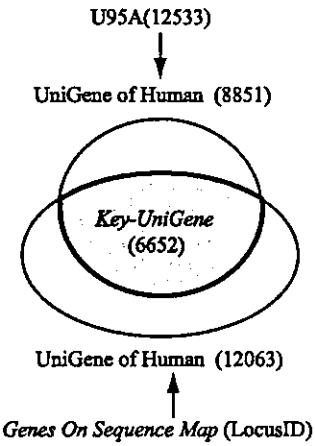


Fig. 1. Correspondence between probeIDs and LocusIDs. To associate gene locus information with gene expression profiles, probeIDs on the Affymetrix U95A oligonucleotide arrays and the LocusIDs on Genes On Sequence Map (*Homo sapiens* build 27) of NCBI were translated into UniGenes. We utilized the 12,063 LocusIDs, which had the corresponding UniGenes, on chromosome 1 to 22 of Genes On Sequence Map. The X and Y chromosomes were excluded, because the gene expression data utilized in this study were obtained from both sexes. Since these 12,063 LocusIDs had one-to-one correspondence with UniGenes, these were translated into 12,063 UniGenes. Out of 12,533 probes on the U95A array, 11,334 were translated into unduplicated 8,851 UniGenes, by referring to the corresponding original GenBank accession number of each probe set. Although the 12,063 UniGenes were obtained from Genes On Sequence Map, only 6,652 of the 12,063 UniGenes were in common with the 8,851 UniGenes translated from the probes on the U95A array. In this article, these 6,652 UniGenes are called "Key-UniGenes."

"bucket" whose length was 100,000 base pairs (100 kbp), and the Key-UniGenes were assigned the corresponding buckets according to their reading position (Fig. 2, *A* and *B*). A reading position indicates the start position for gene transcription and was obtained from Genes On Sequence Map. The number of buckets on chromosome arm *arm* was defined as L_{arm}.

Formation of Locus Cluster

To evaluate the proximity of genes on chromosome arm *arm*, the Key-UniGenes on the *length* neighbor buckets from (*begin*)-th were defined as a cluster $C_{arm\_length\_begin}$ (Fig. 2*A*). Repeating the sufficiently minute changes of *length* and *begin* formed the exhaustive uncertainty cluster sets of Key-UniGenes with chromosomal proximity (Fig. 2*C*). The EIM allows even clusters that overlap each other or include others. Therefore, all neighbor buckets in any area of each chromosome arm were defined as clusters. The number of Key-UniGenes in the cluster $C_{arm\_length\_begin}$ was defined as $n_{arm\_length\_begin}$. $C_{arm\_length\_begin}$ was defined for all

$$arm = 1p, 1q, 2p, 2q, \ldots, 22p, 22q$$

$$length = 2, 3, 4, \ldots [buckets]$$

$$begin = 1, 2, \ldots, (L_{arm} - length + 1)$$

In addition, to avoid considering a region that contains large gaps between genes as "one region," the gaps between the Key-UniGenes that lie next to each other in $C_{arm\_length\_begin}$ were calculated and the maximal gap was defined as $gap_{arm\_length\_begin}$ (Fig. 2*B*). The EIM allows the user to filter out the cluster(s) whose $gap_{arm\_length\_begin}$ is more than gap_{max}, which can be changed interactively. In other words, the user can exclude regions containing large gaps by controlling gap_{max}. When gap_{max} values were 500 kbp, 1 Mbp, 2 Mbp, and 3 Mbp, the percentages of the gaps that were less than gap_{max} were 77.6, 89.4, 96.0, and 98.2%, among all gaps between the Key-UniGenes that lie next to each other.

EIM for Detection of Expression Imbalance Specific To Squamous Cell Carcinomas

Clusters consisting of genes with expression profiles specific to SQs. Probes with expression profiles specific to SQs were extracted as a cluster from 4,083 probes of SQ-NL data sets. Although the EIM does not depend on the type of statistical method used for evaluating the difference between two groups, nonparametric tests such as the Mann-Whitney test have the advantage that no assumption is needed about the distribution of data, compared with parametric tests such as the *t*-test. Thus we explain the case of the Mann-Whitney test as an example.

More specifically, the difference in the level of expression of each gene between two groups (SQs and NLs) was defined using the statistical probability, *P*, of rank sum. Assume that there are two groups (G_a, $n = N_a$; G_b, $n = N_b$) and the rank sums in G_a and G_b are Sum_a and Sum_b, respectively, when all elements ($N_a + N_b$) are sorted in order. For simplicity, assume that Sum_a/N_a is greater than or equal to Sum_b/N_b. *P* is the probability of observing the rank sum of the N_a elements, which are randomly selected from all elements, to be more than Sum_a.

Table 1. *Number of the UniGenes and Key-UniGenes on Genes On Sequence Map*

Chr. Arm	UniGene Number	Key-UniGene Number (L_{arm})	Chr. Arm	UniGene Number	Key-UniGene Number (L_{arm})
1p	715	394	12p	211	107
1q	614	361	12q	488	289
2p	313	179	13p	0	0
2q	485	274	13q	218	127
3p	315	191	14p	0	0
3q	335	171	14q	411	228
4p	111	60	15p	0	0
4q	356	201	15q	379	197
5p	116	61	16p	254	130
5q	472	248	16q	244	123
6p	434	251	17p	218	130
6q	291	158	17q	513	290
7p	180	105	18p	52	34
7q	373	205	18q	135	76
8p	157	95	19p	391	199
8q	262	138	19q	481	249
9p	146	85	20p	122	53
9q	353	193	20q	245	124
10p	104	53	21p	0	0
10q	362	205	21q	137	83
11p	234	129	22p	0	0
11q	502	280	22q	334	176

Distributions of the UniGenes, which were obtained from Genes On Sequence Map (*Homo sapiens* build 27) of NCBI, and Key-UniGenes on each arm of the chromosome. Since the gene expression data utilized in this study were obtained from both sexes, the X and Y chromosomes were excluded. Key-UniGenes are the UniGenes that can be translated into from both the probes on the U95A oligonucleotide arrays and the LocusIDs on chromosome 1 to 22 of the Genes On Sequence Map. The total numbers of the UniGenes and Key-UniGenes are 12,063 and 6,652, respectively. Chr., chromosome; L_{arm}, number of "buckets" on chromosome arm *arm*.

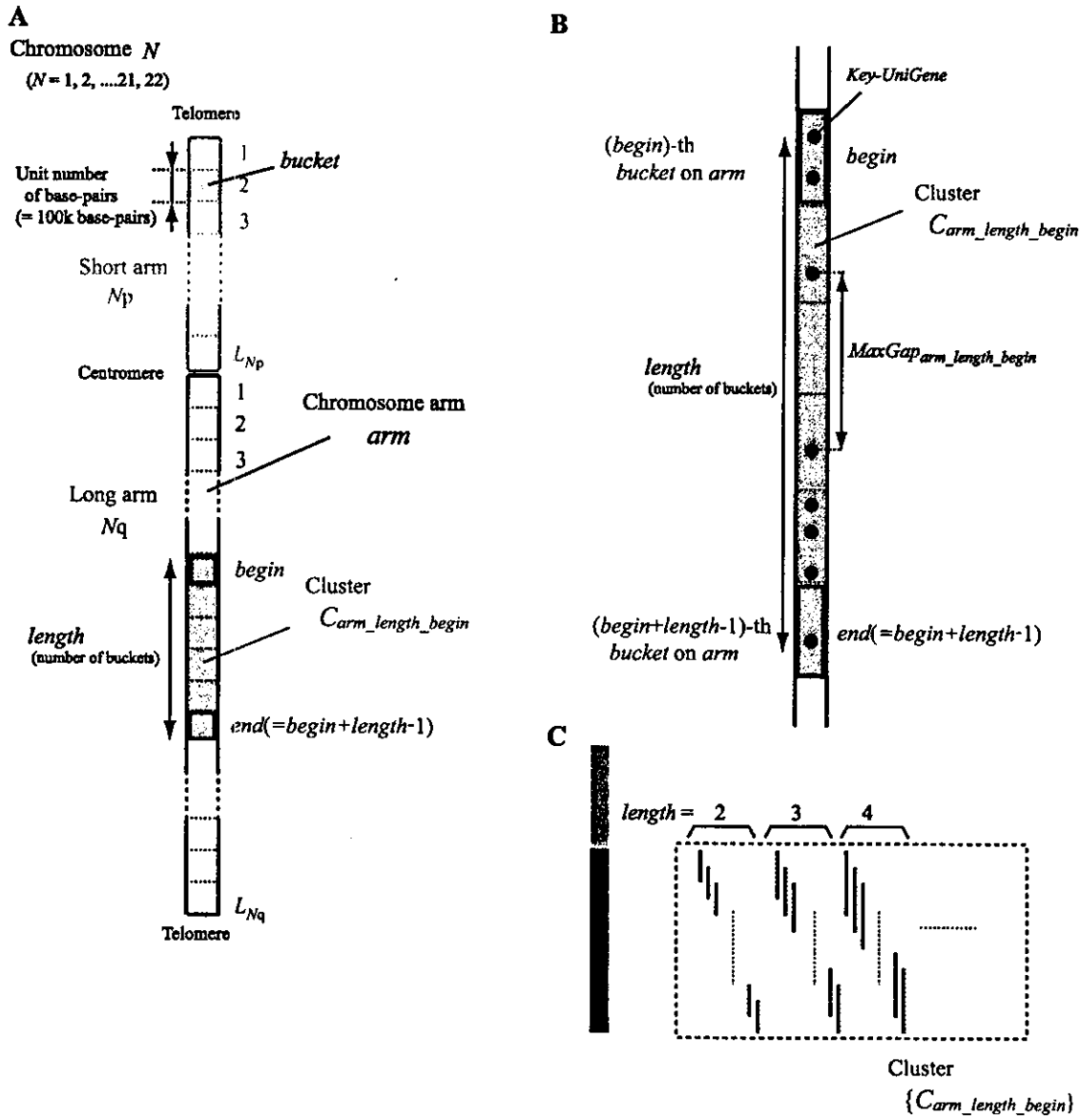


Fig. 2. Formation of clusters of genes with chromosomal proximity. *A*: for easier handling of the gene locus information, each chromosome arm region was quantized by unit region called "bucket" whose length was 100 kbp, and the Key-UniGenes were assigned the corresponding buckets according to their reading positions, which were obtained from Genes On Sequence Map (*Homo sapiens* build 27) of NCBI. The number of buckets on chromosome arm *arm* was defined as L_{arm}. To evaluate the proximity of genes on chromosome arm *arm*, the Key-UniGenes on the *length* neighbor buckets from (*begin*)-th were defined as a cluster $C_{arm\_length\_begin}$. *B*: to avoid considering a region containing large gaps between genes as "one region," the gaps between Key-UniGenes which lie next to each other in $C_{arm\_length\_begin}$ were calculated and the maximal gap was defined as $gap_{arm\_length\_begin}$. The expression imbalance map (EIM) allows the user to filter out the clusters whose $gap_{arm\_length\_begin}$ is more than gap_{max}, which can be changed interactively. In other words, the user can exclude regions containing large gaps by controlling gap_{max}. *C*: repeating the sufficiently minute changes of *length* and *begin* formed the exhaustive uncertainty cluster set of locus information. The EIM allows even the clusters that overlap each other or include others. Therefore, all neighbor buckets in any area of each chromosome arm were defined as clusters.

Based on this P value, the differential level $D_1(g)$ in which g is the probe name was defined as follows

$$D_1(g) = -\log_{10}P \quad (1)$$

Probes whose differential level D_1 was equal to or more than *diff* were defined as a cluster of probes with expression profiles specific to SQs, C_{sign_diff} (Fig. 3). The suffix *sign* indicates a differential direction (+, overexpression; −, underexpression in SQs). Repeating the sufficiently minute changes of *diff* formed the exhaustive uncertainty set of the clusters specific to SQs. C_{sign_diff} was defined for all

$$sign = -, +$$

$$diff = 2, 3, 4, \ldots$$

For example, C_{+3} was a cluster of probes whose differential level $D_1(g)$ of overexpression was 3 or more. The EIM was constructed by all the clusters C_{sign_diff} with *diff* greater than or equal to the minimum acceptable differential level d_{min} (Fig. 3). Since the default value of d_{min} is 2, all the clusters, C_{sign_diff}, would be utilized. The EIM allows the user to control d_{min} interactively for narrowing down the probes, if needed.

The numbers of probes, UniGenes, and Key-UniGenes of each cluster are shown in Table 2; n_{sign_diff} is the number of Key-UniGenes translated from probes of C_{sign_diff}. When multiple probes in a cluster could be mapped to a single UniGene, only the probe with the highest D_1 value was adopted. In addition, Fig. 3 shows probe permutations whose differential levels are 2 or more, arranged in the order of the differential level. Probes with under- and overexpression are arranged on the left and the right of Fig. 3, respectively.

Construction of the EIM. To detect the expression imbalance regions, it is necessary to search for genes with both cancer specificity and chromosomal proximity. The fundamental algorithm of the EIM is to statistically evaluate the overlaps between clusters of genes with cancer specificity and clusters of genes with chromosomal proximity. The clusters specific to the group of SQs, C_{sign_diff}, are arranged on the abscissa, and the locus clusters, $C_{arm\_length\_begin}$, are on the ordinate, as shown in Fig. 4. The variable k is the number of common Key-UniGenes between C_{sign_diff} and $C_{arm\_length\_begin}$.

The variable k could be evaluated using the hypergeometric probability, H, for observing at least k common elements between randomly selected n_1 and n_2 elements among all U elements as follows, where n_1 is n_{sign_diff} and n_2 is $n_{arm\_length\_begin}$.

Table 2. *Clusters of probes with expression profiles specific to the group of squamous cell lung carcinomas*

Differential Direction	Cluster Name (C_{sign_diff})	Probe Number	Key-UniGene Number (n_{sign_diff})
Underexpression (SQ < NL)	C_{-2}	1,007	668
	C_{-3}	844	567
	C_{-4}	642	429
	C_{-5}	448	301
	C_{-6}	283	188
	C_{-7}	83	61
Overexpression (SQ > NL)	C_{+2}	958	613
	C_{+3}	759	480
	C_{+4}	543	329
	C_{+5}	334	205
	C_{+6}	143	95
	C_{+7}	13	8

The probes (on the Affymetrix U95A arrays) whose expression profiles show significant difference between squamous cell lung carcinomas (SQs) and normal lung (NLs) were extracted as clusters, C_{sign_diff}. The suffix *sign* indicates the differential direction ("+" = overexpression; "−" = underexpression in SQs), and *diff* indicates a differential level D_1 in gene expression profiles between SQs and NLs. For example, C_{+3} is a cluster of probes whose differential level of overexpression is 3 or more. Repeating the sufficiently minute changes of *diff* formed the exhaustive set of the clusters consisting of genes with expression profiles specific to SQs. The numbers of probes and Key-UniGenes for each cluster are shown.

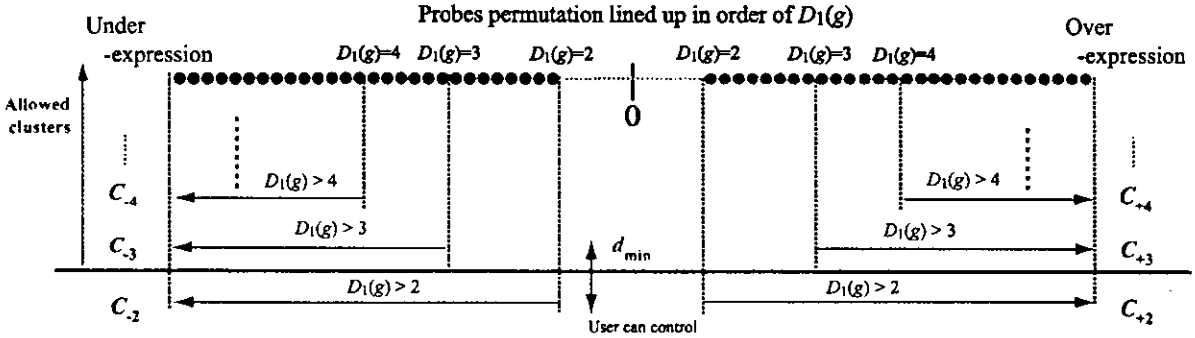


Fig. 3. Probe permutation arranged in order of the difference in gene expression level between squamous cell lung carcinomas (SQs) and normal lungs (NLs). Probes on the U95A arrays are lined up in order of the $D_1(g)$ level, which represents the difference in the gene expression level between SQs and NLs. Only probes with differential levels of 2 or more were arranged. Probes with underexpression and overexpression in SQs are arranged on the *left* and *right* side, respectively. Probes whose differential level $D_1(g)$ is equal to or more than *diff*, are defined as a cluster of probes with expression profiles specific to SQs, C_{sign_diff}. The suffix *sign* indicates the differential direction (+, overexpression; −, underexpression in SQs). Repeating the sufficiently minute changes of *diff* formed the exhaustive uncertainty set of the clusters specific to SQs. The EIM was constructed by all clusters C_{sign_diff} with *diff* that were greater than or equal to the minimum acceptable differential level d_{min}. Since the default value of d_{min} is 2, all the clusters, C_{sign_diff}, would be utilized. The EIM allows the user to control d_{min} interactively for narrowing down the probes, if needed.

$$H(U,n_1,n_2,k) = 1 - \sum_{i=0}^{k-1} \frac{\binom{n_2}{i} \cdot \binom{U-n_2}{n_1-i}}{\binom{U}{n_1}} \quad (2)$$

When the H value is small, the overlap between C_{sign_diff} and $C_{arm\_length\_begin}$ is considered statistically significant. That is, if the H value is small, then the overlap did not occur accidentally. Thus the evaluation value, E, is defined as follows

$$E(U,n_1,n_2,k) = -\log_{10}H(U,n_1,n_2,k) \quad (3)$$

For any combination of C_{sign_diff} and $C_{arm\_length\_begin}$, if both (*begin*)-th and (*begin* + *length* − 1)-th buckets of $C_{arm\_length\_begin}$ have the Key-UniGenes that are included in C_{sign_diff}, then their E values were calculated. This calculation was preprocessing for the EIM. Then, in real-time processing, if both C_{sign_diff} and $C_{arm\_length\_begin}$ met d_{min} and gap_{max}, respectively, then the E value was represented in the intersection area $R_{sign\_diff\_arm\_length\_begin}$ as a gray scale. The user can control d_{min} and gap_{max} interactively. The area where the multiple $R_{sign\_diff\_arm\_length\_begin}$ values overlapped is overwritten at the maximum E value (Fig. 4*B*). A flowchart that details these steps is shown in Fig. 5. The EIM for detecting expression imbalance specific to SQs is shown in Fig. 6. In addition, Fig. 7 shows chromosome 3 of the EIM and the influence of gap_{max} and d_{min} on the detection of the expression imbalance regions specific to SQs.

EIM for Detection of Individual Differences in Expression Imbalance Among SQs

It is effective to extract probes with expression profiles specific to the group of cancers using statistical analyses, such as the Mann-Whitney analysis. However, because this type of analysis treats all specimens with the same pathological diagnosis as one group, the variation in a group is unobservable. This is sometimes a significant problem because cancer specimens generally have a great number of variations. Thus we also developed the EIM for detecting individual differences in expression imbalance among SQs.

Clusters of probes with expression imbalance in each SQ. The first step in the development of the EIM for detecting individual differences in expression imbalance among SQ specimens was to extract probes with under- or overexpression compared with NL specimens, in each SQ specimen independently. Assuming that the expression levels of a certain probe, g, in NL specimens have a lognormal distribution, if the expression level of a SQ specimen, S_i, is included in $100p\%$ of sections on both sides of NL's distributions, its differential level D_2 was defined as follows

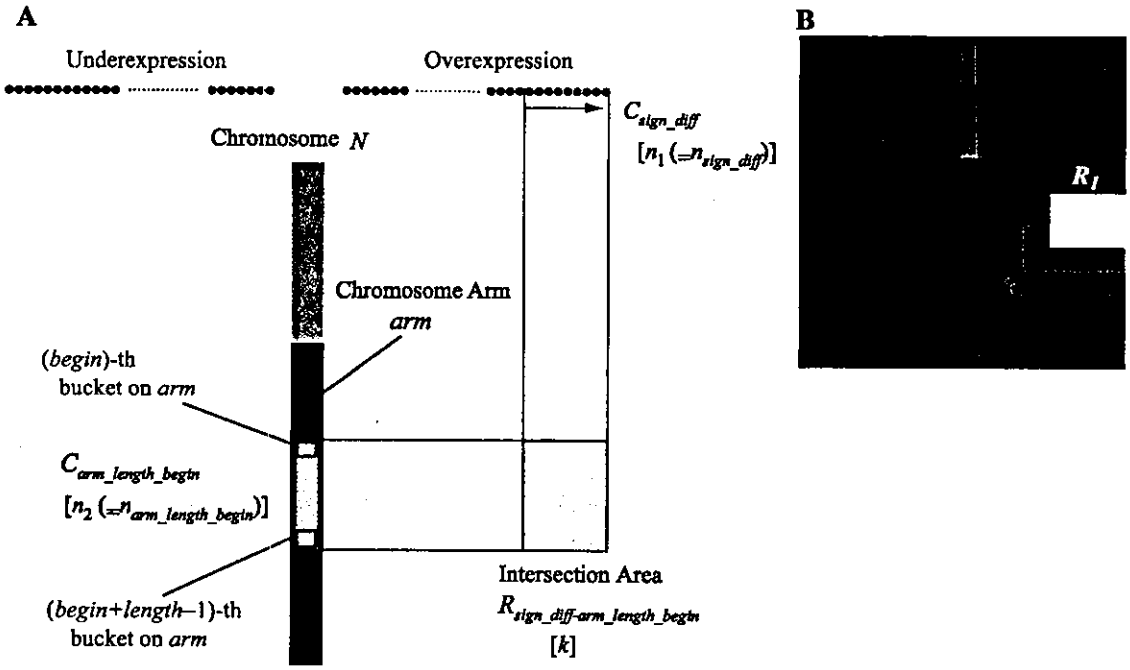


Fig. 4. Clusters of genes specific to the group of SQs vs. clusters of genes with proximity on chromosomes. *A*: to detect expression imbalance regions, it is necessary to search for genes with both cancer specificity and chromosomal proximity. The fundamental algorithm of the EIM is to evaluate statistically the overlaps between clusters of genes with cancer specificity and clusters of genes with chromosomal proximity. The clusters of probes with expression specific to the group of SQ, C_{sign_diff}, are arranged on the abscissa, and those of Key-UniGenes with proximity on chromosomes, $C_{arm\_length\_begin}$, on the ordinate. Among C_{sign_diff} values, the clusters of probes with underexpression and overexpression in SQs are arranged on the *left* and *right* side, respectively. The n_{sign_diff} and $n_{arm\_length\_begin}$ are the numbers of Key-UniGenes in C_{sign_diff} and $C_{arm\_length\_begin}$, respectively; k is the number of common Key-UniGenes both in C_{sign_diff} and $C_{arm\_length\_begin}$. The statistical significance of the overlap between C_{sign_diff} and $C_{arm\_length\_begin}$ was visualized in the intersection area $R_{sign\_diff\_arm\_length\_begin}$ as a gray scale. *B*: the area where the multiple $R_{sign\_diff\_arm\_length\_begin}$ overlapped was overwritten at the maximum E value. Therefore, when the E value of R_1 is higher than that of R_2, the area where R_1 and R_2 overlapped is overwritten at that of R_1.

$$D_2(g,S_i) = -\log_{10}p \quad (4)$$

Regarding each SQ specimen S_i (i = 1, 2,. . ., 21), the probes whose differential levels $D_2(g,S_i)$ were equal to or more than *diff* were defined as the individual-specimen cluster, $C_{sign\_diff\_Si}$, where *sign* is the differential direction (+, overexpression; −, underexpression in each SQ specimen). $C_{sign\_diff\_Si}$ was defined for all

$$sign = -, +$$
$$diff = 2, 3, 4, \ldots$$
$$S_i = 1, 2, \ldots, 21$$

For example, C_{+2_Si} and C_{-2_Si} were clusters of probes whose expression of S_i were included in 1% of sections on both sides of NL's distributions. More specifically, C_{+2_Si} was a cluster of probes whose expression levels were equal to or higher than (ave_{NL} + 2.58 $stddev_{NL}$) in a specimen S_i, where ave_{NL} is the mean and $stddev_{NL}$ is the standard deviation of expression level in NL specimens. In the same manner, C_{-2_Si} was a cluster of probes whose expression levels were equal to or less than (ave_{NL} − 2.58 $stddev_{NL}$); $n_{sign\_diff\_Si}$ is the number of Key-UniGenes in $C_{sign\_diff\_Si}$. If multiple probes in a cluster could be mapped to single UniGene, then only the probe with the highest D_2 value was adopted. The average numbers, $\bar{n}_{sign\_diff}$, of $\{n_{sign\_diff\_Si}\}(i = 1, 2,. . ., 21)$ are shown in Table 3.

Construction of the EIM. In a manner similar to the EIM for detecting expression imbalance of SQ group, that for detecting individual differences in expression imbalance among SQs was also constructed. The individual-specimen clusters, $C_{sign\_diff\_Si}$, were arranged on the abscissa with respect to each S_i, and the locus clusters on the ordinate (Fig. 8). Underexpression clusters were arranged on the left side and overexpression clusters on the right. Since the abscissa represented an array of S_i, it was impossible to represent *diff* on the abscissa like Fig. 4. Therefore, the EIM for individual specimen was visualized by $C_{sign\_diff\_Si}$ with a defined *diff*, and allowed the user to change *diff* interactively.

The number of common Key-UniGenes between $C_{sign\_diff\_Si}$ and $C_{arm\_length\_begin}$, k, could also be evaluated using $E(U, n_1, n_2, k)$ (*Eq. 3*), where n_1 was $\bar{n}_{sign\_diff}$ and n_2 was $n_{arm\_length\_begin}$. If the different specimens have the same number of genes with under- or overexpression on the same local region, then it is necessary to evaluate them as similar. Therefore, $\bar{n}_{sign\_diff}$ instead of $n_{sign\_diff\_Si}$ was used for the evaluation of the overlap between $C_{sign\_diff\_Si}$ and $C_{arm\_length\_begin}$. The E value for any combination of $C_{sign\_diff\_Si}$ and $C_{arm\_length\_begin}$ was calculated,

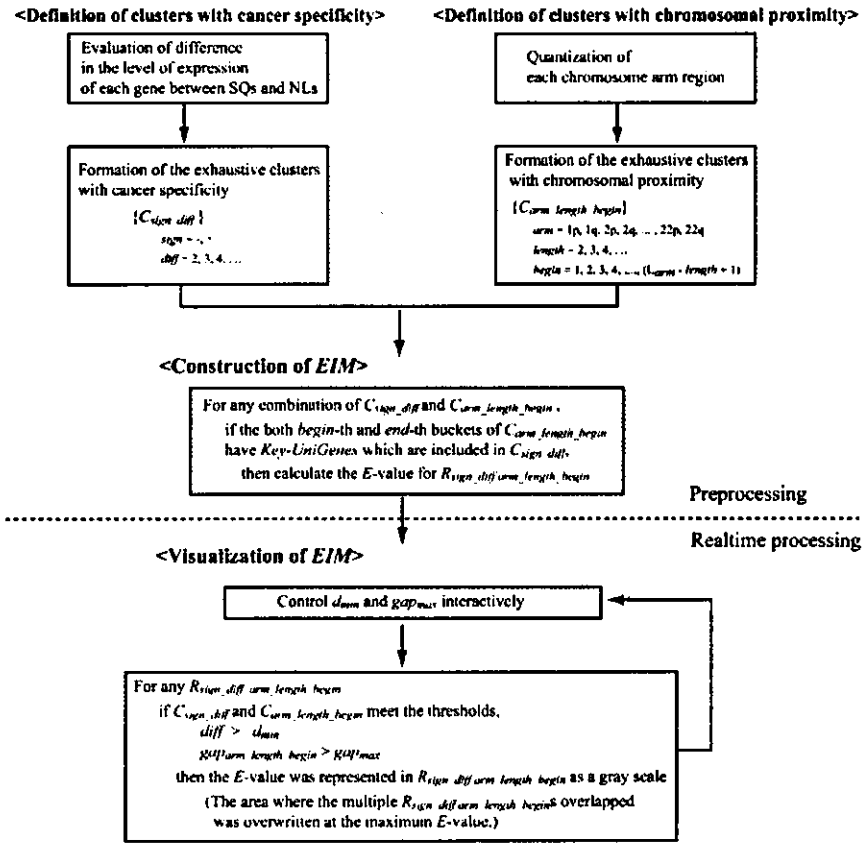


Fig. 5. Flowchart for construction of the EIM for detecting expression imbalance regions specific to SQs. This flowchart provides details of the steps of the EIM for detecting expression imbalance regions specific to SQs. For the steps of "Definition of clusters with cancer specificity," please refer to Fig. 3. For the steps of "Definition of clusters with chromosomal proximity," please refer to Fig. 2. For the steps of "Construction of the EIM" and "Visualization of EIM," please refer to Fig. 4. The user can interactively control the steps in real-time processing by changing gap_{max} and d_{min}.

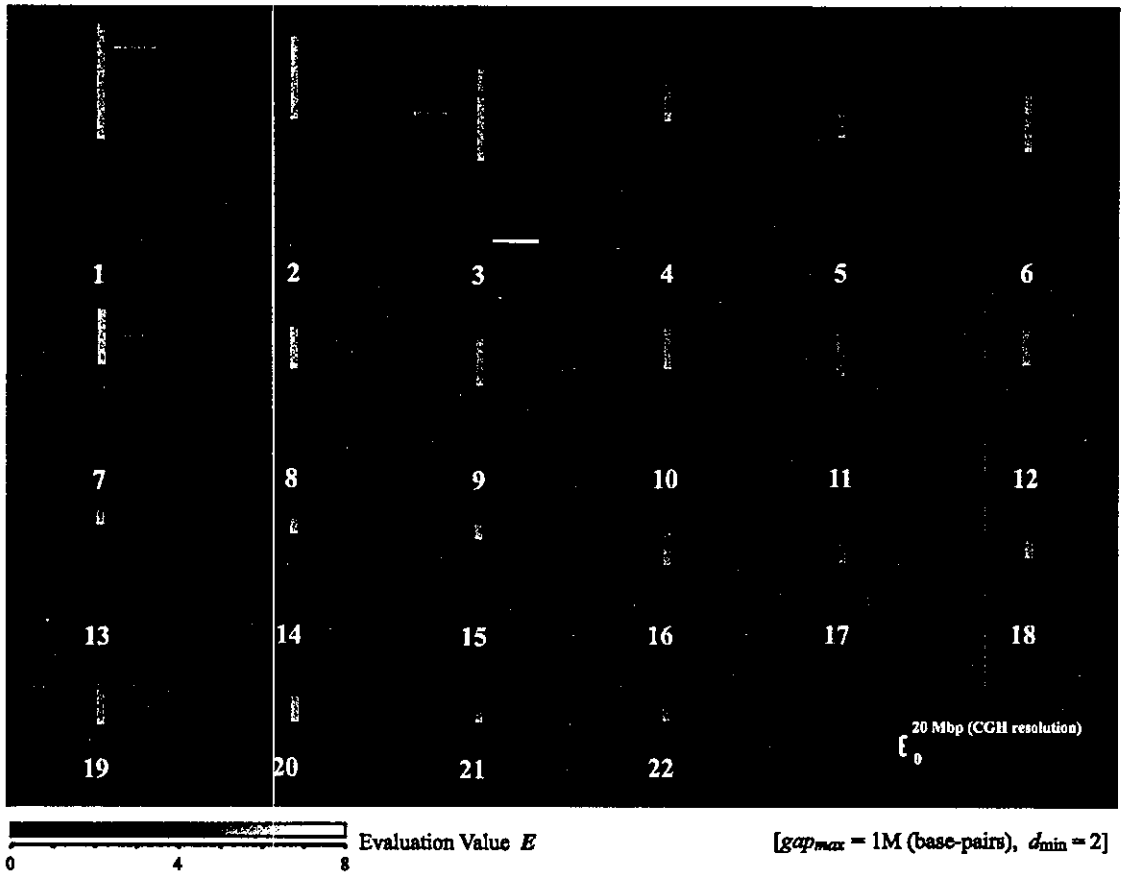


Fig. 6. The EIM applied for detecting expression imbalance regions specific to SQs. The regions of under- and overexpression in SQs were visualized on the *left* and *right* side, respectively, as gray regional signals. All statistical evaluation values of any combinations between the exhaustive uncertainty cluster sets of cancer specificity and chromosomal proximity are visualized on the EIM as the gradation of gray scale simultaneously. Each exhaustive uncertainty cluster set was formed by repetition of the sufficiently minute changes of the threshold of cancer specificity or chromosomal proximity. While the area with high luminance corresponds to the more probable expression imbalance region, the EIM enables the user to search as many genes as possible by referring to more expanded area with lower luminance. The EIM presented the most significant overexpression regions on 3q (the evaluation value E = 7.2), which is a well-known locus with frequent genomic gains, as detected by comparative genomic hybridization (CGH) (6, 8, 9). Note the high resolution of the EIM compared with CGH resolution (~20 Mbp).

Fig. 7. Expression imbalance regions specific to SQs on chromosome 3. *A–I*: chromosome 3 of the EIM and the influence of gap_{max} and d_{min} on the detection of the expression imbalance regions specific to SQs. The EIM represents the E values whose C_{sign_diff} and $C_{arm\_length\_begin}$ meet d_{min} and gap_{max}, respectively. The EIM allows the user to control gap_{max} and d_{min} interactively. The user can narrow down the possible expression imbalance regions by changing gap_{max} and d_{min}. Especially, as is shown in *A–I*, changing gap_{max}, which allows exclusion of regions containing large gaps between genes, markedly affected the detection of expression imbalance regions. *J*: the macrograph of the encircled *region A* from panel *A*. Intersection area $R_{+5\_3q\_1894\_5}$ shows the most significant overexpression region, which is a well-known locus with frequent genomic gains as previously detected by CGH (6, 8, 9). That is, the overlap (k = 6) between C_{+5} and $C_{3q\_1894\_5}$ was statistically the most significant (E = 7.2). C_{+5} was the cluster of probes with overexpression whose differential level $D_1(g)$ was more than 5 and its number of Key-UniGenes, n_{+5}, was 205. $C_{3q\_1894\_5}$ was the region from 189,400 to 189,900 kbp on chromosome 3 and contained 9 Key-UniGenes ($n_{3q\_1894\_5}$ = 9). The maximum gap ($gap_{3q\_1894\_5}$) between Key-UniGenes in $C_{3q\_1894\_5}$ was 146 kbp. In addition, all evaluation values of any combinations between the exhaustive uncertainty cluster sets of cancer specificity and chromosomal proximity are visualized simultaneously on the EIM as gradation of the gray scale. This gradation pattern could convey the distribution of the false balance to the user through visual perception and enabled the detection of as many significant genes as possible. In addition, note the high resolution of EIM compared with CGH resolution (~20 Mbp).

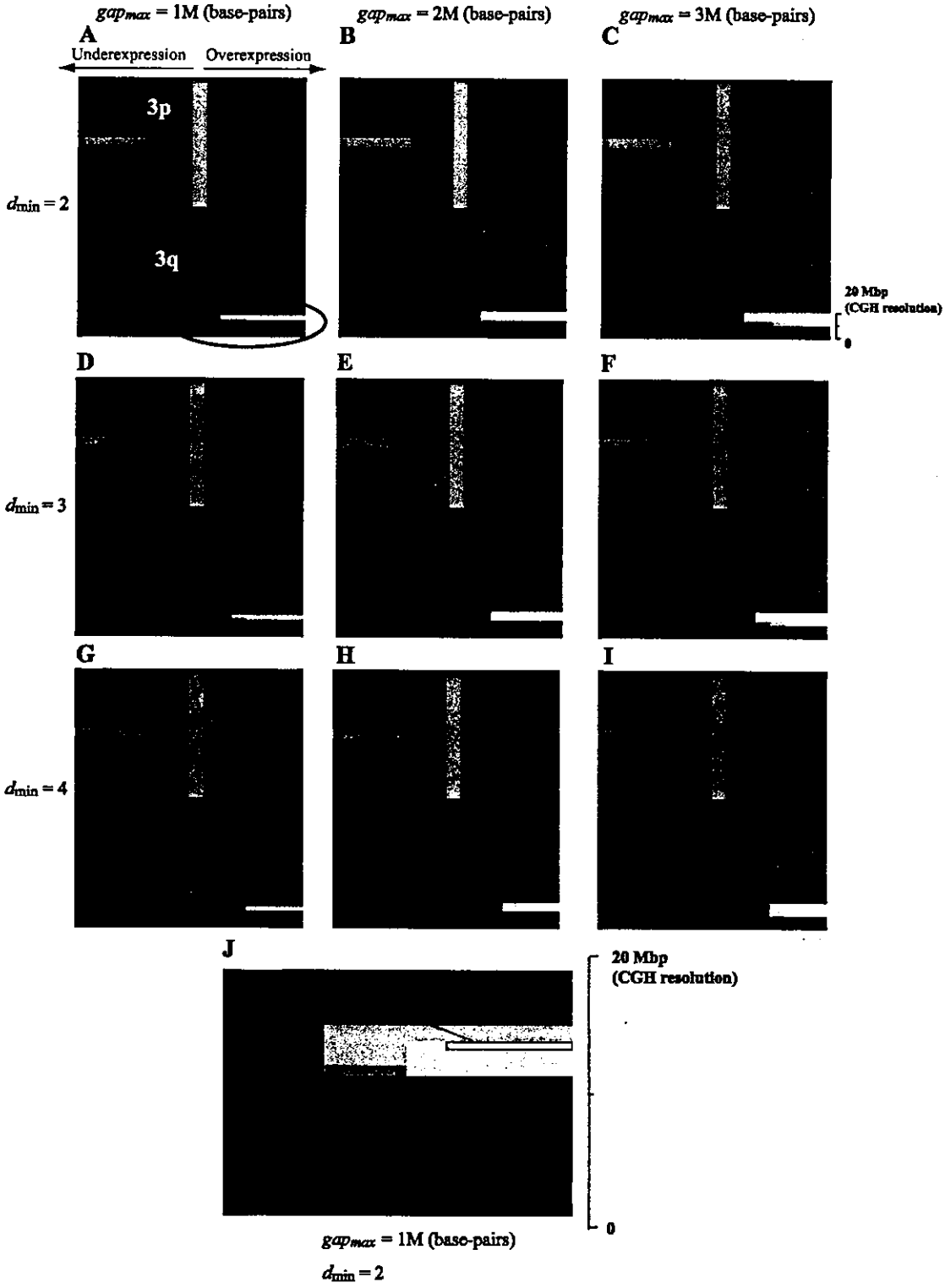


Table 3. *Clusters of probes with under- or overexpression profiles in each squamous cell lung carcinoma*

Differential Direction	Cluster Name ($C_{sign\_diff\_Si}$)	Avg. of Probe Number	Avg. of Key-UniGene Number ($\bar{n}_{sign\_diff}$)	SD of Key-UniGene Number
NL(17) > each SQ	C_{-2_Si}	669	447	103
	C_{-3_Si}	497	331	91
	C_{-4_Si}	387	259	82
	C_{-5_Si}	317	211	76
	C_{-6_Si}	268	181	70
NL(17) < each SQ	C_{+2_Si}	321	208	67
	C_{+3_Si}	188	120	48
	C_{+4_Si}	120	77	35
	C_{+5_Si}	81	50	25
	C_{+6_Si}	58	36	19

To detect individual differences in expression imbalance among 21 SQs, probes (on the U95A array) with under- or overexpression profiles in a SQ specimen, S_i (i = 1,2,. . .,21), compared with NLs were extracted as clusters, $C_{sign\_diff\_Si}$. This extraction was independently performed, regarding each SQ specimen. The suffix *sign* indicates the differential direction (+, overexpression; −, underexpression in each SQ specimen). *diff* indicates a differential level D_2 in gene expression. Shown are the average number of probes and the average and standard deviation (SD) of Key-UniGenes in the 21 clusters with the same differential direction and differential level.

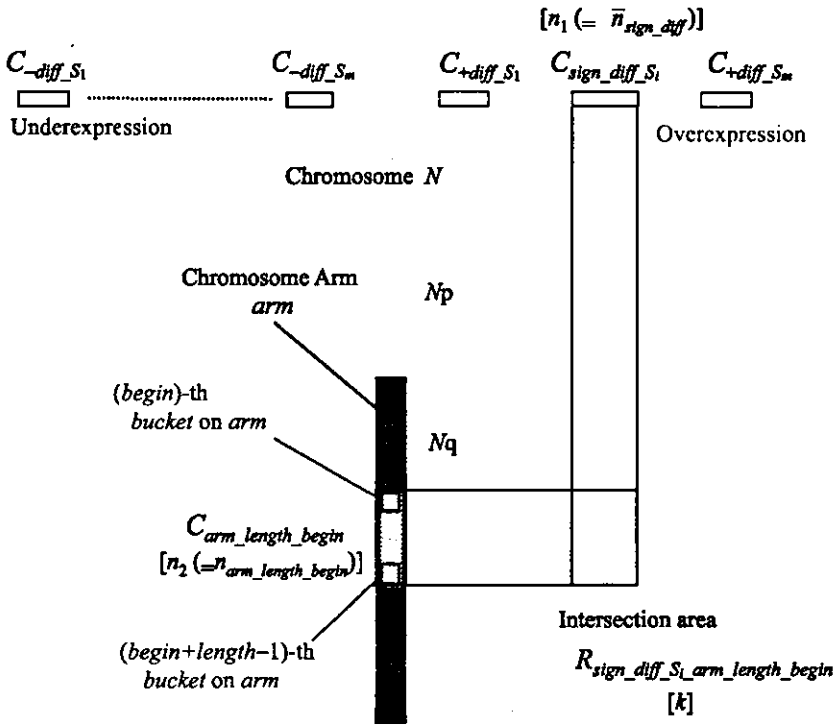


Fig. 8. Individual-specimen clusters vs. locus clusters. In a manner similar to the EIM for detecting expression imbalance of SQ specimen group, that for detecting individual differences in expression imbalance among SQ specimens was also constructed. In a SQ specimen S_i (i = 1, 2,. . ., 21), probes with expression whose differential level $D_2(g,S_i)$ was equal to or higher than *diff* compared with NL specimens were extracted as an individual-specimen cluster, $C_{sign\_diff\_Si}$. This extraction was independently performed with respect to each SQ specimen. The individual-specimen clusters, $C_{sign\_diff\_Si}$ values, were arranged on the abscissa with respect to each S_i, and the locus clusters, $C_{arm\_length\_begin}$ values, on the ordinate. Among $C_{sign\_diff\_Si}$ values, the clusters of under- and overexpression were arranged on the *left* and *right* side, respectively. Since the abscissa represented an array of S_i, it was impossible to represent *diff* on the abscissa like Fig. 4. Therefore, the EIM for individual specimen was visualized by $C_{sign\_diff\_Si}$ with a defined *diff*, and allowed the user to change *diff* interactively; $\bar{n}_{sign\_diff}$ is the average number of Key-UniGenes in $\{C_{sign\_diff\_Si}\}(i = 1, 2,. . ., 21)$; $n_{arm\_length\_begin}$ is the number of Key-UniGenes in $C_{arm\_length\_begin}$; k is the number of common Key-UniGenes between $C_{sign\_diff\_Si}$ and $C_{arm\_length\_begin}$. The significance of overlap between $C_{sign\_diff\_Si}$ and $C_{arm\_length\_begin}$ was visualized in the intersection area $R_{sign\_diff\_Si\_arm\_length\_begin}$ as a gray scale.

when both (*begin*)-th and (*begin* + *length* − 1)-th buckets of $C_{arm\_length\_begin}$ have the Key-UniGenes that are included in $C_{sign\_diff\_Si}$. This calculation was preprocessing for the EIM. Then, in real-time processing, after a certain *diff* was selected, each *E* value was represented in the intersection area, $R_{sign\_diff\_Si\_arm\_length\_begin}$, as a gray scale, if $C_{arm\_length\_begin}$ met gap_{max}. The user can control *diff* and gap_{max} interactively.

A flowchart that details these steps is shown in Fig. 9. The EIM for detecting individual difference of expression imbalance among SQ specimens is shown in Fig. 10. Figure 11 shows chromosome 3 of the EIM and the influence of gap_{max} and *diff* on the detection of the individual differences in expression imbalance among SQs.

RESULTS AND DISCUSSION

Detection of Expression Imbalance Specific to SQs

The EIM showed the distribution of expression imbalance specific to SQs (Fig. 6). It is highly comparable to previous CGH data of lung cancer reported by other investigators (6, 8, 9). There are significant differences among these CGH data because of method variation and sample preparation (especially tumor fraction of clinical samples). So it may be of little importance to compare details with individual CGH experiments. However, the most frequent abnormal loci reported in most of these studies were also detected by the EIM as regional signal images on chromosomes (expression imbalance regions), such as loss of 3p, 4q, 5q, and 8p, and gain of 1q, 3q, and 12p (6, 8, 9). The major difference from the CGH image is that signals are detected in a more confined area, which reflects the high resolution of EIM. Figures 6, 7, 10, and 11 clearly show the high resolution of EIM compared with CGH image. Especially, the intersection area $R_{+5\_3q\_1894\_5}$ showed the most significant overexpression region on 3q (Fig. 7), which is reported to be the most frequent aberration

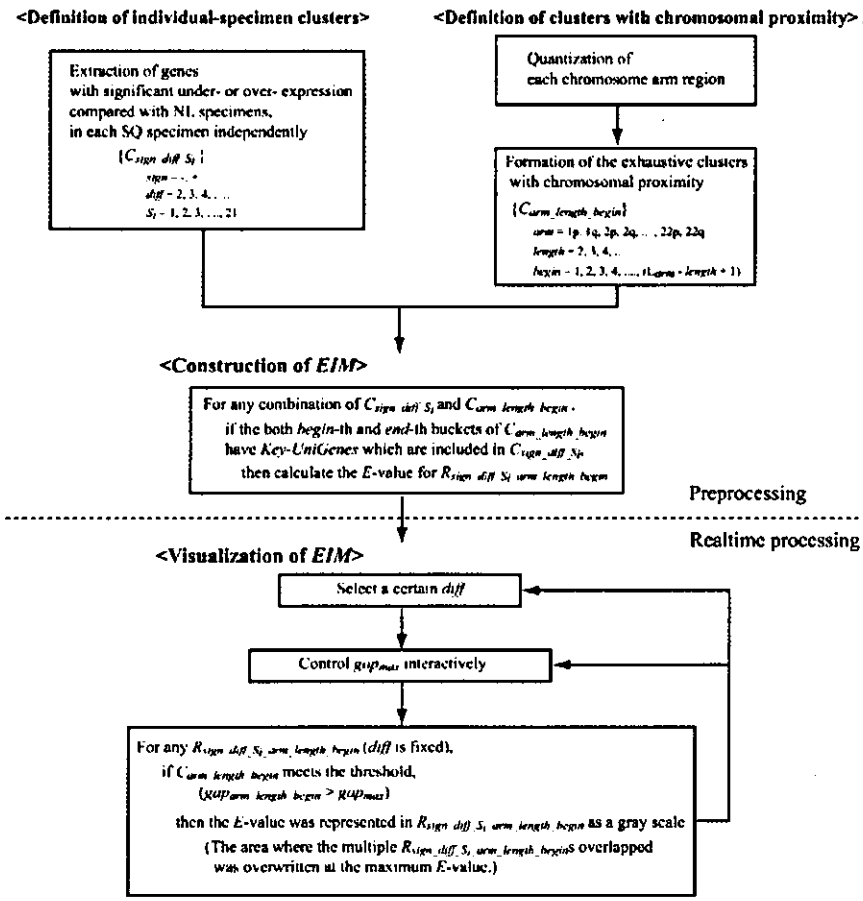


Fig. 9. Flowchart for construction of the EIM for detecting individual differences in expression imbalance among SQs. This flowchart provides details of the steps of the EIM for detecting individual differences in expression imbalance among SQs. For the step of "Definition of clusters with chromosomal proximity," please refer to Fig. 2. For the step of "Construction of the EIM" and "Visualization of EIM," please refer to Fig. 8. In this type of EIM, since the abscissa represented an array of S_i, it was impossible to represent *diff* on the abscissa like Fig. 4. Therefore, the EIM for individual specimen was visualized by $C_{sign\_diff\_Si}$ with a defined *diff*, and allowed the user to change *diff* interactively. In addition, it is possible to exclude regions containing large gaps between genes by changing gap_{max} interactively.

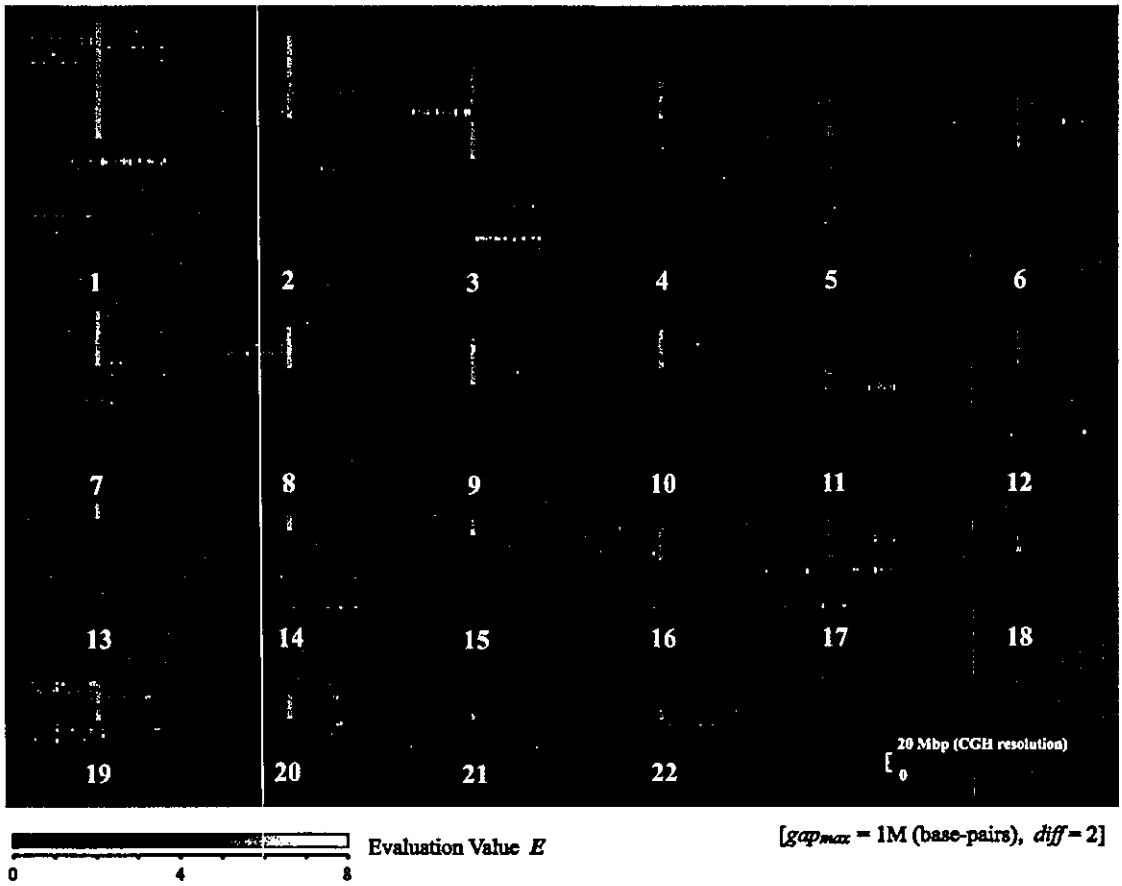
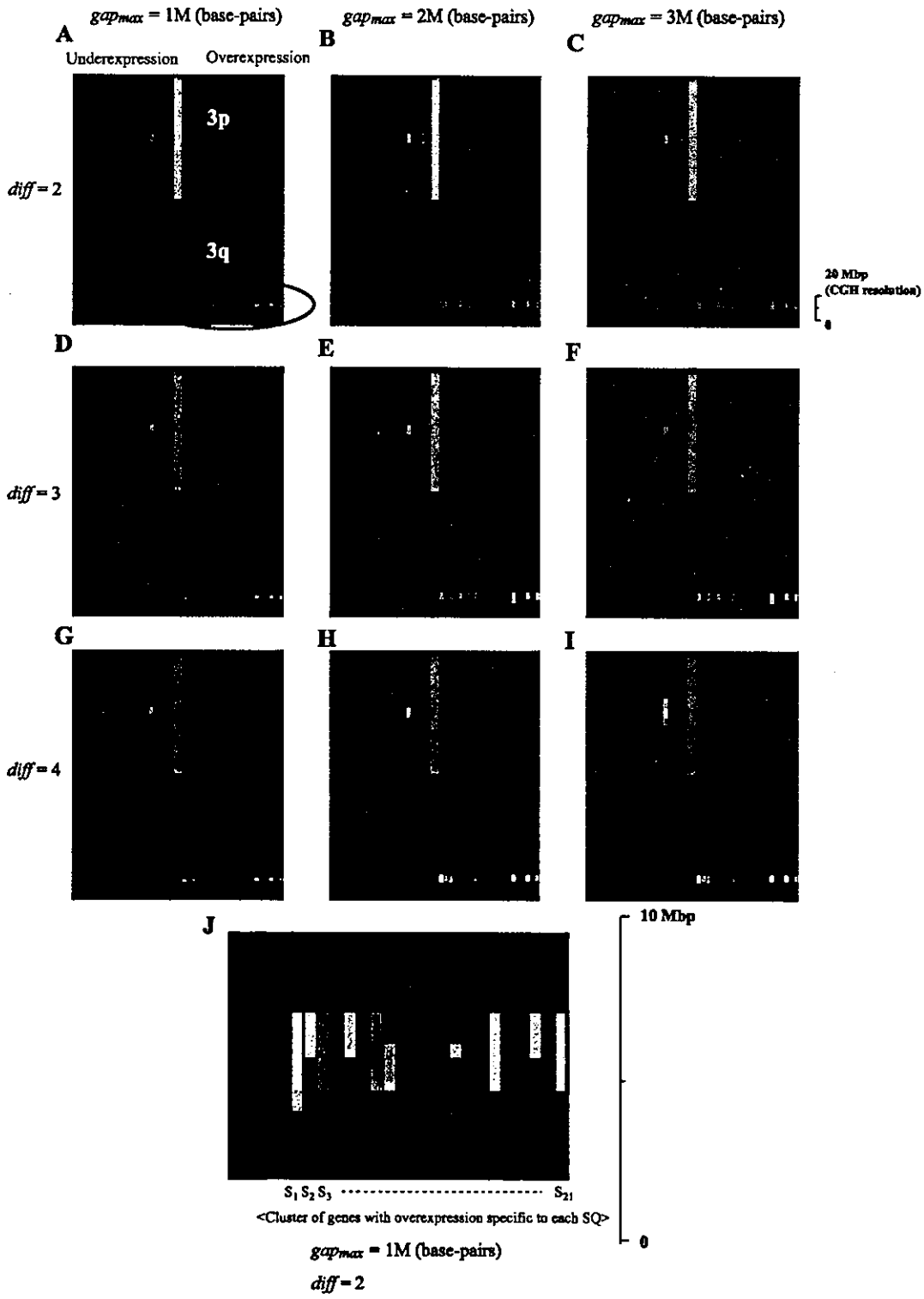


Fig. 10. The EIM for detecting individual difference of expression imbalance among SQs. The EIM was applied for detecting individual differences of expression imbalance among the SQs. Regions of underexpression and overexpression were visualized on the *left* and *right* side, respectively, as gray regional signals. The expression imbalance regions in each SQ were evaluated independently. Note the high resolution of EIM compared with CGH resolution (~20 Mbp).

in SQs by CGH (6, 8, 9). That is, the overlap (k = 6) between C_{+5} (the cluster of probes with overexpression whose differential level $D_1(g)$ is more than 5: n_{+5} = 205) and $C_{3q\_1894\_5}$ (the region from 189,400 to 189,900 kbp on chromosome 3: $n_{3q\_1894\_5}$ = 9, $gap_{3q\_1894\_5}$ = 146 kbp) was statistically the most significant (E = 7.2). Therefore, the overlap was evaluated using the hypergeometric probability for observing at least 6 (=k) common elements between randomly selected 205 (=n_{+5}) and 9 (=$n_{3q\_1894\_5}$) elements among 6,652 (=U) elements. The user can narrow down the possible expression imbalance regions by changing gap_{max} and d_{min} interactively. Especially, as is shown in Fig. 7, *A–I*, changing gap_{max}, which allows exclusion of the regions containing large gaps between genes, markedly influenced the detection of expression imbalance regions. In addition, all evaluation values of any combinations between the exhaustive uncertainty cluster sets of cancer specificity and chromosomal proximity are visualized simultaneously on the EIM as gradation

Fig. 11. Individual difference of expression imbalance on chromosome 3. *A–I*: chromosome 3 of the EIM and the influence of gap_{max} and *diff* on the detection of individual differences in expression imbalance among SQs. With regard to each SQ specimen, the under- and overexpression regions were visualized on the *left* and *right* side, respectively. Since the expression imbalance regions in each SQ were evaluated independently, this type of EIM clarified the individual difference of the overexpression region on 3q, which was detected as the most significant region in the group of SQs by another type of EIM. The user can narrow down the possible expression imbalance regions by changing gap_{max} and *diff*. *J*: macrograph of the encircled *region A* from panel *A*. When gap_{max} was 1 Mbp and *diff* was 2, the EIM showed that 17 of 21 SQs had overexpression regions on 3q, which is comparable to other data sets by CGH (6, 8, 9). In addition, note the high resolution of the EIM compared with CGH resolution (~20 Mbp).



of gray scale, which is clearly shown in Fig. 7*J*. This gradation pattern could convey the distribution of the false balance to the user through visual perception and enabled the detection of as many significant genes as possible.

Table 4 shows the gene list of $C_{3q\_1894\_5}$. Although this overexpression region strongly reflected the known genomic gain detected by CGH, several probes without overexpression were also detected on this region. There may be several reasons for this. First, since several probes with low quality were possibly included in this region, signal intensity does not always reflect their target mRNA expression levels. Improvement of the quality of probes would make it possible to detect the overexpression region more clearly. Second, mRNA expression levels would not completely reflect genomic copy number changes caused by chromosomal gain or loss, although there was strong correlation between them, because they are under various transcriptional control including feedback pathway of lost or gained genes themselves. Mukasa et al. (7) also reported that several genes without reduction of expression were detected in 1pLOH region of oligodendrogliomas. In addition, it should be stated that cancer tissues used here contained significant number of noncancerous stromal or inflammatory cells, which add noisy expression to cancer profiling.

Because of the complex factors discussed above, simple spatial mapping of the microarray expression profiles on chromosomal location gives little information about genomic structure (Fig. 12, *left*). In addition, it is very difficult to define adequate thresholds for cancer specificity and chromosomal proximity, because the distribution of "false balance" is unclear and the risk of overlooking significant genes by arbitrary selection of thresholds is high (i.e., the "threshold problem"). However, the EIM, using a new methodology without arbitrary selection of thresholds in conjunction with hypergeometric distribution-based algorithm, has a high tolerance of these complex factors and controls the risk of overlooking the expression imbalance regions. This advantage of the EIM over the simple spatial mapping is clearly shown in Fig. 12. The EIM detected the underexpression regions, *A* and *B*, and overexpression region, *C*, on chromosome 11, which are known loci with frequent genomic gain or genomic loss (6, 8, 9), although it was difficult to detect it from the simple spatial mapping of D_1 value.

Detection of Individual Difference in Expression Imbalance Among SQ Specimens

The analysis for extraction of probes with expression profiles specific to the group of cancer is very effective and popular. However, this type of analysis sometimes raises a critical problem because the individual difference among a group is unobservable. In this context, the function of the EIM to detect individual difference of expression imbalance in a group is very significant. Figure 11, *A–I*, shows that the user can narrow down the possible expression imbalance regions on chromosome 3 by changing gap_{max} and *diff* interactively. Furthermore, Fig. 11*J* shows the individual difference in the most significant overexpression regions on 3q (gap_{max} = 1 Mbp, *diff* = 2), where 17 of 21 SQs had overexpression regions, a finding comparable with other data sets analyzed by CGH (6, 8, 9).

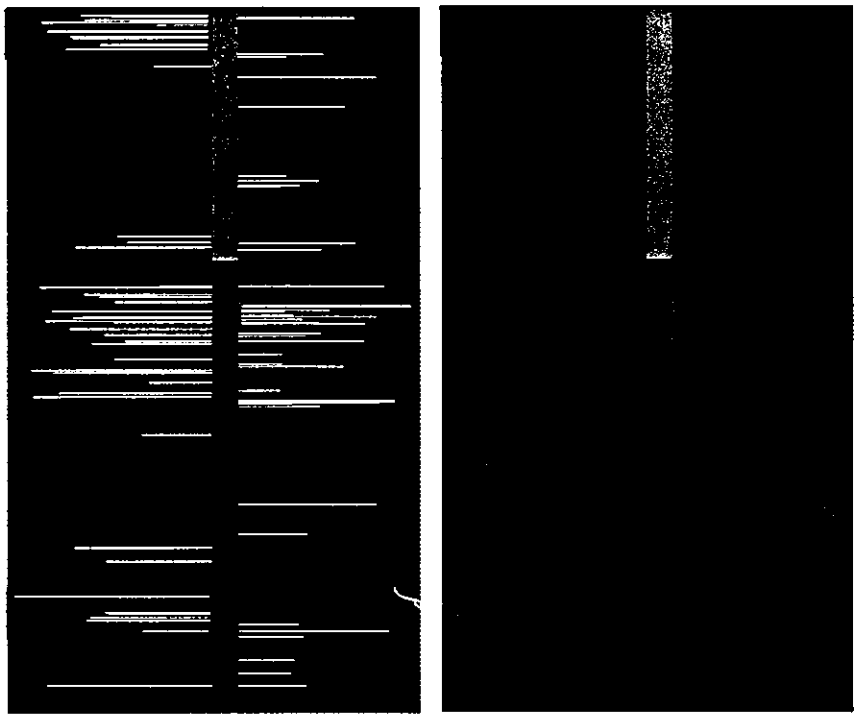
The high-resolution spatial map of expression profiles described in this report, i.e., the EIM, has several significant advantages. Its validity is clearly shown by the fact that many known loci with high frequent genomic losses or gains were detected by regional signals obtained with high resolution by this method.

Recently, several studies have been reported on microarray-based CGH for detecting genome-wide copy number changes (10). However, to our knowledge, no spatial mapping data obtained with such validity and genome-wide coverage have ever been reported previously from this array-CGH method. Experimental difficulty of genome hybridization and limited number of

Table 4. *Gene list of the overexpression region on 3q detected by the EIM*

Cancer Specificity	UniGene	Location, base pairs	Description
*	Hs.108660	189457995	ATP-binding cassette, subfamily C (CFTR/MRP), member_5
?	Hs.343882	189554055	CaM-KII inhibitory protein
x	Hs.129801	189604044	KIAA0604 gene product
x	Hs.1166	189609401	thrombopoietin (myeloproliferative leukemia virus oncogene ligand, megakaryocyte growth and development factor)
*	Hs.74619	189621219	proteasome (prosome, macropain) 26S subunit, non-ATPase, 2
x	Hs.141660	189658124	chloride channel 2
*	Hs.211568	189734699	eukaryotic translation initiation factor 4 gamma, 1
?	Hs.146161	189735389	hypothetical protein MGC2408
*	Hs.153591	189832147	Not56 (*D. melanogaster*)-like protein
*	Hs.174044	189851048	dishevelled 3 (homologous to *Drosophila* dsh)
*	Hs.152936	189862279	adaptor-related protein complex 2, mu 1 subunit

The expression imbalance map (EIM) detected the most significant overexpression regions, $R_{+5\_3q\_1894\_5}$, on 3q in the SQs. This region is a known locus with frequent genomic gains (6, 8, 9). This table shows the gene list of intersection area $R_{+5\_3q\_1894\_5}$. $R_{+5\_3q\_1894\_5}$ evaluated the overlap between C_{+5} (the cluster of probes on the U95A oligonucleotide arrays with overexpression whose differential level are more than 5) and $C_{3q\_1894\_5}$ (the region from 189,400 to 189,900 kbp on chromosome 3: $gap_{3q\_1894\_5}$ = 146 kbp). Differential levels of the genes marked with an asterisk (*) were more than 5, and those of the genes with "x" were less than 5. The genes with "?" were not the Key-UniGenes but the UniGenes that were contained in Genes On Sequence Map.



Simple spatial mapping ***Expression Imbalance Map*** (gap_{max} = 1M (base-pairs), d_{min} = 2)

Fig. 12. Advantages of the EIM over the simple spatial mapping of expression profiles. *Left*: a simple spatial mapping of D_1 value, which was calculated from the expression profiles of SQs, on chromosome 11. *Right*: the EIM of the same region. The EIM allowed detection of the underexpression regions, *A* and *B*, and overexpression region, *C*, on chromosome 11, which are known loci with genomic gain or genomic loss (6, 8, 9), although it is difficult to detect it by simple spatial mapping.

probes on CGH array could be major problems for it. There may be several reasons for the successful result of our alternative approach, calculation of genomic structure from expression profile. The first reason is the use of the Affymetrix-type GeneChip. The large number of probes (12,533) available enables detection of a relatively short abnormal region (chromosomal loss can frequently affect areas as short as a few hundred kbp), although this method can be easily applied to other types of microarrays. The second reason, which is most important, is that the EIM is a visualization method using a new methodology without arbitrary selection of thresholds in conjunction with hypergeometric distribution-based algorithm. By processing the complex factors and the threshold problems which hinder user's visual perception of essential information, the EIM presents to the user a comprehensive visual image of whole genome-wide information, clearly indicating where expression imbalance regions are and which genes are to be examined. It has an obvious advantage over simple spatial mapping of the expression profiles. For further curation by the user, simple clicking of a selected expression imbalance region on the EIM image leads to a direct link to a file that contains the actual gene names of the region, their expression scores, and other biological information. In addition, if the user input the UniGene number of genes of interest, the EIM indicates its position on the chromosome. Therefore, the EIM can be a broadband interface that enables user's visual perception of complex data and further curation.

Using the EIM, we might be able to detect regional under- or overexpressions independent of copy number changes, such as gene methylation silencing and/or imprinting abnormality (11). In addition, by using the Kruskal-Wallis test (4), which is a rank sum test to deal with three or more data groups instead of Mann-Whitney test, the EIM can easily extend to multiple phenotypes.

In conjunction with the microdissection technique, which can isolate only tumor-cell-specific RNA (2), our EIM can more precisely detect potential genomic structural changes, which offer more diagnostic and therapeutic impact.

Conclusion

In this report, we describe the development of the expression imbalance map, or EIM, a visualization method without arbitrary selection of thresholds, in conjunction with hypergeometric distribution-based algorithm, for detecting expression imbalance regions. By using this method, many known as well as potential loci with high frequent genomic losses or gains were detected as regional signals with much higher resolution than conventional methods, such as CGH. The EIM can be a broadband interface which enables user's visual perception of complex data and further curation,

and its advantage is obvious over simple spatial mapping of the expression profiles on chromosomal location. Therefore, the EIM would provide the user with further insight into the genomic structure through mRNA expression.

This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1. **Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, and Meyerson M.** Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98: 13790–13795, 2001.
2. **Bonner RF, Emmert-Buck M, Cole K, Pohida T, Chuaqui R, Goldstein S, and Liotta LA.** Laser capture microdissection: molecular analysis of tissue. *Science* 278: 1481–1483, 1997.
3. **Fujii T, Dracheva T, Player A, Chacko S, Clifford R, Strausberg LS, Buetow K, Azumi N, Travis WD, and Jen J.** A preliminary transcriptome map of non-small cell lung cancer. *Cancer Res* 62: 3340–3346, 2002.
4. **Hayter AJ.** *Probability and Statistics for Engineers and Scientists* (2nd ed.). Florence, KY: Duxbury Press, 2002.
5. **Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, and Pinkel D.** Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258: 818–821, 1992.
6. **Lu YJ, Dong XY, Shipley J, Zhang RG, and Cheng SJ.** Chromosome 3 imbalances are the most frequent aberration found in non-small cell lung carcinoma. *Lung Cancer* 23: 61–66, 1999.
7. **Mukasa A, Ueki K, Matsumoto S, Tsutsumi S, Nishikawa R, Fujimaki T, Asai A, Kirino T, and Aburatani H.** Distinction in gene expression profiles of oligodendrogliomas with and without allelic loss of 1p. *Oncogene* 21: 3961–3968, 2002.
8. **Pei J, Balsara BR, Li W, Litwin S, Gabrielson E, Feder M, Jen J, and Testa JR.** Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* 31: 282–287, 2001.
9. **Petersen S, Aninat-Meyer M, Schluns K, Gellert K, Dietel M, and Petersen I.** Chromosomal alterations in the clonal evolution to the metastatic stage of squamous cell carcinomas of the lung. *Br J Cancer* 82: 65–73, 2000.
10. **Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, and Brown PO.** Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23: 41–46, 1999.
11. **Reik W and Walter J.** Imprinting mechanisms in mammals. *Curr Opin Genet Dev* 8: 154–164, 1998.
12. **Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, and Krahe R.** Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci USA* 98: 1124–1129, 2001.

Heat Shock Transcription Factor 1 Protects Cardiomyocytes From Ischemia/Reperfusion Injury

Yunzeng Zou, MD, PhD*; Weidong Zhu, MD, PhD*; Masaya Sakamoto, MD; Yingjie Qin, MD; Hiroshi Akazawa, MD; Haruhiro Toko, MD, PhD; Miho Mizukami, MD; Norihiko Takeda, MD; Tohru Minamino, MD, PhD; Hiroyuki Takano, MD, PhD; Toshio Nagai, MD, PhD; Akira Nakai, MD, PhD; Issei Komuro, MD, PhD

Background—Because cardiomyocyte death causes heart failure, it is important to find the molecules that protect cardiomyocytes from death. The death trap is a useful method to identify cell-protective genes.

Methods and Results—In this study, we isolated the heat shock transcription factor 1 (HSF1) as a protective molecule by the death trap method. Cell death induced by hydrogen peroxide was prevented by overexpression of HSF1 in COS7 cells. Thermal preconditioning at 42°C for 60 minutes activated HSF1, which played a critical role in survival of cardiomyocytes from oxidative stress. In the heart of transgenic mice overexpressing a constitutively active form of HSF1, ischemia followed by reperfusion-induced ST-segment elevation in ECG was recovered faster, infarct size was smaller, and cardiomyocyte death was less than wild-type mice. Protein kinase B/Akt was more strongly activated, whereas Jun N-terminal kinase and caspase 3 were less activated in transgenic hearts than wild-type ones.

Conclusions—These results suggest that HSF1 protects cardiomyocytes from death at least in part through activation of Akt and inactivation of Jun N-terminal kinase and caspase 3. (***Circulation.*** **2003;108:3024-3030.)**

Key Words: ischemia ■ reperfusion ■ survival

Because the loss of functional cardiomyocytes causes heart failure, it is important to find the molecules that protect cardiomyocytes from death. The death trap is a useful method to identify cell-protective genes.[1] After transfection of the cDNA library constructed using cardiac mRNA in the mammalian expression vector, COS7 cells were cultured with a lethal dose of H_2O_2. We isolated several cDNAs from the surviving cells, and one of them was the heat shock transcription factor 1 (HSF1). The HSF family (HSF1-4) regulates the transcription of heat shock protein (HSP) genes.[2,3] In higher eukaryotes, expression of HSP genes is regulated primarily by HSF1 and HSF3 in response to various stresses and by HSF2 during development, whereas HSF4 seems to lack the activity as a positive transactivator. As a classical stress-responsive factor, HSF1 binds to heat shock element (HSE), which is present upstream of many HSP genes, and activates transcription of HSP genes under stress conditions. HSPs have been reported to be induced in various cardiovascular diseases and to have protective roles against various stresses.[4–6] Although HSF1 has also been reported to be expressed in hearts,[7] its role remains unknown. In the present study, using the transgenic mice expressing the active form of HSF1 (ΔHSF1),[8] we examined the role of HSF1 in the heart subjected to ischemia/reperfusion injury.

Methods

Materials

[γ-^{32}P]ATP was purchased from Du Pont-New England Nuclear Co. DMEM and FBS were from GIBCO BRL Co. pCMV SPORT heart cell expression cDNA library was from Life Technologies. The enhanced chemiluminescence reaction system was from Amersham. Other reagents were from Sigma.

Cloning of Cardioprotective Genes by Death Trap Method

COS7 cells cultured in DMEM supplemented with 10% FBS were resuspended in serum-free DMEM at 2×10^7 cells/mL immediately before transfection. Transfection of pCMV SPORT heart cell expression cDNA library was performed by electroporation (220 V, 960 μFD) at 50 μg of plasmid cDNA per mL. Forty-eight hours later, the transfected cells were stimulated with 1 mmol/L H_2O_2 for 16 hours in serum-free DMEM to induce cell death and cDNA was recovered from the survived colonies. This procedure was repeated 4 times.[1]

Received May 28, 2003; revision received August 4, 2003; accepted August 7, 2003.

From the Department of Cardiovascular Science and Medicine (Y.Z., W.Z., M.S., Y.Q., H.A., H.T., M.M., T.M., H.T., T.N., I.K.), Chiba University Graduate School of Medicine, Chiba; Department of Cardiovascular Medicine (N.T.), University of Tokyo Graduate School of Medicine, Tokyo; and Department of Bio-Signal Analysis, Applied Medical Engineering Science (A.N.), Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan.

*These authors contributed equally to this study.

Correspondence to Issei Komuro, MD, PhD, Department of Cardiovascular Science and Medicine, Chiba University, Graduate School of Medicine, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan. E-mail komuro-tky@umin.ac.jp

Circulation **is available at http://www.circulationaha.org** **DOI: 10.1161/01.CIR.0000101923.54751.77**

Cell Preparation and DNA Transfection

Primary cultures of cardiomyocytes were prepared from ventricles of 1-day-old Wistar rats, and 0.5 to 10 μg of HSF1 cDNA per dish was transfected into COS7 cells with the 1.5 to 20 μg of green fluorescence protein (GFP)-expressing vector by the standard calcium phosphate method.[9,10]

ΔHSF1 Transgenic Mice and Murine Ischemia/Reperfusion Model

Construction of the ΔHSF1 transgene and generation of the ΔHSF1 transgenic mice have been previously described.[8] Ischemia/reperfusion injury was produced in 12-week-old male transgenic mice and their littermate wild-type mice by transiently ligating the left coronary artery.[11] All protocols were approved by the guidelines of Chiba University.

Gel Mobility Shift Assay

DNA binding activities of HSF1 was examined as previously described using a self-complementary consensus HSE oligonucleotide (5'-CTAGAAGCTTCTAGAAGCTTCTAG-3') (Sigma) as a probe.[8]

Antisense Experiment

Phosphothionate antisense oligonucleotides (5'-CTAGAAGCTTCTAGAAGCTTCTAG-3') of HSF1 or scramble oligonucleotides (5'-AGTCACGATCTATAGATCTGAGTC-3') (Sigma) were prepared and applied to the culture medium (10 μmol/L) before the thermal preconditioning treatment of cardiomyocytes.

Apoptosis Analysis

Apoptotic death of cardiomyocytes was determined by TUNEL and by DNA ladder analysis, as previously described.[10,12]

Western Blot Analysis

Total protein extracts from the heart or immunoprecipitates were electrophoresed on an SDS-polyacrylamide gel (SDS-PSGE) and transferred to Immobilon-p membrane (Millipore). The blotted membranes were incubated with antibodies to HSF1, HSP27, *HSP70*, *HSP90*, *HSP110*, caspase3, Apaf1, Akt, phospho-Akt (Ser 473), Jun N-terminal kinase (JNK), phospho-JNK (Thr 183/Tyr 185), and α-actin (Santa Cruz Biotechnology), respectively. Immunoreactivity was detected using an enhanced chemiluminescence reaction system according to the manufacturer's instructions.

Statistics

Data are shown as mean±SE. Multiple group comparison was performed by one-way ANOVA followed by the Bonferroni procedure for comparison of means. A 2-tailed Student *t* test was used to compare transgenic with nontransgenic specimens under identical conditions. $P<0.05$ was considered statistically significant.

Results

Cloning of Cardioprotective Genes

COS7 cells were transfected with a heart cDNA library and then exposed to lethal dose of H_2O_2. Six cDNA clones were isolated from independent colonies of surviving cells, and 3 of them were identical to HSF1. To confirm the protective role of HSF1, we transfected the isolated HSF1 cDNA into COS7 cells. Overexpression of the HSF1 protected COS7 cells from H_2O_2-induced death (Figure 1A), whereas other cDNA, such as GFP, had no effect (data not shown).

To examine whether HSF1 protects cardiomyocytes against H_2O_2, cultured cardiomyocytes were exposed to H_2O_2 after thermal preconditioning at 42°C for 60 minutes followed by additional culture at 37°C for 24 hours. Electrophoretic mobility shift assay using HSF1 binding sequence revealed that the shifted band was enhanced after the thermal preconditioning (Figure 1B), indicating that HSF1 was activated in cultured cardiomyocytes by thermal preconditioning. Agarose gel electrophoresis showed that DNA ladder formation was observed in the cells treated with H_2O_2, whereas the DNA fragmentation was barely detectable in control cells and cells pretreated with thermal preconditioning (Figure 1C). The antisense oligonucleotides of HSF1 but not the scramble oligonucleotides abolished the protective effect of thermal

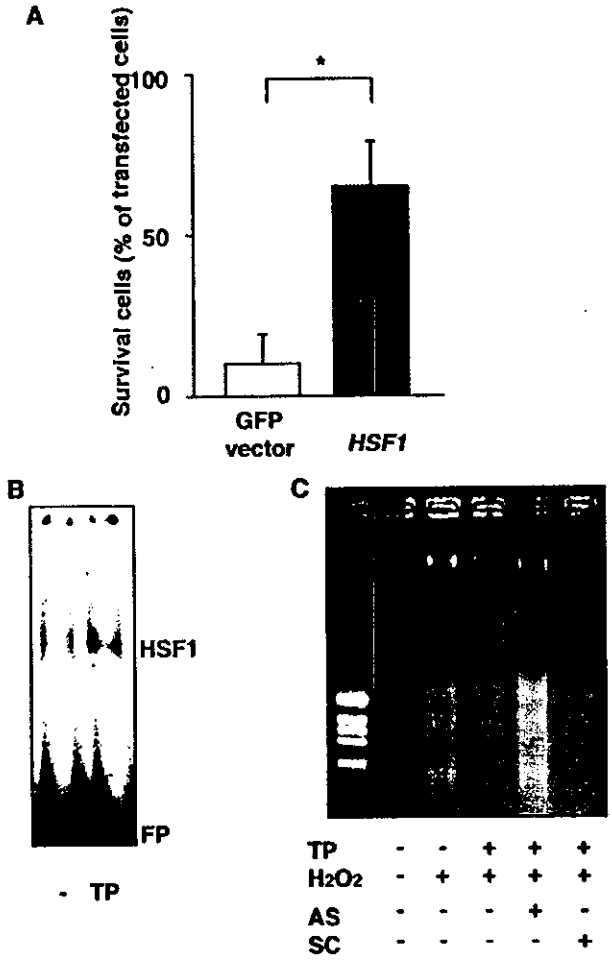


Figure 1. Protective effects of HSF1 in cultured cells. A, Survival of COS7 cells from H_2O_2 stress. COS7 cells were transfected with 2.5 μg of isolated *HSF1* cDNA with 7.5 μg of GFP expressing vector or GFP vector alone and additionally cultured with 1 mmol/L H_2O_2 after 24 hours. Survived cells are expressed as a percentage of transfected cells. Data are presented as mean±SE from 3 independent experiments. *$P<0.05$. B, DNA-binding activity of *HSF1*. Cultured cardiomyocytes were subjected to thermal preconditioning (TP) at 42°C for 60 minutes. A gel mobility shift assay was performed with whole-cell extracts using a DNA fragment containing HSE. A representative photograph from 3 independent experiments is shown. FP indicates free probe. C, DNA fragmentation. Cardiomyocytes with or without TP were exposed to 100 μmol/L H_2O_2 for 24 hours. Some cells were incubated with antisense (AS) or scrambled (SC) oligonucleotides of *HSF1* for 18 hours before TP. Genomic DNA was separated in 1.5% agarose gels and stained by ethidium bromide.

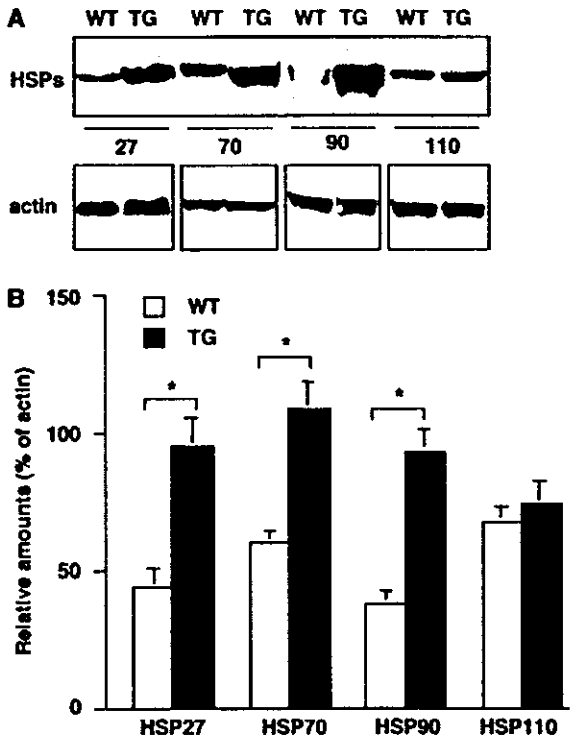


Figure 2. Expression of HSPs in the heart. A, Expression of HSPs levels in hearts from wild-type (WT) and transgenic (TG) mice were examined by Western blot analysis using each antibody. A representative photograph from 3 independent experiments is shown. α-actin blot was also presented as a loading control. B, Quantitative analysis of HSP protein expression. Intensities of HSPs and α-actin bands were measured by densitometric scanning of the autoradiograms. Relative amounts of HSPs are expressed as percentage of α-actin. Data are presented as mean±SE from 3 independent experiments. *P<0.05.

preconditioning (Figure 1C), suggesting that the thermal preconditioning protects cardiomyocytes from H_2O_2-induced cell death through activation of HSF1.

HSF1 Transgenic Mice

To elucidate the protective role of HSF1 in the heart, we examined the transgenic mice that overexpress *ΔHSF1*.[8] The transgenic mice were apparently healthy, and there were no significant differences in body weight, heart weight, blood pressure, and heart rate between the transgenic mice and littermate wild-type mice (data not shown).

Western blot analysis of heart extracts revealed that ΔHSF1 protein was expressed only in the adult transgenic mice, and there was no difference in expression levels of endogenous HSF1 between the transgenic and wild-type mice (data not shown).[8] The expression of HSPs 27, 70, and 90 was markedly upregulated in the transgenic heart compared with wild-type heart (Figures 2A and 2B).

Ischemia/Reperfusion Injury

Mice were subjected to cardiac ischemia for 40 minutes followed by reperfusion for 120 minutes. There were no abnormalities in ECG before ischemia in both the transgenic mice and wild-type mice (Figure 3). When the left coronary

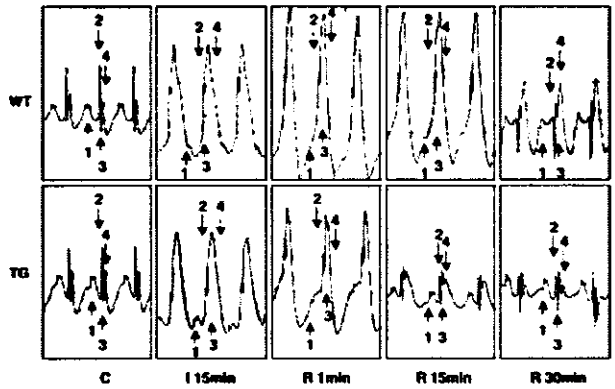


Figure 3. ECG recordings. ECG was recorded before and during ischemia/reperfusion. Representative ECG recordings before ischemia (C), at 15 minutes after ischemia (I 15min), and at 1, 15, and 30 minutes after reperfusion (R 1min, R 15min, and R 30min, respectively). Arrows 1 through 4 indicate P wave, QRS wave, ST segment, and T wave, respectively.

artery was occluded, the ST-segment was rapidly elevated in both types of mice. There was no significant difference in ECG change during ischemia between the transgenic and wild-type mice. When the suture was released to allow reperfusion, the ST-segment returned to baseline within 15 minutes in the transgenic mice whereas the ST-segment elevation remained elevated over 30 minutes in wild-type

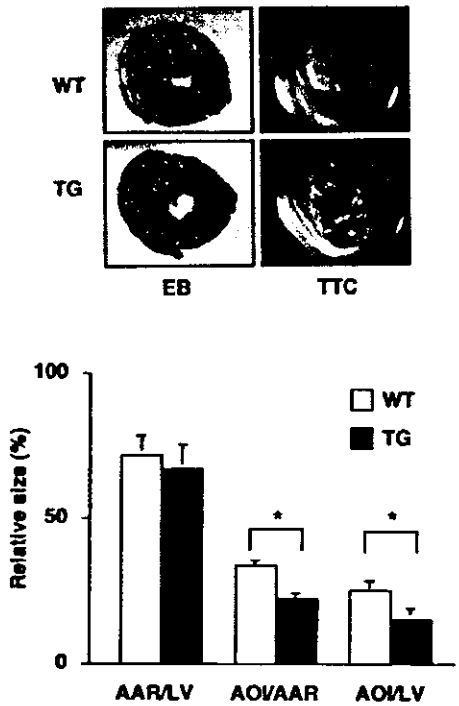


Figure 4. AAR and AOI. A, Representative photographs of wild-type (WT) and transgenic (TG) hearts stained by Evans blue dye (EB) and triphenyltetrazolium chloride (TTC) after ischemia/reperfusion (I/R). Red areas in left photographs indicate AAR, and pale white areas in right photographs indicate AOI. B, AAR is presented as a percentage of whole LV mass. AOI is expressed as a percentage of AAR and of whole LV mass. Values are mean±SE of 3 independent experiments. *P<0.05.

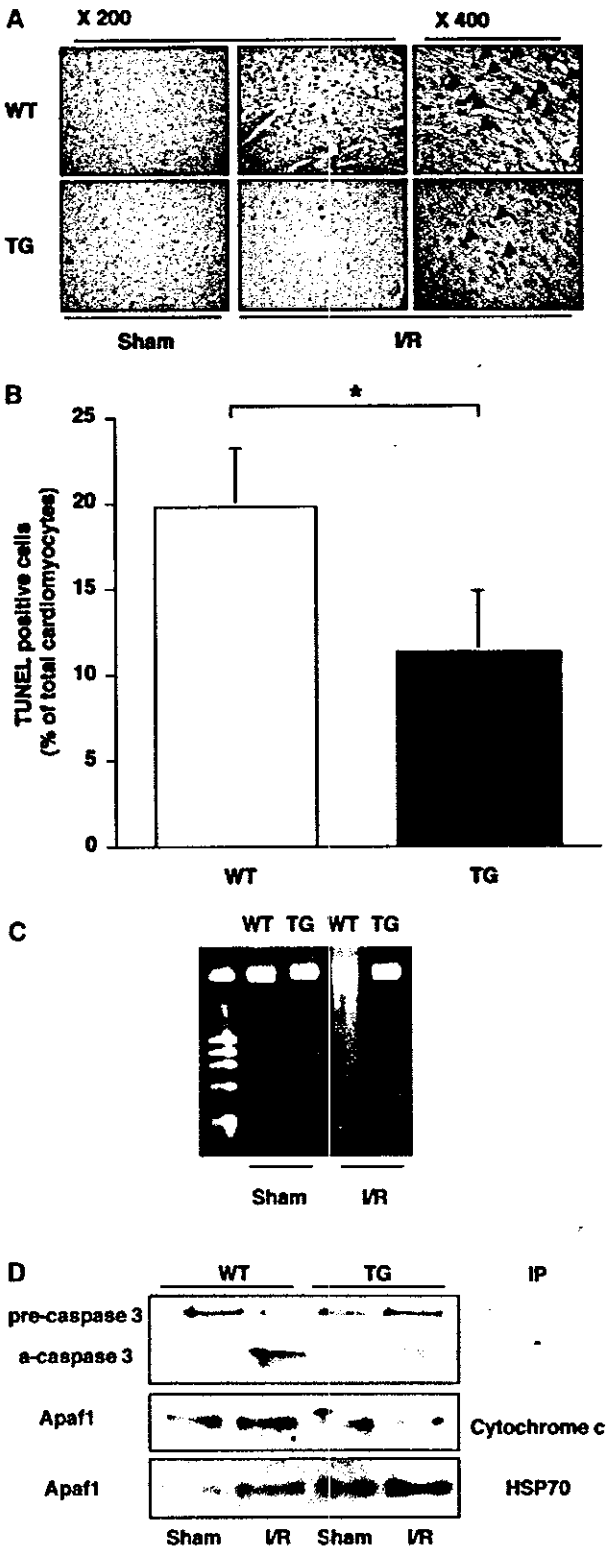


Figure 5. Apoptosis in the heart. A, TUNEL staining. Representative photographs in the edge of ischemic area after ischemia/reperfusion (I/R) or in similar area after sham operation (Sham) from wild-type (WT) and transgenic (TG) hearts are shown (original magnification ×200 and ×400). Arrows indicate TUNEL-positive cells. B, Numbers of TUNEL-positive cells. Fifty fields in ischemic area from 5 sections of each heart were counted. The number is expressed as percentage of total cardiomyocytes. *$P<0.05$. C, DNA ladders. Genomic DNA from the heart was separated in 1.5% agarose gels and stained by ethidium bromide. D, Activation of caspase 3 and Apaf1. WT and the TG mice were subjected to I/R. Total cell extracts from the heart or immune complexes with cytochrome C or HSP70 were subjected to SDS-PAGE. Western blot analysis was performed using antibodies against caspase 3 or Apaf1. Pre-caspase 3 indicates precursor of caspase 3; a-caspase 3, activated caspase 3; and IP, immunoprecipitation. Representative autoradiograms from 3 independent experiments are shown.

mice. These results suggest that *ΔHSF1* exerts its protective effect on the electrical activity of myocardium against ischemia/reperfusion injury.

We also measured areas at risk (AAR) and areas of infarction (AOI) after ischemia/reperfusion. Because we ligated the left coronary artery at the most proximal portion, the occlusion consistently created a large ischemic area (AAR, red myocardium in Figure 4A, left), and there was no difference in the ischemic area between both groups. However, there was a significant difference in infarct size (AOI, pale white areas in Figure 4A, right). Both areas of infarction/left ventricle (LV) and areas of infarction/ischemic area were significantly smaller in the transgenic group than those in the nontransgenic mice (Figure 4B).

Apoptosis in Mice

Ischemia/reperfusion injury has been reported to induce apoptosis in cardiomyocytes.[13] Apoptotic death of cardiomyocytes was examined in the heart by TUNEL and DNA ladder analysis. Ischemia/reperfusion induced apoptosis in many cardiomyocytes (Figures 5A and 5B) and produced marked DNA ladder formation (Figure 5C) in the ischemic area of wild-type heart. The number of apoptotic cells and the DNA fragmentation were significantly less in the *ΔHSF1* transgenic heart than the wild-type one (Figures 5A, 5B, and 5C). Caspase 3 was activated and Apaf1-cytochrome C complex was significantly increased in wild-type mice but less in the transgenic mice after ischemia/reperfusion (Figure 5D). Instead, more Apaf1-*HSP70* association was observed in the transgenic heart than in the wild-type heart after ischemia/reperfusion (Figure 5D).

Activation of Protein Kinases

It has been reported that many stresses, including ischemia/reperfusion, activate several protein kinases, including Akt/protein kinase B and JNK,[14] and that activation of Akt induces survival of cells whereas activation of JNK usually triggers death-signaling pathways. In the basal state, there was no difference in the activity of Akt and JNK between the wild-type and the transgenic heart (Figures 6A through 6C). Ischemia/reperfusion induced a significant activation of Akt and JNK in the wild-type heart (Figures 6A through 6C), suggesting that both survival and death signaling were activated in response to ischemia/reperfusion. In the transgenic heart, activation of Akt in response to ischemia/reperfusion was more prominent whereas activation of JNK was weaker than in the wild-type heart (Figures 6A through 6C). Although there was no change in protein

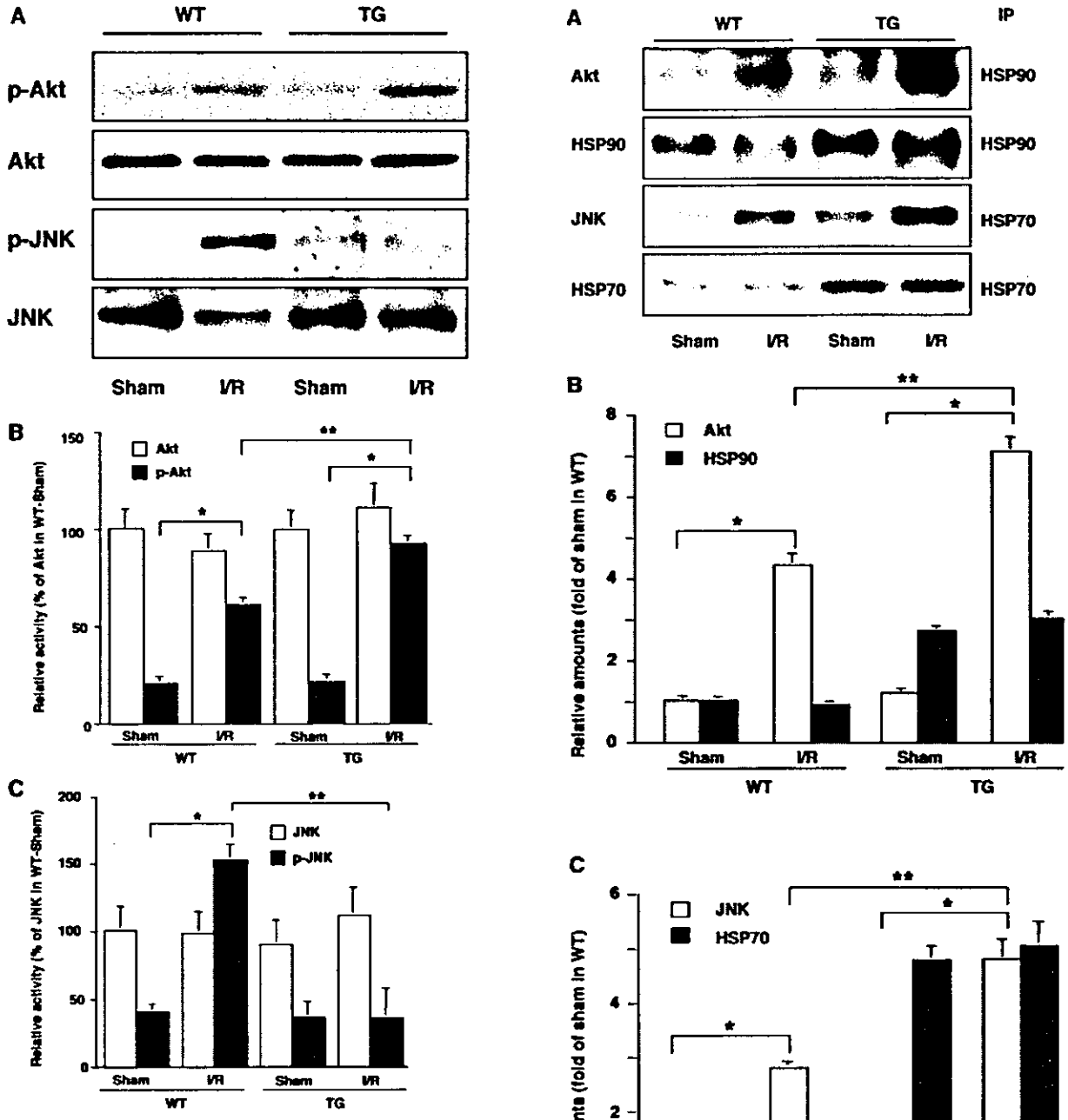


Figure 6. Activation of Akt and JNK in the heart. A, Protein extracts from the total heart were subjected to SDS-PAGE, and Western blot analysis was performed. p-Akt indicates phosphorylated Akt; p-JNK, phosphorylated JNK. B and C, Quantitative analysis of Akt (B) and JNK (C) activities. The intensities of Akt or p-Akt and JNK or p-JNK bands were measured by densitometric scanning of the autoradiograms. Relative amounts of Akt or p-Akt and JNK or p-JNK are expressed as percentage of Akt and JNK, respectively, in sham-operated wild-type (WT) mice. TG indicates transgenic mice; I/R, ischemia/reperfusion. Data are mean±SE from 3 independent experiments. *,**$P<0.05$.

levels of Akt and JNK after ischemia/reperfusion (Figure 6A), the amount of Akt and JNK that bound to *HSP90* and *HSP70*, respectively, was more markedly increased in the transgenic heart compared with the wild-type heart after ischemia/reperfusion (Figures 7A through 7C).

Figure 7. Association of Akt and JNK with HSPs. A, Protein extracts of the heart were immunoprecipitated (IP) using antibodies against HSP70 or HSP90. Immune complexes were subjected to SDS-PAGE, and the membranes were incubated with antibodies as indicated. B and C, Quantitative analysis of Akt (B) and JNK (C) bound to HSPs. Intensities of Akt, JNK, and HSP bands were measured by densitometric scanning of the autoradiograms. Relative amounts of Akt, JNK, and HSP are expressed as fold of Akt, JNK, and HSPs, respectively, in sham-operated wild-type (WT) or transgenic (TG) mice. I/R indicates ischemia/reperfusion. Data are mean±SE from 3 independent experiments. *,**$P<0.05$.

Discussion

The functional cloning system, death trap, is designed to isolate molecules that protect cells from death.[1] Using this method, *HSF1* gene was isolated from heart cDNA library. *HSF1*, a major heat stress-responding factor, upregulates many *HSP* genes, including *HSP110*, *HSP90*, *HSP70*, and small *HSPs*.[2,3] Expression levels of *HSP90*, *HSP70*, and *HSP27* were significantly elevated in the heart of transgenic mice overexpressing Δ*HSF1* compared with wild-type mice. Although *HSP70* has been reported to have a cardioprotective function,[15-17] the role of other HSPs in the heart is largely unknown. The Δ*HSF1* transgenic mouse provides a good model to examine the role of activated *HSF1* and of multiple HSPs in the heart.

Hearts of the transgenic mice were more resistant to ischemia/reperfusion injury, as indicated by faster recovery of ST-segment elevation in ECG and smaller infarct size. There are several potential mechanisms underlying the cardioprotective effect of HSPs.[4,5] HSPs are generally believed to help the correct folding of many proteins and restore their functional structures or target denatured proteins to the lysosome for degradation as molecular chaperones.[4,5] These functions of HSPs as molecular chaperones play important roles in maintaining the normal cell functions and promoting cell survival. Functions of HSPs other than molecular chaperones have recently attracted much attention in many organs, including the heart. It has been reported that accumulation of HSPs in myocardium shows an ATP-sparing effect that enhances preservation of high-energy phosphates.[4,5] Repeated coronary occlusions, called ischemic preconditioning, have been reported to induce rapid normalization of the elevated ST-segment during reperfusion by improving regional acidosis and hypokalemia.[18] HSPs may be involved in the faster normalization of ST-segment after reperfusion through restoration of the metabolic and ionic balance.

We showed here that ischemia/reperfusion induced less apoptosis of cardiomyocytes in the Δ*HSF1* transgenic mice than wild-type mice. Apoptosis is induced by defined biochemical mechanisms, including release of cytochrome C from mitochondria to cytoplasm and activation of Apaf1 through forming an apoptosome complex.[19,20] The apoptosome complexes cleave and activate caspase 3, leading to the inevitable fate of cell death. It has been reported that overexpression of caspase 3 in the heart significantly increases infarct size[21] and that the treatment with a caspase inhibitor conversely reduces infarct size.[22] HSPs have been reported to inhibit the formation of the apoptosome complexes through forming cytosolic complexes with Apaf1 or cytochrome C and then inhibit activation of caspases.[19,20] In this study, Apaf1 formed more complexes with *HSP70* and less with cytochrome C in the transgenic heart than the wild-type one, and ischemia/reperfusion activated Apaf1 and caspase 3 more weakly in the transgenic heart than the wild-type one. These results and observations suggest that HSPs protect cardiomyocytes from ischemia/reperfusion-induced cell death at least in part through forming complexes with Apaf1 and inhibiting the activation of caspase 3.

Both Akt and JNK were activated in the heart after ischemia/reperfusion, as reported previously.[23,24] Activation of Akt has been reported to protect cardiomyocytes against various stresses, such as oxidative stress[25] and the ischemia/reperfusion injury,[26] whereas JNK has been indicated to induce apoptosis during ischemia/reoxygenation in rat cardiomyocytes.[27] In the present study, ischemia/reperfusion activated more Akt and less JNK in the transgenic heart than the wild-type one. HSP90 has been reported to bind to Akt and promote activation of Akt through inhibition of protein phosphatase 2A.[28] Another recent report also showed that overexpression of *HSP90* leads to an increased phosphorylation of Akt.[29] On the other hand, Hsp70 family can suppress stress-activated signaling by directly binding to JNK.[14,30] We here observed that association between both *HSP90* and Akt and *HSP70* and JNK was more enhanced in the *HSF1* transgenic heart than in the wild-type heart. These results collectively suggest that HSPs exert cardioprotective effects through activation of Akt and suppression of JNK. Additional study is needed to elucidate which mechanism plays a major role in *HSF1*/HSP-induced protection of cardiomyocytes.

Acknowledgments

This work was supported by a Grant-in-Aid for Scientific Research, Developmental Scientific Research, and Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, and Culture and by a grant for research on life science from Uehara Memorial Foundation, Japan (to I. Komuro).

References

1. Vito P, Lacana E, D'Adamio L. Interfering with apoptosis: Ca^{2+}-binding protein ALG-2 and Alzheimer's disease gene ALG-3. *Science*. 1996;271: 521–525.
2. Benjamin IJ, McMillan DR. Stress (heat shock) proteins: molecular chaperones in cardiovascular biology and disease. *Circ Res*. 1998;83: 117–132.
3. Pirkkala L, Nykanen P, Sistonen L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J*. 2001;15:1118–1131.
4. Gray CC, Amrani M, Yacoub MH. Heat stress proteins and myocardial protection: experimental model or potential clinical tool? *Int J Biochem Cell Biol*. 1999;31:559–573.
5. Latchman DS. Heat shock proteins and cardiac protection. *Cardiovasc Res*. 2001;51:637–646.
6. Pockley AG. Heat shock proteins, inflammation, and cardiovascular disease. *Circulation*. 2002;105:1012–1017.
7. Knowlton AA, Sun L. Heat-shock factor-1, steroid hormones, and regulation of heat-shock protein expression in the heart. *Am J Physiol Heart Circ Physiol*. 2001;280:H455–H464.
8. Nakai A, Suzuki M, Tanabe M. Arrest of spermatogenesis in mice expressing an active heat shock transcription factor 1. *EMBO J*. 2000; 19:1545–1554.
9. Zou Y, Komuro I, Yamazaki T, et al. Cell type-specific angiotensin II-evoked signal transduction pathways: critical roles of Gbetagamma subunit, Src family, and Ras in cardiac fibroblasts. *Circ Res*. 1998;82: 337–345.
10. Zhu W, Shiojima I, Hiroi Y, et al. Functional analyses of three Csx/Nkx-2.5 mutations that cause human congenital heart disease. *J Biol Chem*. 2000;275:35291–35296.
11. Harada K, Komuro I, Hayashi D, et al. Angiotensin II type 1a receptor is involved in the occurrence of reperfusion arrhythmias. *Circulation*. 1998; 97:315–317.
12. Gu Y, Zou Y, Aikawa R, et al. Growth hormone signalling and apoptosis in neonatal rat cardiomyocytes. *Mol Cell Biochem*. 2001;223:35–46.
13. Fliss H, Gattinger D. Apoptosis in ischemic and reperfused rat myocardium. *Circ Res*. 1996;79:949–956.
14. Gabai VL, Sherman MY. Interplay between molecular chaperones and signaling pathways in survival of heat shock. *J Appl Physiol*. 2002;92: 1743–1748.