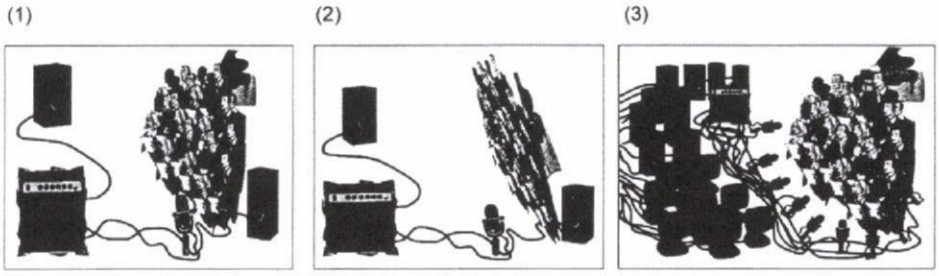

Figure 3.3 *Information from gene chip and/or microarray analyses is multidimensional. A microarray consisted of 34 000 genes behaving in the 30 000-dimensional Euclidian universe; thus, proper computational analysis requires appropriate multidimensional analytical power. See text.*

Figure 3.3a shows a cartoon where 30 000 persons are expressing (dispatching) their individual information. Depending on how a receiver receives the information, such information might be distorted because of the low-dimensional accepting system; the information might be distorted solely by expression intensities, as shown in Figure 3.3b for example. The possible reason why the intended information is not obtained from the microarray data is not always because of the poorly normalized raw data or the lack of significant data. One may have to realize that the reason may be the insufficient computational incorporation of multidimensional aspects of microarray data. Consequently, it may be appropriate when a possible multidimensional analysis of microarray data can be carried out to recover the data multidimensionally, as shown in Figure 3.3c.

3.6 Reproducibility of Microarray Data and Comparison among the Data Obtained from Different Platforms

Reproducibility, homogeneity and stability during data sampling are critical so that global gene expressions can be compared, because RNAs sampled from tissues are labile in general. Thus, the proper robotic administration of test compounds and the standardization of test materials are of special importance, the technical advancement of which has contributed markedly to the present microarray technologies (Schadt *et al.*, 2001; Churchill, 2002; Kroll and Wolfl, 2002; Astrand, 2003; Stoyanova *et al.*, 2004). Furthermore, spotting and spike gene incorporations also contributed to the development of a hardware system that supports qualitative data evaluation and semiquantitative gene expression profiles (Hill *et al.*, 2001). Consequently, newly introduced journals in the field of computational sciences publish a number of informational studies. Data from different platforms are also actively compared and reported because of minimum requirements in the field of microarray study (Brazma *et al.*, 2001). Actual sample evaluation data are not presented in this paper; however, the similarities of those obtained data with those obtained from different platforms are often confirmed by PCA. This is not only because of equivalences at each technical level of data processing, but also because it is more likely that those gene expression data are sequentially linked together. Current gene expression data obtained by

microarray and gene chip technologies are largely comparable, unless specific spike genes or other additional modifiers are incorporated in the system (Petersen *et al.*, 2005).

3.7 Pathological and Toxicological Endpoints

In this section, let us consider the general data mining of observed outcomes obtained by gene chip and microarray technologies. From Figure 3.4, which shows different routes from the bottom to the mountain top, one can assume that the top where the routes merge is a pathological endpoint where common genes are mostly expressed, whereas the other enclosed area at the foot of the mountain can be considered a toxicological endpoint, which is assumed to be based on the different routes to the summit represented by stochastic and probabilistic gene expression clusters. The former is considered to include a group of specific and deterministic gene expression profiles, i.e. an 'essential leukemogenic gene' profile, which could be used as possible biomarker genes for diagnosis, whereas the latter clusters represent various probabilistic uncertainties whose profiles are considered to be different from one cluster (route) to another, i.e. 'stochastically necessary genes.' In this regard, toxicological endpoints do not seem to provide definitive information of gene expression; however, toxicological endpoints represented by 'stochastically necessary genes'

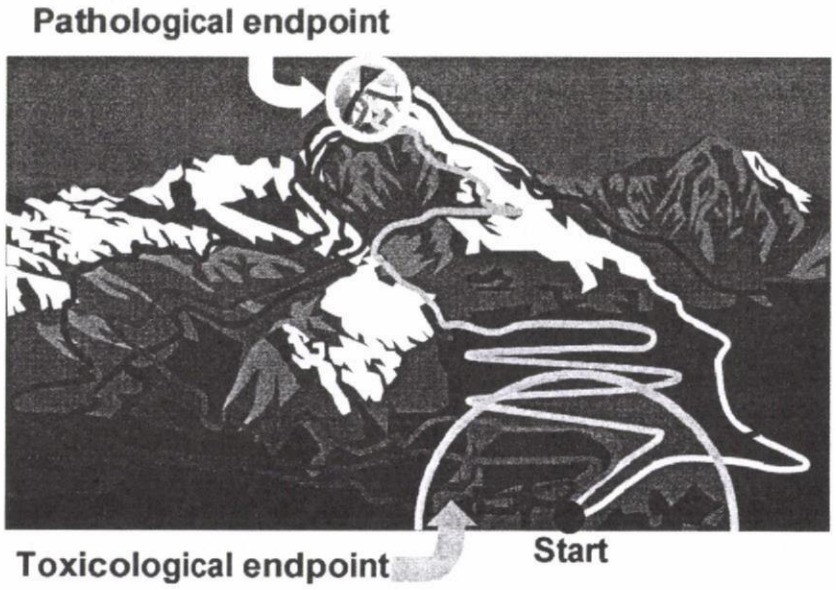

Figure 3.4 *'Pathological endpoint' and 'toxicological endpoint'. Each route from the start to the summit represents an individual probabilistic variety of different gene expression profiles. Depending on the characteristics of toxicological impacts, responders may show different toxicological endpoints in different clusters, even if one uses an identical and homogeneous experimental protocol with a highly purified inbred strain (a probabilistic quantum effect based on the uncertainty principle; see text).*

can provide probabilistic but cluster-specific predictability among various independent clusters. The latter clusters may not be statistically determined unless hundreds of a relatively large number of quantum cases are examined. Both an 'essential leukemogenic gene' profile and a 'stochastically necessary gene' profile are required for the early prediction of radiation-induced myelogenous leukemogenesis.

3.7.1 Pathological Endpoints

The strain C3H/He mouse develops a similar myelogenous leukemia both spontaneously and upon irradiation. The average incidence of the latter is 35% after 3 Gy irradiation, whereas that of the former is 1% (Seki *et al.*, 1991; Yoshida *et al.*, 1997). The line configuration in Figure 3.5 shows six cases of spontaneous leukemias on the left and six cases of radiation-induced myelogenous leukemias on the right. In this figure, the line configuration

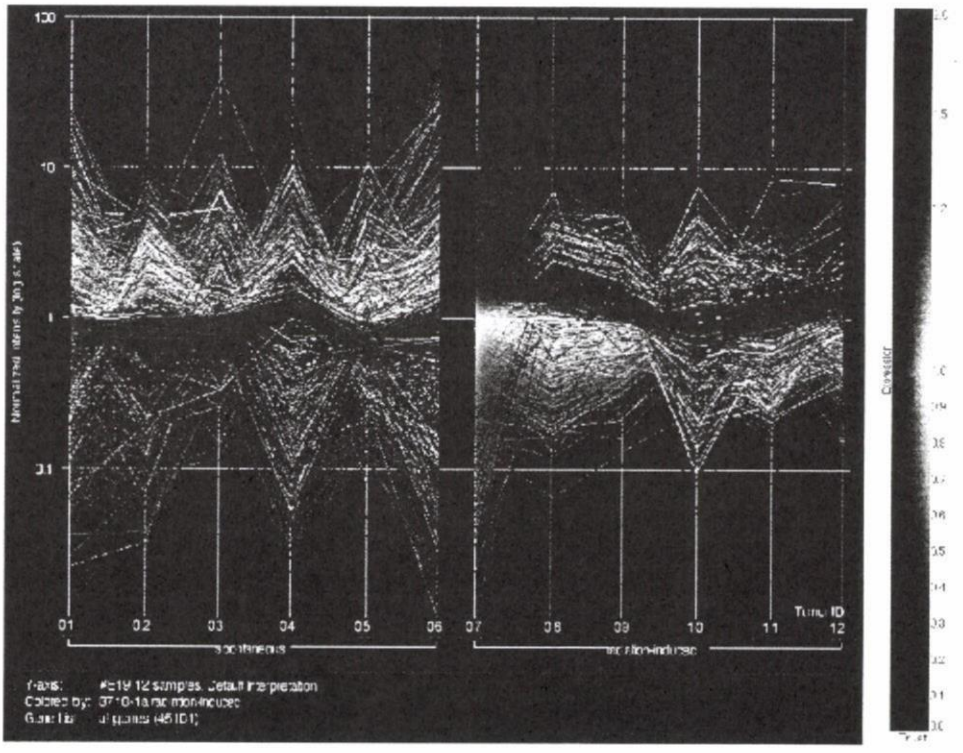

Figure 3.5 *Linear configurations of spontaneous and radiation-induced myelogenous leukemias. Six individual data on the left are from spontaneously developed myelogenous leukemias in C3H/He mice. The other six individual cases on the right are from radiation-induced myelogenous leukemias in C3H/He mice after 3 Gy X-ray exposure (Seki et al. 1991; Yoshida et al. 1997). Along the gene expression intensity from the highest (red) to the lowest (green) of group #07, the same gene in the other groups was connected and designated with the same color. Accordingly, overexpressed genes in the radiation-induced myelogenous leukemia groups are largely repressed in the spontaneous myelogenous leukemias. See text.*

associated with the spontaneous leukemias on the left shows very prominent and wide divergence in expression intensities from one individual mouse to another, and individual differences in gene ordering are also significant among each other. In contrast, each case of radiation-induced myelogenous leukemia on the right shows relatively homogeneous expression intensities compared with the former cases of spontaneous leukemias. When one compares the linear configurations of spontaneous and radiation-induced myelogenous leukemias, genes associated with radiation-induced myelogenous leukemias are not expressed in the spontaneous leukemias similarly but rather diversely. These findings are compatible with the observation by real time RT-PCR (Applied Biosystems 7900 Sequence Detection System, ABI, Foster, CA) shown in Figure 3.6, in which the former shows diverged *Sh2d1a* expressions from spontaneous myelogenous leukemias (*Sh2d1a* top) and the latter, relatively homogeneous ones from radiation-induced myelogenous leukemias (*Sh2d1a* bottom). Similarly, another gene, *Gngt1*, depicted from radiation-induced myelogenous leukemia shows relatively homogeneous expressions in most of radiation-induced myelogenous leukemias (*Gngt1* bottom), whereas the expressions of *Gngt1* were not detected in spontaneous myelogenous leukemias (*Gngt1* top). These expression profiles are further analyzed by PCA and the gene cluster associated with radiation-induced myelogenous leukemia can be discriminated from that associated with spontaneous leukemia (Figure 3.7). Representative genes that can be used to differentiate between the two types of myelogenous leukemia can be determined by PCA. The list of genes that differentiate both leukemias include *Met* (met proto-oncogene), *Fosl2* (fos-like antigen2), *Fancd2* (Fanconi anemia, complementation group D2), and *Fmr2* (fragile X mental retardation 2 homolog), among others. It is important that the expression intensities of these discriminant genes described above are not always high in all radiation-induced myelogenous leukemias, but sometimes low in a stochastic manner. Therefore, it is assumed that these genes cannot be the 'essential leukemogenic genes,' but 'stochastically necessary genes' for radiation-induced myelogenous leukemias. As described later, these genes, however, still seem to be less discriminant for radiation-induced myelogenous leukemias, although the function of each gene is linked to radiation injury. Why are radiation-induced leukemia-specific gene expression profiles not determined by the above computational analysis?

Regardless of the difference between spontaneous and radiation-induced myelogenous leukemias, it took nearly a lifetime for both myelogenous leukemias to develop fatally. Thus, the profile of each type of myelogenous leukemia may be associated with an age-related gene expression profile. Such an age-related gene expression profile may overlap to some extent with the gene expression profiles associated with spontaneous and radiation-induced myelogenous leukemias. The relationship of this factor with the above-mentioned differentiation will be considered later.

3.7.2 Toxicological Endpoints

Toxicological endpoints are different from pathological endpoints in terms of the nature of probabilistic clusters, as discussed above. One such frequent epigenetic modification, e.g. DNA methylation, constitutes a post-replicative modification, in which a methyl group is added covalently to a DNA residue. As it is known that cancer diagnosis before the cancer reaches the 'point of no return' is considered 'logically' difficult, toxicologists consider the possibility of identifying gene repertoires that are possibly associated with

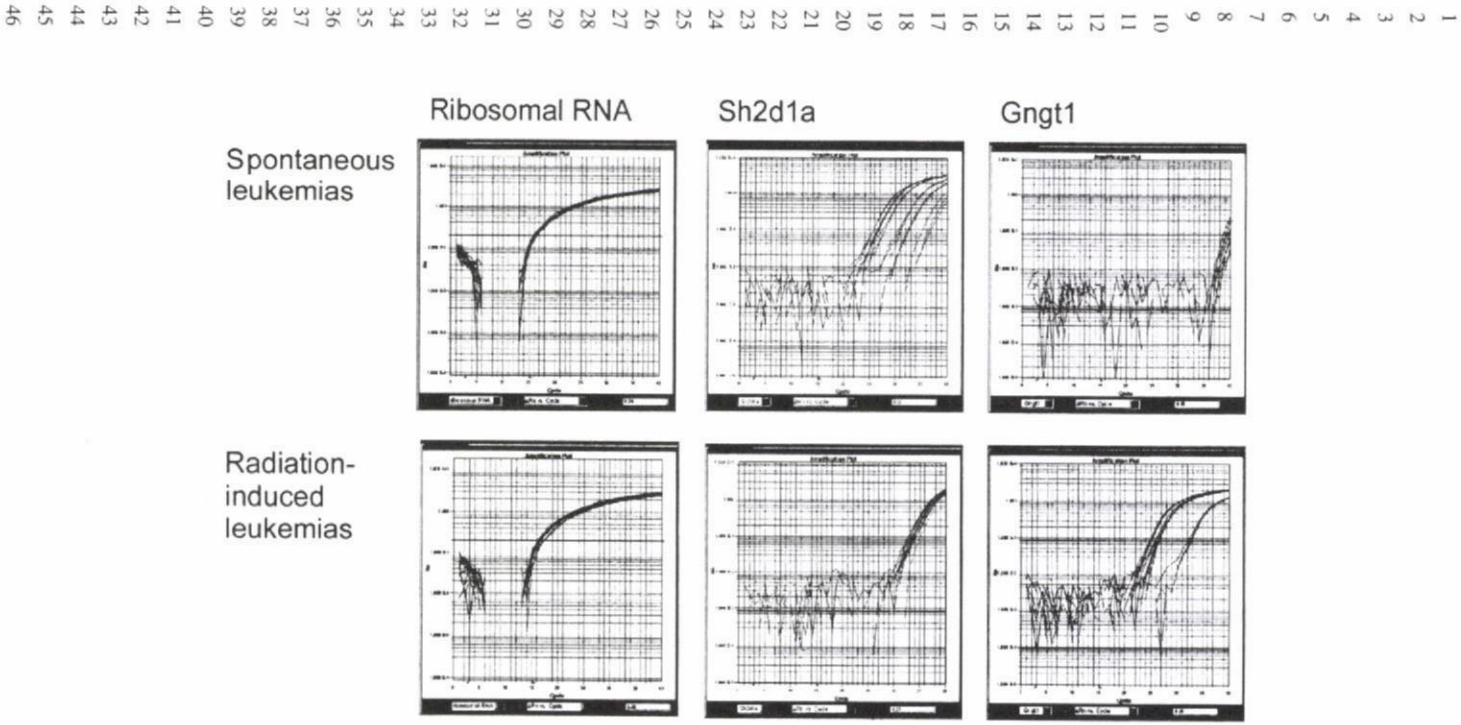

Figure 3.6 *Specific messenger RNA products of genes detected by fluorescent probes (TaqMan™ probes, ABI). Ordinate axes represent relative fluorescence (VIC™, 6-FAM™) and horizontal axes the number of cycles amplified. Ribosomal RNAs as control for RT-PCR on the left; a sample gene from spontaneous leukemia, Sh2d1a, in the middle; and a sample gene from radiation-induced leukemia Gngt, on the right are evaluated in both spontaneous and radiation-induced myelogenous leukemias (six cases each, upper and lower row respectively). Sh2d1a expressions in spontaneous leukemias (Sh2d1a top) are diverged in expression intensities, whereas those from radiation-induced leukemias are relatively homogeneous (Sh2d1a bottom). Similarly, another gene, Gngt1, shows relatively homogeneous expressions in most of radiation-induced leukemias (Gngt1 bottom), whereas the expressions were not detected in spontaneous leukemias (Gngt1 top). Quantitative real-time PCR triplicates are shown in each figure.*

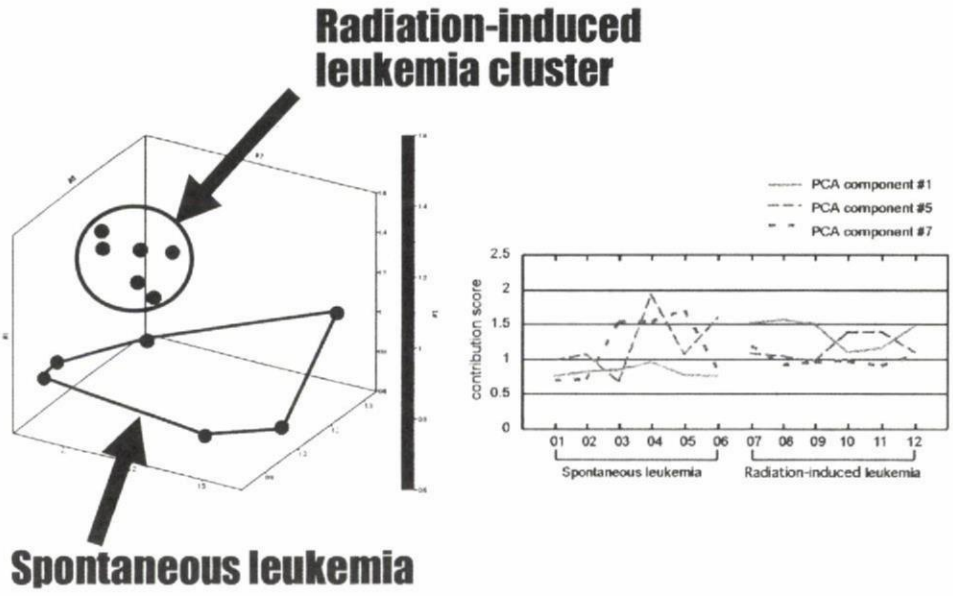

Figure 3.7 *PCA of six each of the spontaneous and radiation-induced myelogenous leukemias is shown in the three-dimensional contribution scores for components #1, #5 and #7, which discriminate the radiation-induced myelogenous leukemia cluster from the spontaneous myelogenous leukemia cluster. The line graph on the right shows actual contribution scores converted from each eigenvector value, which were used for the three-dimensional expression on the left. Note that the contribution scores of the spontaneous leukemias, except for component #1, are relatively divergent in comparison with those of radiation-induced leukemias.*

carcinogenesis at the early stage. There seems to be no definitive theoretical understanding of this issue. The most discouraging reason concerning this issue is that previous results obtained by the study of the International Life Science Institute (ILSI) consortium showed that the gene expression profiles associated with genophilic and nondirect genophilic pharmaceutical compounds were clearly differentiated, but both expression profiles from mice 4 h and 24 h after treatment showed up-regulation and down-regulation respectively (data not shown; Hu *et al.*, 2004). The possible diagnostic profiles, therefore, are considered to change with observation time or depending on the strain or treatment dose. Where can one find an appropriate discriminant axis? This is the question that needs to be answered.

The experimental results obtained by the ILSI consortium, however, can be interpreted differently. Those profiles may change with observation time, dose and strain; however, those profiles associated with genophilic compounds are always found to change in a direction opposite to that of the change of profiles associated with nondirect genophilic compounds; it is plausible that discriminant biomarkers for both genophilic and nondirect genophilic compounds would show this trend in both groups. In this regard, some sample trials carried out to determine these discriminant biomarker genes are discussed next.

The effect of age on experimental animals during toxicological experiments is one of the interesting factors associated with toxicological endpoints. Two groups of linear configurations represented by five mice each for the 2-month-old and 21-month-old groups are shown in Figure 3.8a. These two different age groups show different expression diversities; the aged group (21 months old) shows a much wider range of expression intensities than the younger (2 months old) group. Furthermore, the order of gene expression intensity between the two groups is reversed. The wider expression intensities of the aged group suggest that senescent changes appear stochastically and are probabilistically different from one individual to another. Note that the wider expression intensities do not indicate differences in the quality of individual animals but the different probabilistic responses of individual animals during aging. These individual differences are also observed at the cellular level, if each cellular function is affected by gene expression (Bahar *et al.*, 2006).

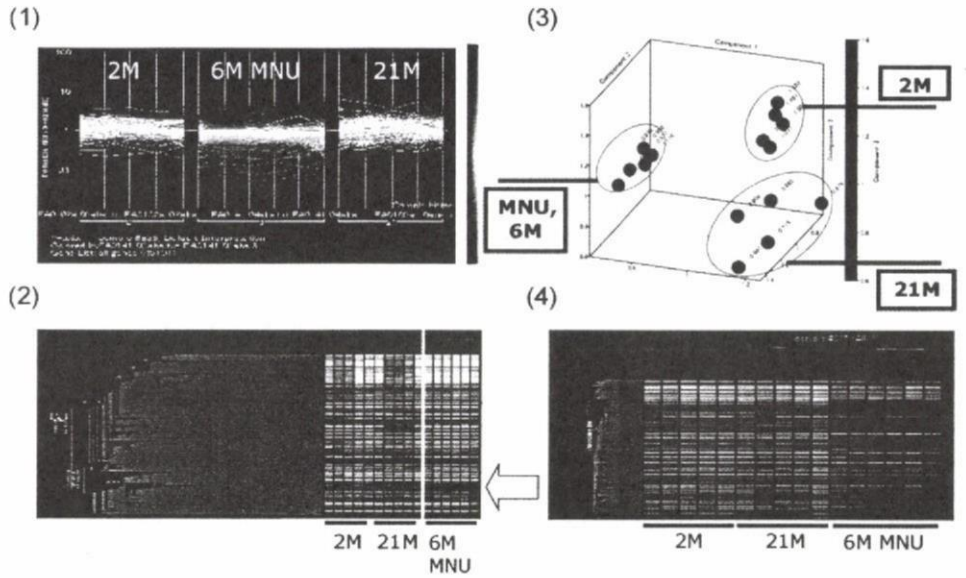

Figure 3.8 *(a) Age-related gene expression profiles determined from the bone marrow of 2-month-old and 21-month-old mice are shown in the line configuration. See text. (Six gene expression profiles of bone marrow obtained from mice 6 months after treatment with a single dose of methyl-nitroso-urea (MNU) at 50 mg kg^{-1} b.w.) (b) Two-dimensional dendrographic diagram of the same expressed genes from the bone marrow of 2-month-old and 21-month-old mice. (Six gene expression profiles of bone marrow obtained from mice 6 months after treatment with a single dose of MNU. Arrow indicates the gene cluster specifically up-regulated in the MNU-treated groups; an expanded view of the profiles is shown in (d).) (c) PCA of age control group, five mice each for the 2-month-old and 21-month-old groups, is shown in the three-dimensional contribution scores for components #1, #2 and #3, which discriminate between the clusters from the 2-month-old and 21-month-old groups. Six gene expression profiles of bone marrow obtained from mice 6 months after MNU treatment belong to another separate cluster. (d) Expanded view of profiles indicated by the arrow in (b) showing the gene cluster specifically up-regulated in the MNU-treated groups.*

Such differences between the two age groups cannot be defined using a dendrogram, as observed in the 10 columns on the left of Figure 3.8b, in which the aged group and the young group are not clearly separated. The reason that these different age groups are not clearly separated in the dendrogram is because of the relatively small number of genes with weak expressions that may define these different age groups. To pinpoint the possible responsible genes that define both groups, PCA was applied. Results show that there are discriminant components defined by components #1, #2 and #3, as shown in Figure 3.8c.

When one compares the gene expression profiles of these two age groups with those of mice treated with methyl-nitrosourea (MNU), a direct genophilic leukemogenic compound, the expression profiles of six bone marrow tissues from MNU-treated mice are clearly defined in the six columns on the right in Figure 3.8d. The discrimination of the clusters shown in the three-dimensional expressions is clearly separated in Figure 3.8c. The genes responsible for these discriminations can be autogenerated (data not shown).

3.8 Radiation-specific Probabilistic Union Genes after Subtracting Age-specific Gene Expressions

As mentioned previously, regardless of the difference between spontaneous and radiation-induced myelogenous leukemias, it took nearly a lifetime for both leukemias to develop fatally. Furthermore, radiation-induced myelogenous leukemias express stochastically divergent profiles, as observed in the line configuration in Figure 3.5. Therefore, each gene expression profile from a total of six radiation-induced myelogenous leukemia cases was analyzed by PCA to elucidate each unique gene list for nontreated mice and compared with the gene expression profiles of bone marrow cells from five 21-month-old nontreated mice. To obtain the union gene list, the gene expression profile of the 21-month-old group was compared separately with each individual expression profile of radiation-induced myelogenous leukemias by PCA followed by the selection of genes with a contribution score of over 1.0 from components #2 or #3 and #6 from the result of PCA (Table 3.1). In this PCA, the 287 union genes obtained were subtracted by 45 genes that overlapped with another 249 union genes analyzed by the PCA combination between the profile of the 21-month-old bone marrow group and that of spontaneous myelogenous leukemias (data not shown). Consequently, the final number of union genes obtained was 128 genes after the subtraction of 114 expression sequence tags (ESTs), which were generated in an unsupervised manner; however, most of them showed gene functions consistent with the response to radiation exposure. Specifically, five radiation-damage-related genes, including *Hus1*, *Eef1a2*, *Vegfc*, 13 cell cycle/cell-growth-related genes, and 12 apoptosis/cell-death-related genes were observed. Notably, 42 tumorigenesis-related genes (i.e. 33 %) were observed in which the down-regulation of tumor suppressor genes and the up-regulation of tumor promoter genes were commonly observed. The expressions of genes for cytoskeleton and cell adhesion molecules and those of genes for oxidative stress and inflammatory cytokines were also observed and included in the list; the expressions of these genes are also considered consequences of the xenobiotic responses to radiation exposure, because these genes were not included in the list of genes associated with spontaneous myelogenous leukemia (data not shown). As a reference to the functions of the autogenerated genes,

Table 3.1 *Union gene list for radiation-induced leukemias*

Affymetrix systemic name		Common name	Genbank ID	Description
1415874_at		*Spry1*	NM_011896	sprouty homolog 1 (*Drosophila*)
1416001_a_at		*Cotl1*	NM_028071	coactosin-like 1 (*Dictyostelium*)
1417097_at		*Nrbf1*	NM_025297	nuclear receptor binding factor 1
1417194_at		*Sod2*	NM_013671	superoxide dismutase 2, mitochondrial
1417602_at		*Per2*	AF035830	period homolog 2 (*Drosophila*)
1417623_at	*	*Slc12a2*	BG069505	solute carrier family 12, member 2
1417851_at		*Cxcl13*	AF030636	chemokine (C–X–C motif) ligand 13
1418062_at		*Eef1a2*	NM_007906	eukaryotic translation elongation factor 1 alpha 2
1418094_s_at		*Car4*	NM_007607	carbonic anhydrase 4
1418450_at		*Islr*	NM_012043	immunoglobulin superfamily containing leucine-rich repeat
1418547_at		*Tfpi2*	NM_009364	tissue factor pathway inhibitor 2
1418597_at		*Top3a*	NM_009410	topoisomerase (DNA) III alpha
1418666_at		*Ptx3*	NM_008987	pentaxin related gene
1418697_at	*	*Temt*	NM_009349	thioether *S*-methyltransferase
1418712_at		*Cdc42ep5*	NM_021454	CDC42 effector protein (Rho GTPase binding) 5
1418713_at		*Pcbd*	NM_025273	6-pyruvoyl-tetrahydropterin synthase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1)
1418764_a_at		*Bpnt1*	BB412311	bisphosphate 3′-nucleotidase 1, metal-dependent lithium-inhibited phosphomonoesterase protein family
1419196_at		*Hamp1/Hepc*	NM_032541	hepcidin antimicrobial peptide
1419353_at		*Dpm1*	NM_010072	dolichol-phosphate (beta-D) mannosyltransferase 1
1419365_at		*Pex11a*	NM_011068	adaptor-related protein complex 3, sigma 2 subunit
1419417_at		*Vegfc*	NM_009506	vascular endothelial growth factor C
1419561_at	*	*Ccl3*	NM_011337	chemokine (C–C motif) ligand 3

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1419664_at		*Srr*	BC011164	serine racemase
1419714_at		*Pdcd1lg1*	NM_021893	programmed cell death 1 ligand 1
1419967_at		*Seh1l*	AW540070	SEH1-like
1419970_at		*Slc35a5*	C86506	solute carrier family 35, member A5
1420034_at		*Ppp2r2d*	AU019644	protein phosphatase 2, regulatory subunit B, delta isoform (AU019644 Mouse eight-cell stage embryo cDNA *Mus musculus* cDNA clone J0520E06 3-, mRNA sequence)
1420052_x_at	*	*Psmb1*	C81484	proteasome (prosome, macropain) subunit, beta type 1
1420090_at	*	*Raf1*	AA990557	v-raf-1 leukemia viral oncogene 1
1420688_a_at		*Sgce*	NM_011360	sarcoglycan, epsilon
1420843_at	*	*Ptprf*	BF235516	protein tyrosine phosphatase, receptor type, F
1420872_at	*	*Gucy1b3*	BF472806	guanylate cyclase 1, soluble, beta 3
1421251_at		*Zfp40*	NM_009555	zinc finger protein 40
1421462_a_at		*Lepre1*	NM_019783	leprecan 1
1421619_at	*	*Kcnh3*	NM_010601	potassium voltage-gated channel, subfamily H (eag-related), member 3
1422025_at	*	*Mitf*	NM_008601	microphthalmia-associated transcription factor
1422218_at		*P2rx7*	NM_011027	purinergic receptor P2X, ligand-gated ion channel, 7
1423070_at		*Rpl21*	BG922742	general transcription factor III A
1423259_at		*Idb4*	BB121406	inhibitor of DNA binding 4
1423499_at		*Sncaip*	AK017012	synuclein, alpha interacting protein (synphilin)
1423677_at		*Fkbp9*	AF279263	FK506 binding protein 9
1424041_s_at		*C1s*	BC022123	complement component 1, s subcomponent
1424228_at		*Polr3h*	AK019868	polymerase (RNA) III (DNA directed) polypeptide H
1424295_at		*Dppa3*	AY082485	*Mus musculus* stella mRNA, complete cds
1424322_at		*Apex2*	AB072498	apurinic/apyrimidinic endonuclease 2

(Continued)

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1424586_at		*Ehbp1*	AF424697	EH domain binding protein 1
1424651_at		*BC021611*	BC021611	hypothetical protein LOC257633
1424893_at		*Ndel1*	BC021434	nuclear distribution gene E-like homolog 1 (*A. nidulans*)
1425198_at		*Ptpn2*	BG076152	protein tyrosine phosphatase, non-receptor type 2
1425278_at		*Ube4a*	BC021406	ubiquitination factor E4A, UFD2 homolog (*S. cerevisiae*)
1425366_a_at		*Hus1*	AF076845	Hus1 homolog (*S. pombe*)
1425555_at	*	*Crk7/Crkrs*	BG070845	Cdc2-related kinase, arginine/serine-rich (RIKEN cDNA 1810022J16 gene)
1425597_a_at		*Qk*	AW060288	quaking
1425608_at		*Dusp3/VHR*	BC016269	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)
1425750_a_at		*Jak3*	L40172	Janus kinase 3
1425865_a_at		*Lig3*	U66057	ligase III, DNA, ATP-dependent
1425918_at		*Egln3*	BC022961	EGL nine homolog 3 (*C. elegans*)
1427558_s_at	*	*Alg12*	AJ429133	asparagine-linked glycosylation 12 homolog (yeast, alpha-1,6-mannosyltransferase)
1427595_at		*Acac*	BE650741	acetyl-Coenzyme A carboxylase alpha
1427833_at		*Spi16/mBM17*	U96702	serine protease inhibitor 16
1427843_at		*Cebpb*	AB012278	CCAAT/enhancer binding protein (C/EBP), beta
1428386_at		*Acsl3*	AK012088	acyl-CoA synthetase long-chain family member 3
1430148_at		*Rab19*	BM241400	RAB19, member RAS oncogene family
1430391_a_at		*Siat8d*	AK003690	sialyltransferase 8 (alpha-2,8-sialyltransferase) D
1430483_a_at		*Tmem79*	AK010144	transmembrane protein 79 (RIKEN cDNA 2310042N02 gene)
1430651_s_at		*Zfp191*	AI504586	zinc finger protein 191
1431066_at		*Fut11*	BB626220	fucosyltransferase 11

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1432072_at	*	*Kif2a*	AK016720	kinesin family member 2A
1432115_a_at		*Pign*	AK014165	phosphatidylinositol glycan, class N
1433509_s_at		*Reep1*	BQ174328	receptor accessory protein 1(D6Ertd253e)
1433992_at	*	*Apxl*	BQ176992	apical protein, *Xenopus laevis*-like
1434349_at		*Vars2l*	AV258022	valyl-tRNA synthetase 2-like
1434369_a_at		*Cryab*	AV016515	crystallin, alpha B
1435132_at		*Disp1*	AI505698	dispatched homolog 1 (*Drosophila*)
1435557_at	*	*Fhod1*	AV298805	formin homology 2 domain containing 1
1435962_at		*Rps6*	BG089974	ribosomal protein S6 (Transcribed sequence with strong similarity to protein sp:P10660 (*H. sapiens*) RS6_HUMAN 40S ribosomal protein S6)
1436429_at		*Zfp606*	BB198855	zinc finger protein 606 (BB198855 RIKEN full-length enriched, 0 day neonate thymus *Mus musculus* cDNA clone A430007N09 3-, mRNA sequence)
1436521_at		*Slc36a2*	AI596194	solute carrier family 36 (proton/amino acid symporter), member 2
1436623_at		*Entpd7*	AV381133	ectonucleoside triphosphate diphosphohydrolase 7 (RIKEN cDNA 2900026G05 gene)
1436682_at		*Tmsb10*	AW259435	thymosin, beta 10 (up29e07.x1 NCI_CGAP_Mam2 *Mus musculus* cDNA clone IMAGE:2655780 3′ similar to gb:S54005 THYMOSIN BETA-10 (HUMAN), mRNA sequence)
1436895_at		*Centd1*	BB182934	centaurin, delta 1
1436904_at		*Thrap1*	BB667559	hypothetical protein D030023K18

(Continued)

Table 3.1 *(Continued)*

Affymetrix systemic name	Common name	Genbank ID	Description
1436993_x_at	*Pfn2*	BB560492	profilin 2 (BB560492 RIKEN full-length enriched, 10 days neonate olfactory brain *Mus musculus* cDNA clone E530111B09 3′ similar to AL096719 *Homo sapiens* mRNA; cDNA DKFZp566N043 (from clone DKFZp566N043), mRNA sequence.)
1437059_at	*Sox21*	BB046776	SRY-box containing gene 21 (BB046776 RIKEN full-length enriched, 11 days embryo *Mus musculus* cDNA clone 6230417M22 3-, mRNA sequence)
1437106_at	*Jarid1a*	BM246184	jumonji, AT rich interactive domain 1A (Rbp2 like) (K0734F05–3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit-/Sca-1-) cDNA Library (Long) *Mus musculus* cDNA clone NIA:K0734F05 IMAGE:30076864 3-, mRNA sequence)
1437123_at	*Mmrn2*	BB038352	multimerin 2
1437307_at	*Senp8*	BG069815	SUMO/sentrin specific protease family member 8
1437473_at	*Maf*	AV284857	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog (RIKEN cDNA A230108G15 gene)
1437789_at	*Birc6*	BB527646	baculoviral IAP repeat-containing 6 (BB527646 RIKEN full-length enriched, 15 days embryo head *Mus musculus* cDNA clone D930041P14 3-, mRNA sequence)
1437863_at	*Bche*	BB667762	butyrylcholinesterase (BB667762 RIKEN full-length enriched, adult male liver tumor *Mus musculus* cDNA clone C730038G20 3-, mRNA sequence)

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1438463_x_at		*Zdhhc6*	AV142865	AV142865 *Mus musculus* C57BL/6J 10–11 day embryo *Mus musculus* cDNA clone 2810427C08, mRNA sequence.
1438825_at		*Calm3*	AV047570	calmodulin 3 (Similar to calmodulin – rabbit (tentative sequence) (LOC384465), mRNA (AV047570 *Mus musculus* adult C57BL/6J testis *Mus musculus* cDNA clone 1700069D17, mRNA sequence))
1438857_x_at		*Irak1*/pelle-like	BB058253	Irak1(interleukin-1 receptor-associated kinase 1)/pelle-like (BB058253 RIKEN full-length enriched, 2 days neonate sympathetic ganglion *Mus musculus* cDNA clone 7120478B17 3- similar to U56773 *Mus musculus*
1439247_at		*Dock10*	BB763030	dedicator of cytokinesis 10 (BB763030 RIKEN full-length enriched, B16 F10Y cells *Mus musculus* cDNA clone G370018M23 3-, mRNA sequence)
1440180_x_at		*Zbtb3*	AV258279	zinc finger and BTB domain containing 3 (AV258279 RIKEN full-length enriched, adult male testis (BNN132) *Mus musculus* cDNA clone 4923101A10 3′, mRNA sequence)
1440871_at		*Baiap1*	AI835038	BAI1-associated protein 1
1441272_at	*	*Matr3*	BI249188	matrin 3 (602994742F1 NCI_CGAP_Mam5 *Mus musculus* cDNA clone IMAGE:5150530 5-, mRNA sequence)
1442100_at		*Inpp5f*	BB619843	inositol polyphosphate-5-phosphatase F

(Continued)

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1443229_at		Atad2	AV319821	ATPase family, AAA domain containing 2 (RIKEN cDNA 2610509G12 gene (AV319821 RIKEN full-length enriched mouse cDNA library, C57BL/6J testis male 13 days embryo *Mus musculus* cDNA clone 6030413I17 3-, mRNA))
1443493_at		Dhx37	BB766805	DEAH (Asp–Glu–Ala–His) box polypeptide 37
1443952_at		Nr1d1	BI525006	nuclear receptor subfamily 1, group D, member 1 (602924093F1 NCI_CGAP_Lu33 *Mus musculus* cDNA clone IMAGE:5056607 5-, mRNA sequence)
1445195_at		C77631	C77631	expressed sequence C77631 (Mouse 3.5-dpc blastocyst cDNA *Mus musculus* cDNA clone J0035A08 3′ similar to Mouse T-cell receptor (TCR V-alpha 16.1) gene exons 1–2, mRNA, mRNA sequence)
1447753_at		Cdc37l	BB391093	cell division cycle 37 homolog (*S. cerevisiae*)-like (BB391093 RIKEN full-length enriched, 0 day neonate cerebellum *Mus musculus* cDNA clone C230073C03 3-, mRNA sequence)
1447897_x_at		Anapc11	AV019615	AV019615 *Mus musculus* 18-day embryo C57BL/6J *Mus musculus* cDNA clone 1190010L24, mRNA sequence.
1448169_at		Krt1-18	NM_010664	keratin complex 1, acidic, gene 18
1448443_at		Serpini1	NM_009250	serine (or cysteine) proteinase inhibitor, clade I, member 1
1448986_x_at	*	Dnase2a	NM_010062	deoxyribonuclease II alpha

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1449481_at		Slc25a13	BC016571	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 13
1449493_at		Insl5	NM_011831	insulin-like 5
1449700_at		Igbp1	C81413	immunoglobulin (CD79A) binding protein 1
1449789_x_at		Ly6g6c	AV088850	lymphocyte antigen 6 complex, locus G6C (*Mus musculus* tongue C57BL/6J adult *Mus musculus* cDNA clone 2310040E07, mRNA sequence)
1449851_at	*	Per1	AF022992	period homolog 1 (*Drosophila*)
1450046_at		Tmem59/ORF18	NM_019801	transmembrane protein 59 thymic dendritic cell-derived factor 1
1450135_at		Fzd3	AU043193	frizzled homolog 3 (*Drosophila*)
1450173_at		Ripk2	NM_138952	receptor (TNFRSF)-interacting serine-threonine kinase 2
1450199_a_at		Stab1	NM_138672	stabilin 1
1450208_a_at		Elmo1	NM_080288	engulfment and cell motility 1, ced-12 homolog (*C. elegans*)
1450296_at	*	Klrb1a	NM_010737	killer cell lectin-like receptor subfamily B member 1A
1450297_at		Il6	NM_031168	interleukin 6
1450424_a_at		Il18bp	AF110803	interleukin 18 binding protein
1451541_at		Bcs1l	BC019781	RIKEN cDNA 1700112N14 gene
1451583_a_at		BC025076	BC025076	hypothetical protein LOC216829 membrane magnesium transporter 2
1451592_at		P42pop	AF364868	Myb protein P42POP
1451768_a_at		Slc20a2	AF196476	solute carrier family 20, member 2
1451950_a_at		Cd80	D16220	CD80 antigen
1451996_at		Bbp	AF353993	beta-amyloid binding protein precursor
1452253_at		Crim1	AK018666	cysteine-rich motor neuron 1
1452905_at		Gtl2	AV015833	GTL2, imprinted maternally expressed untranslated mRNA

(Continued)

Table 3.1 *(Continued)*

Affymetrix systemic name		Common name	Genbank ID	Description
1453055_at		Sema6d	BB462688	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D
1453227_at		Rhobtb3	BG801497	Rho-related BTB domain containing 3
1453481_at		Zdhhc2	BB342242	zinc finger, DHHC domain containing 2
1453690_at	*	Mpp7	AV292557	membrane protein, palmitoylated 7 (MAGUK p55 subfamily member 7) (RIKEN full-length enriched, 6 days neonate head *Mus musculus* cDNA clone 5430426E14 3′, mRNA sequence)
1454414_at	*	Btbd7	AK017755	BTB (POZ) domain containing 7
1455158_at		Itga3	BI664675	integrin alpha 3
1455297_at		SPIN-2	BG070258	Similar to Spindlin-like protein 2 (SPIN-2) (LOC278240), mRNA
1455404_at		Jph2	BG870711	junctophilin 2
1455717_s_at	*	Daam2	BM206030	dishevelled associated activator of morphogenesis 2
1455985_x_at		Shmt2	AV213251	serine hydroxymethyltransferase 2 (mitochondrial) (AV213251 RIKEN full-length enriched, ES cells *Mus musculus* cDNA clone 2410126G07 3′, mRNA sequence)
1456975_at		Taok1	BM238077	TAO kinase 1(RIKEN cDNA 2810468K05 gene)
1457040_at		Lgi2	BE947711	leucine-rich repeat LGI family, member 2
1457311_at		Camk2a	AW490258	calcium/calmodulin-dependent protein kinase II alpha
1457451_at		Acvr2	BB199213	activin receptor IIA
1458047_at		Tnfsf13b	BB667811	tumor necrosis factor (ligand) superfamily, member 13b
1458381_at		Clic5	BB028501	chloride intracellular channel 5
1458641_at		Braf	BM217816	Braf transforming gene

Table 3.1 *(Continued)*

Affymetrix systemic name	Common name	Genbank ID	Description
1459597_at	Mtpn	BG074849	myotrophin
1459868_x_at	Il11ra1	AV313111	interleukin 11 receptor, alpha chain 1 (RIKEN full-length enriched, adult male thymus *Mus musculus* cDNA clone 5830408C01 3′ similar to X74953 *M. musculus* ETL-2 mRNA, mRNA sequence.)
1460170_at	Ext2	NM_010163	exostoses (multiple) 2
1460666_a_at	Ebf3	NM_010096	early B-cell factor 3

*: overlapped in both lists of union genes for radiation-induced and spontaneous leukemias.

10 genes that represent the 128 genes mentioned above are described as follows: *Eef1a2* (eukaryotic translation elongation factor 1 alpha 2), expressed in tumors of the ovary, breast and lung (Amiri *et al.*, 2007; Tomlinson *et al.*, 2007) and plays a role in the resistance to apoptosis induced by oxidative stress (Chang and Wang, 2007); *Top3a* (topoisomerase (DNA) III alpha), required for accurate DNA replication (Oh *et al.*, 2002) and related to telomere–telomere recombination (Tsai *et al.*, 2006), cancer, and aging (Laursen *et al.*, 2003); *Ppp2r2d* (protein phosphatase 2, regulatory subunit B, delta isoform), expression biomarker for blast crisis (Liu *et al.*, 2007; Neviani *et al.*, 2007); *Lepre1* (leprecan 1), basement-membrane-associated proteoglycan that functions in growth suppression and is a potential suppressor gene (Wassenhove-McCarthy and McCarthy, 1999); *Idb4* (inhibitor of DNA binding4), promotes neuronal stem cell proliferation (Yun *et al.*, 2004) and induction of leukemia cell apoptosis (Yu *et al.*, 2005); *Hus1* (hydroxyurea sensitive1), DNA damage checkpoint (Harris *et al.*, 2006), required for telomere maintenance and functions as Rad9–Hus1–Rad1 checkpoint; *Dusp3/VHR* (dual-specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)), induces expression of cyclin D1 in breast cancer (Hao *et al.*, 2007) and arrests the cell cycle in VHR (Rahmouni *et al.*, 2006); *Igbp1/alpha 4* (immunoglobulin (CD79A) binding protein 1), apoptosis inhibitor via dephosphorylation of c-Jun and p53 (Kong *et al.*, 2004) and biomarker for acute myelogenous leukemia (Cruse *et al.*, 2005; Bhargava *et al.*, 2007); *Itga3* (integrin alpha 3), inhibitor of caspase 3 activity (Manohar *et al.*, 2004), which is up-regulated in adenocarcinoma of the lung (Boelens *et al.*, 2007), esophagus (Hourihan *et al.*, 2003) and stomach (Varis *et al.*, 2002); *Il11ra1* (Interleukin 11 receptor alpha 1), functions in carcinogenesis associated with up-regulation of PI3K and p44/p42 MAPK in gastric cancer (Nakayama *et al.*, 2007) and colon cancer (Yoshizaki *et al.*, 2006) associated with STAT3 in the prostate cancer (Zurita *et al.*, 2004), and constitutively activated in myeloma/B-CLL (Tsimanis *et al.*, 2001). The functions of these genes may satisfy the characteristics of cluster-specific gene expression profiles linked to radiation-induced leukemias.

Among the 149 genes, 21 genes, including *raf1*, *Mitf*, and *Crkrs*, overlap in both lists of union genes for both radiation-induced and spontaneous leukemias (see asterisks in

Table 3.1). These overlapped genes observed in both union gene lists imply that not all the radiation-specific union genes (i.e. 'stochastically necessary genes') are always required for the development of radiation-induced myelogenous leukemias, but a combination of the 'stochastically necessary genes' in the list is required in addition to the common leukemogenic genes. An essential rule for the combination of the 'stochastically necessary genes' for radiation-induced leukemogenicity is not yet identified.

3.9 New Risk Evaluation Strategy Using Gene Expression Profiles

These toxicological endpoints obtained by gene chip and microarray technologies can be used to evaluate their quantitative and qualitative differences in gene expression for risk evaluation by PCA. Figure 3.9a shows sample expression data between the two groups, one for bone marrow tissues exposed to 0.6 Gy and the other for nonirradiated controls, which is shown by two-dimensional expressions from PCA components #1 and #2. No observed effect level (NOEL) or no observed adverse effect level (NOAEL) can be statistically calculated by the 95 % confidence areas of the two clusters, so that the differences between the two groups can be considered as the risk evaluation parameters. Furthermore, when one attempts to evaluate the possible differences between the two groups by gene chip and microarray methods, one can use any component(s) with low differential contribution. However, when one attempts to evaluate homogeneity, one should evaluate possible

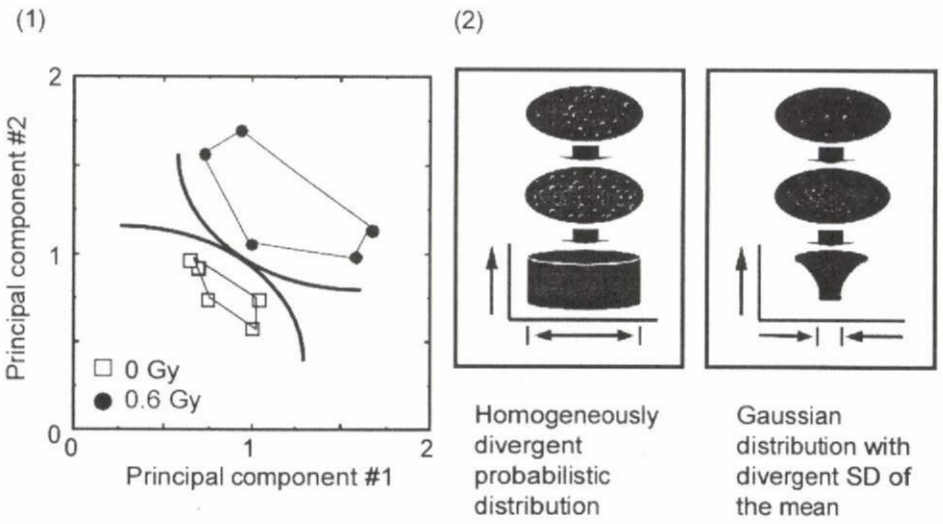

Figure 3.9 *(a) Sample gene expression clusters associated with the bone marrow after 0.6 Gy whole-body irradiation and those associated with the control are shown in the two-dimensional principal component diagram plotted along the contribution factors. Areas of each cluster are statistically defined along with the confidence levels of NOEL and/or NOAEL. See text. (b) Two types of data distribution pattern: homogeneously divergent probabilistic distribution on the left and divergent distribution due to Gaussian distribution with the error of the mean on the right.*

consistencies not only for the major component, but also for other components with low contribution, and determine whether the differences can be ignored. Because minor components may sometimes play an important toxicological role, careful examination of the genomic repertoire is required to determine whether the expressions of responsible genes are identical in a case-by-case manner.

Lastly, it is very important to recognize the characteristics of each data point. When one calculates NOEL or NOAEL, one should note the different characteristics and diversity of data between the two types, such as whether the data show probabilistic homogeneously diverse distribution or a Gaussian normal equivalent distribution due to error/deviation, as shown on the left and right sides of Figure 3.9b. Interestingly, data from developmental toxicology and the growth parameters tend to show the latter convergent distribution. However, data from toxicological changes in relation to senescence tend to show the former scattered distribution.

In this chapter, murine spontaneous and radiation-induced myelogenous leukemias were used as experimental models to discuss an essential principle of data-mining in toxicogenomics. The examinations were focused only on the bone marrow. A possible reason for the predictability observed in tissues other than the bone marrow may be attributable to the characteristics of radiation-induced tissue injury, of which the general rule may be stochastic but generally applicable to other tissues. In the case of chemicals, the predictability of the results of a single tissue examination may be limited owing to the possible tissue-specific interactions of such test chemicals.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Albanese J, Martens K, Karkanitsa LV, Dainiak N (2007) Multivariate analysis of low-dose radiation-associated changes in cytokine gene expression profiles using microarray technology. *Exp Hematol* **35**, 47–54.

Amiri A, Noei F, Jeganathan S, Kulkarni G, Pinke DE, Lee JM (2007) eEF1A2 activates Akt and stimulates Akt-dependent actin remodeling, invasion and migration. *Oncogene* **26**, 3027–3040.

Astrand M (2003) Contrast normalization of oligonucleotide arrays. *J Comput Biol* **10**, 95–102.

Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dolle ME, Calder RB, Chisholm GB, Pollock BH, Klein CA, Vijg J (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011–1014.

Bhargava P, Kallakury BV, Ross JS, Azumi N, Bagg A (2007) CD79a is heterogeneously expressed in neoplastic and normal myeloid precursors and megakaryocytes in an antibody clone-dependent manner. *Am J Clin Pathol* **128**, 306–313.

Bloomfield VA, Crothers DM, Tinoco I (2000) Nucleic Acids: Structures, Properties, and Functions. University Science Books: Sausalito, CA.

Boelens MC, van den Berg A, Vogelzang I, Wesseling J, Postma DS, Timens W, Groen HJ (2007) Differential expression and distribution of epithelial adhesion molecules in non-small cell lung cancer and normal bronchus. *J Clin Pathol* **60**, 608–614.

Borlak J (2005) *Handbook of Toxicogenomics*. Wiley–VCH Verlag: Weinheim.