6. Wilding P, Senior MB, Inubushi T, Ludwick ML. Assessment of proton nuclear magnetic resonance spectroscopy for detection of malignancy. *Clin Chem* 1988, **34**, 505–511.
7. Sträuli P, Haemmerli G. The V2 carcinoma of the rabbit as an integrated model of tumor invasion. *Bull Cancer (Paris)* 1984, **71**, 447–452.
8. Vargha G, Miltényi L, Endes J, Endes P, Kelemen J, Szücs L. Über eine neue Transplantationsmethode des VX2-Nierencarcinoms von Kaninchen. *Res Exp Med* 1974, **164**, 77–82.
9. Göthlin JH, Barbaric ZL. Tissue adhesive in surgical and crush renal trauma. An experimental angiographic investigation in the rabbit. *Scand J Urol Nephrol* 1978, **12**, 145–149.
10. Bell JD, Sadler PJ, Macleod AF, Turner PR, La Ville A. ^{1}H NMR studies of human blood plasma. Assignments of resonances for lipoproteins. *FEBS Lett* 1987, **219**, 239–243.
11. Herring FG, Phillips PS, Pritchard PH. Proton magnetic resonance spectroscopy of plasma from patients with dyslipoproteinemia: identification of factors governing methyl and methylene proton line widths. *J Lipid Res* 1989, **30**, 521–528.
12. Gadeholt G. Effects of intra-arterial and systemic infusion of doxorubicin treatment in rabbits with implanted renal VX-2 carcinoma. *J Interv Radio* 1989, **4**, 67–71.
13. Nydegger UE, Butler RE. Serum lipoprotein levels in patients with cancer. *Cancer Res* 1972, **32**, 1756–1760.
14. Barclay M, Skipski VP, Terebus-Kekish O, Greene EM, Kaufman RJ, Stock CC. Effect of cancer upon high-density and other lipoproteins. *Cancer Res* 1970, **30**, 2420–2430.
15. Wold S, Espensen K, Geladi P. Principal component analysis. *Chemometrics Intell Lab Syst* 1987, **2**, 37–52.
16. Kvalheim OM, Karstang TV. A general-purpose program for multivariate data analysis. *Chemometrics Intell Lab Syst* 1987, **2**, 235–237.

Acknowledgement—This work was supported with grants from The Norwegian Cancer Society.

Eur J Cancer, Vol. 26, No. 5, pp. 615–618, 1990.
Printed in Great Britain

0277–5379/90$3.00 + 0.00

Detection of Malignant Tumours by Multivariate Analysis of Proton Magnetic Resonance Spectra of Serum

Einar Sletten, Olav M. Kvalheim, Svetlana Kruse, Mikael Farstad and Odd Søreide

Proton magnetic resonance spectra of blood serum have been subjected to multivariate data analysis to discriminate between samples from cancer patients and from controls. The main feature was the use of digitally defined resonance profiles. The methyl and methylene lipoprotein signals centred at 1.3 and 0.9 parts per million are non-lorentzian composite peaks that cannot be described properly by the line width at half-height. Instead 71 and 76 data points were used to describe the methylene and methyl peak profiles, respectively. These data points were used as input to a principal component analysis to distinguish between malignant (n = 29) and control samples (n =55). At a probability level of 0.01 (F-test) modelling classified all patients except 2 correctly, while 1 control was slightly above the predictive level for malignancy.
Eur J Cancer, Vol. 26, No. 5, pp. 615—618, 1990.

INTRODUCTION

CANCER may influence lipid metabolism in man [1, 2]. Interest in proton magnetic resonance (NMR) as a diagnostic tool for cancer was revived by Fossel *et al.* [3]. The mean line widths of the methyl and methylene resonances of lipoprotein lipids in plasma correlated with the presence of disease, values being lower in patients with malignant tumours. Several research groups have failed to produce the same excellent descrimination [4–10]. Fossel [11] pointed out several methodological factors to account for this discrepancy: field strength, recording temperature, field homogeneity and storage conditions.

In our opinion the major problem with the Fossel test is found in the univariate description of the resonance profiles. The lipoprotein proton NMR spectra from serum represent several molecular species, including cholesterol, triacylglycerols, phospholipids and free fatty acids. The major parts of the methyl and methylene signals [5] represent low density lipoprotein (LDL), high density lipoprotein (HDL), very low density lipoprotein (VLDL) and chylomicrons. A non-lorentzian composite signal cannot be properly defined by a single value of the line width at half-height. In addition, variable amounts of lactate, manifested as two satellite peaks on the low-field side of methylene, may affect line width measurements.

To circumvent this problem we suggest that digitally defined profiles of the methyl and methylene peaks should be used in the analysis. We have applied multivariate principal component analysis [12, 13] with the major part of the spectral profile as input variables. We have successfully used a similar method to analyse diffuse reflectance infrared spectral profiles of coals [14].

Correspondence to: E. Sletten, Department of Chemistry, University of Bergen, Allegt. 41, N-5007 Bergen, Norway.
E. Sletten, Olav Kvalheim and S. Kruse are at the Department of Chemistry and M. Farstad and O. Søreide are at the Laboratory for Clinical Biochemistry, University of Bergen, Bergen, Norway.

THEORY

Principal component modelling is now well-established for the interpretation and classification of multivariate data [12, 13]. Here we describe the features necessary for understanding supervised modelling [15, 16].

Principal component model

Principal components can be modelled at two levels: (1) unsupervised, to reveal groups of similar samples, and (2) supervised, to model preconceived groups of samples. The first step in both types is the extraction of the major eigenvectors from the variance–covariance matrix. These uncorrelated (orthogonal) vectors, called principal components, minimize the residual variation in the multivariate data. A principal components model can be written in matrix form as follows:

$$X_{NM} = U_{NA} G^{\frac{1}{2}}_{AA} P'_{AM} + E_{NM} \quad (1)$$

The subscripts show the dimensions of the matrices. The data matrix X contains the values of M variables measured for N samples. The variables are assumed to be centred around the mean. The matrices U and P contain the orthonormal sample scores and variable loadings, respectively, for the A principal components with largest eigenvalues. The diagonal matrix G contains the norms (squared eigenvalues) of the A principal components. If the measured variables are highly correlated the first few principal components account for most of the variation in the data. The residual matrix E contains the unexplained variation in the data.

Dimension of model

The aim of the supervised approach is to obtain the best possible model with respect to classification of future samples with unknown groupings. Thus, the predictive ability of the principal components model needs to be optimized, which is done by cross-validation.

A much-used cross-validation technique, which is especially suited for small sample sets, is the 'leave-out-one procedure' [17]. A modification, which we used is to keep out one block of samples at a time. The kept-out samples are then fitted to the respective principal components models and their squared residuals estimated. We start from models with no principal components, continue with models with one principal component and successively include the next principal component until the sum of squared residuals is no longer reduced. This is the model with the optimum predictive ability. Finally, a model is computed with A principal components and all the samples included. By this procedure, all the samples are utilized both for modelling and for validating the model.

Classification criterion in supervised modelling

The average lack-of-fit obtained for the left-out controls in the respective models can be used as a criterion for determining whether a sample belongs to the group of controls or is different (e.g. a cancer patient). By an F-test, a sample's fit to the model is measured [15, 16]. The squared residual of a fitted sample is compared with the squared average residual obtained from the validation procedure. The degrees of freedom used for the F-test are $f = M-A$ and $f = (M-A) \times (N-A-1)$.

MATERIALS AND METHODS

Patients and blood samples

Blood was collected in Vacutainer tubes with citrate as anticoagulant, centrifuged immediately and transferred to NMR tubes. Proton NMR spectra were recorded in less than 8 h after blood sampling. Sera from 29 patients with malignancy of different types and in different stages were analysed. Diagnoses included malignancy of: prostate, ovary, pancreas, testes, colon and ventricle. All the patients were undergoing treatment for the disease at the time of sampling. The age range was 18–82. The control group was 58 volunteers aged 25–67. The amount of total cholesterol, HDL and triglyceride was measured for all samples. Three control samples were discarded, one because of diagnosed severe hypertriglyceridaemia and two because of experimental problems.

NMR analysis

No chemicals were added to the samples before analysis on a Bruker AM-400 MHz Fourier-transform spectrometer. Samples of 0.5 ml serum were prepared in 5 mm NMR tubes. The spectra were recorded without spinning the samples, and the instrument was run in the unlocked mode with adequate field stability. Thus no locking substance had to be added to the serum. Water suppression was accomplished by presaturation of the water resonance with the decoupler on for 6 s. Typical spectra consist of sixteen transients, with 32K data points and collected with an r.f. detection pulse of 9.3 μs (corresponding to a flip angle of 60°). All spectra were recorded at 25°C. However, this may not be the optimal temperature for spectral resolution.

Multivariate analysis

The resonance profiles centred around 1.3 parts per million (ppm) and 0.9 ppm were described by 147 data points, normalized separately and used as input for principal component analysis [12, 13]. In the supervised modelling the controls were treated separately.

Cross-validation was used to check the predictive power of the model [17] and thus to determine dimensionality. The controls were divided into five groups and one group was left out at a time. Six predictive principal components were found for the control group, explaining 99% of the variance in the data. By including a seventh and eighth component no significant improvement was obtained. The mean residual standard deviation estimated for the control group (0.013) was used in an F-test to calculate the screening criterion of 0.018 at a 0.01 probability level. All calculations were done on a MICROVAX-II with the SIRIUS program [18].

RESULTS

The methyl and methylene region of a water suppressed proton spectrum of serum is composed of signals from LDL, VLDL and chylomicrons, progressively shifted to higher frequencies [5]. In addition, methyl resonances from varying amounts of lactate are seen as two distinct signals on the high frequency side of the broad methylene peak. The general appearance of the spectrum did not change when remeasured after storage of the serum for several days at 4°C.

All methylene peaks from the controls had a distinct shoulder. The relative position of the shoulder resonance influences the measured line width. The corresponding signal in the sample from the patient group lacks the shoulder, but the variable

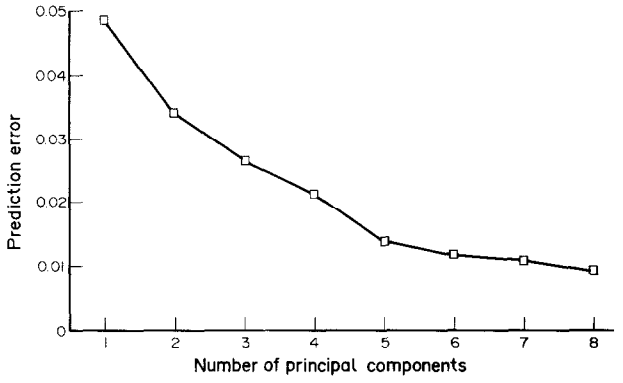

Fig.1. Standard error of prediction versus number of principal components used in modelling control group.

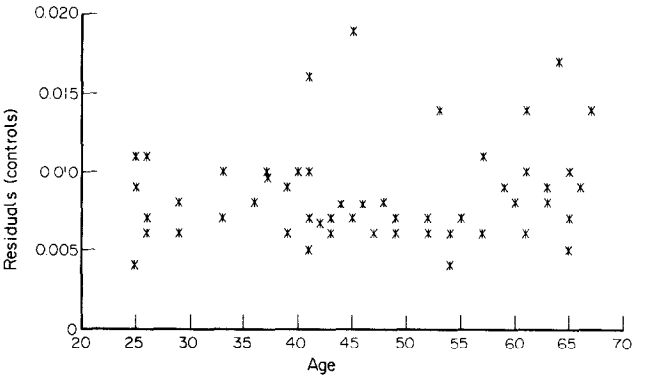

Fig.3. Distribution of residuals versus age in controls.

amount of lactate is found to render line width measurements practically useless.

In multivariate modelling six principal components gave optimum prediction (Fig. 1). Adding one more component did not improve fit significantly. The distribution of residuals calculated in the analysis was plotted for patients and controls (Fig. 2). The cut-off at a probability of 0.01 (F-test) was 0.018. Only 2 patient samples fell in the 'wrong' category. However, 1 of these patients had a rare, atypical tumour growing in neural epithelium. One sample in the control group was slightly above the cut-off and would be classified as positive in a screen.

DISCUSSION

Fossel's method of line width measurements at half-height as a screening criterion was not suitable for our data. About a third of our controls would be classified as cancer patients by the simple line width test. Other groups have arrived at similar conclusions [9, 10]. The presence of malignancy is obviously not the only factor that significantly affects the relative amount of the lipid fractions and thus influence spectral profiles. Nutritional status, storage conditions and, most importantly, age and sex are among the factors that should be considered. The multivariate approach is suited for this type of N-dimensional covariance. Any systematic variation in the data that is non-specific to malignancy will, in principle, be accounted for by the multivariate model, thus enabling the extraction of pertinent information from background 'noise'. A plot of residuals versus age of controls (Fig. 3) showed no functional dependency, indicating that any 'age' factor in the data has been accounted for by multivariate fitting. We assume that each of the six predictive principal components for the control data is somehow related to variation in specific biophysical properties and/or experimental artifacts. These variations are observed as changes in chemical shift and spectral intensities, which again are a function of the relative concentrations in lipoprotein constituents (HDL, LDL and VLDL). It is sometimes possible to assign a principal component to a specific factor (e.g. age or triglyceride content) by target rotation [18]. Due to the strong correlation between several of the factors involved, this procedure was not possible for our data. Fortunately, the discriminatory power of principal components analysis does not rely on such assignments.

To improve the multivariate analysis a larger control group is needed as a basis for modelling. Also controls with non-malignant diseases should be included. The problem of hypertriglyceridaemia as a source of false positives still remains [7]. However, this rare condition is easily diagnosed and should not cause undue concern.

At which stage is the method able to detect malignancy? We are doing a series of experiments implanting VX-2 carcinoma in rabbits and monitoring changes in lipoprotein spectra during progressive cancerous growth [19]. A preliminary multivariate data analysis indicated a functional dependency between the magnitude of residuals and number of days after VX-2 implantation.

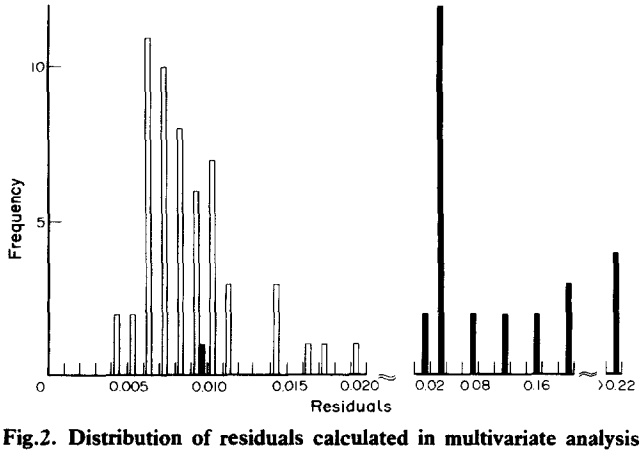

Fig.2. Distribution of residuals calculated in multivariate analysis for controls (□) and patients (■) Cut-off at probability of 0.01 is 0.018.

1. Nydegger UE, Butler RE.Serum lipoprotein levels in patients with cancer. *Cancer Res* 1972, **32**, 1756–1760.
2. Barclay M, Skipski VP, Terebus-Kekish O, Greene EM, Kaufman RJ, Stock CC. Effect of cancer upon high-density and other lipoproteins. *Cancer Res* 1970, **30**, 2420–2430.
3. Fossel ET, Carr JM, McDonagh J. Detection of malignant tumors: water suppressed proton nuclear magnetic resonance spectroscopy of plasma. *N Engl J Med* 1986, **315**, 1369–1376.
4. Small DM, Hamilton JA. *N Engl J Med* 1987, **316**, 1412–1413.
5. Bell JD, Sadler PJ, Macleod AF, Turner PR, La Ville A. ^{1}H NMR studies of human blood plasma. Assignments of resonances for lipoproteins. *FEBS Lett* 1987, **219**, 239–243.
6. Nicholson JK, Nicholson F. Proton spectroscopy of plasma and testing for malignancy. *Lancet* 1987, 280–281.
7. Wilding P, Senior MB, Inubushi T, Ludwick ML. Assessment of proton nuclear magnetic resonance spectroscopy for detection of malignancy. *Clin Chem* 1988, **34**, 505–511.
8. Buchthal SD, Hardy MA, Brown TR. Assessing the value of identifying the presence of malignant disease in human plasma by proton NMR spectroscopy. *Am J Med* 1988, **85**, 528–532.

9. Herring FG, Philiphs PS, Pritchard PH. Proton magnetic resonance spectroscopy of plasma from patients with dyslipoproteinemia: identification of factors governing methyl and methylyene proton line widths. *J Lipid Res* 1989, **30**, 521–528.
10. Berger S, Pflüger K-H, Etzel WA, Fischer J. Detection of tumors with nuclear magnetic resonance spectroscopy of plasma. *Eur J Cancer Clin Oncol* 1989, **25**, 535–543.
11. Fossel ET, Commentary on 'Detection of tumors with nuclear magnetic resonance spectroscopy of plasma' by S Berger, K-H Pflüger, WA Etzel, J Fischer. *Eur J Cancer Clin Oncol* 1989, **25**, 925–927.
12. Wold S, Espensen K, Geladi P. Principal component analysis. *Chemometrics Intel Lab Syst* 1987, **2**, 37–52.
13. Jolliffe IT. *Principal Component Analysis*. Berlin, Springer, 1986.
14. Christy AA, Velapoldi RA, Karstand TV, Kvalheim OM, Sletten E, Telnaes N. Multivariate calibration of diffuse reflectance infrared spectra of coals as an alternative to rank determination by vitrinite reflectance. *Chemometrics Intel Lab Syst* 1987, **2**, 199–207.
15. Kvalheim OM. On-line classification and updating of disjoint principal component models. *Chemometrics Intel Lab Syst* 1988, **3**, 243–247.
16. Wold S. Pattern recognition by means of disjoint principal components models. *Pattern Rec* 1976, **8**, 127–139.
17. Allen DM. Mean square error of prediction as a criterion for selecting variables. *Technometrics* 1971, **13**, 469–475.
18. Kvalheim OM, Karstang TV. A general-purpose program for multivariate data analysis. *Chemometrics Intel Lab Syst* 1987, **2**, 235–237.
19. Gadeholt G, Kruse S, Halsteinslid L, Sletten E. Proton nuclear resonance spectroscopy of serum lipoproteins in rabbits with implanted VX-2 carcinoma. *Eur J Cancer* 1990, **26**, 611–615.

Acknowledgements—This work was supported by grants from the Norwegian Cancer Association and the Norwegian Research Council for Science and the Humanities. We thank Miss A. Smaabrekke at the University Health Center for providing the blood samples for the controls.

Eur J Cancer, Vol. 26, No. 5, pp. 618–622, 1990.
Printed in Great Britain

0277–5379/90$3.00 + 0.00

Chromosomal Abnormalities in Patients with Non-cutaneous T-cell Non-Hodgkin's Lymphoma

Harry C. Schouten, Warren G. Sanger, Dennis D. Weisenburger and James O. Armitage for the Nebraska Lymphoma Study Group

In contrast to non-Hodgkin's lymphomas (NHL) with a B-cell phenotype, almost no data have been reported dealing with correlations between chromosomal abnormalities and characteristics of the disease in patients with T-cell NHL. In a retrospective analysis we studied all patients with a non-cutaneous T-cell NHL and chromosomal abnormalities that were evaluated at our institution; 20 patients could be identified. Numerical abnormalities involving chromosomes 3, 4, 5, 22 and X were observed most frequently. Structural abnormalities involved mainly the breakpoints 1q22–25, 6q23 and 11q13. There appeared to be an association between +7, breakpoints 2p23–24, 4p14–15, 8q21 and the presence of extranodal disease. All patients with +7 had a diffuse mixed histology. Patients with +2, +3, +11, +17, +18, +20 or breakpoint 1q22–25 had an immunoblastic lymphoma and patients with breakpoints 9q32–34 or 14q12 had a lymphoblastic lymphoma. No correlations were observed between chromosomal abnormalities and response to therapy, survival or phenotypic markers. Abnormalities involving the chromosomes containing the T-cell receptor genes and T-cell markers were infrequent. Several breakpoints were identified that correlate with already described oncogenes.
Eur J Cancer, Vol. 26, No. 5, pp. 618—622, 1990.

INTRODUCTION

SEVERAL STUDIES describing chromosomal abnormalities in patients with non-Hodgkin's lymphoma (NHL) have been published. They not only suggest that some abnormalities occur frequently, but also that particular karyotypes are correlated with characteristics of the disease [1–11]. The overwhelming majority of patients studied have NHL with a B-cell phenotype. The number of patients with cytogenetically studied T-cell NHL is much smaller [6, 9, 12–16] and few correlations between chromosomal abnormalities and characteristics of the disease have been reported.

Therefore, in order to analyze the particular chromosomal abnormalities that are related with non-cutaneous T-cell NHL, we performed a retrospective analysis of all patients with non-cutaneous T-cell NHL who had an abnormal cytogenetic analysis and were treated by the Nebraska Lymphoma Study Group.

Correspondence to: H.C. Schouten M.D., Department of Internal Medicine, Section Oncology–Hematology, University Hospital Maastricht, PO Box 1918, 6201 BX Maastricht, The Netherlands.
H.C. Schouten and J.O. Armitage are at the Department of Internal Medicine, W.G. Sanger is at the Department of Pediatrics and W.G. Sanger and D.D. Weisenburger are at the Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, U.S.A. H.C. Schouten is a Clinical Fellow, Netherlands Cancer Foundation Koningin Wilhelmina Fonds.