AD_____

Award Number DAMD17-95-1-5062

TITLE: Assessment of Genotoxic Injury by Induced Environmental Chemical Exposure: Biomarkers for Cancer Risk Assessment

PRINCIPAL INVESTIGATOR: Donald C. Malins

CONTRACTING ORGANIZATION: Pacific Northwest Research Foundation
Seattle, Washington 98122

REPORT DATE: February 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1999	Final (21 Aug 95 - 1 Feb 99)

4. TITLE AND SUBTITLE
Assessment of Geneotoxic Injury Induced by Environmental Chemical Exposure: Biomarkers for Cancer Risk Assessment

5. FUNDING NUMBERS
DAMD17-95-1-5062

6. AUTHOR(S)
Donald C. Malins

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Pacific Northwest Research Institute
Seattle, Washington 98122

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Models based on Fourier transform-infrared spectrometry and gas-chromatograpohy-mass spectrometry of DNA from the human prostate, breast, and ovary have been developed. The application of the models to various stages in the progression of normal tissue to the metastatic state has revealed that DNA undergoes identifiable changes in structure in the development of malignancy. The structural changes in DNA allow for the creation of prediction models (e.g., for cancer) that have promising clinical applications. Moreover, preliminary studies of the hepatic DNA from control and exposed (i.e., to dimethylnitrosamine) Japanese Medaka suggest that the same models would potentially have application to identifying changes in the DNA of aquatic organisms exposed to toxic chemicals.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Prostate, Breast, Ovary and Japanese Medaka	53
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

( ) Where copyrighted material is quoted, permission has been obtained to use such material.

(N/A) Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

(X) Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

(N/A) In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

(N/A) For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 32 CFR 219 and 45 CFR 46.

(N/A) In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

Principal Investigator's Signature

2-24-99
Date

TABLE OF CONTENTS

I. INTRODUCTION

In prior studies sponsored by USABRDL we revealed for the first time a variety of base lesions in the DNA of the female breast arising from reactions of the hydroxyl radical (•OH) (1). The base lesions included 8-hydroxyguanaine (8-OH-Gua), 8-hydroxyadenine (8-OH-Ade) and the putatively non-genotoxic ring-opened (Fapy) structures, 2,6-diamino-4-hydroxy-5-formamidopyrimidine (Fapy-G) and 4,6-diamino-5-formamidopyrimidine (Fapy-A). The •OH is believed to arise from the metal-catalyzed decomposition of H_2O_2 produced from redox cycling of estrogen metabolites (2) and possibly certain xenobiotics (e.g., aromatic hydrocarbons) (3,4). The base lesion concentrations were substantial in both microscopically normal and breast cancer tissues. In some cases, they represented greater than 1 base modification in 1000 normal bases. The concentrations of the 8-OH derivatives increased significantly in breast tumors, whereas the concentrations of the Fapy derivatives declined substantially. For example, the 8-OH-Gua concentrations in invasive ductal carcinoma (IDC) increased two-fold compared to those of the normal breast and the Fapy-A values declined over ten-fold. The GC-MS studies demonstrated that the •OH is capable of substantially modifying DNA in the progression of normal breast tissue to cancer. This work and our other prior findings (5) were the first to demonstrate the importance of •OH-induced DNA base damage in the etiology of breast cancer and provide a means for assessing the likelihood of tumor development on the basis of cancer probability-risk relationships.

In the present reporting period (the 5 year project was terminated by the Army after three years), we used Fourier transform-infrared (FT-IR)/statistics models to show structural alterations in DNA in relation to prostate (6), ovary and breast carcinogenesis (7) and conducted FT-IR spectral

analyses on the liver DNAs of control Japanese Medaka and Medaka exposed to dimethyenitrosamine at the USABRDL laboratories, Fort Detrick, MD.

FT-IR/statistics models, employing principal components analysis (PCA), allowed spectra to be expressed as points in space, each point being a highly discriminating measure of DNA structure. Changes in the cellular environment (e.g., increased free radicals) potentially lead to alterations in DNA structure and, hence, the vibrational and rotational motion of functional groups, thus changing the spatial location of the points. We showed that clusters of points, each representing prostate (6), ovarian and breast (7) DNAs had different spatial locations, sizes, or both, depending on whether the DNA was from normal tissue, primary tumor, or a primary tumor that had metastasized (7). The DNA alterations were linked to tumor formation and the probability of cancer was assessed using logistic regression of data that described the FT-IR wavenumber-absorbance associations between individual specimens in relation to the base and phosphodiester-deoxyribose structure. To gain additional insight into differences in spectra that reflect DNA changes in the transformation of normal tissue to cancer, we used models based on multivariable normal analysis of PC scores. This provided a unique potential for correctly classifying the DNA of normal tissues, primary tumors and metastasizing primary tumors (having disseminated metastases) (6-8). Prior to these findings, it was virtually impossible to determine whether a primary tumor had metastasized in a cancerous tissue without first identifying metastases elsewhere in the body.

Using the powerful FT-IR/statistics technology, groups (clusters) of PC points representing normal prostate DNAs were well separated from groups representing prostatic adenocarcinoma and benign prostatic hyperplasia (BPH) (6). This was the first evidence showing that prostatic DNAs of

healthy and transformed tissues were structurally different and could be discriminated on the basis of subtle changes in spectral properties. This advance opened up the possibility for constructing cancer probability relationships that are a potential basis for prostate cancer prediction. Logistic regression or discriminant analyses were employed to estimate a prostate specimen's "cancer probability" between 0.0 (non-cancer) and 1.0 (cancer), based on its PC scores. The predicted cancer probabilities were plotted vs. calculated risk scores. The derived probability values between those of normal and transformed prostate tissues represented various degrees of cancer risk. The probability-risk relationships were a promising basis for screening and prognostic trials in studies of prostate cancer by virtue of their high sensitivity and specificity. Moreover, the findings suggested that the predictive models could be applied to cancer risk in other circumstances, such as with animals exposed to environmental carcinogens (9).

In our most recent studies (7), FT-IR/statistics models demonstrated that the malignant transformation of morphologically normal human ovarian and breast tissues involves the creation of a high degree of structural modification (disorder) in DNA, prior to restoration of order in distant metastases. Order-disorder transitions were revealed by methods including PC analysis of IR spectra in which DNA samples were represented by points in two-dimensional space. Differences between the geometric sizes of clusters of points and between their locations revealed the magnitude of the order-disorder transitions. IR spectra provided evidence for the types of structural changes involved. Normal ovarian DNAs formed a tight cluster comparable to those of DNA from normal human blood leukocytes (HBL) with respect to spatial location and diversity. The DNAs of ovarian primary carcinomas, including those that had given rise to metastases, had a high degree of disorder, whereas the DNAs of distant metastases from ovarian carcinomas were relatively ordered. However, the spectra of the distant metastases were more diverse than those of

normal ovarian DNAs in regions assigned to base vibrations, implying increased genetic changes. DNAs of normal female breasts were substantially disordered (e.g., compared to the HBL), as were those of the primary carcinomas, whether or not they had metastasized. The DNAs of distant breast cancer metastases were relatively ordered. These findings evoked a unified theory of carcinogenesis in which the creation of disorder in DNA structure is an obligatory process followed by the selection of ordered, mutated DNA forms that ultimately give rise to metastatic cells.

Studies of Medaka liver DNA provided by the Fort Detrick laboratory involved the determination of IR spectra to provide a basis for establishing differences between the DNA of control groups and those exposed to toxic.

II. BODY

A. EXPERIMENTAL METHODS

Tissue Acquisition and DNA Isolation.

Human Tissues. Samples of human prostate, ovary and breast were obtained from the NCI Cooperative Human Tissue Network. Eighteen samples of BPH and 8 samples of adenocarcinoma served as test samples, each comprising a portion of the histologically identified lesion. Human blood leukocyte samples were obtained from 5 healthy individuals. Ovary samples were obtained from 13 morphologically normal tissues (O_n), 6 primary adenocarcinomas (AC), 9 metastasized primary adenocarcinomas (AC_m) and 7 distant metastases to the colon (AC_{dm}). Breast samples were obtained from 19 reduction mammoplasty tissues (RMT) of patients who had undergone hypermastia surgery, 10 invasive ductal carcinomas (IDC), 23 metastasized IDCs (IDC_m) and

seven samples of distant metastases to axillary nodes (IDC_{dm}). Eight samples of prostate tissue obtained from individuals who died by accidents were examined histologically and were found to be normal. These served as controls. No extraneous histological data were evident in any of the tissues employed in these studies. DNA was isolated from tissues as described and purity was established spectroscopically (1).

Medaka Tissues. Liver samples from control and exposed Japanese Medaka were excised by USABRDL, Fort Collins, MD and sent to PNRI in dry ice for DNA isolation and FT-IR spectral analysis. DNA purity was established as described (1).

FT-IR Spectral Analysis of Prostate DNA.

FT-IR spectral analyses were carried out essentially as described (10). The procedure involved the use of a FT-IR microscope spectrometer. A thin film of DNA is placed on a BaF_2 window and an IR beam is focused on it. The interferogram recorded in the detector is then Fourier-transformed into an absorbance spectrum that is baselined and normalized to an absorbance of 1.0 in the range of interest (e.g., 1750-770 cm^{-1}). To develop a common basis for plotting, PC scores for the entire sample database were calculated, giving equal weight to each group. The difference between two DNA spectra or between the centroids (mean spectra) of two groups was defined as the Euclidean distance. This was expressed as a percentage by dividing it by the square root of the number of wavelengths (e.g., 1750 to 770 cm^{-1}), then dividing by the mean normalized absorbance and multiplying by 100. All hypothesis testing and plotting was carried out using the SAS and S-PLUS statistical packages.

For FT-IR/PCA spectral analysis, each spectrum was normalized across the range of 1750 to 770 cm^{-1}, as described previously (11). This yielded a relative absorbance value for each wavenumber, with a mean of 1.0. Euclidean distance was used to define the difference between a pair of spectra, either for the entire spectrum or for a sub-region (11,12). This standard distance measure is defined as the square root of the sum of squared absorbance differences between spectra at each of the wavenumbers considered (e.g., 1051 for the entire spectral region 1750 - 770 cm^{-1}). The Euclidean distance can also be expressed in a more descriptive form as a percent. The numerator of the percent is the Euclidean distance divided by the square root of the number of wavenumbers for a region. The denominator used here for the percent for any region is the mean normalized absorbance between 1750 - 770 cm^{-1}, which is 1.0 for every case.

PC analysis was used to identify a few variables (components) that capture most of the information in the original, long list of variables (the spectral absorbances at each wavenumber). This reduction in the number of variables is analogous to the process in educational testing whereby many individual test scores, such as in reading and arithmetic, are combined into a single academic performance score. Four PC scores (i.e., four dimensions) were found to be sufficient to describe the 1051 dimensions of the normalized spectra. PC scores were calculated with the grand mean of all spectra subtracted from each spectrum. The nonparametric Spearman correlation coefficient was used to assess the association of PC scores with patient ages and Gleason scores. The nonparametric analysis was used because some of the distributions are skewed or are not normal ("bell-shaped"), which can lead to a bias in statistical significance when estimated from the Pearson correlation coefficient.

Two cases, which were outliers, were omitted from these analyses, leaving 29 cases. The omitted BPH sample and the omitted cancer sample had spectra very different from the included cases. Their Euclidean distances from the most similar spectra were 52% and 41%, respectively. All other spectra differed from their "nearest neighbor" spectrum by at most 21%, with a majority of spectra differing by less than 11%. A plot of the two outlier spectra also showed drastically reduced absorbance in the region around 1650 cm^{-1} representing vibrations of the nucleic acids. The Kruskal-Wallis and Mann-Whitney tests were used to determine if the three groups had similar diversity, defined as the mean distance of a spectrum to its group centroid. A permutation test was used to determine whether the three groups tended to cluster separately (representing an internal similarity of spectral properties in a group). The distance of each spectrum to its nearest neighbor in its own group (either normal, BPH, or cancer) was calculated, and the mean of these nearest neighbor distances for of all of the spectra was the test statistic. The test was carried out by randomly permuting group membership labels 1000 times and recalculating the test statistic each time. A smaller observed distance to the nearest neighbor than that obtained by random re-labeling of groups is an indication of clustering. A nonparametric, rank-based version of this test was carried out by expressing each distance as a rank. For each spectrum, the distances to other spectra were ranked and the permutation test was carried out as described above, but with distances replaced by ranks. The test statistic was a mean rank. Again, a smaller observed mean rank than the mean obtained from random permutation is an indication of clustering. Both the test using distance and the test using ranks were carried out for the entire spectrum, 1750 - 770 cm^{-1}, and for several subregions.

Finally, discriminant analysis was used as a model to determine if PC scores could be used to discriminate between pairs of DNA groups (normal vs. BPH, normal vs. cancer and BPH vs.

cancer). The discriminant analysis yields a risk score, which is a linear combination of PC scores, and a predicted probability of a sample being in one of the two groups considered (e.g., probability of being BPH when BPH is compared to normal). These predicted probabilities, along with a chosen probability cut point, can be used to classify samples and provide estimates of sensitivity and specificity, or percent of samples correctly classified. For each analysis a cut point was chosen that jointly maximized sensitivity and specificity.

FT-IR spectral Analysis of Ovarian and Breast DNA (7).

Spectra and PC plots were obtained as described for the prostate DNA. We used the permutation test ($5x10^2$ permutations) to test the null hypothesis that the distance between centroids of two groups is zero (i.e., that the mean spectra are the same for the two groups) and that the observed distance arises by chance. We also used a two-sided unequal variance t-test for the null hypothesis that the mean absorbance at a given frequency is equal between groups. The t-test, carried out at each frequency, yields a plot of P-values vs. frequency. Re-sampling (with 10^3 samples) was used to test the null hypothesis that the distance between states (e.g., between the centroid for normal tissue and that for primary tumor tissue) is the same for the ovary and breast. The same re-sampling procedure was used to compare the "base" region (1750-1315 cm^{-1}) to the "phosphodiester-deoxyribose" region (1314-770 cm^{-1}). The P-value for these re-sampling tests is defined as twice the proportion of re-sampled observations that are on the opposite side of zero from the observed differences, with a maximum of $P = 1.0$. We tested for differences in diversity between two groups based on the ratio of group variances at each wavenumber using a two-sided F-test. Differences in PC cluster size and/or location were determined using a test for the equality of covariance matrices of PC scores. Statistical analysis of age vs. PC scores was used to test whether age played a role in determining spectral characteristics.

FT-IR Spectral Analysis of Medaka DNA.

Data were obtained representing 107 FT-IR scans of 86 Medaka fish samples. Some samples represented multiple fish, and some samples had replicate scans. (See Table 1 for tank and treatment values.)

Table 1. Tank time – treatment values.

	6 Week	3 Month	6 Month	9 Month
Control	4 tanks, 4 samples	4 tanks, 6 samples	4 tanks, 11 samples, 11 fish	4 tanks, 11 samples, 11 fish
	1 sample per tank	tanks 1, 2=1 sample	tanks 1, 2, 4 = 3 samples	tanks 1, 2, 3 = 3 samples
		tanks 3, 4 = 2 samples	tank 3 = 2 samples	tank 4 = 2 samples
Low	2 tanks, 2 samples	4 tanks, 5 samples	4 tanks, 9 samples, 9 fish	4 tanks, 9 samples, 9 fish
	1 sample per tank	tanks 5, 7, 8 = 1 sample	tank 5 = 3 samples	tanks 5, 6, 8 = 2 samples
		tank 6 = 2 samples	tanks 6, 7, 8 = 2 samples	tank 7 = 3 samples
High	3 tanks, 3 samples	4 tanks, 5 samples	4 tanks, 11 samples, 11 fish	4 tanks, 10 samples, 10 fish
	1 sample per tank	tanks 10,11,12=1 sample	tanks 9, 10 = 2 samples	tanks 9, 12 = 2 samples
		tank 9 = 2 samples	tank 11 = 4 samples, tank 12 = 3 samples	tanks 10, 11 = 3 samples
# of Samples	9	16	31	30
# of Scans	9	12+11 rep.+4 other	31 + 1 replicate	30 +9 replicates
# of Fish	?	60 (5 per tank)	31	30

All FT-IR absorbance spectra were processed as in earlier studies (10,11) to yield a mean normalized absorbance of 1.0 in the range of interest 1750-770 cm^{-1}. Multiple scans of the same sample were averaged to yield one scan per sample.

Unweighted principal component (PC) scores were calculated for 86 spectra in a manner analogous to our previous studies (without removing the mean). Based on cluster analysis, showing the distance of each sample to its nearest neighbor, six samples were determined to be outliers compared to the rest of the group (Fig. 1; see addendum) These six samples had distances

to the nearest neighbor of about seven to more than 20 (Y-axis of Fig. 1), which translates to a percentage difference of 22% to over 60% between these samples and the most similar spectra in the rest of the group. (The conversion of spectral distances to percentages was used, based on 100% multiplied by the root-mean squared difference in normalized absorbance for two spectra across the 981 wavenumbers between 1750-770 cm^{-1}, and divided by the mean normalized absorbance, which is 1.0 in this range.) In previous studies, we have excluded samples that were this different (or more different) from the other samples studied.

All of the six outliers occurred in the six and nine-month samples, but there was no tendency for particular treatments or tanks to occur more frequently among the outliers.

The principal component scores were re-computed without the six outliers, and the balance of the analyses were carried out on the 80 samples. As noted later, there was little difference in the spectra among the three treatment groups (control, low and high), so that most analyses are confined to the differences between samples collected at different times.

The mean distance of each sample to the centroid of its time group (6 weeks, 3 months, 6 months, 9 months) was calculated and compared to the distances between pairs of groups using the Mann-Whitney test. Discriminant analysis was used to determine if the spectra—as represented by the PC scores—could be used to classify the samples by time or by treatment. Analysis of variance (ANOVA) was used on the normalized absorbances at each wave number to determine those frequencies that showed more differences among time groups.

B. RESULTS

Prostate Studies. FT-IR/PCA studies of prostate DNA (normal, cancer and BPH) yielded four components (four PC scores per case) which explained a total of 90% of the spectral variation over 1051 wavenumbers. That is, most of the features of the 29 prostate spectra could be described by four PC scores (labeled PC1, PC2, PC3, PC4). The first two PC scores explained 76% of the variation and were adequate for a two-dimensional representation. Figure 2 shows that the three groups were distinctly clustered. The two outliers omitted from the analysis are also represented on this plot and appear to the right of the main clusters.

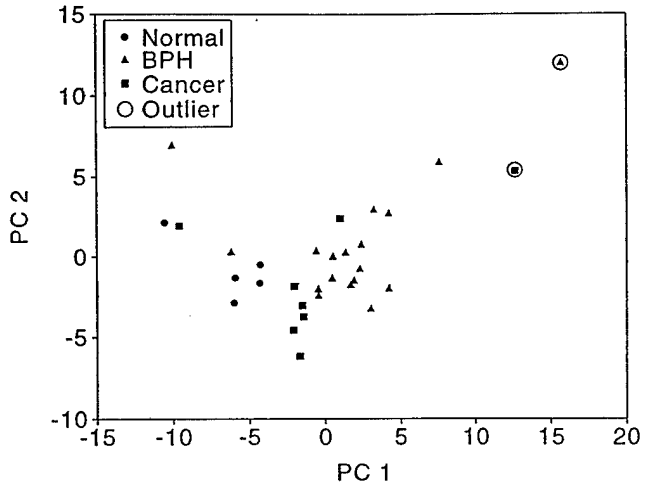

Fig. 2. Two-dimensional PC plot derived from FT-IR/PCA spectral analysis showing distinct clustering of normal, BPH and prostate cancer points. Notably, both of the groups of prostate lesions occur to the right of the points for the DNA of normal prostate.

The actual distance of the outlier points to other points is larger than that shown in this two-dimensional plot due to differences represented by other dimensions. The permutation test for clustering of groups (1750 - 770 cm^{-1}) yielded P = 0.1 based on the distance measure and P = 0.01 using the nonparametric ranking technique (Table 2). The greater significance obtained by the ranking method arises from the relative isolation of one or two cases from the core of their group

(Fig. 2), a configuration which influences the distance measure more than the ranking measure. Using these techniques, significant clustering was observed for two regions of the spectrum: 1174 - 1000 cm^{-1} (assigned to strong stretching vibrations of the PO_2^- and C-O groups of the phosphodeister- deoxyribose structure) and 1499 - 1310 cm^{-1} (assigned to weak NH vibrations and CH in-plane deformations of the nucleic acids) (13-15). The P-values for mean distance and mean rank for these regions ranged from 0.02 to < 0.001 (Table 2). The significance levels obtained strongly reject the null hypothesis that the observed clustering of the three groups occurred by chance.

Table 2. Mean distance to nearest neighbor of same group and permutation test for nonrandom clustering.

Spectral region, cm^{-1}	Mean distance*			Mean rank†		
	Observed	Random permutation	*P* value	Observed	Random permutation	*P* value
1750–700	12.2	12.8	0.1	2.0	3.0	0.01
1750–1500	12.3	12.3	0.5	2.4	3.0	0.09
1499–1310	5.9	6.5	0.02	1.6	3.0	<0.001
1309–1175	6.7	6.5	0.7	3.0	3.0	0.5
1174–1000	13.2	15.0	0.02	2.0	3.0	0.01
999–700	6.9	7.4	0.1	2.3	3.0	0.05

Distance is expressed as a percent difference between spectra; 1000 permutations were performed for each spectral subregion.
*Mean Euclidean distance to nearest neighbor in the same group expressed as a percent.
†Mean rank of Euclidean distance of each spectrum to nearest neighbor in the same group.

Detailed comparisons were made between the spectra of pairs of groups: normal vs. cancer, normal vs. BPH and BPH vs. cancer. The statistical significance of differences in mean normalized absorbance between groups was assessed for each wavenumber between 1750 - 770 cm^{-1}, using the unequal variance t-test (Fig. 3). The plot shows the comparison of the mean spectrum for each of the two groups, as well as the P-value from the t-test. The regions with $P \leq 0.05$ represent differences between groups (e.g., normal vs. cancer), which are much less likely to be due to chance than regions with $P > 0.05$.

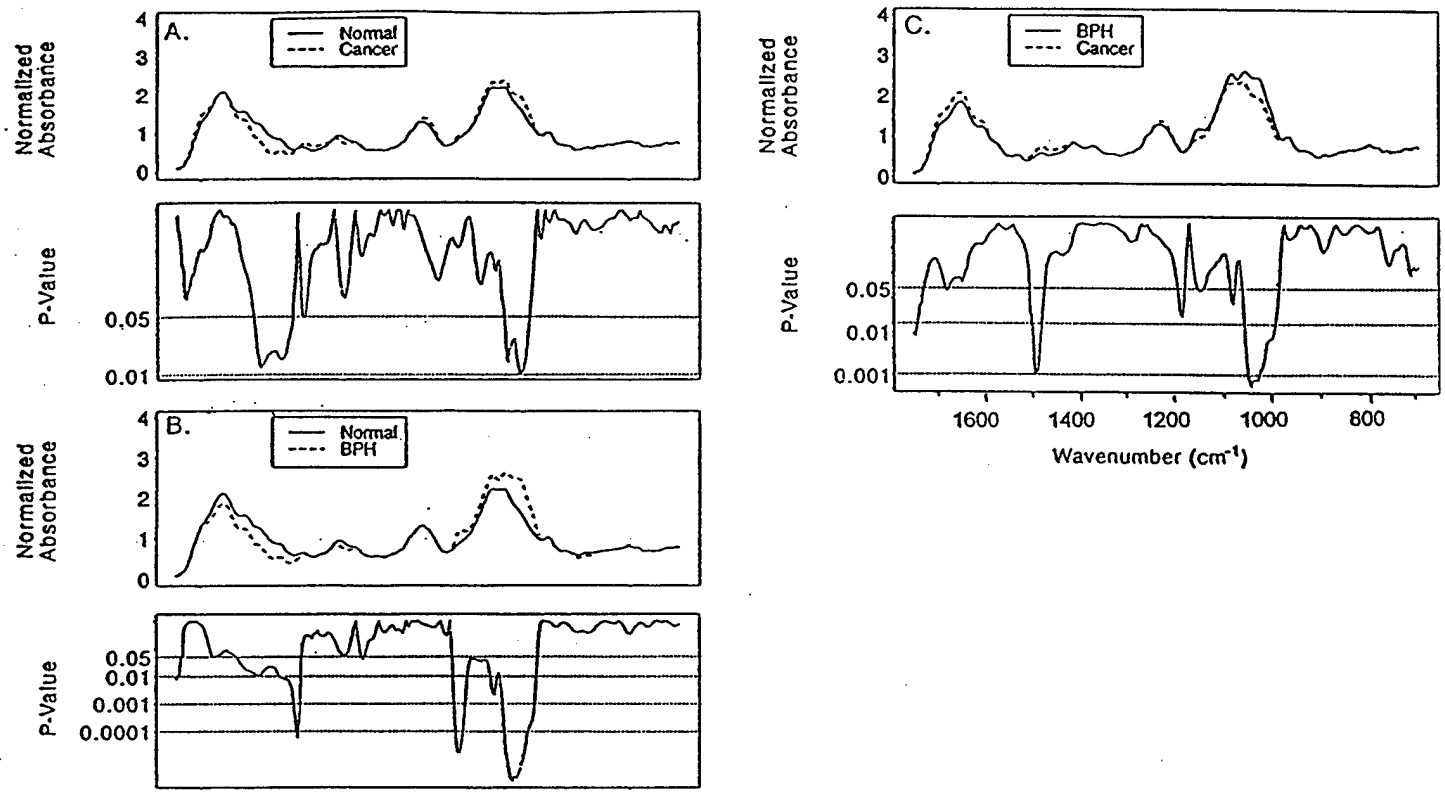

Fig. 3. Comparison of the mean spectrum of cancer vs. normal prostate tissue (A), BPH vs. normal tissue (B) and cancer vs. BPH (C). The lower plot of each panel shows the statistical significance of the difference in mean absorbance at each wavenumber, based on the unequal variance t-test. P-values are plotted on the $\log_{10}$ scale.

The spectral regions with significant differences in absorbance for the phosphodiester-deoxyribose structure are similar ($\approx$ 1050 - 1000 cm^{-1}); however, absorbances associated with the bases vary among the groups. That is, for the normal vs. cancer comparison, the region of significant difference is primarily $\approx$ 1475 -1400 cm^{-1}, whereas for the normal vs. BPH comparison it is $\approx$ 1600 - 1500 cm^{-1}. The comparison for BPH vs. cancer is focused at $\approx$ 1500cm^{-1}. For the normal vs. BPH and BPH vs. cancer comparisons, significant differences are shown between $\approx$ 1175 to 1120 cm^{-1}, a region that likely includes symmetric stretching vibrations of the PO_2 group (13-15). The difference in means at all of these spectral regions is apparent from the plots of mean spectra per group in Figure 3. The structural modifications are pivotal in the spatial distribution of points in the PC plot (Fig. 2) and in the pronounced discrimination between clusters (Table 2).

Cluster diversity. The diversity of the three groups, expressed as the mean distance to the group centroid, did not differ significantly (p = 0.8). However, the normal prostate group was slightly less diverse (mean distance = 11.7%) than was the BPH group (mean distance = 14.5%) or prostate cancer group (mean distance = 13.9%). In comparable studies of the female breast, the diversity was not different between normal breast and primary breast cancer DNA. However, significant differences did exist between normal breast and primary tumors with disseminated metastases and primary cancer and metastatic cancer (12,16).

Group Classification. PC scores can be readily used to classify subjects into groups, when pairs of groups are compared using discriminant analysis. The discriminant analysis (Table 3) is an equation which yields a risk score, R, when the values of the PC scores are inserted into the equation. R is transformed to a probability by the following statistical equation: probability = exp(R)/[1+exp(R)]. A cut point is chosen and if the probability exceeds this cut point, the case would be classified as BPH. The actual cut points are noted below.

Table 3. Logistic Regression models for probability of BPH (vs. normal), cancer (vs. normal), and cancer (vs. BPH).

	Coefficients ± SE					Correct classification rate		
Model	Intercept	PC1	PC2	PC3	PC4	By group, %	Overall, %	*P* value*
Normal vs. BPH	24.9 ± 0.1	5.2 ± 0.2	5.8 ± 0.04	3.9 ± 0.03		Normal, 100; BPH, 100	100	<0.001
Normal vs. cancer	34.3 ± 0.1	12.0 ± 0.04			−21.0 ± 0.1	Normal, 100; cancer, 100	100	<0.001
BPH vs. cancer	−14.5 ± 8.1	−4.5 ± 2.6	−3.7 ± 2.0		−11.1 ± 6.3	BPH, 88; cancer, 100	92	<0.001

Normal, $n = 5$; BPH, $n = 17$; cancer, $n = 7$. *P* values are based on the null hypothesis that each model is not predictive of group membership. *P* values are calculated from a χ^2 test on change in deviance.
**P* value for the null hypothesis that the probability of a case falling into a specified group is unrelated to the PC scores.

As shown in Table 3, the model for normal vs. cancer and normal vs. BPH correctly classifies each group 100% and 100% overall (P-values in each case were < 0.001). The correct classification rate for cancer vs. BPH was close to 90%, based on a designation of "cancer" for a predicted

probability of ≥ 0.1. (Probability cut-points of 0.15 to 0.41 achieve the same correct classification rates in the BPH vs. cancer comparison). The predicted probabilities based on the models in Table 3 are given in Figure 4. The individual risk score is based on the appropriate PC model (Table 3) and the predicted probability is a mathematical function of the risk score, as noted above. All of the BPH and cancer cases have predicted probabilities extremely close to 1.0 and all of the normal cases have predicted probabilities of ≤ 0.002 when BPH or cancer are compared to normal cases. These marked distinctions in predicted probabilities confirm the clear separation of groups, as shown in Figure 2. When cancer is compared to BPH, predicted cancer probabilities ranged from 0.42 to 1.00 and predicted BPH probabilities ranged from 0.00 to 0.65.

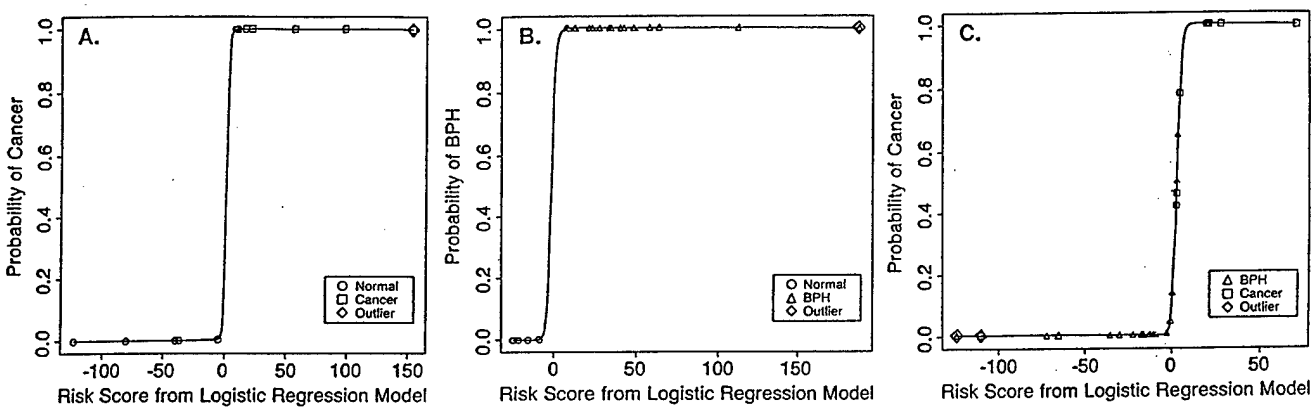

Fig. 4. Sigmoid curves depicting the probability of DNA being classified as normal tissue vs. cancer (A), normal tissue versus BPH (B), and BPH versus cancer (C). The curves are based on the logistic regression models depicted in Table 3. The predicted probabilities rise very rapidly over a narrow range, which reflects a high degree of discrimination among groups and a precipitous change in DNA structure associated with the normal-BPH and normal cancer progressions. Each sample is plotted at its predicted probably.

The two outliers omitted from the analyses tend to support the findings. The outlier BPH and cancer points lie to the right in the PC plot (Fig. 2). When the models shown in Table 3 were used to classify the two outliers, the BPH outlier was correctly classified, using the normal vs. BPH model, with a predicted BPH probability close to 1.0. The cancer outlier is also correctly classified in the normal vs. cancer model with a predicted cancer probability close to 1.0. In the BPH vs. cancer model, the BPH outlier is correctly classified with a predicted cancer probability close to

zero; however, the cancer outlier is incorrectly classified as a BPH with a cancer probability close to zero. Overall, the findings suggest that the •OH likely plays a major role in the transformation of normal prostate tissue to the cancer state, which is consistent with its proposed role in breast cancer (1,5,11,12,16,17).

Age and Gleason Score relationships. Age does not appear to be a factor in creating the pronounced distinctions among groups, although the incidence of prostate cancer increases significantly over the age of 50 years (18). The age ranges for the three groups were 16 - 73 years for normal (n = 5); BPH, 58 - 73 (n = 17); and cancer, 61 - 76 (n = 7). Among the Spearman correlations of age with each of the four PC scores, none were statistically significant. In all, 28 correlations were considered, consisting of age correlated with each PC score in each of the three groups, as well as in all pairs of groups (e.g., age correlated with each PC scores in normal and BPH tissue combined) and in the entire pooled set of 29 cases. Spearman correlations ranged in magnitude from 0.01 to 0.59 with P = 0.09 to P = 1.0. The most significant correlation was r = -0.51 between age and PC4 in the combined normal and cancer groups (P = 0.09). When PC4 was omitted from the logistic regression analysis and models were based on PC1 - PC3, the P-values corresponding to those in Table 3 were, top to bottom, $P < 0.001$, $P < 0.001$ and P = 0.005, again supporting a non-random distinction among the groups. These results based on PC4 and the weak or nonsignificant correlations between age and other PC scores do not support any role for age in the ability to use spectra to distinguish among the groups.

The Ovary and Breast: Development of a Unified Theory of Carcinogenesis Based on Order-Disorder Transitions in DNA Structure (7). Change in mean distance from the centroid (diversity) and/or change in mean spectra (PC location) are both measures of alterations in the order-disorder status of cellular DNA. As an example, differences between mean spectra of the normal ovary (O_n) and primary adenocarcinoma (AC) groups are illustrated in Fig. 5A; $P<0.05$ delineates wavenumber regions in which the more significant differences exist (Fig. 5B). These differences also are consistent with the substantial change in centroid location between the O_n and AC groups (Table 4). A significant change was not found between the mean spectra of the AC and metastasized primary adenocarcinoma (AC_m) groups (Fig. 5C); however, a significant spectral change was evident in the transition from the AC_m to the distant metastases AC_{dm}. There was not a significant difference between the mean spectra of the O_n and the AC_{dm}. The O_n was a relatively tight group, whereas the AC was highly diverse. The diversities of the AC and AC_m groups were similar, and the AC_{dm} was a substantially tighter cluster than the relatively diverse AC_m group. Table 4 shows that the O_n and the AC_{dm} had similar diversities and comparable mean spectra. However, Fig. 5D-E shows that the O_n and the AC_{dm} groups had substantially different patterns of diversity (different standard deviations; Fig. 5D) at a number of wavelengths, particularly in the left area of the spectrum (base vibrations above $\approx 1315\ cm^{-1}$). Many of these differences in diversity yielded $P<0.05$ (Fig. 5E). The null hypothesis that the two groups have the same diversity pattern (identical wavenumber-by-wavenumber standard deviations across the spectrum) is rejected with $P = 0.02$, based on the covariance matrices for PC scores 2-6.

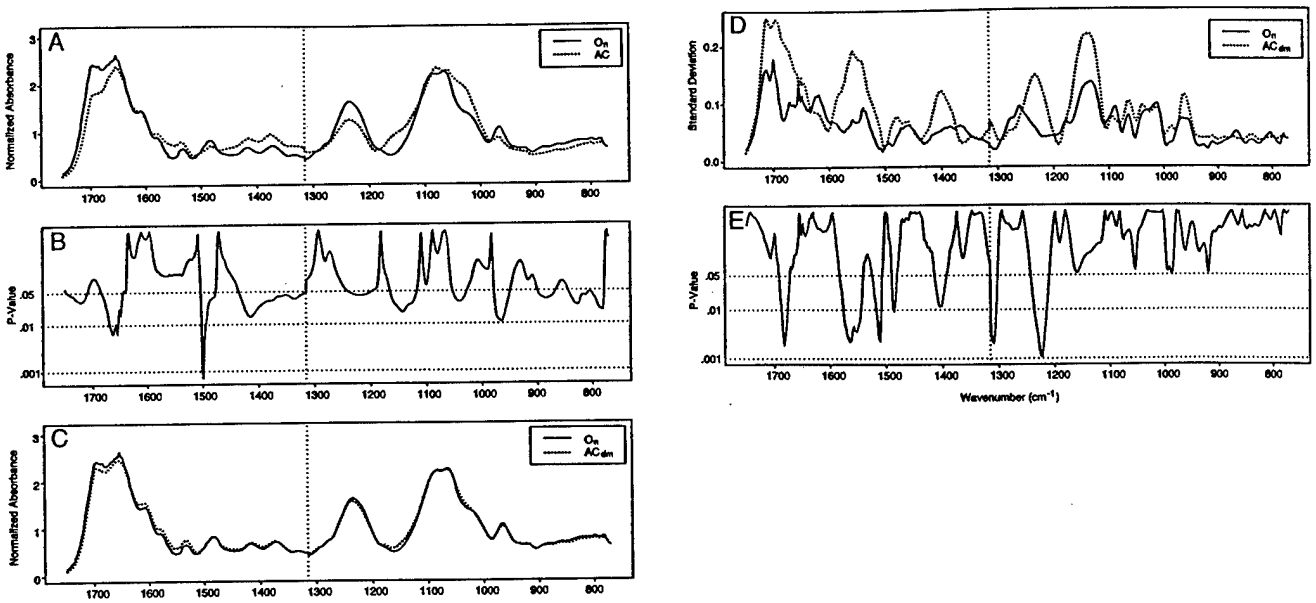

Figure 5. DNA spectral comparisons: (*A*) grand mean spectra of morphologically normal ovarian tissue (O_n) and primary ovarian adenocarcinoma (AC); (*B*) P-values for spectral comparison in (A); (*C*) grand mean spectra of O_n with AC metastases to colon (AC_{dm}); (*D*) standard deviations between spectral comparisons in (*C*); and (*E*) P-values for standard deviation comparisons for (*D*).

Table 4. Order-disorder comparisons between DNA structures in the neoplastic tranformation of morphologically normal breast and ovarian tissues.

Groups compared		Difference between grand mean spectra as percentage		Diversity: mean difference from group mean spectrum as percentage		
1	2	Percentage	*P* value*	Group 1: mean ± SD	Group 2: mean ± SD	*P* value†
Ovary						
O_n	AC	23.5	<0.002	8 ± 3	20 ± 8	0.001
AC	AC_m	6.2	0.7	20 ± 8	16 ± 6	0.4
AC_m	AC_{dm}	16.1	0.02	16 ± 6	9 ± 4	0.008
O_n	AC_{dm}	6.2	0.1	8 ± 3	9 ± 4	0.4
Breast						
RMT	IDC	9.3	<0.002	10 ± 5	8 ± 3	0.4
IDC	IDC_m	7.0	0.09	8 ± 3	13 ± 5	0.003
IDC_m	IDC_{dm}	16.1	0.002	13 ± 5	10 ± 4	0.1
RMT	IDC_{dm}	16.3	<0.002	10 ± 5	10 ± 4	0.9

*One sided *P* values based on permutation test.
†Two sided *P* values based on Mann–Whitney test.

The differences in mean spectra and diversity (Table 3) of the $O_n \rightarrow AC$, $AC \rightarrow AC_m$ and $AC_m \rightarrow AC_{dm}$ transitions are graphically illustrated in PC plots using the second and third PC scores (Fig. 6). The differences in diversity and locations of the O_n and AC clusters are evident in Fig. 5A in which the AC cluster is more diverse and shifted to the left of the O_n. In Fig. 6B, the AC and AC_m clusters occupy about the same PC location and are equally diverse, reflecting their similar mean spectra and diversity (Table 4). Fig. 6C shows that the AC_m and the AC_{dm} differ both in location

and diversity. In Fig. 6D, the AC_{dm} and O_n samples overlap considerably and are about equal in size, each representing a tight cluster. However, as indicated, the two groups differ in their standard deviations at certain wavenumbers (Fig. 6D-E). That is, disorder in different DNA structures (as represented by the spectral properties) distinguishes the two groups.

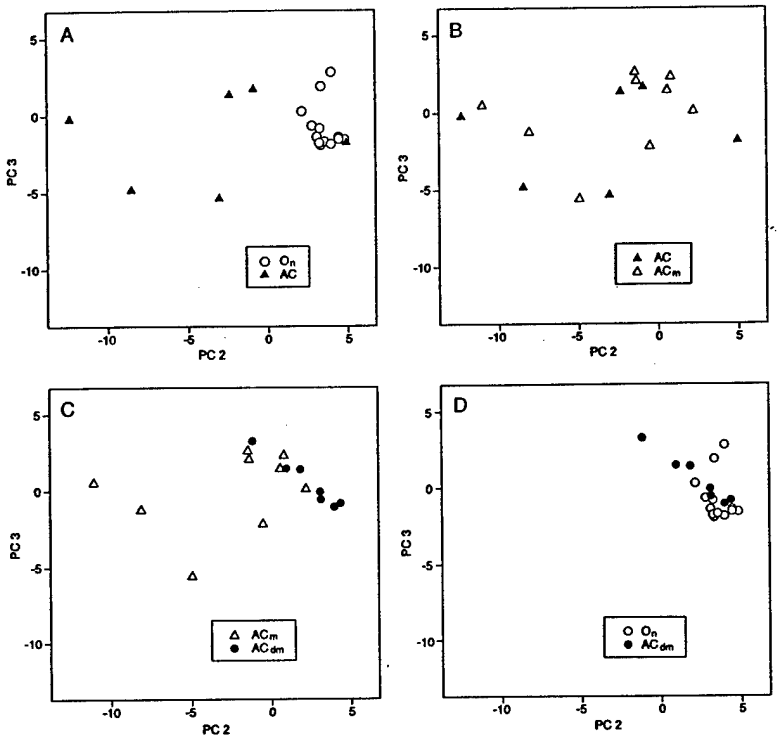

Figure 6. PC plots comparing spectra of ovarian DNAs from (*A*) morphologically normal tissue (O_n) and primary adenocarcinoma (AC); (*B*) AC and metastasized primary AC (AC_m); (*C*) AC_m with AC metastases to the colon (AC_{dm}); and (*D*) a comparison of O_n with AC_{dm}. See text and Table 2 for statistical comparisons of order-disorder status between groups.

Cluster analysis showed that the various stages of ovarian tumor progression comprise a mixture of sub-groups (i.e., disorder mixed with relative order). Figure 7 shows a cluster analysis of FT-IR spectra depicting the Euclidean distance, expressed as a percentage, between each spectrum and its "nearest neighbor." The O_n shows a fairly tight cluster with no nearest neighbor distances beyond about 10% (Fig. 7A). The AC cluster (Fig. 7B) shows a wide range between spectra, with some as close as about 6% and some as distant as 30%. The AC_m cluster (Fig. 7C) appears to be a mixture between a tight sub-group of DNAs (lower in the panel) that have no more than a 10% nearest

neighbor distance and a second relatively diverse sub-group (higher in the panel) with 18-19% nearest neighbor distance. All spectra in the second sub-group are at least 25% distant from spectra in the first sub-group. An individual spectrum (sample 72) appears at an intermediate distance from the two sub-groups. The AC_{dm} group (Fig. 7D) is relatively tight, with nearest neighbor distances not exceeding 14%.

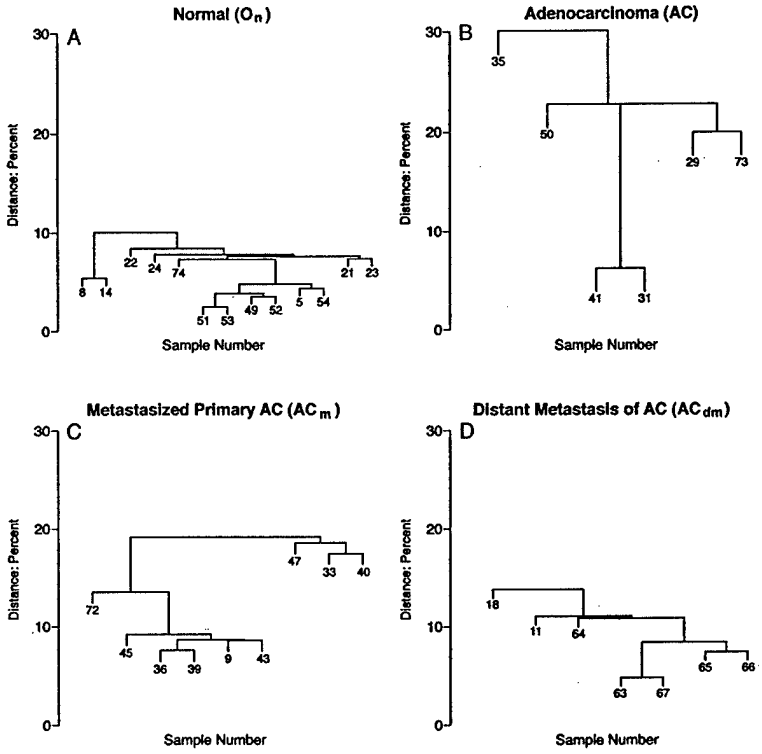

Figure 7. Cluster analysis of spectra of ovarian DNAs. This analysis is based on the distance of each sample to its nearest neighbor. The y-axis shows the percent difference between spectra (e.g., 31 is ≈6% different from 41 in panel *B*).

The breast samples also show substantial differences between groups both in PC cluster diversity and mean spectra, although the range of differences in mean spectra between groups and the range of diversities are not as great as among the ovary samples (Table 4). The differences between groups range from 9-16% for the breast samples (compared to 6%-24% for the ovary samples), and the mean distance from the centroid varies from 8%-13% (compared to 8%-20% for the ovary). The transition RMT → IDC is characterized by a moderate, but highly significant,

difference between mean spectra; however, there was not a significant change in the mean difference from the centroid. The transition IDC → IDC_m shows a marginally significant mean change, but a notably significant change in diversity. The difference in the mean spectra of IDC_m → IDC_{dm} is also significant, with no significant change in diversity. The IDC_{dm}, the terminal stage in the sequence of transitions, has a significantly different mean spectrum from the RMT with very little difference in diversity. Comparisons between the RMT, IDC_m and IDC_{dm} clusters are shown in the PC plots of Figure 8. Figure 8A shows a substantial overlap between the IDC_m and the IDC_{dm} clusters; however, the IDC_{dm} cluster is more compact, representing a more ordered state as reflected by its smaller mean distance to the centroid. Figure 8B shows little overlap between the RMT and IDC_{dm} clusters, indicating different mean spectra.

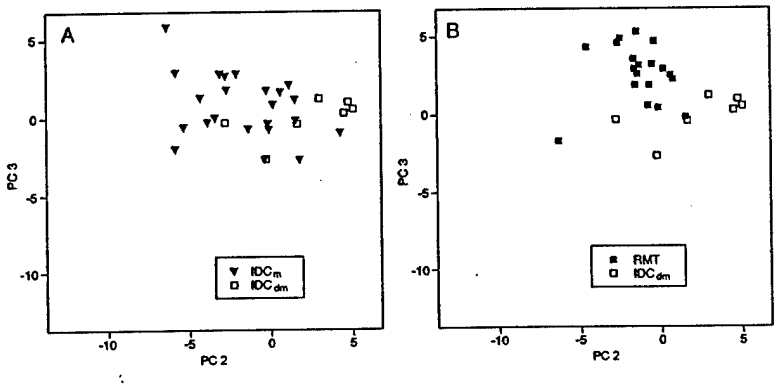

Figure 8. PC plots comparing spectra of human breast DNAs from (*A*) metastasized primary adenocarcinoma (IDC_m) and IDC metastases to axillary nodes (IDC_m); and (*B*) morphologically normal reduction mammoplasty tissue (RMT) and IDC_{dm}. See text and Table 2 for statistical comparisons of order-disorder status between groups.

The sequence of ovary and breast DNA transitions is graphically illustrated in the PC plot of Fig. 9, which depicts the centroids of each cluster. The ovary DNAs show a substantial "leap" from the centroid of the O_n group (in the "order" region) to the AC centroid, a short step back to the AC_m centroid, and finally a shift to the AC_{dm} centroid located close to that of the O_n (6% distance). The breast centroids proceed along a different path, but ultimately converge on the order region, as

occurs with the ovary (Fig. 9). The RMT, IDC, and IDC_m centroids are located in the "disorder" region. The final stage of the progression, represented by the IDC_{dm} centroid, is close to that of the O_n, the AC_{dm} and the HBL centroids. Also included is a hypothetical normal tissue DNA (HNT) centroid that is the mean of the O_n, AC_{dm}, IDC_{dm} and HBL centroids. The centroid of the HNT is intended to serve as a reference point and speculative origin for essentially unmodified normal breast DNAs.

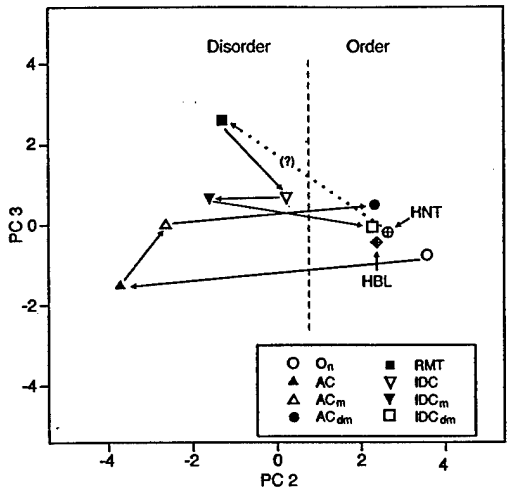

Figure 9. Tumor progression pathways are depicted in a PC plot of human ovarian and breast DNA centroids (derived from groups of spectra). The centroid representing the DNAs of a hypothetical normal tissue (HNT) and human blood leukocytes (HBL) are also included (see text for details). The vertical line (⋮) broadly distinguishes the centroids of the relatively ordered and disordered groups.

Comparisons of changes in the base (1750-1315 cm^{-1}) and phosphodiester-deoxyribose (1314-770 cm^{-1}) regions are given in Table 5, plus the sum of transitions for both of these spectral regions in relation to the breast and ovary DNAs. On the basis of the percentage change in spectra between states, the total cancer progression involves remarkably large structural changes in DNA, as indicated by the total path length of 46% for the ovarian base region and 39% for the corresponding phosphodiester-deoxyribose region (Table 5). Comparable results for the breast were also substantial: 42% for the base region and 21% for the phosphodiester-deoxyribose region. Considering that the RMT is reported to be significantly modified (1,11), the path length between

the possible starting point of the breast cancer process (the HNT centroid) and the RMT centroid (Fig. 9) contributes substantially to the total path distance of the breast. This distance was 25% for the base region and 9% for the phosphodiester-deoxyribose region. The total path length, if the HNT to RMT path were included, would be 67% for the base region and 30% for the phosphodiester-deoxyribose region. Patient age appears to play a negligible role in determining spectral differences between the ovarian and breast groups. This result was established on the basis of regression analyses of age in relation to PC scores.

Table 5. Length of centroid-to-centroid transitions pathways, as percent, for ovary and breast carcinogenesis

	1,750 to 1,315 cm^{-1} Base region		1,314 to 770 cm^{-1} Deoxyribose region		Difference Base–deoxyribose regions		
Transition	Distance, %	SE*	Distance, %	SE*	Difference	SE*	*P* value*
Ovary							
O_n to AC	24.1	8.3	20.8†	6.5	3.3	4.6	0.5
AC to AC_m	3.9	5.3	4.4	4.7	−0.5	4.4	1.0
AC_m to AC_{dm}	17.8	5.7	13.3	4.2	4.5	3.5	0.2
Total path	45.8	11.7	38.5	10.4	7.3	7.5	0.2
Breast							
RMT to IDC	11.9	2.2	6.2†	1.3	5.7	1.8	<0.001
IDC to IDC_m	8.7	3.0	4.4	1.5	4.3	2.0	0.02
IDC_m to IDC_{dm}	21.2	4.4	10.0	1.9	11.2	3.0	0.01
Total path	41.8	5.7	20.6	2.7	21.2	4.9	<0.001

*From resampling.
†P = 0.03 for ovary transition distance compared with corresponding breast transition distance; based on resampling.

Medaka.

Clustering. The lack of outliers and the obvious clustering of samples is shown in Fig. 10[1], where the Y-axis now represents the distance of each sample to its nearest neighbor, expressed as a percentage. The samples are labeled by their time group (1 = earliest, six weeks; 4 = latest, 9 months.) Of note, there is some tendency for samples to cluster together by time, particularly for time 2, where all samples appear in one particular branch of the cluster tree.

[1] Figures 10-19 are in addendum (pages 40-50)

Figure 11 shows cluster tree labeling by treatment group (1 = control, 2 = low dose, 3 = high dose), and there is a tendency for treatment 2 samples (low) to appear lower in the tree, indicating that they are more likely to occur mixed with other samples, rather than isolated samples, and treatments 1 and 3 supply most of the large distances (top left of exhibit), suggesting that they tend to occur in isolation. There was no obvious or strong tendency for particular tanks of fish (samples) to occur together in the spectral plots, as shown in Figure 12.

Classification of Samples. We carried out a stepwise discriminant analysis using the first 15 PC scores as independent variables and, in separate analyses, time or treatment as the grouping variables. In the treatment analysis, none of the PC scores could be selected into the discrimination model with $p < 0.05$. This implies that the FT-IR spectra do not discriminate among these treatment groups. Only by accepting non-significant ($p > 0.05$) PC scores could we develop a discrimination model, presented here only for the purpose of graphical display. Figures 13 and 14 show the samples labeled by treatment group and plotted by first and second (non-significant) discriminant scores. Each discriminant score is a linear combination of PC scores. The overlap of the treatment groups in these plots is obvious, and the centroids of the groups (here defined as the mean values of the discriminant scores in the group) are very close together.

In contrast, time has a *dramatic effect* on DNA spectra as shown by discriminant analysis. The group separations are shown in Figures 15-17, the same type of scatterplots used earlier. In the two-dimensional plots, there is some overlap between groups 3 and 4, otherwise, the groups are quite distinct. The overlap is likely to be even smaller than shown in the displays, given that the discrimination process yielded six dimensions: discrimination is based on PC scores 2, 4, 5, 6, 8

and 9, the first six PC scores chosen in a stepwise analysis. Additional PC scores, through statistically significantly related to time groups, added little to the discrimination among groups.

The results of the discriminant analysis classification are shown in Table 6, where 78% of the cases were correctly classified (62/80), and 74% (59/80) were correctly classified by cross-validation, omitting one case at a time. (The cross-validation is somewhat non-conservative, because the same six PC scores from a single stepwise analysis are used in the cross-validation, rather than using potentially different sets of PC scores, depending on the omitted case at each cross-validation step.)

Table 6. Discriminant analysis classification of Medaka findings.

			Predicted Group Membership				
		TIME	1.00	2.00	3.00	4.00	Total
Original	Count	1.00	9	0	0	0	9
		2.00	0	15	0	1	16
		3.00	0	0	19	8	27
		4.00	0	2	6	20	28
	%	1.00	100.0	.0	.0	.0	100.0
		2.00	.0	93.8	.0	6.3	100.0
		3.00	.0	.0	70.4	29.6	100.0
		4.00	.0	7.1	21.4	71.4	100.0
Cross-validated[a]	Count	1.00	9	0	0	0	9
		2.00	0	15	0	1	16
		3.00	0	2	15	10	27
		4.00	0	3	7	18	28
	%	1.00	100.0	.0	.0	.0	100.0
		2.00	.0	93.8	.0	6.3	100.0
		3.00	.0	7.4	55.6	37.0	100.0
		4.00	.0	10.7	25.0	64.3	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 78.8% of original grouped cases correctly classified.

c. 71.3% of cross-validated grouped cases correctly classified.

Diversity. Except for the three-month time point, the treatment groups have comparable diversity (defined as the mean distance to the common group centroid), as shown in Table 7. The three-

month group is significantly different in its diversity from all other time groups, and this is the only significant difference in diversity among the group comparisons. A global test for equality of covariance matrices for these six PC scores yields $p < 0.001$, leading us to reject the null hypothesis that all of the groups have the same size and shape of clusters. Though not tested, it is likely that times 1, 3, and 4 do not all have the same shape (though they have about the same size as shown in Table 7), based on the orientation of clusters in Figure 17.

Table 7. 80 unique Medaka samples (6 outliers removed): Diversity of time groups.

Diversity: Mean Difference from Group Mean Spectrum as percentage							
Group 1	Group 2	Mean Group 1	SD Group 1	Mean Group 2	SD Group 2	T Test P-Value	Mann-Whitney Test P-Value
6 Weeks	3 Months	24.1	6.9	11.5	5.7	0.0003	0.0002
6 Weeks	6 Months	24.1	6.9	29.2	11.7	0.1243	0.2184
6 Weeks	9 Months	24.1	6.9	28.2	13.5	0.2418	0.7149
3 Months	6 Months	11.5	5.7	29.2	11.7	<0.0001	<0.0001
3 Months	9 Months	11.5	5.7	28.2	13.5	<0.0001	<0.0001
6 Months	9 Months	29.2	11.7	28.2	13.5	0.768	0.5756

Spectral Regions. Finally, the spectral differences among the groups occur in several regions, as shown in the "p-value plots" of Figures 18-19. The upper panel shows the mean spectrum for each time group and the lower pane shows the p-value (per wavenumber) for the null hypothesis that all groups have the same mean normalized absorbance at the specific wavenumber, based on ANOVA (Fig.18) or the nonparametric Kruskal-Wallis test (Fig. 19). The plot shows very small p-values for a number of regions, particularly around 1680-1490 (mostly base vibrations), 1290 and 1170-990 cm^{-1} (regions identified by eye; mostly phosphodiester – deoxyribose vibrations) It is also obvious from the plots that time 2 (3 months) differs strikingly from the other groups. In Figures

15-17 and Table 6, this group also shows a different orientation than the other groups, and a different size (Table 7.)

III. CONCLUSION AND IMPLICATIONS

Prostate Studies. Overall, the prostate findings indicate that DNA is altered in ways that produce clustering and, consequently, discrimination between normal prostate, BPH and prostate cancer DNA (Fig. 2; Tables 2 and 3). The •OH is known to produce mutagenic base lesions, such as 8-OH-Gua and 8-OH-Ade (1,5,11,12,16,17,19-23), and also cause damage to deoxyribose by abstracting hydrogens from one or more positions associated with the furanose ring (24). These events can ultimately lead to broadly based genomic instability and strand breaks (25). Recent evidence also indicates that decreased antioxidant levels and increased base modifications occur in BPH tissue compared with adjacent normal prostate (26). Moreover, FT-IR spectral analysis of calf thymus and normal breast DNA exposed for various times to •OH-generating systems (Fe^{++}/H_2O_2) revealed substantial alterations in areas of the spectrum assigned to vibrations of the nucleic acids and the phosphodiester-deoxyribose moiety (D.C.M., S.J.G., and J. Cramer, unpublished results). Collectively, this evidence supports the proposition that the •OH is intimately involved in altering the structure of DNA, thus contributing to clustering and discrimination between clusters; however, it is recognized that other factors (e.g., hypermethylation) (27) may also contribute to these alterations. The prostate findings closely resemble those obtained with the female breast (11,12,16) in which the cancer-related •OH-modification of DNA was termed radical-induced DNA disorder (RIDD) (12). RIDD also appears to be significant in the etiology of BPH and prostate cancer and constitutes a formidable barrier to overcome in cancer prevention and treatment. The findings with the normal prostate and primary

prostate cancer are consistent with those obtained for the breast, although additional studies are necessary to determine whether DNA from metastatic prostate cancer cells will show the increased diversity found with metastatic breast cancer. Increased structural diversity generated in primary tumors is likely an important factor in selecting malignant DNA forms that potentially give rise to malignant cell populations, as previously suggested (12).

The Gleason score, which uses microscopically evinced architectural changes to classify tumor status (18), had little association with the prostate DNA PC scores, although based on the n = 7 cancer cases, there was limited power to detect other than strong associations. Spearman correlations of PC scores 1 - 4 with the Gleason score ranged from - 0.49 to + 0.26, with P = 0.2 to 0.8. Further studies will be required to establish whether the DNA alterations are correlated with changes at the cellular level; however, both the Gleason and PC scores reflect complex suites of underlying biological changes that are not easily identified.

BPH is not known to be etiologically related to prostate cancer; however, it is of interest that the BPH vs. prostate cancer curve (Fig. 4C) shows several cases having intermediate probabilities. The configuration of cases in Fig. 2 also provides some insight into the controversial view that BPH is a direct precursor of prostate cancer (18). The findings do not support this concept in that the BPH group lies "beyond" the cancer group, starting from the normal group. This positioning suggests that a transition from BPH to cancer would involve a reversal of some of the spectral transitions shown to be associated with cancer, or that there are additional changes to the BPH DNA that mimic a reversal in the progression to cancer. Alternatively, modifications may result in DNA structures that lead to a variety of nonneoplastic lesions, including BPH. Support for this concept comes from studies of English sole exposed to environmental chemicals (9) which showed

significant correlations between 8-OH-Ade and 8-OH-Gua and five nonneoplastic hepatic lesions (including the putatively preneoplastic lesion, basophilic foci). The results suggested that the •OH is likely a common factor in the etiology of both the modified bases and the nonneoplastic lesions. Although BPH may not be a direct precursor of prostate cancer, FT-IR/PCA spectral analysis may provide a promising means of predicting the occurrence of prostate cancer, based on the structural status of BPH DNA.

The absence of transition states in the normal to cancer and normal to BPH curves is of interest. This is likely due to the fact that "transition" tissues having DNA values between zero and 100% probability (Fig. 4A-C) were not part of this study. Clearly, additional research with a larger number of samples is necessary to obtain information on the ability of the PCA/FT-IR technology to detect transition states associated with the normal to BPH and normal to prostate cancer progressions. Additional studies also seem warranted to test the important hypothesis that the PCA/FT-IR spectral analysis of DNA from prostate tissue ($\approx 20\mu g$ is required) will provide a sensitive means for screening and predicting prostate cancer.

Evidence with the prostate, suggests that DNA structure is progressively altered in response to factors in the microenvironment, notably •OH concentrations, that are likely etiologically related to the development of prostate tumors (adenocarcinoma) and BPH. It is suggested that intervention to forestall or correct the genetic instability of these tissues and likely increase in cancer risk should focus on controlling cellular redox status and •OH concentrations. The approaches may include control of the iron-catalyzed conversion of H_2O_2 to the •OH (28); regulation of •OH production through redox cycling of hormones (29) and environmental xenobiotics (30); and antioxidant/reluctant therapy (31,32).

The Ovary and Breast: Development of a Unified Theory of Carcinogenesis Based on Order-Disorder Transitions in DNA Structure. In previous studies (10-12,16), FT-IR/statistics models provided the first evidence showing that cellular transformations in breast tumor formation (e.g., RMT $\rightarrow$ IDC $\rightarrow$ IDC_m) involve order-disorder transitions in DNA structure. As described above, prior studies using these models (6), showed that the transformation of morphologically normal prostate (normal $\rightarrow$ adenocarcinoma) and (normal $\rightarrow$ BPH) also produces discernible changes in the order-disorder status of DNA. These initial studies raised the important question of whether changes in DNA, as determined using FT-IR/statistics, represent critical events on which cancer progression depends in order to reach the stage of distant metastases.

In the ovary, the transition $O_n \rightarrow$ AC represents a major change in the DNAs from a relatively ordered to a substantially disordered state (Table 4; Fig. 6A and Fig. 9). Pronounced alterations in areas of the spectra assigned to both base and phosphodiester-deoxyribose structures (Table 5) reflected the global nature of these alterations. The order-disorder status was virtually unchanged in the transition AC$\rightarrow$$AC_m$ (Table 4; Fig. 6B); however, the transition to the AC_{dm} resulted in a major change toward the reinstatement of order comparable to that of the O_n, as indicated by the differences in mean spectra and cluster diversities (Table 4; Fig. 5C-D). The data on standard deviations of spectra (Fig. 5D-E) further demonstrated that, despite the comparable mean spectra of the O_n and AC_{dm}, differences exist between these groups in vibrations associated with the base and phosphodiester structures. This is consistent with the presence of abundant mutations that characterize metastases (33). Moreover, the highly significant difference in the PO_2^- structure ($\approx$ 1250 cm^{-1}) (Fig. 5E) likely arose from alterations in base pairing, which would be expected to disrupt the arrangement of the phosphate groups along the DNA backbone, thus altering the

vibrational properties of the PO_2^- group. A comparable analysis of the breast data was not appropriate since the RMT samples were disordered.

Cluster analysis (Fig. 7) provided additional insight into the nature of the changes in the ovarian DNAs showing, for example, that the disordered AC (Fig. 7B) and AC_m (Fig. 7C) each comprise a mixture of sub-groups. Of interest is the appearance of a sub-group within the AC_m (samples 47, 33 and 40) that may represent remnants of the AC group, and another sub-group (samples 9-72) that exhibits a relatively ordered state similar to that of the AC_{dm} (Fig. 7D). These data, together with those in Fig. 5E, support the hypothesis that there is a selection of ordered, mutated DNAs for the next stage of cancer progression (AC_{dm}) arising from the pronounced degree of disorder found in the AC_m. The magnitudes of the order-disorder transitions in the ovarian DNAs are substantial, as indicated by the path length data (Table 5): 46% for the base region and 39% for the phosphodiester-deoxyribose region. We suggest that the great number of different DNAs produced in these transitions provide a pool from which viable molecular structures can be selected, consistent with the ultimate attainment of metastases.

In the breast, the creation of disordered DNAs in the transitions RMT $\rightarrow$ IDC $\rightarrow$ IDC_m was reported previously (10-12,16). The inclusion of data on the IDC_{dm} in the present study afforded the opportunity to explore the nature of tumor progression in the breast to the stage of axillary node metastases. The disorder in the RMT (1,6,11) contrasts with the relatively ordered status of the O_n, (Table 4; Fig. 9). The magnitude of RMT disorder is substantial based on the path length between this group and the HNT. The 19 RMT samples analyzed had a location distinct from that of the ordered forms, such as the O_n and the HBL, whose centroids are shown on the right side of Figure 9. A possible explanation for the difference in disorder is that the morphologically normal

breast is under greater oxidative stress than the ovary (e.g., from •OH), notably due to estrogen metabolism (2,33,34). Previous studies have shown that substantial base modifications exist in the normal breast DNAs (1,11), reaching as high as one base modification in 10^3 normal bases (2). We are unaware of comparable data on the ovary.

The RMT → IDC transition involves structural changes (disorder) as reflected in a significant distance between the centroids (Table 4; Fig. 9), without a significant change in cluster diversity. Order-disorder transitions of this type may be mostly intramolecular, involving vertical base residue stacking interactions, for example, that are known to produce significant changes in DNA spectra (14,15). The IDC → IDC_m transition involves a substantial increase in diversity (in contrast to the AC → AC_m transition) (Table 4). The IDC_m → IDC_{dm} transition showed a major shift toward order (Table 4), as shown by the fact that the IDC_{dm} cluster was spatially close to that of the HBL, O_n and the AC_{dm} clusters (Fig. 9). The diversities of the RMT and IDC_{dm} clusters are similar; however, they have different PC locations (Table 4; Fig. 9). The initially formed IDC would be expected to be relatively ordered, prior to being progressively damaged by micro-environmental factors, such as •OH, that may be produced from H_2O_2 reported to be "constitutively" generated in primary cancer cells (36). In the developing tumor, the damaged forms of DNA would obscure the detection of the initially formed DNA structures. In this context, the progression of morphologically normal breast tissue to distant metastases may not be fundamentally different from that of the comparable ovarian progression, assuming that the disordered RMT was produced from ordered DNAs (e.g., HNT) at some earlier stage in life, possibly shortly after puberty. We recognize the possibility that ordered breast DNAs may exist in certain human populations, notably those from Asia that have a low incidence of breast cancer

(37). The virtual lack of relationship between patient age and the present results is consistent with previous studies of radical-induced changes in DNA of human tissues (12,16,38,39).

The transition from disorder in primary tumors, whether or not they had metastasized, to order in distant metastases may involve a significant change in the cellular redox status of the DNA. Prior studies of the IDC $\rightarrow$ IDC_m transition (12) suggested that a shift toward reductive conditions take place in metastasized primary breast tumors. The evidence was based on a change in the model $\log_{10}$ (Fapy Ade/8-OH-Ade) reflecting an increase in Fapy Ade as the size of the metastasized primary tumor increased [Fapy Ade is reported to be preferentially synthesized under reductive conditions (40)]. An additional factor consistent with this apparent shift in redox status is the reported development of hypoxia in transformed tissues (41). The proposed shift toward reductive conditions in the metastasized tumor cells would be expected to suppress the progression of oxidative DNA damage, thus helping to preserve (stabilize) DNA structures that ultimately become part of the ordered IDC_{dm} group.

The vertical transfer of electrons from base to base along the helix has been reported to extend to 25 base pairs so that a structural change at one point would likely trigger structural changes far afield (40). Recent evidence for the long-range oxidative repair of thymine dimers further demonstrates this unique property of DNA (42). The characteristics of DNA raise the distinct possibility that the *overall structure* of some forms of DNA (e.g., resulting from disrupted base stacking) in a disordered system alter protein expression and function well beyond changes associated with the coded information inherent in the linear sequence of bases.

The creation of disorder in DNA out of a relatively ordered system and the ultimate restoration of order may be regarded as a prime example of chaos theory (43). A salient feature of complex biological systems is that chaos created at one level of activity can give rise to order at another level: that is, order arises out of chaos and certain dynamic factors are responsible for its emergence (deterministic chaos) (43). In most complex biological systems, the dynamic processes are elusive; however, several factors may be influential in the present order-disorder transitions. These include the reported preferential attack of the •OH on the base structures compared to the attack on deoxyribose (yielding DNA forms with mutated bases and intact deoxyribose moieties) (24) and the preference shown in DNA polymerization for intact substrates (44). Regardless of the processes involved, it is reasonable to assume that the creation of disorder, prior to the attainment of order in the DNAs of metastases, is pivotal in tumor development. We find no inconsistency between prior findings relating mutations in growth controlling genes, such as proto-oncogenes and tumor suppressor genes, to carcinogenesis (45) because the creation of disorder in DNA would be expected to lead to a large number of genetic changes that would increase cancer risk.

The disruption of the disordered status of DNA through intervention is an attractive possibility for reducing cancer risk. This might be accomplished using therapeutic agents that reduce cellular •OH concentrations, or through diets rich in antioxidants (46). Alternatively, the possibility exists to increase the severity of DNA damage in tumor tissues by using DNA-cleaving molecules having selective anti-cancer activity (46,47).

Medaka Studies. These studies show pronounced statistical differences based on time of exposure and obvious differences in spectra were found between groups. This suggests that the

FT-IR/statistics technology is a potentially sensitive tool for assessing changes in DNA structure in laboratory and field studies involving the exposure of organisms to environmental pollution.

OVERALL ACCOMPLISHMENTS

During the three years prior to the termination of this proposed 5-year project, two major papers were published in PNAS: [Malins, et al., Proc. Natl. Acad. Sci. USA (1997) 94, 259-264 and Proc. Natl. Acad. Sci. USA (1998) 95, 7637-7642]. Also, a review of the potential of the FT-IR/statistics technology for biology and medicine was published, as requested by the editors of Nature Medicine [Malins, et al., Nat. Med. (1997) 3, 927-930]. Overall, we believe that the FT-IR/statistics technology developed under USARBDL sponsorship will potentially have broad application to understanding, diagnosing and predicting diseases, such as cancer Alzheimer's disease, diabetes mellitus, heart disease, Parkinson's disease, other neurodegenerative disorder, infertility, radiation effects, aging, pharmacokinetic, evaluations of drugs, genetic alterations in cultured cells and the effects of environmental contaminants on terrestrial and aquatic animals.

IV. References

1. Malins, D. C., Holmes, E. H., Polissar, N. L. & Gunselman, S. J. (1993) *Cancer* **71**, 3036-3043.
2. Liehr, J. G. (1997) *Environ. Health Perspect.* **105**, 565-569.
3. Cavalieri, E. L., Stack, D. E., Devanesan, P. D., Todorovic, R., Dwivedy, I., Higginbotham, S., R. L. & Rogan, E. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10937-10942. Johansson, S. L., Patil, K. D., Gross, M. L., Gooden, J. K., Ramanathan, R., Cerny.
4. Frenkel, K., Wei, L. & Wei H. (1995) *Free Radical. Biol. Med.* **19**, 373-380.

5. Malins, D. C. & Haiminot, R. (1991) *Cancer Res.* **51**, 5430-5432.

6. Malins, D. C., Polissar, N. L. & Gunselman, S. J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 259-264.

7. Malins, D.C., Polissar, N.L., Schaefer, S., Su, Y. & Vinson, M. (1998) *Proc Natl. Acad. Sci. USA* **95**, 7637-7642.

8. Malins, D.C., Polissar, N.L., Schaefer, S., Su, Y. & Vinson, M. (1999) *J. Natl Cancer Inst.* In Press.

9. Malins, D. C., Polissar, N. L. & Gunselman, S. J. (1997) *Proc. Natl. Acad. Sci. USA* ***94***, 3611-3615.

10. Malins, D. C., Polissar, N. L., Su, Y., Gardner, H. S. & Gunselman, S. J. (1997) *Nature Medicine* **3**, 927-930.

11. Malins, D. C., Polissar, N. L., Nishikida, K., Holmes, E. H., Gardner, H. S. & Gunselman, S. J. (1995) *Cancer* **75**, 503-517.

12. Malins, D. C., Polissar, N. L. & Gunselman, S. J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 2557-2563.

13. Parker, F. S. (1983) In *Application of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry* (Plenum, New York), pp. 349-398.

14. Tsuboi, M. (1969) *Appl. Spectrosc. Rev.* **3**, 45-90.

15. Tsuboi, M. (1974): in *Basic Principles in Nucleic Acid Chemistry*, ed. Ts'o, P. O .P. (Academic, New York), pp. 399-452.

16. Malins, D. C., Polissar, N. L. & Gunselman, S. J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14047-14052.

17. Malins, D. C. (1993) *J. Toxicol. Environ. Health* **40**, 247-261.

18. Kirby, R. S., Christmas, T. J., & Brawer, M. (1996) in *Prostate Cancer*, (Mosby, London), pp. 1-170.

19. Halliwell, B. & Aruoma, O. I. (1991) *FEBS Lett.* **281**, 9-19.

20. Feig, D. I., Reid, T. M., & Loeb, L. A. (1994) *Cancer Res.* **54**, 1890s-1894s.

21. Kuchino, Y., Mori, F., Kasai, H., Inoue, H., Iwai, S., Miura, K., Ohtsuka, E., & Nishimura, S. (1987) *Nature* **327**, 77-79.

22. Kamiya, H., Miura, H., Murata-Kamiya, N., Ishikawa, H., Sakaguchi, T., Inoue, H., Sasaki, T., Masutani, C., Hanaoka, F., Nishimura, S., & Ohtsuka, E. (1995) *Nucleic Acids Res.* **23**, 2893-2899.

23. Cheng, K. C., Cahill, D. S., Kasai, H., Nishimura, S., & Loeb, L. A. (1992) *J. Biol. Chem.* **267**, 166-172.

24. von Sonntag, C., Hagen, U., Schon-Bopp, A. & Schulte-Frohlinde, D. (1981) *Adv. Radiat. Biol.* **9**, 110-142.

25. Morgan, W. F., Day, J. P., Kaplan, M. I., McGhee, E. M., & Limoli, C. L. (1996) *Radiat. Res.* **146**, 247-258.

26. Olinski, R., Zastawny, T. H., Foksinski, M., Barecki, A., & Dizdaroglu, M. (1995) *Free Radic. Biol. Med.* **18**, 807-813.

27. Isaacs, W. B., Bova, G. S., Morton, R. A., Bussemakers, M. J. G., Brooks, J. D., & Ewing, C. M. (1995) *Cancer Surv.* **23**, 19-32.

28. Imlay, J. A., Chin, S. M., & Linn, S. (1988) *Science* **240**, 640-642.

29. Han, X. & Liehr, J. G. (1995) *Carcinogenesis* **16**, 2571-2574.

30. Bagchi, D., Bagchi, M., Hassoun, E. A., & Stohs, S. J. (1995) *Toxicology* **104**, 129- 140.

31. Ames, B. N., Shigenaga, M. K. & Hagen, T. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7915-7922.

32. Bast, A., Haenen, G. R. M. M., & Doelman, C. J. A. (1991) *Am. J. Med.* **91**, Suppl. 3C, 2S-13S.

33. Fidler, I. J. & Nicolson, G. L. (1991) in *The Breast : Comprehensive Management of Benign and Malignant Diseases,* eds. Bland, K. I. & Copeland, E. M. III (Saunders, Philadelphia), pp. 262-291.

34. Yager, J. G. & Liehr, J.G. (1996) *Ann. Rev. Pharmacol. Toxicol.* **36**, 203-232.

35. Liehr J. G. (1997) *Eur. J. Cancer Prev.* **6**, 3-10.

36. O'Donnell-Tormey J., De Boer, C. J., & Nathan, C. F. (1985) *J. Clin. Invest.* **76**, 80-86.

37. Armstrong B. & Doll, R. (1975) *International J. Cancer* **15**, 617-631.

38. Musarrat, J., Arezina-Wilson, J. & Wani, A. A. (1996) *Eur. J. Cancer* **32A**, 1209-1214.

39. Sanchez-Ramos, J. R., Overvik, E. & Ames, B. N. (1994) *Neurodegeneration* **3**, 197-204.

40. Steenken, S. (1989) *Chem. Rev.* **89**, 503-520.

41. Höckel, M., Schlenger, K., Aral, B., Mitze, M., Schäffer, U. & Vaupel, P. (1996) *Cancer Res.* **56**, 4509-4515.

42. Dandliker, P.J., Homlin, R.E. & Barton, J.K. (1997) *Science* **275**, 1465-1468.

43. Kauffman, S. A. (1993): *The Origins of Order, Self-Organization and Selection in Evolution* (Oxford University Press, New York), pp. 33-67 & 175-182.

44. Joyce, C. M. (1997) *Proc. Natl. Acad. Sci. USA* **94** 1619-1622.

45. Cooper, G. M. (1995) *Oncogenes,* (Jones and Bartlett, Boston), pp. 67-177.

46. Schwartz, J. L. (1996) *Symposium: Prooxidant Effects of Anitoxidant Vitamins* (American Institute of Nutrition), pp. 1221S-1227S.

47. Hiramoto, K., Fujino, T. & Kikugawa, K. (1996) *Mutat. Res.* **360**, 95-100.

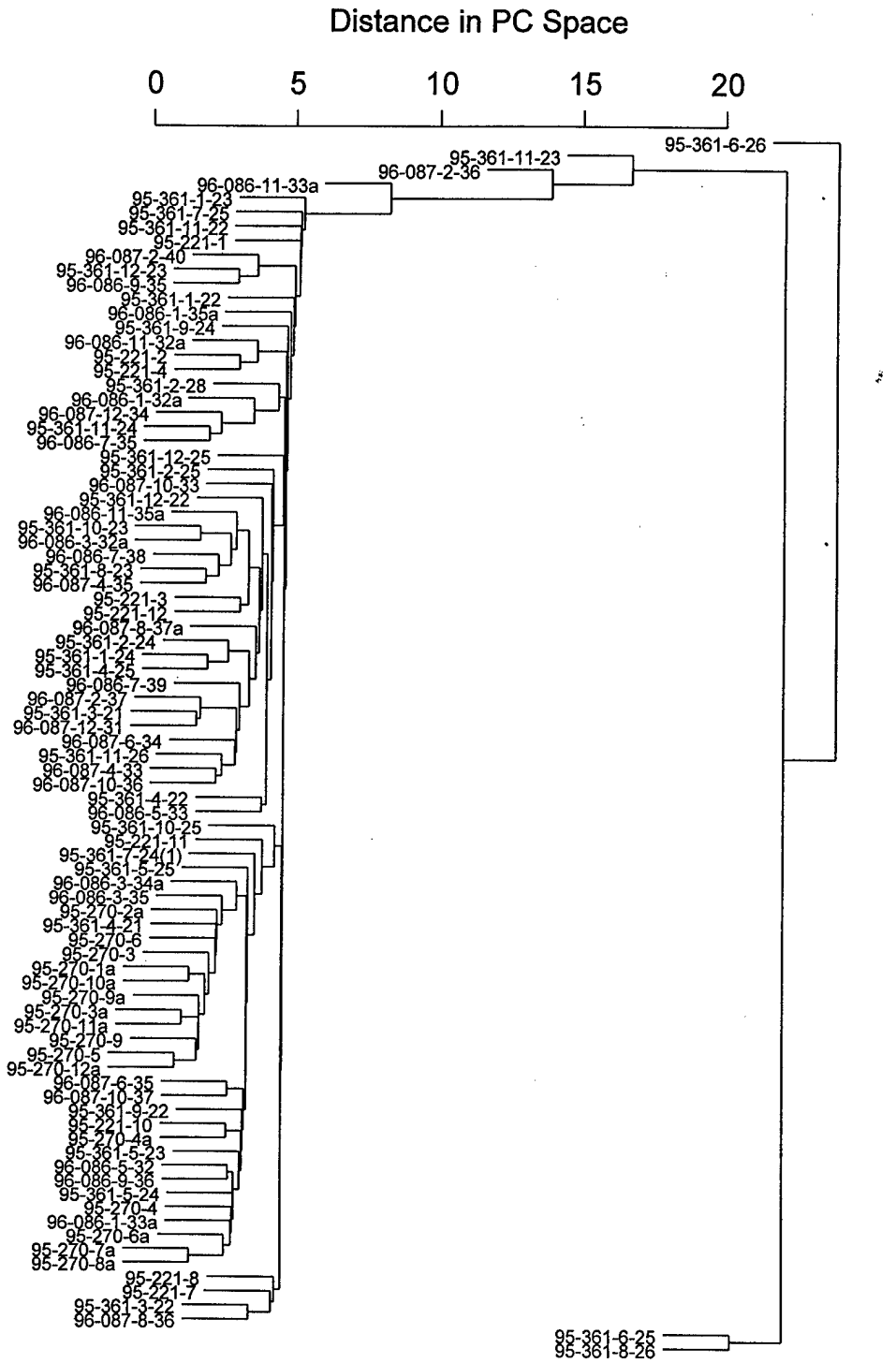

Fig. 1. 86 unique Medaka samples: Data on nearest neighbor clustering of samples. (See text for details)

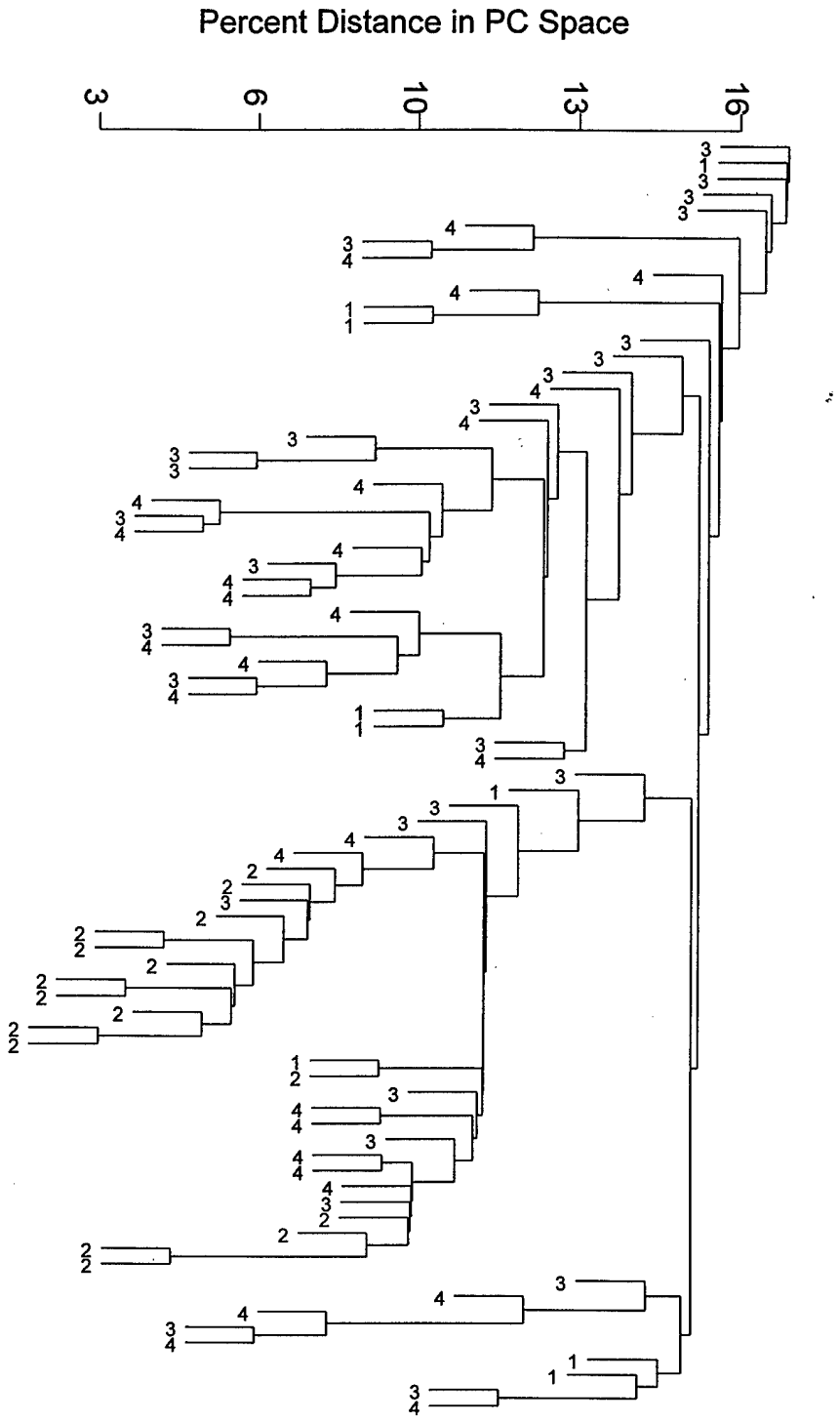

Fig. 10. 80 unique Medaka samples: Time group labels; nearest neighbor clustering of samples.

(See text for details)

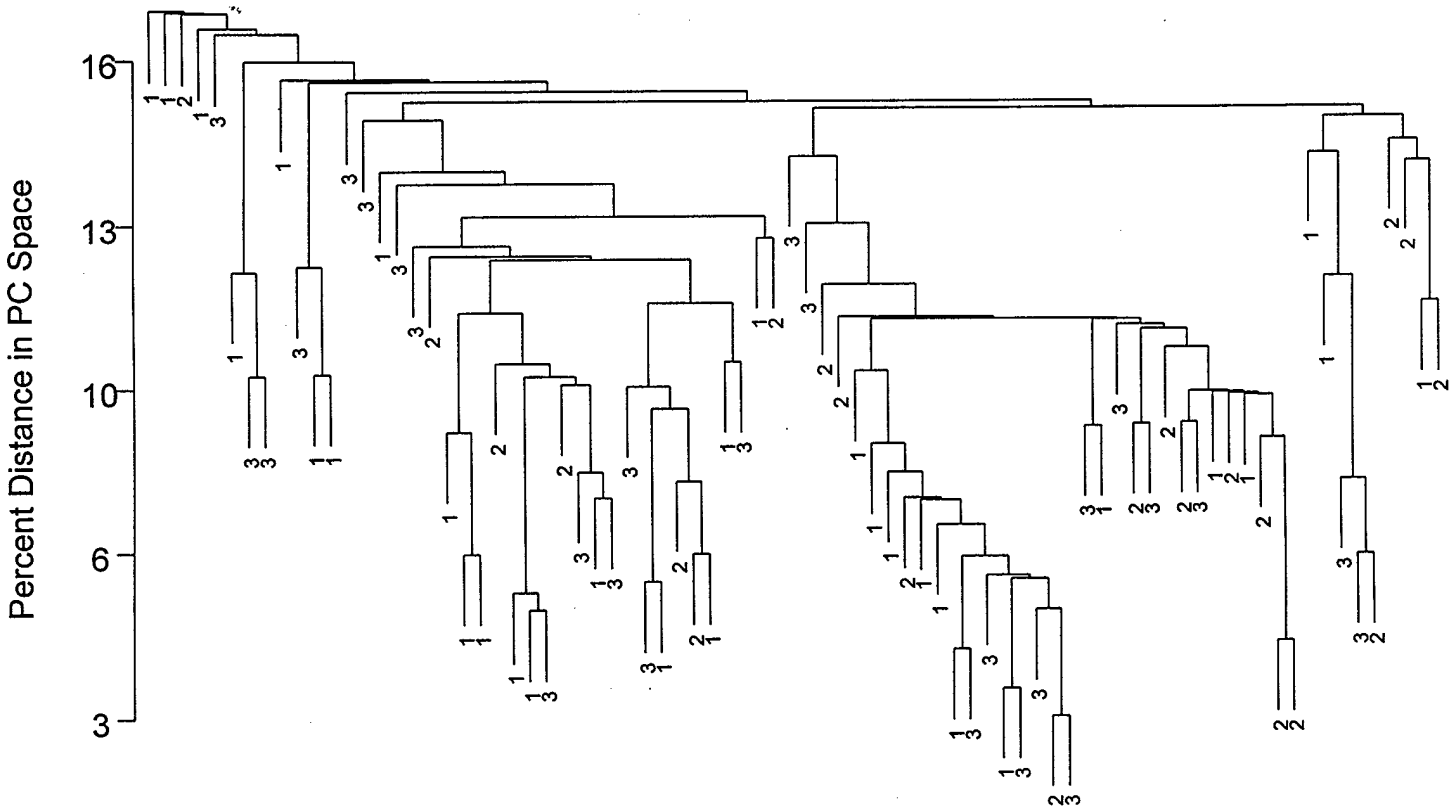

Fig. 11. 80 unique Medaka samples: Treatment group labels; nearest neighbor clustering of samples. (See text for details)

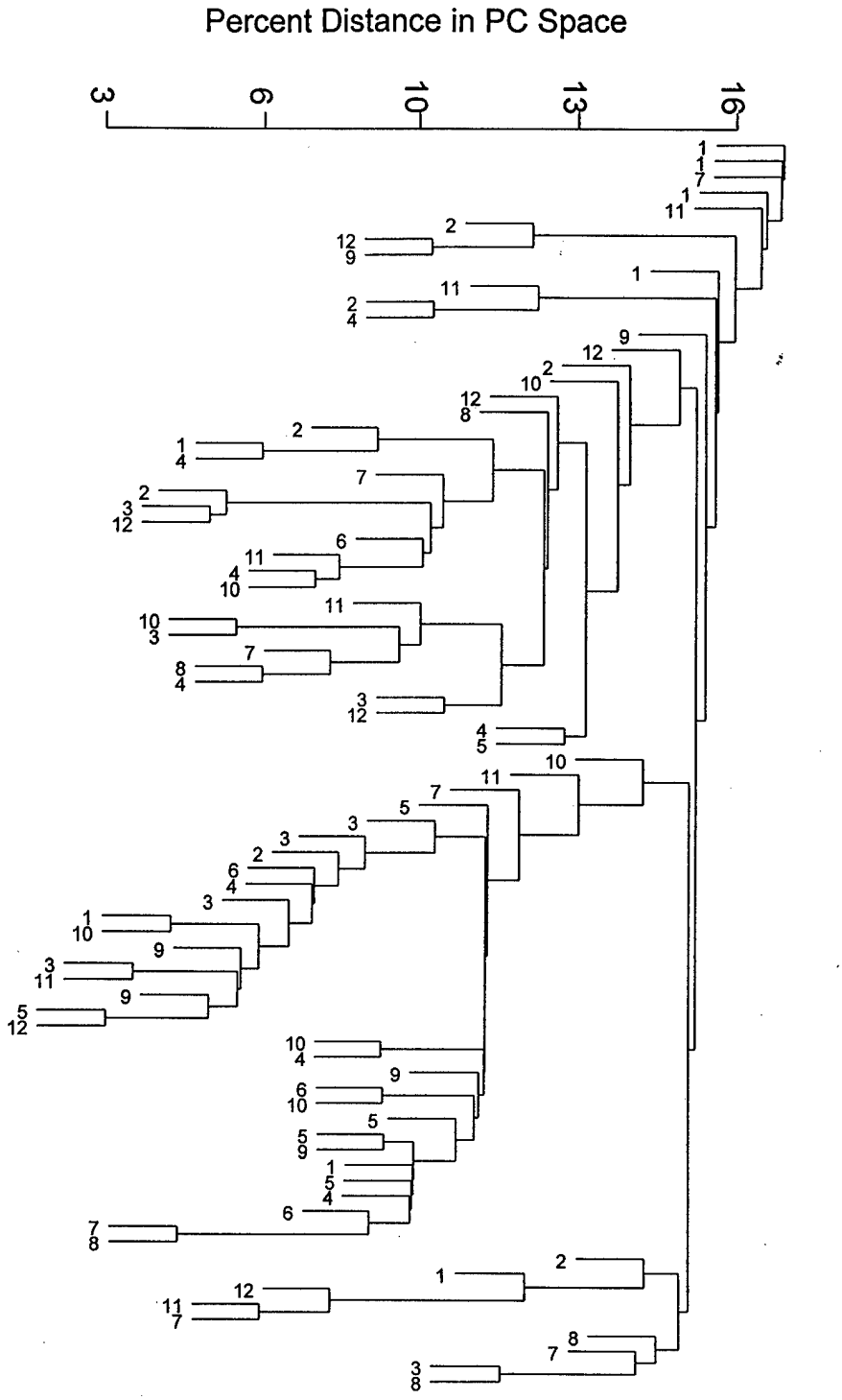

Fig. 12. 80 unique Medaka samples: Tank group labels; nearest neighbor clustering of samples.

(See text for details)

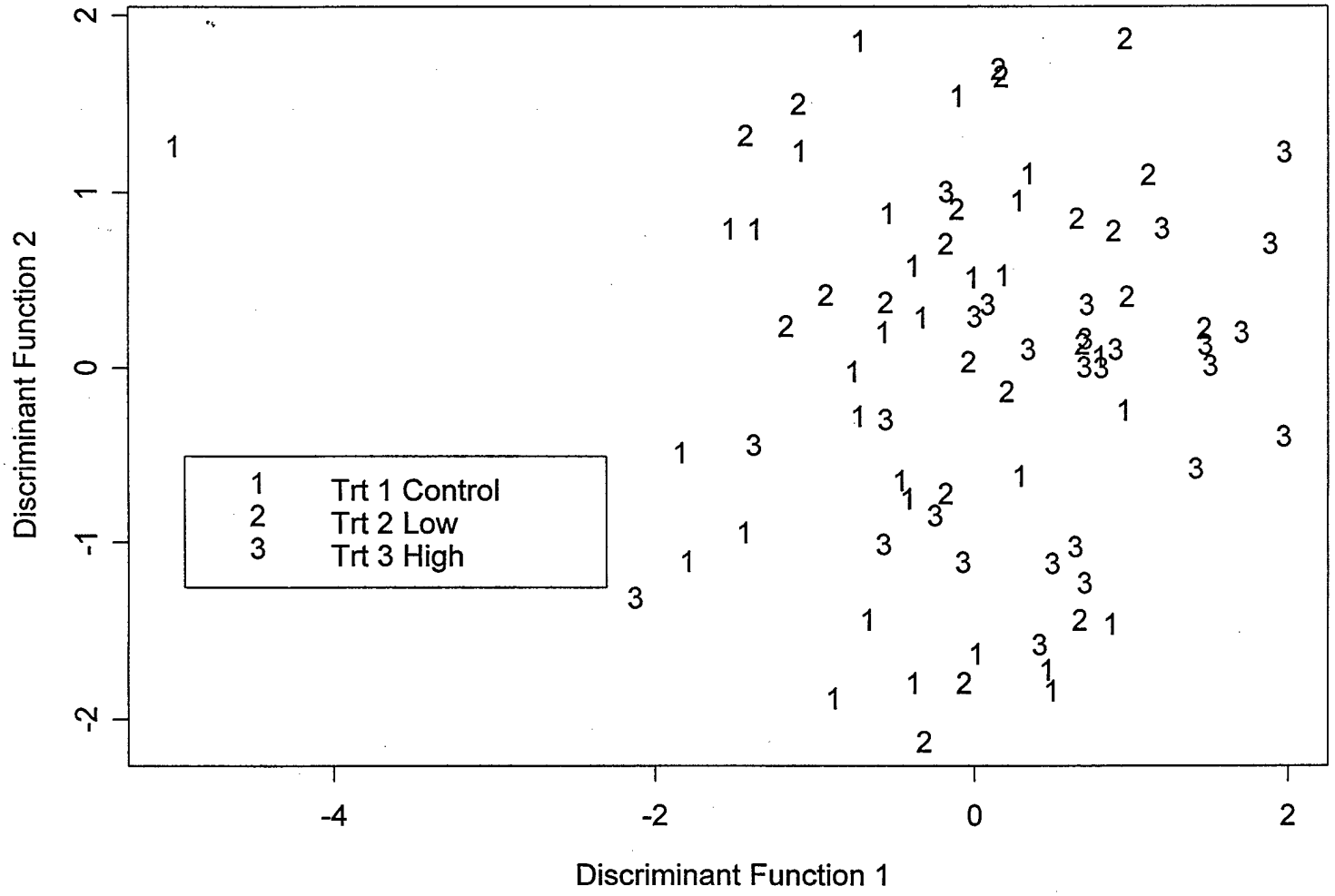

Fig. 13. 80 unique samples: Discriminant function plot for treatment variable. (See text for details)

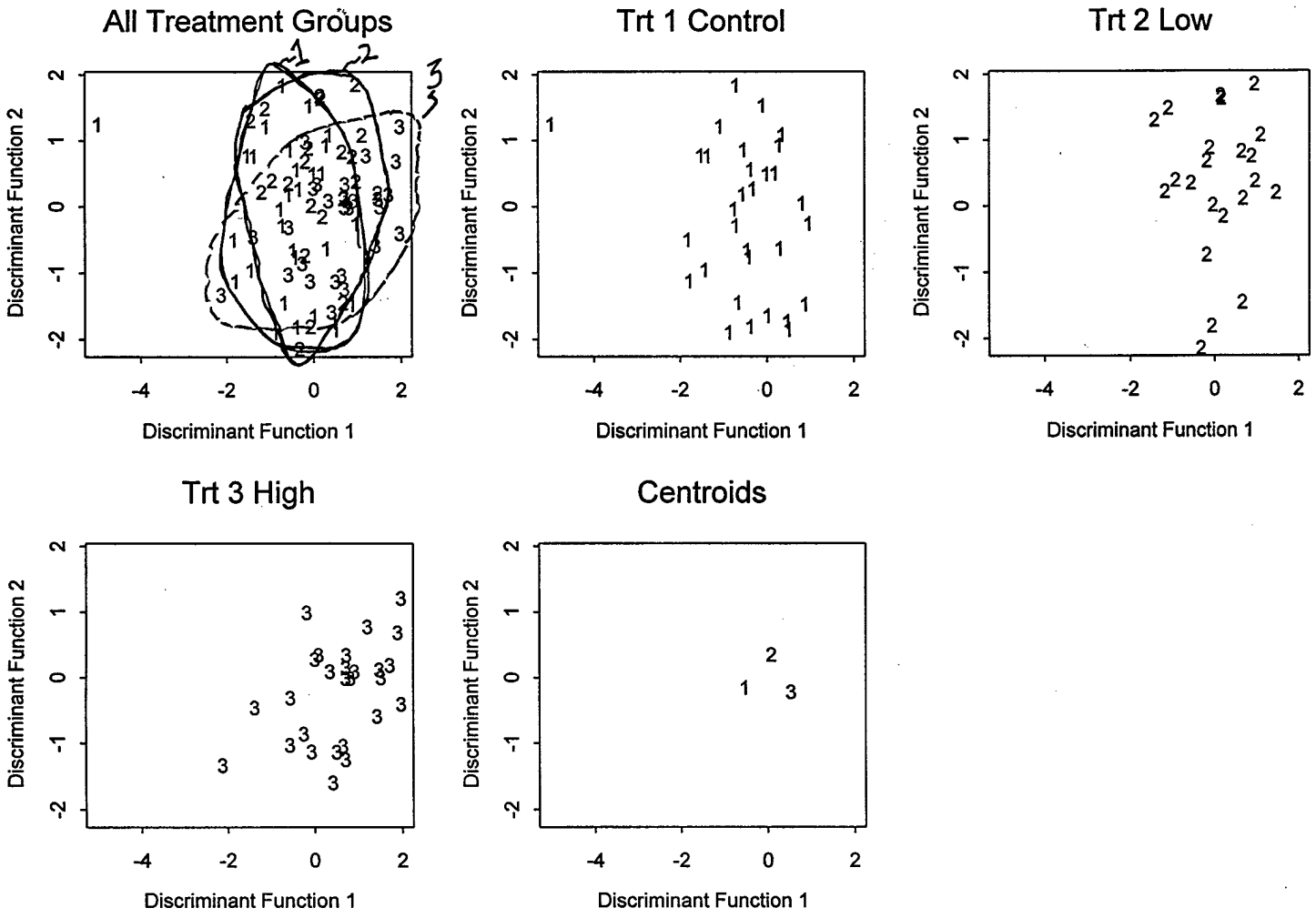

Fig. 14. Samples labeled by treatment group and plotted by first and second (non-significant) discriminant scores. (see text for details)

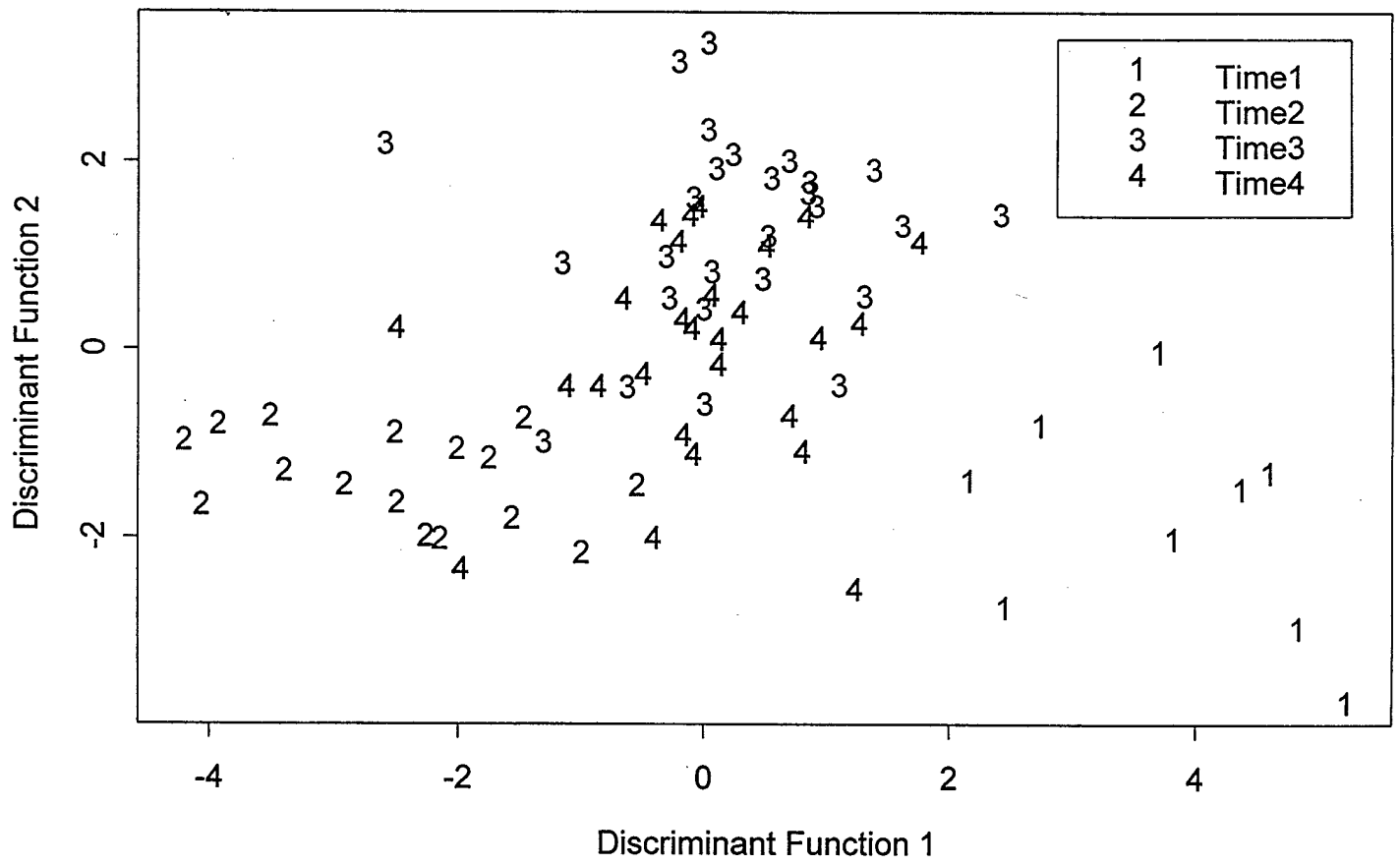

Fig. 15. 80 unique Medaka samples: Discriminant function plot for time variables. (See text for details)

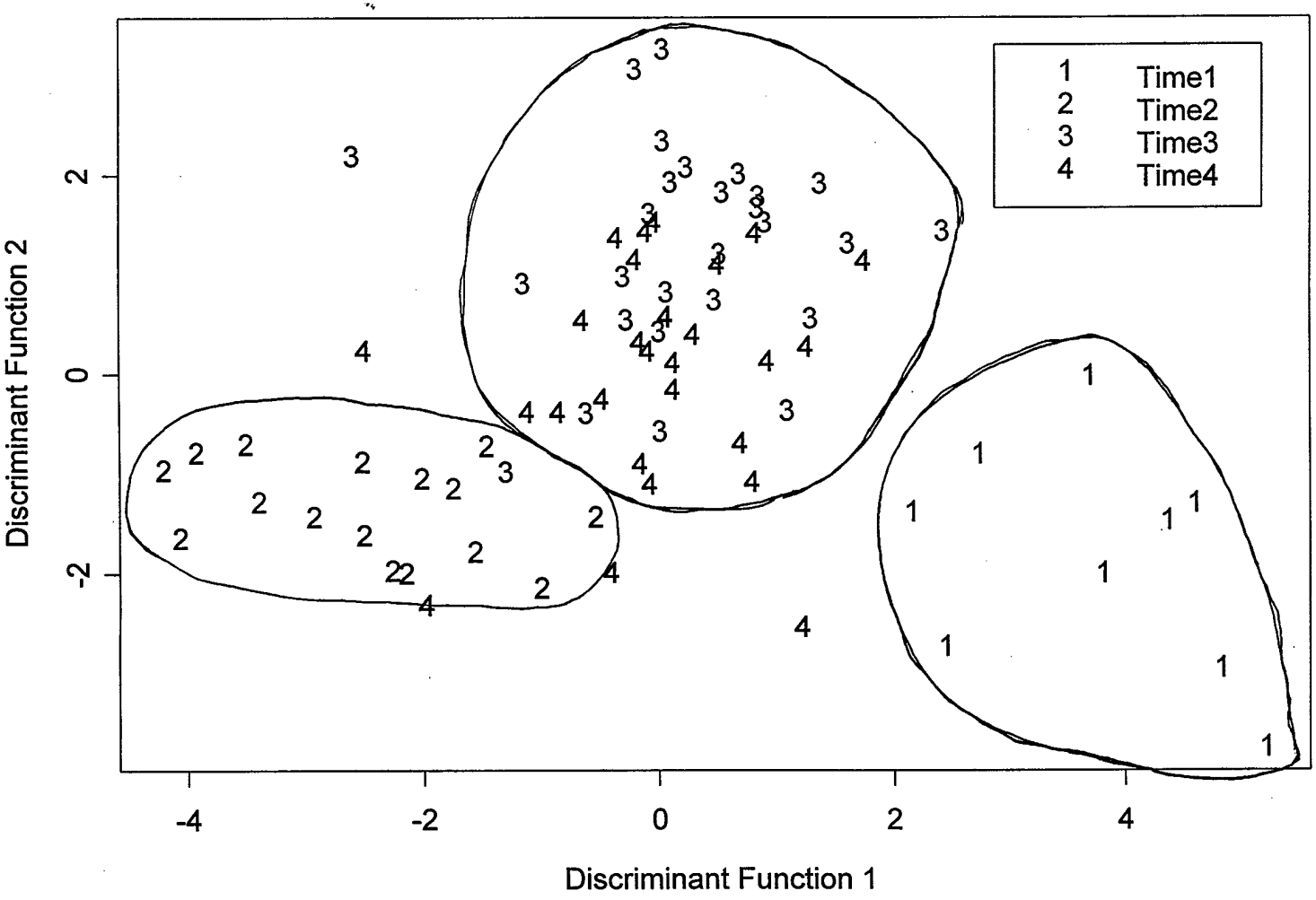

Fig. 16. 80 unique Medaka samples: Discriminant functions 1 and 2 (time variables). (See text for details)

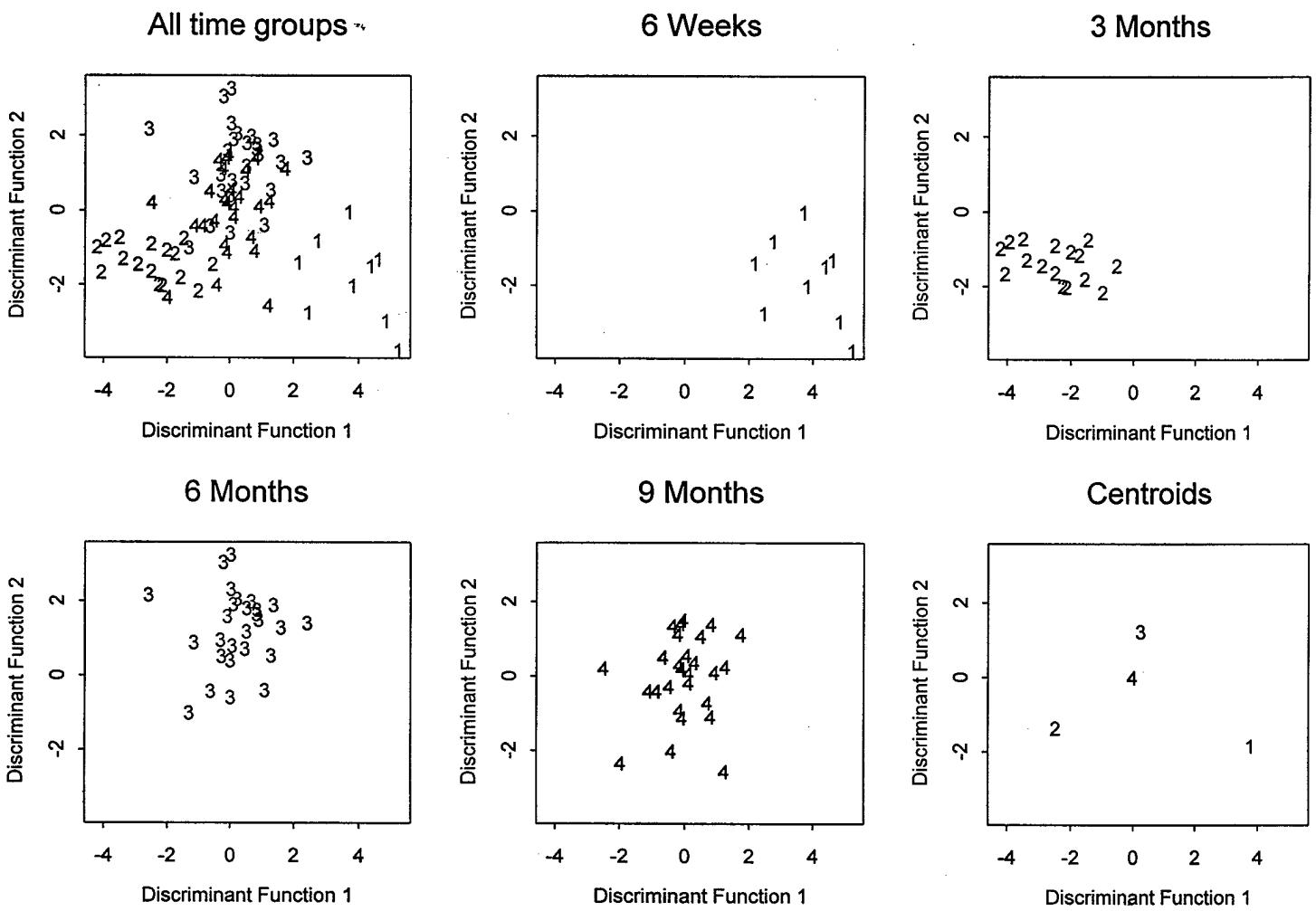

Fig. 17. 80 unique Medaka samples: Discriminant function plots for time variable. (See text for details)

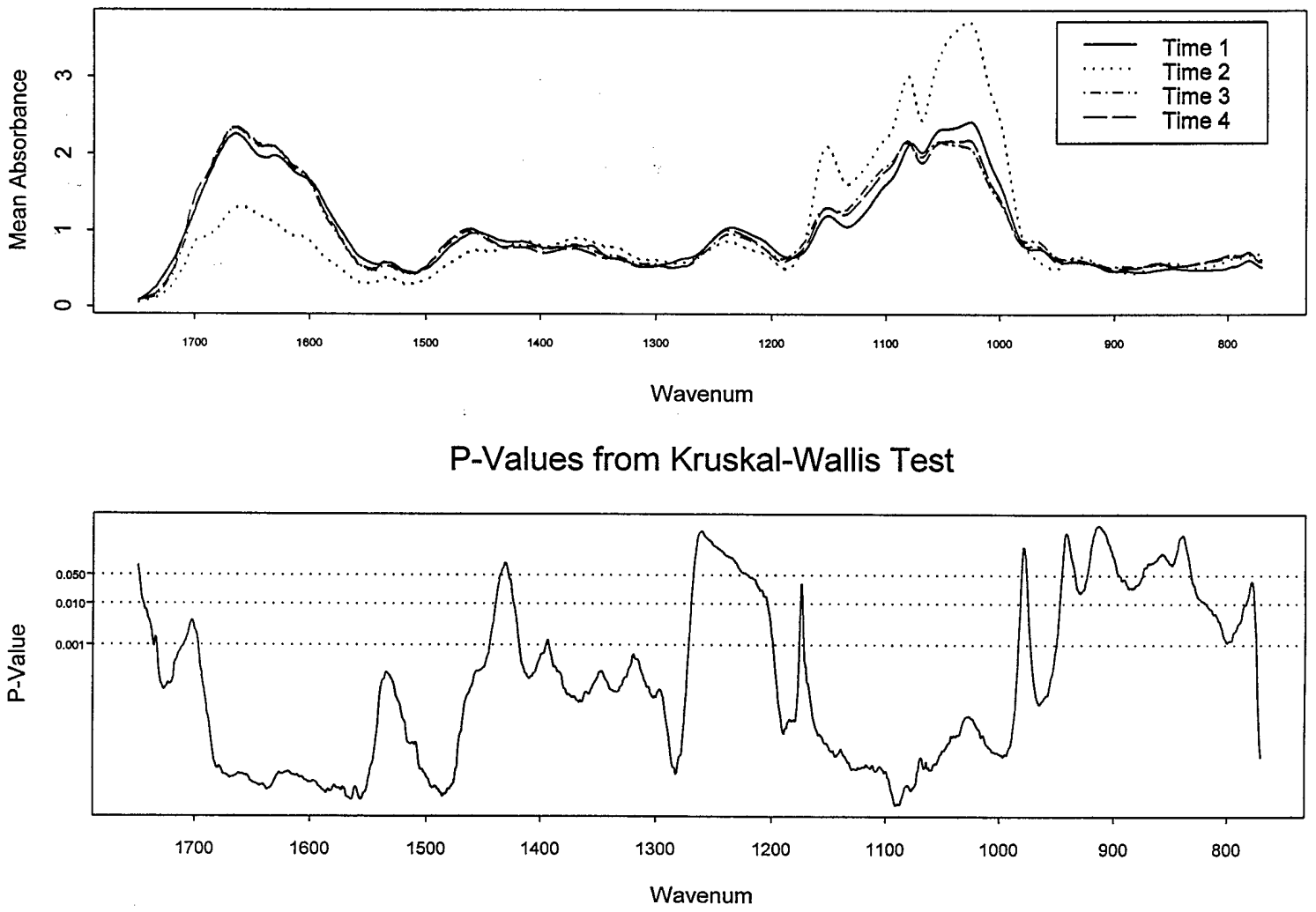

Fig. 18. 80 unique Medaka samples: Mean absorbance per time group. (See text for details)

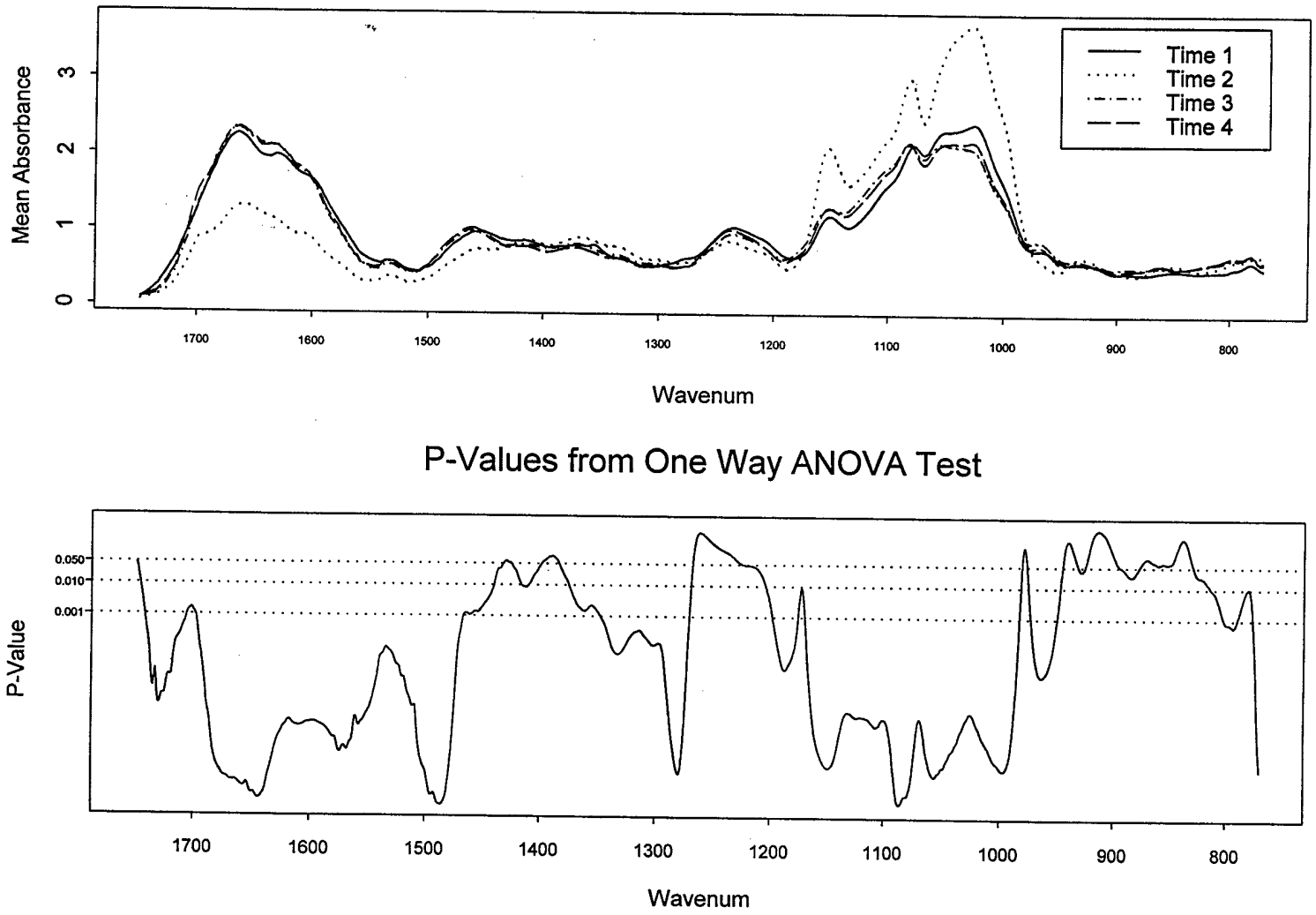

Fig. 19. 80 unique Medaka samples: Mean absorbance per time group. (See text for details)