036	0.023	0.013	0.022	0.040	0.011	0.018	
<i>CDKN2A</i>	interleukin 6 (interferon, beta 2)	9p21	0.003	0.000	0.011	0.003	0.022	0.009	0.000	0.001	0.001	0.003	0.004	0.015	0.009	0.008	0.000	0.000	0.011	0.004	0.008	0.013	0.006	0.002	0.000	0.019	
<i>CHEK1</i>	CHK1 checkpoint homolog (<i>S. pombe</i>)	11q24-q24	0.002	0.007	0.008	0.003	0.018	0.004	0.004	0.002	0.001	0.008	0.001	0.007	0.007	0.003	0.021	0.000	0.000	0.003	0.003	0.003	0.003	0.001	0.002	0.008	0.004
<i>CHEK2/RAD53</i>	CHK2 checkpoint homolog (<i>S. pombe</i>)	22q11 22q12.1	0.000	0.010	0.002	0.003	0.013	0.002	0.004	0.000	0.001	0.000	0.002	0.000	0.002	0.006	0.011	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.001	0.014	
<i>CSPG2/VCAN</i>	Versican	5q14.3	0.003	0.003	0.018	0.040	0.000	0.012	0.002	0.027	0.014	0.003	0.004	0.000	0.000	0.000	0.004	0.000	0.006	0.027	0.000	0.007	0.001	0.017	0.016	0.006	
<i>CX3CL1</i>	Chemokine (C-X3-C motif) ligand 1	16q13	0.000	0.000	0.008	0.008	0.000	0.002	0.002	0.000	0.008	0.002	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.003	0.001	0.001	0.002	0.001	0.004	
<i>DAPK1</i>	Death-associated protein kinase 1	9q34.1	0.002	0.000	0.006	0.016	0.000	0.003	0.002	0.000	0.010	0.007	0.005	0.000	0.000	0.002	0.007	0.000	0.003	0.001	0.013	0.000	0.002	0.005	0.001	0.006	
<i>DAXX</i>	Death-associated protein 6	6p21.3	0.007	0.014	0.014	0.019	0.004	0.007	0.026	0.000	0.010	0.011	0.016	0.007	0.009	0.004	0.000	0.000	0.011	0.008	0.008	0.000	0.010	0.013	0.010	0.013	
<i>DDR1</i>	Discoidin domain receptor family, member 1	6p21.3	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.007	0.000	0.000	0.000	0.000	
<i>E2F1</i>	APBA2BP, Amyloid beta (A4) precursor	20q11.22	0.012	0.000	0.013	0.005	0.045	0.008	0.009	0.000	0.003	0.005	0.011	0.011	0.016	0.009	0.007	0.000	0.014	0.023	0.010	0.010	0.011	0.013	0.006	0.009	
<i>E2F3</i>	E2F transcription factor 3	6p22	0.007	0.007	0.002	0.000	0.007	0.007	0.002	0.002	0.003	0.001	0.001	0.004	0.002	0.004	0.011	0.000	0.001	0.001	0.000	0.003	0.007	0.003	0.004	0.001	
<i>ERK/EPHB2</i>	EPHB2, EPH receptor B2	1p36.1-p35	0.009	0.000	0.008	0.003	0.000	0.026	0.002	0.002	0.005	0.002	0.007	0.000	0.000	0.004	0.007	0.000	0.007	0.001	0.000	0.000	0.001	0.006	0.000	0.021	
<i>FADD</i>	Fas (TNFRSF6)-assoc. via death domain	11q13.3	0.009	0.000	0.006	0.003	0.013	0.006	0.009	0.006	0.009	0.007	0.011	0.007	0.009	0.009	0.000	0.000	0.005	0.030	0.010	0.001	0.014	0.003	0.005	0.006	
<i>FANCA</i>	Fanconi anemia, complement. group A	16q24.3	0.010	0.000	0.003	0.000	0.031	0.015	0.009	0.001	0.003	0.008	0.006	0.004	0.027	0.005	0.000	0.000	0.003	0.001	0.005	0.004	0.002	0.014	0.004	0.009	
<i>GADD45A</i>	Growth arrest/DNA-damage-inducible	1p31.2-p31.1	0.010	0.000	0.004	0.000	0.004	0.004	0.006	0.001	0.003	0.002	0.002	0.000	0.002	0.002	0.007	0.000	0.000	0.008	0.000	0.001	0.000	0.003	0.002	0.006	
<i>HDAC1</i>	Histone deacetylase 1	1p34	0.022	0.007	0.009	0.019	0.011	0.037	0.004	0.118	0.020	0.007	0.014	0.011	0.058	0.091	0.007	0.000	0.016	0.032	0.010	0.018	0.014	0.336	0.012	0.026	

<i>IL6</i>	Interleukin 6 (interferon, beta 2)	7p21	0.002	0.000	0.001	0.030	0.002	0.001	0.000	0.000	0.004	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	
<i>JUN</i>	V-jun sarc. virus 17 oncogene homolog	1p32-p31	0.009	0.003	0.023	0.019	0.011	0.021	0.004	0.003	0.019	0.005	0.026	0.004	0.000	0.008	0.011	0.000	0.028	0.012	0.000	0.026	0.008	0.010	0.005	0.029
<i>KRAS</i>	V-Ki-ras2 rat sarc. oncogene homolog	12p12.1	0.005	0.010	0.000	0.011	0.007	0.006	0.000	0.007	0.003	0.002	0.004	0.004	0.002	0.014	0.004	0.000	0.019	0.000	0.008	0.001	0.007	0.003	0.015	0.004
<i>LIG4</i>	Ligase IV, DNA, ATP-dependent	13q33-q34	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.000	0.004	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.002	0.001
<i>MDM2</i>	Mdm2/p53 binding protein	12q14.3-q15	0.000	0.000	0.000	0.000	0.002	0.008	0.004	0.004	0.000	0.002	0.000	0.000	0.000	0.003	0.014	0.000	0.001	0.000	0.013	0.000	0.000	0.002	0.002	0.002
<i>MDM4</i>	Mdm4/p53 binding protein	1q32	0.002	0.007	0.004	0.005	0.000	0.009	0.002	0.015	0.005	0.003	0.004	0.015	0.000	0.010	0.000	0.000	0.004	0.000	0.000	0.001	0.003	0.003	0.002	0.008
<i>MSH2</i>	MutS homolog 2 nonpolyposis type 1	2p22-p21	0.002	0.007	0.001	0.005	0.002	0.006	0.015	0.004	0.004	0.011	0.002	0.004	0.007	0.001	0.007	0.000	0.002	0.000	0.005	0.001	0.007	0.003	0.040	0.004
<i>MYC</i>	V-myc myelocytomatosis homolog	8q24.12-q24.13	0.007	0.100	0.008	0.011	0.018	0.016	0.000	0.010	0.005	0.006	0.006	0.011	0.018	0.005	0.004	0.000	0.007	0.003	0.010	0.015	0.004	0.017	0.007	0.009
<i>P300/ EP300</i>	EP300, E1A binding protein p300	22q13.2	0.002	0.000	0.008	0.005	0.002	0.012	0.013	0.016	0.003	0.008	0.006	0.011	0.002	0.015	0.000	0.000	0.002	0.001	0.000	0.001	0.005	0.006	0.002	0.009
<i>P38/ MAPK1</i>	MAPK1, Mitogen- activated prot. Kinase 1	22q11.2 22q11.21	0.022	0.014	0.012	0.019	0.013	0.016	0.009	0.059	0.007	0.014	0.009	0.029	0.000	0.033	0.004	0.000	0.014	0.019	0.003	0.016	0.006	0.013	0.027	0.015
<i>PCAF</i>	P300/CBP-associated factor	3p24	0.002	0.000	0.010	0.013	0.000	0.001	0.000	0.003	0.005	0.001	0.002	0.000	0.000	0.004	0.000	0.000	0.002	0.003	0.005	0.000	0.004	0.005	0.000	0.002
<i>RB1</i>	Retinoblastoma 1	13q14.2	0.005	0.007	0.004	0.003	0.002	0.004	0.002	0.012	0.008	0.005	0.001	0.004	0.002	0.009	0.004	0.000	0.002	0.000	0.005	0.006	0.005	0.000	0.008	0.009
<i>REF3L</i>	REV3-like, catal. subun. of DNA pol zeta	6q21	0.000	0.007	0.002	0.005	0.002	0.005	0.004	0.003	0.007	0.001	0.002	0.004	0.002	0.002	0.000	0.000	0.005	0.000	0.000	0.007	0.002	0.006	0.005	0.007
<i>SHC1</i>	SHC (Src homol. 2 transforming prot. 1	1q21	0.014	0.003	0.033	0.013	0.029	0.005	0.013	0.044	0.013	0.020	0.026	0.015	0.002	0.024	0.018	0.000	0.004	0.058	0.028	0.035	0.022	0.006	0.036	0.023
<i>SIRT1</i>	Sirtuin 1	10q21.3	0.002	0.000	0.000	0.000	0.002	0.002	0.002	0.001	0.000	0.001	0.001	0.011	0.002	0.003	0.000	0.000	0.000	0.001	0.003	0.009	0.001	0.002	0.005	0.002
<i>STAT1</i>	Sign. Transduc.+activ. of transcript. 1	2q32.2	0.012	0.014	0.006	0.011	0.016	0.016	0.004	0.014	0.028	0.023	0.006	0.025	0.004	0.013	0.004	0.000	0.012	0.015	0.010	0.007	0.029	0.094	0.021	0.066
<i>STAT5A</i>	Sign. Transduc.+activ. of transcript.5A	17q11.2	0.009	0.000	0.002	0.016	0.009	0.006	0.000	0.000	0.002	0.006	0.002	0.044	0.000	0.001	0.000	0.000	0.002	0.005	0.000	0.001	0.000	0.008	0.002	0.002
<i>TNF</i>	TNF (superfamily, member 2)	6p21.3	0.000	0.000	0.000	0.003	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>TP53</i>	Tumor protein p53 (Li-Fraumeni)	17p13.1	0.005	0.000	0.014	0.000	0.011	0.016	0.004	0.049	0.012	0.009	0.008	0.029	0.013	0.028	0.004	0.000	0.007	0.001	0.010	0.022	0.002	0.021	0.021	0.014
<i>TP53BP2</i>	Tumor protein p53 binding protein, 2	1q42.1	0.003	0.000	0.003	0.005	0.000	0.002	0.000	0.005	0.003	0.003	0.003	0.011	0.000	0.003	0.004	0.000	0.002	0.000	0.003	0.003	0.002	0.003	0.010	0.004
<i>TP73</i>	Tumor protein p73	1p36.3	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.005	0.000	0.000	0.002	0.000	0.000	0.001	0.014	0.000	0.000	0.001	0.008	0.000	0.000	0.000	0.000	0.001
<i>TP73L</i>	TP73L, Tumor protein p73-like	3q28	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.071	0.000	0.000	0.001	0.004	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.000	0.000	0.001
<i>TRADD</i>	TNFRSF1A- associated via death domain	16q22	0.002	0.000	0.003	0.000	0.000	0.005	0.000	0.001	0.007	0.002	0.009	0.000	0.002	0.004	0.000	0.000	0.002	0.004	0.000	0.000	0.001	0.000	0.000	0.001
<i>WRN</i>	Werner syndrome	8p12-p11.2	0.003	0.003	0.000	0.000	0.002	0.001	0.000	0.004	0.001	0.000	0.001	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.003
<i>WT1</i>	Wilms tumor 1	11p13	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.006	0.001	0.000	0.000	0.000	0.000	0.001	0.002
<i>XRCC4</i>	XRCC4 DNA repair protein	5q13-q14	0.000	0.021	0.002	0.000	0.000	0.001	0.026	0.001	0.000	0.002	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.003	0.000	0.004	0.003	0.002	0.003	0.003
	Sum up		0.328	0.372	0.391	0.427	0.526	0.421	0.374	0.591	0.333	0.319	0.423	0.482	0.365	0.498	0.923	0.000	0.312	0.375	0.374	0.313	0.345	0.720	0.440	0.537

The columns give the following information (from left to right): short name of the gene, gene description, expression for 24 cancers.

which there is no column indicated, no corresponding data are available in dbEST.

In normal tissue, the expression levels of *TP53* range from 0.033 in cartilage to 0.001 in liver, i.e., over the 24 tissues *TP53* expression varies by a factor of 33. Also the effects of cancer are very different. In cervix, ovary, skin, and stomach cancer *TP53* is down regulated while it is up regulated in other cancers by up to a factor of 25 in head and neck cancers. Thus, the expression levels of *TP53* and the tested 63 gene products of related pathways do not show any consistent trend.

The situation changes dramatically when 63 *TP53*-related genes are analyzed. Fig. 2 shows, for the example of normal bone and ovary tissues, an interesting and consistently observed principle. The columns in Fig. 2, which represent the sums of the expression levels of the 63 genes in two tissues as different as bone and ovary, have similar heights, i.e., their cumulated expression levels are similar, despite the fact that their compositions, i.e., the contributions of individual genes, are very different. For example, in bone, *CDKN1A/P21* (dotted slices) has an expression level sixfold that in ovary. This difference is compensated for by decreased expression levels of other genes. Such a behavior defines the system of 63 p53-interacting genes as a candidate for a functional gene ensemble as mentioned in the introduction.

Fig. 3 shows that the observation of Fig. 2 is not specific for these two cancers but can be more generalized. Five tissues are shown, now in comparison with their corresponding cancers.

The tissues have been selected because their cumulated expression levels in cancer are similar to or only slightly higher than those of their normal tissues. Again, similar cumulated levels are achieved with very different contributions of single genes.

Fig. 4 gives the corresponding data for all tissues. The columns are plotted as pairs, with the normal tissue on the left and the corresponding cancer tissue on the right. Again, the compositions of the different columns vary vastly. All columns, except those of normal cervix and pancreas tissue and those of muscle and stomach cancer, have heights between 0.2 and 0.6, i.e., the variation of a factor of up to 33 for individual genes has been reduced to a factor of 3. A major part of this reduction just reflects the better statistics upon averaging. However, since data for normal and cancer tissue are treated identically, a remaining difference in the mean value and the variation can be isolated from these data that reflects a real biomedical effect.

When the cumulated expression levels in cancer are plotted as a function of the expression levels in normal tissues, a linear regression yields a straight line with a slope close to zero; thus, cumulated expression levels in cancer are independent of those in normal tissue.

The message of Fig. 5 is that there is a tendency, in cancer, to unify the cumulated expression levels. Note that this statement still holds when the two extreme cases of muscle cancer and cancer of the nervous tissue (either one of them or both) are removed from the regression analysis.

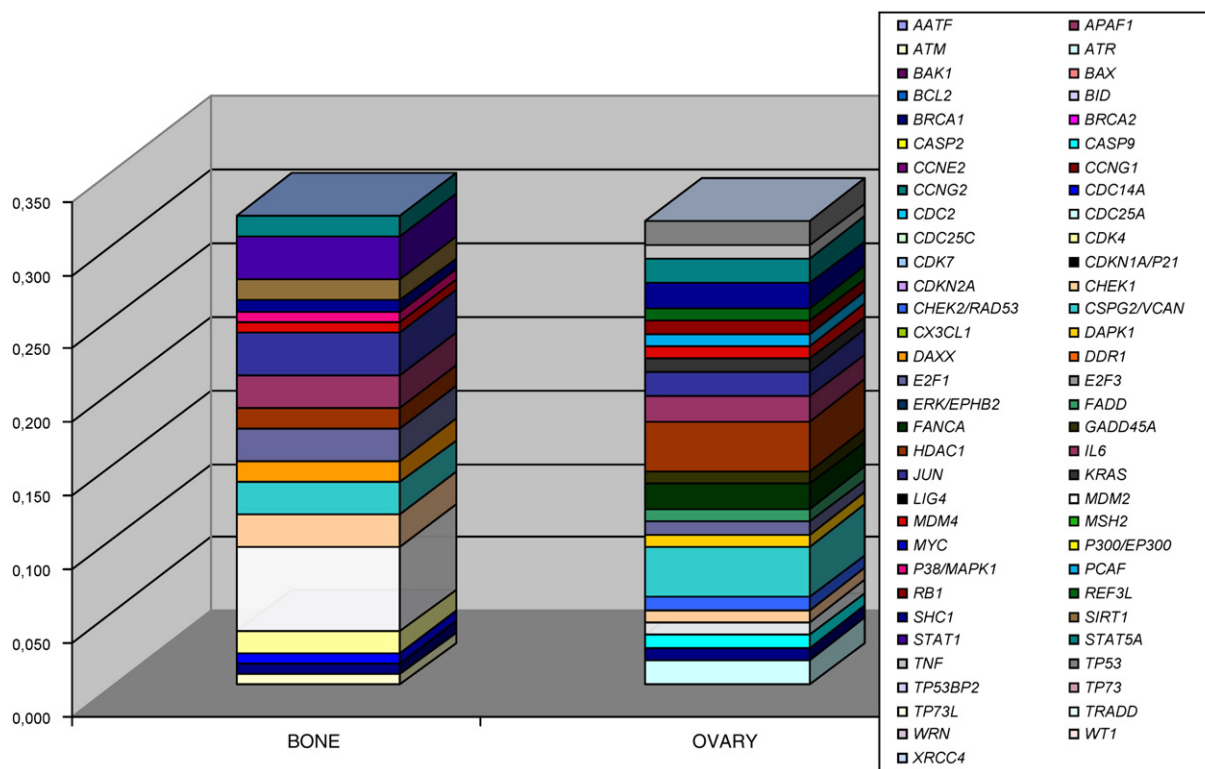


Fig. 2. Cumulated, nonweighted expression levels of all 63 genes for normal bone and ovary tissue. Both columns are similar in height, i.e., the cumulated expression levels are similar. However, individual genes contribute very differently to the sum. For example, *CDKN1A/P21* (white dotted slice, fifth from bottom) contributes sixfold more in bone than in ovary (fourth from bottom).

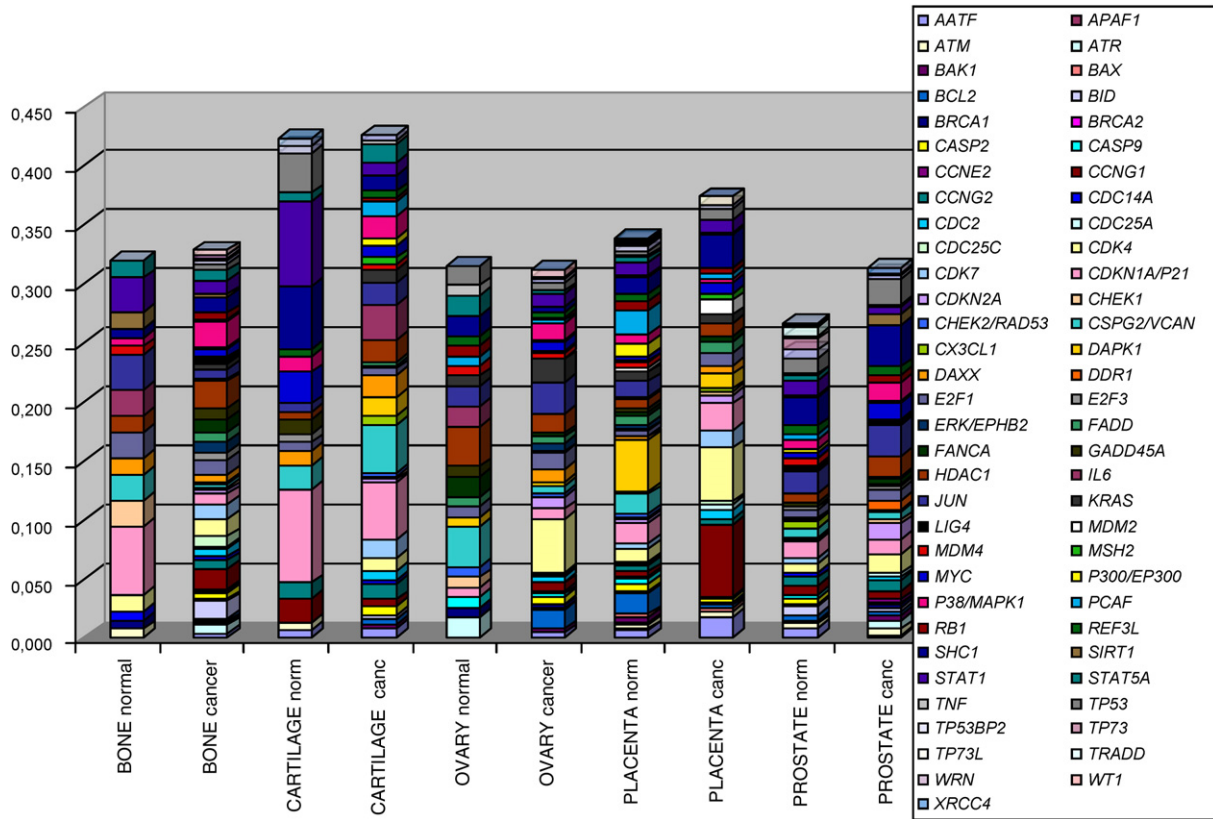


Fig. 3. Cumulated, nonweighted expression levels for bone, cartilage, ovary, placenta, and prostate together with their cancers.

The discussion so far has treated each tissue separately. While such an approach is sufficient to get some plausible results, it treats a given gene in different tissues as a completely

independent quantity. However, genes of, for example, a classical pathway should behave similarly in different tissues, i.e., they are not really independent. Therefore, all 63 genes in

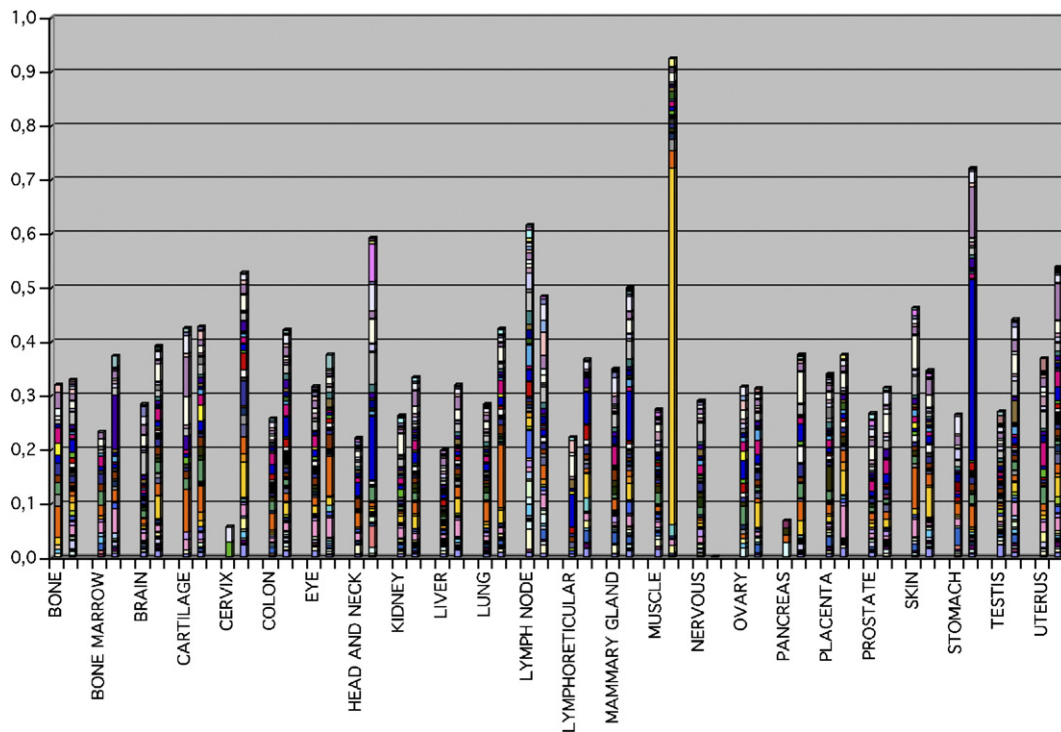


Fig. 4. Cumulated, nonweighted expression levels in all 24 tissues for normal tissue and cancer with the single genes as slices in the cumulated column.

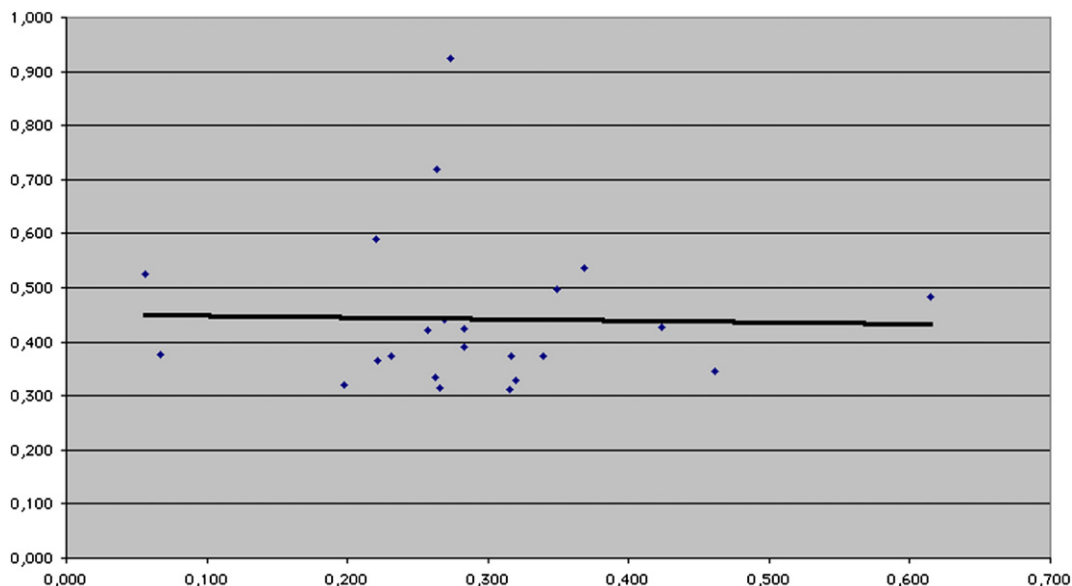


Fig. 5. Cumulated, nonweighted expression levels sorted according to the corresponding value in normal tissue. Cervix, with the lowest expression in normal tissue, is in the leftmost position. The solid line is the regression line.

all 24 normal tissues and all 24 cancer tissues have to be regarded as one large ensemble. A first hint comes from Fig. 5. There, for most cancers, the cumulated gene expression is higher for cancer than for normal tissue, but for some tissues the situation is reverse. Therefore, cumulated gene expression does not completely qualify as a statistical classifier.

Weighted sums, instead of cumulated gene expressions, are better suited to separate the values for normal and cancer tissue. Regarding the weights as variables, we define a system of 48 linear equations, one for each tissue, the weighted sum being +1 for normal and -1 for cancer tissue, respectively. As described under Materials and methods, the system can be solved for the weights. In fact, there is a 15-dimensional solution space. In Table 2, we present only one extreme solution for the weights, which have minimal moduli with respect to a certain minimality criterion (Materials and methods).

It is difficult to characterize the statistical significance of this linear classifier, as it should incorporate the significance levels of the gene expression levels themselves, which are not precisely known. Instead, we compute the maximal simultaneous deviation tolerance for all gene expression levels. For the given weighted sum a simultaneous deviation of 5.18% from all norm expression levels is tolerable and will still give a true positive. If all deviations are higher and have the right sign, a false positive will result. But as the particular deviations need to have the same sign (opposite sign, respectively) for cancer (normal, respectively) tissue as the associated weight, this particular event has a probability of less than 2^{-63} to give a false positive. To describe the statistical relevance completely, all possible deviation combinations would have to be taken into account, of course.

So, in practice, individual deviations of gene expression levels can be much higher than 5.18%, as others will be lower or point to the opposite direction. Altogether, the prediction of the

weighted sum is “normal tissue” for a value larger than -0.0558 and “cancer tissue” for a value smaller than -0.0558 .

Discussion

The practical use of the classifier defined in this work is that for a sample from an arbitrary test tissue, for example from an

Table 2
Weights for the 63 genes calculated for the data of Table 1 to obtain a robust statistical classifier

<i>AATF</i>	-26.6	<i>CDKN1A/P21</i>	-4.6	<i>MDM4</i>	37.8
<i>APAF1</i>	129.9	<i>CDKN2A</i>	-79.3	<i>MSH2</i>	-31.7
<i>ATM</i>	22.6	<i>CHEK1</i>	-88.7	<i>MYC</i>	0.4
<i>ATR</i>	-5.3	<i>CHEK2/RAD53</i>	196.4	<i>P300</i>	13.8
<i>BAK1</i>	7.3	<i>CSPG2</i>	-14.6	<i>P38/MAPK1</i>	1.3
<i>BAX</i>	-21.5	<i>CX3CL1</i>	43.5	<i>PCAF</i>	56.8
<i>BCL2</i>	27.2	<i>DAPK1</i>	7.5	<i>RB1</i>	-40.7
<i>BID</i>	-43.0	<i>DAXX</i>	29.7	<i>REF3L</i>	-8.1
<i>BRCA1</i>	30.8	<i>DDR1</i>	82.0	<i>SHC1</i>	24.6
<i>BRCA2</i>	-26.7	<i>E2F1</i>	16.0	<i>SIRT1</i>	118.7
<i>CASP2</i>	14.7	<i>E2F3</i>	93.5	<i>STAT1</i>	-0.2
<i>CASP9</i>	-28.2	<i>ERK/EPHB2</i>	-23.9	<i>STAT5A</i>	25.8
<i>CCNE2</i>	-80.1	<i>FADD</i>	-54.6	<i>TNF</i>	-46.4
<i>CCNG1</i>	13.0	<i>FANCA</i>	69.8	<i>TP53</i>	-28.6
<i>CCNG2</i>	-34.9	<i>GADD45A</i>	64.5	<i>TP53BP2</i>	55.3
<i>CDC14A</i>	182.8	<i>HDAC1</i>	-4.3	<i>TP63</i>	16.2
<i>CDC2</i>	-13.2	<i>IL6</i>	-10.8	<i>TP73</i>	88.7
<i>CDC25A</i>	-55.5	<i>JUN</i>	-40.5	<i>TRADD</i>	-26.0
<i>CDC25C</i>	-79.5	<i>KRAS2</i>	-15.6	<i>WRN</i>	-109.8
<i>CDK4</i>	-3.9	<i>LIG4</i>	12.7	<i>WT1</i>	65.2
<i>CDK7</i>	-87.0	<i>MDM2</i>	1.7	<i>XRCC4</i>	-96.3

With these weights, all 24 normal tissues, without exception, have a weighted sum of +1; those of all 24 cancer tissues, without exception, have a sum of -1. When we apply this classifier to 13 additional tissues and their cancers, which were, due to poor data quality, not included into the construction of the classifier, 19/26 cases are correctly classified. Of the remaining 7 cases (4 normal and 3 cancer tissues), 5 have an exceedingly high expression of the oncogene JUN. If one corrects for this fact, only 2 cases are finally predicted incorrectly.

individual patient, with uncertain diagnosis, it can be robustly decided, on the basis of the expressions of the selected 63 genes and with the determined 63 weight factors, if it is cancerous (then its weighted sum should be close to -1) or normal (then its weighted sum should be close to $+1$). Different subtypes of cancers, different tumor stages, possible ethnic variability, and epidemiological aspects can then be determined as deviations of details in the set of 63 *TP53*-related genes.

A finding of more principal importance is that despite gross differences in the expression of single genes the cumulated expression of the 63 varies only by a factor of 3 in all 24 healthy and cancer tissues, with very few exceptions. Such a statement would be almost trivial if the ensemble of investigated genes were the whole set of genes in a cell, since on the one hand, one would expect that the expression of individual genes is massively dysregulated to make a cell cancerous, but on the other hand, the cumulated expression and use of all genes in a cell cannot be grossly different from that of a normal cell, since even in aggressive cancer cells the metabolism as a whole is increased at most a few fold. That such an observation is made already at the level of an ensemble of only 63 genes, i.e., less than 1% of all genes in a cell, is remarkable, though it will require additional experiments to understand this in molecular detail.

There may be the objection that the effects we have seen are just due to an improved statistics upon “averaging” over many genes. This is, however, not the case: With 63 genes randomly selected from genes of stress response, we find no increase in the average cumulative expression between normal and cancer tissue (even a slight, though nonsignificant decrease) and no significant narrowing of the relative standard deviation over the 24 tissues. Thus, an increase in overall expression and a narrowing of expression levels are real.

Materials and methods

Sixty-three genes were selected from the pathways involving the *TP53* tumor-suppressor gene. They are located up- or downstream of *TP53* and were selected from the following pathways according to the KEGG (www.genome.jp/kegg): ko04310, ko04210, ko05030, and ko05040. We have not only chosen genes that are direct binding partners of p53, but have included further genes (mostly tumor-suppressor or oncogenes of the pathways MAP kinase signaling, cell cycle, and apoptosis and transcription factors).

Expression data for these genes as well as the classification of tissues have been obtained via the Virtual Northern function of NIH’s database dbEST [14,15,20]. This database searches the literature for data from DNA or oligonucleotide chip experiments, normalizes them, and lists the results gene by gene for a set of 51 tissues (plus three pooled values). Table 3 gives a typical output, which one can obtain via <http://cgap.nci.nih.gov/Genes/GeneFinder> for the example of the *TP53* gene.

The data in this table are already preevaluated and imply sophisticated statistical considerations, which have been accepted by a wide range of users of this database as being adequate for describing gene expression in normal and in cancer tissue [14,20]. For example, to get an idea whether the statistical quality of the data is sufficient to state overexpression in cancer, let us first assume that the Gaussian statistics is valid. If one finds 16 counts for cancer, the Gaussian error is $\sqrt{16}=4$. If for normal tissue 9 ± 3 counts are found, the ratio may be formally calculated as 1.77. The calculated error of the difference is then 7 counts, as the measured difference. For some applications one may accept this as “just significant.” One should, however, realize that, according to the Gaussian theory, approximately 32% of all ratios in that case indicate “overexpression,”

Table 3

A typical output obtained from dbEST/Virtual Northern for *TP53*

Tissue	ESTs normal	ESTs cancer	ESTs <i>p</i>
All tissues	286/3,383,090	347/2,493,099	0
Adrenal cortex		1/10,398	
Adrenal medulla	0/563		
Bone	0/16,327	3/58,481	0.32
Bone marrow	0/15,569	0/29,006	
Brain	46/281,828	29/205,624	0.27
Cartilage	5/15,115	0/37,250	0.02
Cerebellum	1/85,605	0/0	
Cerebrum	11/201,756	1/3,409	0.3
Cervix	1/3,587	5/44,447	0.37
Colon	3/28,048	27/170,009	0.27
Ear	0/16,691		
Embryonic tissue	18/206,103		
Endocrine	3/16,599	0/3,196	0.36
Esophagus	0/87	2/16,534	0.5
Eye	2/123,703	2/46,738	0.3
Gastrointestinal tract	6/27,888	1/12,896	0.24
Genitourinary	5/8,868	5/27,005	0.11
Head and neck	1/42,364	50/101,855	0
Heart	2/76,379		
Kidney	3/105,997	11/91,306	0.02
Limb			
Liver	1/108,116	9/100,640	0.02
Lung	11/132,698	14/178,796	0.46
Lymph node	4/67,732	8/27,585	0.02
Lymphoreticular	0/18,077	6/44,599	0.15
Mammary gland	16/55,033	32/115,515	0.45
Muscle	1/81,895	1/28,054	0.38
Nervous	14/230,648	0/0	
Ovary	2/11,727	7/94,538	0.29
Pancreas	0/7,427	1/75,431	0.47
Pancreatic islet	2/101,122	0/30,585	0.37
Parathyroid		0/20,837	
Peripheral nervous system	0/15,784	0/908	
Pineal gland	0/6,855		
Pituitary gland	0/13,819	0/1,576	
Placenta	11/254,569	4/39,851	0.17
Pooled tissue	6/293,822	0/0	
Prostate	11/84,150	15/68,285	0.12
Retina	2/33,149		
Salivary gland	1/2,589	1/10,358	0.35
Skin	5/53,904	2/125,563	0.06
Spleen	20/52,423		
Stem cell			
Stomach	9/28,058	13/62,881	0.21
Synovium	0/259	0/1,875	
Testis	14/146,275	26/121,114	0.01
Thymus	1/4,775	0/179	0.49
Thyroid	4/13,493	6/38,730	0.25
Uncharacterized tissue	24/162,630	46/306,017	0.46
Uterus	0/38,584	19/140,465	0.02
Vascular	20/45,617		
White blood cells			
Whole body	0/45,346		

For a set of 51 tissues (plus four pooled data sets) values for expression in normal tissue and in the corresponding tumor tissue are given (for details see text). The *p* values in the column on the right indicate the statistical quality of the data for each cancer. This value has to be taken into account when a single gene for a single tissue is discussed; only data with *p* values <0.05 are completely reliable. In the discussion of whole ensembles of data, larger *p* values are acceptable.

whereas this statement is not true but a result of statistical chance. If the same ratio ($=1.77$) had been obtained with $160 \pm \sqrt{160}$ and $90 \pm \sqrt{90}$ counts, the error of the difference would be $12.7 + 9.5 = 20.2$ counts at a measured difference of 60 counts. Here, the risk of giving a false statement “overexpression” is much lower, a few percent. Also, low count data allow, in contrast to the first example above, reliable statements, if the ratio is high. For example, for 25 ± 5 and 9 ± 3 the ratio is 2.8, its calculated error difference of 8 at a measured difference of 16 is 2 standard deviations. The probability that a statement with more than 2 standard deviations is obtained by chance is approximately 5% or 1:20, i.e., only in every 20th data set would a wrong statement on overexpression result for purely statistical reasons. The database dbEST gives these percentages, divided by 100, as p values, for each data pair “cancer versus normal,” and allows also for non-Gaussian statistics where necessary, for example, at count rates below 10, where Poisson statistics replaces Gaussian statistics. These measures of statistical quality can also be applied to any other pair of data, for example, when results have to be compared between two (or pair-wise between many) tissues.

Currently, the database contains information on the expression of some 4 million genes or ESTs. For a given tissue, typically some 10,000 to 100,000 pieces of data are available. To understand this in more detail, one may imagine a large chip for each tissue with a number of hybridization signals for each gene or EST. The number “ h ” of hybridization signals (typically a few up to a few tens) divided by the number “ t ” of hybridization targets on the chip is used as the expression level, usually a number much smaller than 1. Note that the way to evaluate data given by dbEST is not a result of the present work, but essentially given by the providers of the database at NIH. With the remarks above we just try to make the complex dbEST data comprehensible to a reader who does not wish to go into the details of the database (for additional details see also [3,8]).

For example, in the case of colon cancer, Table 3 gives $h=27$ hybridization signals of p53 on $t=170,009$ targets, i.e., $h/t=0.000159$. Since we are interested only in relative expression levels, we multiply h/t by 100 for convenience, i.e., the value 0.0159 will be used in Table 1.

Once one has complete expression data for many genes and many tissues, one may attempt to get a quantitative measure to distinguish between normal and cancer tissue. Such a measure is called a statistical classifier. In a quite general sense, a classifier is the result of a mathematical procedure, which can be defined more or less arbitrarily. In our case it is just a suitably chosen sum of expression levels. In the first part of the Results we used the direct (*nonweighted*) sums of the expression levels of the 63 genes to discuss differences and similarities of gene expression in cancer and normal tissues. This explained a few basic results of our investigation but would not be suitable as a classifier. Therefore, we used *weighted* sums of the gene expression levels a_{ji} for normal and cancer tissues. We looked for 63 a priori unknown weights g_i as the solution of a system of 48 linear equations. There the expression levels, a_{ji} for the 63 genes in the 48 ($=24$ normal + 24 cancer) tissues as they are listed in Table 1, are the coefficients of the corresponding equation system. To *construct* the classifier, we have used just the 24 tissues and their cancers, which are discussed throughout this paper. We go, however, a step further and *apply* that classifier to 13 additional tissues, which has a fundamental drawback: since they are in the literature less well investigated tissues, the statistical quality of their data is poor and would usually not be suitable for distinction between normal and cancer tissue. Despite that, otherwise, serious problem, a result of our study is that, with the classifier obtained here, it is possible to classify tissues even when the statistical quality of the data is poor.

We define the weighted sums for all normal tissues to be $b_j=1$ for normal tissue and $b_j=-1$ for cancer tissue, resulting in the equation

$$\sum a_{ji}g_i = b_j (i = 1, \dots, 63; j = 1, \dots, 48).$$

In the first part of the Results, in which we discuss “cumulated expression levels,” all g_i are set to 1, i.e., all expression levels will have the same weight of 1. A disadvantage of this nonweighted approach is that genes with a low expression level are underestimated in their importance. In the second part of the Results the weights are different for each gene and thus allow for the relative importance of each given gene.

“Full pivoting Gauss elimination” [21] showed that the system has maximal rank yielding a 15 ($=63-48$)-dimensional solution space. Accordingly, 15 parameters can be chosen freely and the other 48 are “affine” combinations of these. We chose several objective functions to find optimal solutions in this

parameter space. The solution we present in Table 2 is minimal with respect to the weighted (weights h_i) sum of the 63 squared weights g_i . The weights h_i here are calculated as the square roots of the maximum of the associated expression levels of the respective gene in all tissues. Choosing other weights h_i in objective functions does not alter the character of the minimal solution substantially. The minimal solution was found by usual weighted least-squares fit routines [22].

To get an impression of the statistical relevance of such an “all tissue” classifier, we compute the maximal possible derivation of all norm expression values, such that, if all expression level measurements deviate less than this value, the prediction “normal” or “cancer” is correct (true positive). For normal tissue, the value of the weighted sum is decreased if, for positive weights g_i , the expression level is decreased, e.g., by a factor of $1-p$. For negative weights g_i , it is increased by a factor of $1+p$. For cancer tissue, the respective factors are $1+p$ and $1-p$, respectively. Equating the maximal deviations for normal and cancer tissue, we find the maximal tolerable deviation from the norm level in the case in which all deviations combine in the decreasing (for normal tissue) or increasing (cancer tissue) direction (worst case analysis). The tolerance level we find for our example is $p=0.0518$ or 5.18% and the corresponding weight sum for the disturbed expression levels is $s=-0.0558$.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft, Grant Po 284/8–9, and the Thüringer Ministry of Technology and Labour, Grant B 309-05001. B.A. thanks Dr. Toby Gibson for enabling her to perform this work at the European Molecular Biology Laboratory.

References

- [1] O. Warburg, Über den Stoffwechsel der Karzinomzelle (On the metabolism of the carcinoma cell), *Naturwissenschaften* 12 (1924) 1131–1137.
- [2] K. Garber, Energy boost: the Warburg effect returns in a new theory of cancer, *J. Natl. Cancer Inst.* 96 (2004) 1805–1806.
- [3] B. Altenberg, K.O. Greulich, Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes, *Genomics* 84 (2004) 1014–1020.
- [4] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D.S. Goldberg, L.V. Zhang, S.L. Wong, G. Franklin, S. Li, J.S. Albalá, J. Lim, C. Fraughton, E. Llamas, S. Cevik, C. Bex, P. Lamesch, R.S. Sikorski, J. Vandenhaute, H.Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M.E. Cusick, D.E. Hill, F.P. Roth, M. Vidal, Towards a proteome-scale map of the human protein–protein interaction network, *Nature* 437 (2005) 1173–1178.
- [5] J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, J.S. Weissman, Single cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise, *Nature* 441 (2006) 840–844.
- [6] J. Lotem, D. Netanel, E. Domany, L. Sachs, Human cancers overexpress genes that are specific to a variety of normal human tissues, *Proc. Natl. Acad. Sci. USA* 102 (2005) 18556–18561.
- [7] C.F. Basil, Y. Zhao, K. Zavaglia, P. Jin, M.C. Panelli, S. Voiculescu, S. Mandruzzato, H.M. Lee, B. Seliger, R.S. Freedman, P.R. Taylor, N. Hu, P. Zanovello, F.M. Marincola, E. Wang, Common cancer biomarkers, *Cancer Res.* 66 (2006) 2953–2961.
- [8] B. Altenberg, C. Gemuend, K.O. Greulich, Ubiquitous cancer genes: multipurpose molecules for protein micro-arrays, *Proteomics* 6 (2006) 67–71.
- [9] T. Soussi, K. Dehouche, C. Beroud, p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis, *Hum. Mutat.* 15 (2000) 105–113.
- [10] B. Vogelstein, D. Lane, A.J. Levine, Surfing the p53 network, *Nature* 408 (2000) 307–310.

- [11] S.B. Gruber, G.M. Petersen, K.W. Kinzler, B. Vogelstein, Cancer, crash sites, and the new genetics of neoplasia, *Gastroenterology* 116 (1999) 210–212.
- [12] B. Vogelstein, K.W. Kinzler, Cancer genes and the pathways they control, *Nat. Med.* 10 (2004) 789–799.
- [13] M. Oren, A. Damalas, T. Gottlieb, D. Michael, J. Taplick, J.F. Leal, R. Maya, M. Moas, R. Seger, Y. Taya, A. Ben Ze'Ev, Regulation of p53: intricate loops and delicate balances, *Ann. N. Y. Acad. Sci.* 973 (2002) 374–383.
- [14] A. Lal, A.E. Lash, S.F. Altschul, V. Velculescu, L. Zhang, R.E. McLendon, M.A. Marra, C. Prange, P.J. Morin, K. Polyak, N. Papadopoulos, B. Vogelstein, K.W. Kinzler, R.L. Strausberg, G.J. Riggins, A public database for gene expression in human cancers, *Cancer Res.* 59 (1999) 5403–5407.
- [15] M.S. Boguski, T.M. Lowe, C.M. Tolstoshev, dbEST—database for “expressed sequence tags”, *Nat. Genet.* 4 (1993) 332–333.
- [16] C. Schaefer, L. Grouse, K. Buetow, R.L. Strausberg, A new cancer genome anatomy project web resource for the community, *Cancer J.* 7 (2001) 52–60.
- [17] R.L. Strausberg, S.F. Greenhut, L.H. Grouse, C.F. Schaefer, K.H. Buetow, In silico analysis of cancer through the Cancer Genome Anatomy Project, *Trends Cell Biol.* 11 (2001) S66–S71.
- [18] R.L. Strausberg, The Cancer Genome Anatomy Project: new resources for reading the molecular signatures of cancer, *J. Pathol.* 195 (2001) 31–40.
- [19] M.S. Boguski, C.M. Tolstoshev, D.E. Bassett Jr., Gene discovery in dbEST, *Science* 265 (1994) 1993–1994.
- [20] L. Zhang, X.L. Ma, Q. Zhang, C.L. Ma, P.P. Wang, Y.F. Sun, Y.X. Zhao, H. Zhang, Expressed sequence tags from a NaCl-treated Suaeda salsa cDNA library, *Gene* 267 (2001) 193–200.
- [21] E. Isaacson, H.B. Keller, *Analysis of Numerical Methods*, Wiley, New York, 1966.
- [22] W.K. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes—The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge, UK, 1986.