Cluster analysis of comparative genomic hybridization (CGH) data using self-organizing maps: Application to prostate carcinomas

Torsten Mattfeldt [a,*], Hubertus Wolter [a],
Ralf Kemmerling [b], Hans-Werner Gottfried [c]
and Hans A. Kestler [d]

[a] *Department of Pathology, University of Ulm, Ulm, Germany*
[b] *Department of Medical Genetics, University of Ulm, Ulm, Germany*
[c] *Department of Urology, University of Ulm, Ulm, Germany*
[d] *Department of Neural Information Processing, University of Ulm, Ulm, Germany*

Comparative genomic hybridization (CGH) is a modern genetic method which enables a genome-wide survey of chromosomal imbalances. For each chromosome region, one obtains the information whether there is a loss or gain of genetic material, or whether there is no change at that region. Usually it is not possible to evaluate all 46 chromosomes of a metaphase, therefore several (up to 20 or more) metaphases are analyzed per individual, and expressed as average. Mostly one does not study one individual alone but groups of 20–30 individuals. Therefore, large amounts of data quickly accumulate which must be put into a logical order. In this paper we present the application of a self-organizing map (Genecluster) as a tool for cluster analysis of data from pT2N0 prostate cancer cases studied by CGH. Self-organizing maps are artificial neural networks with the capability to form clusters on the basis of an unsupervised learning rule, i.e., in our examples it gets the CGH data as only information (no clinical data). We studied a group of 40 recent cases without follow-up, an older group of 20 cases

with follow-up, and the data set obtained by pooling both groups. In all groups good clusterings were found in the sense that clinically similar cases were placed into the same clusters on the basis of the genetic information only. The data indicate that losses on chromosome arms 6q, 8p and 13q are all frequent in pT2N0 prostatic cancer, but the loss on 8p has probably the largest prognostic importance.

Keywords: Artificial neural networks, chromosome aberrations, cluster analysis, comparative genomic hybridization, multivariate analysis, prognostic factors, prostatic cancer, self-organizing maps, tumor suppressor genes

Figure 2 on http://www.esacp.org/acp/2001/23-1/mattfeldt.htm.

1. Introduction

Modern molecular biological methods may produce large amounts of data which are difficult to survey. This statement applies particularly to array techniques, where the expression of thousands of genes is measured. Here the problem may arise to find clusters of genes which behave in a similar manner [29]. To a smaller extent, analogous problems are found during evaluation of comparative genomic hybridization (CGH) data.

CGH is a method which allows screening of the whole genome for gains and losses of the genetic material. Genomic DNA of tumor tissue as well as the DNA of normal tissue are isolated, differentially stained and hybridized to normal metaphase chromosomes. When the tumor-DNA is stained green and the normal DNA is stained red, for example, this leads to a green stain at locations with a gain of tumor DNA, whereas a red stain is obtained at locations with losses of tumor DNA because here the normal DNA dominates (Fig. 2). The results are quantitated by digital image analysis. This leads to a series of results for the 24 chromosomes (22

*Corresponding author: Prof. Dr. T. Mattfeldt, Department of Pathology, Oberer Eselsberg M23, D-89081 Ulm, Germany. Fax: +49 731 58738; E-mail: torsten.mattfeldt@medizin.uni-ulm.de.

autosomes and 2 sex chromosomes). For convenience, during this paper each chromosome arm was taken as a unit, regardless of its different size. Since the short arm of the acrocentric chromosomes and the Y chromosome are uninformative, 42 chromosome arms as chromosomal regions were taken into account during this analysis. For each chromosome arm one of the alternatives 'unchanged', 'loss' or 'gain' (or equivalently 1, 0 or 2) is noted. In short, one case is reduced (theoretically) to a row of 42 numbers, in which each element can assume the value 0, 1 or 2. Our task consists in the formation of a certain number of clusters to which the cases are assigned in a biologically meaningful manner. This task has to be fulfilled without knowing further variables, as is usual for clustering methods, solely on the basis of the CGH data. The present paper is an example to achieve this for prostatic cancer, which has been intensively studied by CGH [1,7,24].

The required grouping can be principally obtained by all kinds of clustering techniques. For example, hierarchical cluster analysis, k-means clustering or fuzzy c-means clustering may be used [3,8,17]. Hierarchical clustering forces data points into a strict hierarchy of nested subsets and has been noted to suffer from a lack of robustness [22]. K-means and fuzzy c-means clustering are completely unstructured approaches proceeding in an entirely local fashion. SOMs allow to impose a partial structure on the clusters (chain, grid) easing visualization [13,15,26]. Furthermore, they have been applied to a variety of problems and have been extensively studied empirically [19]. Here, we used Kohonen's SOM as an easy to apply tool of analysis [13–16]. Recently such nets were successfully used for cluster analysis in gene expression [29]; our group has applied related networks of the same authors with a supervised learning rule for predictive purposes in prostate carcinoma research [20,21].

2. Materials and methods

2.1. Patient population

Group I. This material consisted of 40 recently obtained primary uncultured prostate carcinomas. All cases were adenocarcinomas, the pTNM classification was pT2N0 [23,28]. The primary tumor specimens were prostatectomy specimens. Small tissue blocks of tumor material and normal seminal vesicle tissue from the same patient were flash-frozen in liquid nitrogen immediately after surgical removal. Five μm sections were cut from freshly frozen tumor and normal sem-

inal tissue blocks and stained with hematoxylin and eosin to ensure the histological representativeness of the samples. Based on microscopic evaluation the tumor region was selected and removed for DNA extraction with a scalpel [27]. Qiagen-Blood & Cell Culture-Kit (Qiagen GmbH, Hilden, Germany), following the instructions of the supplier.

Group II. The archive of the Urological Department of the University of Ulm from 1985–1995 was searched for patients with prostatic cancer in stage pT2N0 in whom a radical prostatectomy with pelvic lymphadenectomy had been performed (228 cases). The current TNM classification according to the UICC was used, and the series included cases in stages pT2a and pT2b [20]. When at least one postoperative PSA level was found to be above 0.5 ng/ml serum, or if a local relapse or a metastasis was found, a case was defined as tumour with progression, otherwise as tumour without progression. All cases of the group with progression ($n = 27$), from which technically acceptable CGH data could be obtained, were selected for the study (10 cases). From the large group of cases without progression ($n = 201$), 10 cases matched by age, duration of follow-up, and preoperative PSA level were selected as control group [20]. The serum level of PSA was measured in all patients with the Hybritech kit. It was necessary to go back to 1985 to find enough of the rare pT2N0 cases with postoperative progression, hence the preoperative PSA levels were not available in a few cases.

Pathology. The prostatectomy specimens were evaluated histopathologically at the Department of Pathology of the University of Ulm. The specimens were step sectioned at 3–5 mm slice thickness, and at least 2 additional sections from the resection margins and at least 1 additional section from each seminal vesicle were taken. The tumor-bearing slides of all prostatectomy specimens were reevaluated by the first author with respect to Gleason score, WHO grade, maximum diameter of the cut tumour tissue on section, and various further parameters not considered in this paper. While reevaluating the specimens, the investigator was blinded with respect to postoperative progression.

2.2. Comparative genomic hybridization (CGH)

CGH was performed as described previously [2,9, 10] with minor modifications. We currently use standard nick translation for labeling genomic DNAs with biotin-16-dUTP (tumor DNA) and digoxigenin-11-

dUTP (normal DNA). The fragment length of our genomic probes after nick translation was 500 to 1600 bp. One μg of labeled tumor and normal DNA were precipitated together with 70 μg Cot-1 DNA. The mixture was dissolved in 12 μl of hybridization buffer: 50% formamide, 10% dextran sulfate, and 2 × SSC (1 × SSC: 0.15 M NaCl, 0.015 M sodium citrate, pH 7). This hybridization mixture was denatured at 75°C for 6 minutes and pre-annealed for 30 minutes at 37°C and hybridized to a slide with normal male lymphocyte (46,XY) metaphase spreads, denatured separately in a formamide solution (70% formamide, 2 × SSC, pH 7–7.2) for 2 minutes at 70°C and dehydrated through a series of graded solutions of 70%, 90% and 100% ethanol. The hybridization was performed for 2 to 3 days at 37°C in a moist chamber. After hybridization, the slides were washed 3 times in 50% formamide, 2 × SSC, pH 7–7.2 for 5 minutes at 42°C and 3 times in 0.1 × SSC, pH 7–7.2 for 5 minutes at 60°C. The tumor DNA was detected with a single layer of avidin-conjugated fluorescein isothiocyanate (FITC) (Vector Laboratories, Inc., Burlingame, CA), and the normal DNA was detected with anti-digoxigenin antibody conjugated to rhodamine (Boehringer Mannheim GmbH, Mannheim, Germany) for 45 minutes at 37°C. The chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) and embedded in antifade solution (Vector Laboratories, Inc., Burlingame, CA).

2.3. Digital image analysis

Three single-color images (matching DAPI = blue, FITC = green and rhodamine = red) were acquired from 15–20 metaphases using a Zeiss fluorescence microscope (Carl Zeiss, Oberkochen, Germany) and a Hamamatsu chilled charge-coupled-device (CCD) camera (Hamamatsu Photonics K.K., Tokyo, Japan) interfaced to a computer workstation. The ISIS digital image analysis system (Metasystem GmbH, Altlussheim, Germany) was used with CGH analysis software (Version 3.02). Fluorescence ratio (green : red) for each chromosome type were derived for these metaphase cells. All ratio profiles from each chromosome were averaged, and the standard deviation of the profile was calculated at each point. For all the profiles, losses of DNA sequences are defined as chromosomal regions in which the mean green : red ratio is below 0.8 whereas gains are defined as chromosomal regions in which this ratio is above 1.25. These threshold values are symmetric cutoff values, 1.25 and its reciprocal value, 0.8. Interpretation of CGH-results followed previously described protocols [11]. Hybridization of FITC-labeled normal DNA with rhodamine-labeled normal DNA was used as negative control. Some chromosomal regions have been shown to give unreliable results. The distal part of chromosome arm 1p(1p34-pter) and chromosome arm 16p are difficult to evaluate. Heterochromatin blocks such as the distal long arm of the Y chromosome or the centromeres, and the near-centromere heterochromatic regions of chromosomes 1, 9, 16 were excluded from CGH analysis as well as centromeres and short arms of the acrocentric chromosomes 13, 14, 15, 21, and 22.

2.4. Short introduction to self-organizing maps

Artificial neural networks (ANN) are information processing systems consisting of a number of simple units (neurons), communicating with each other by connections. Such systems 'learn' by processing external information; according to the learning rule, they are classified into ANNs with supervised learning and with unsupervised learning. Here, the unsupervised SOM has been used, as good teacher signals like tumor progression were only available in group II (20 cases). Considering the small number of samples (60: group I + group II) with a feature number of 30, a classification using training and test data sets in terms of prediction of tumor status would be inappropriate. Hence, we centered on a purely exploratory approach.

Self-organizing maps (SOMs) belong to the ANNs with an unsupervised learning rule. That means, we have an ANN, to which only the input vectors (input data, input information) are presented, and no output vectors. The task of a typical SOM consists in finding clusters of the input data, with similar vectors in the same clusters.

The fundamental structure of a SOM is a layer of neurons. This layer has a thickness of 1 neuron and usually has a simple geometric shape, e.g., a plane, or a line (chain) in the plane [15]. These neurons have weight vectors, that lie in the n-dimensional space of the input vectors. The basic active neurons are called Kohonen neurons, the layer is the Kohonen layer. The Kohonen neurons are locally connected with each other, for example, by a quadratic or hexagonal lattice when the Kohonen layer is a plane. During the learning process the weight vectors are moved in the n-dimensional space until they have moved as close as possible to the input vectors. That neuron whose

weight vector has come nearest to an input vector is called a winner neuron. This event leads to a second effect: the weight vectors of all neurons from the vicinity of the winner neuron are modified according to a predefined neighbourhood function. On the whole, the learning process has the effect that properties of the n-dimensional input space are projected to the low-dimensional space of the Kohonen neurons ($d = 1$ or 2). In particular, input vectors lying close to each other will form a cluster in the Kohonen layer.

The classical implementation of a SOM is the package SOMPAK, generated by Kohonen and coworkers. It can be obtained as free academic software by anonymous ftp from the server cochlea.hut.fi under /pub/som_pak and runs under Unix (Linux) and DOS. The input variables are fed as ASCII file into the system, and the user has to enter a number of system parameters such as the number of neurons (nodes) in the Kohonen plane, the neighbourhood function and others.

An alternative to SOMPAK is the recently published SOM Genecluster, presented in [29]. The input data are also given as an ASCII data set. For the calculations only very few parameters have to be indicated to the system, most parameters are preset to standard values. For details of our application see Results section. The result consists in tables where each data point is assigned to one cluster; here we selected to ascribe the data to three clusters. The motivation was that it is most convenient to perform the grading of malignancy in three steps, e.g., I–III, indicating low, intermediate and high malignancy. Genecluster is free academic software and runs under Windows NT. It can be obtained by internet under http://genome.wi.mit.edu.

3. Results

First we report on the basic findings in the two groups. In group I (40 cases) 18 cases showed chromosomal alterations, whereas a normal gene dosage was detected with CGH in 22 cases. The most frequent changes were loss on 8p and 13q (8 cases). Mean Gleason score was 5.6, mean WHO grade 1.87.

An overview of the CGH-findings in group II is shown in Fig. 1b. The most frequent findings in this group were 8p loss (8 cases) and 8q gain (4 cases). As there are two subgroups with 10 patients each (no progression and progression), we have displayed the data in two columns in Table 1.

Evaluation with the nonparametric U-test (see legend to Table 1) showed significant results. Cases with progression had significantly more losses, more gains, a higher Gleason score and a higher WHO grade.

The third group results from pooling groups I and II (60 cases). The results are somewhere intermediate between these groups: mean Gleason score 6.2, mean WHO grade 2.0, mean number of losses: 1.0, mean number of gains: 1.15. The most frequent change by far was 8p loss (16 cases), followed by 11 cases of 13q loss and 8 cases of 6q loss.

Finally, SOM Genecluster was applied to our data. The program was applied to the groups I and II separately, and then to the pooled dataset III. We consider the chromosomal findings as the only known input variables and a number of prognostically important biological variables, such as grading, progression, Gleason-score, etc. as dependent variables. In all these 3 studies, there were 30 input variables each, corresponding to all chromosome arms that could be studied by CGH and in which at least one imbalance occurred (when no case in any group showed an imbalance for a chromosome arm, this arm was considered as noninformative). In the following learning process, the SOM acts solely on the input variables, as described above. When operating with Genecluster, standard settings were used. The number of clusters was set to 3 corresponding to a chain of 3 units length. Number of iterations per run was set to 5000, as thereafter one could not see significant changes of the error. As the neighbour function of the SOM the step function (bubble) was used. The initial and final learning rate values were $\alpha_i = 0.1$ and $\alpha_f = 0.005$, and the initial and final radius values of the step function were $r_i = 5$ and $r_f = 0.2$. The map was initialized using random vectors. We performed 20 repeated runs for each network. The results were sometimes the same in repetitions, or they changed slightly (no more than 1 case per group). When the clustering had been finished, we recorded the result by giving the number of the clusters, the number of cases, progression (if applicable), the mean number of gains, the mean number of losses, the mean Gleason score and the mean WHO grade, by combining the case numbers again with the dependent variables. Moreover, the frequencies of the most abundant changes – losses of 8p, 6q and 13q – were recorded per cluster. First, we show the result for group I (40 cases) in Table 2.

How do we have to interpret this outcome? The SOM has created cluster 3 with the least deviation from normality; this is by far the largest cluster. There is no
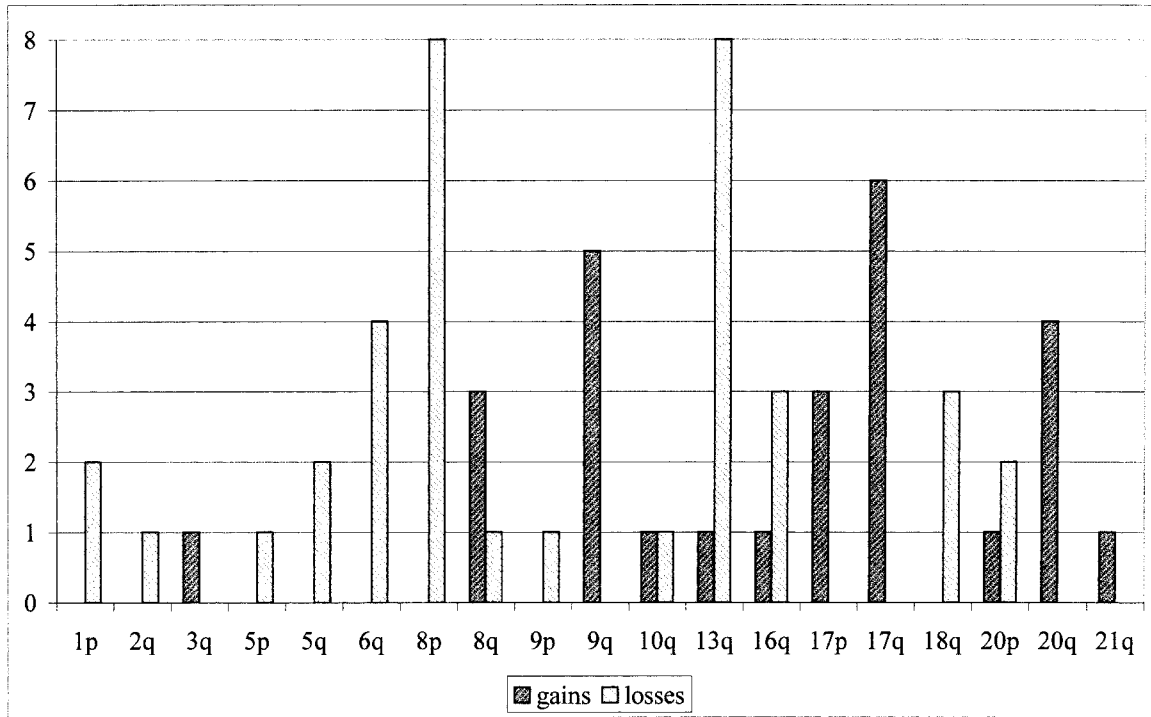
(a)
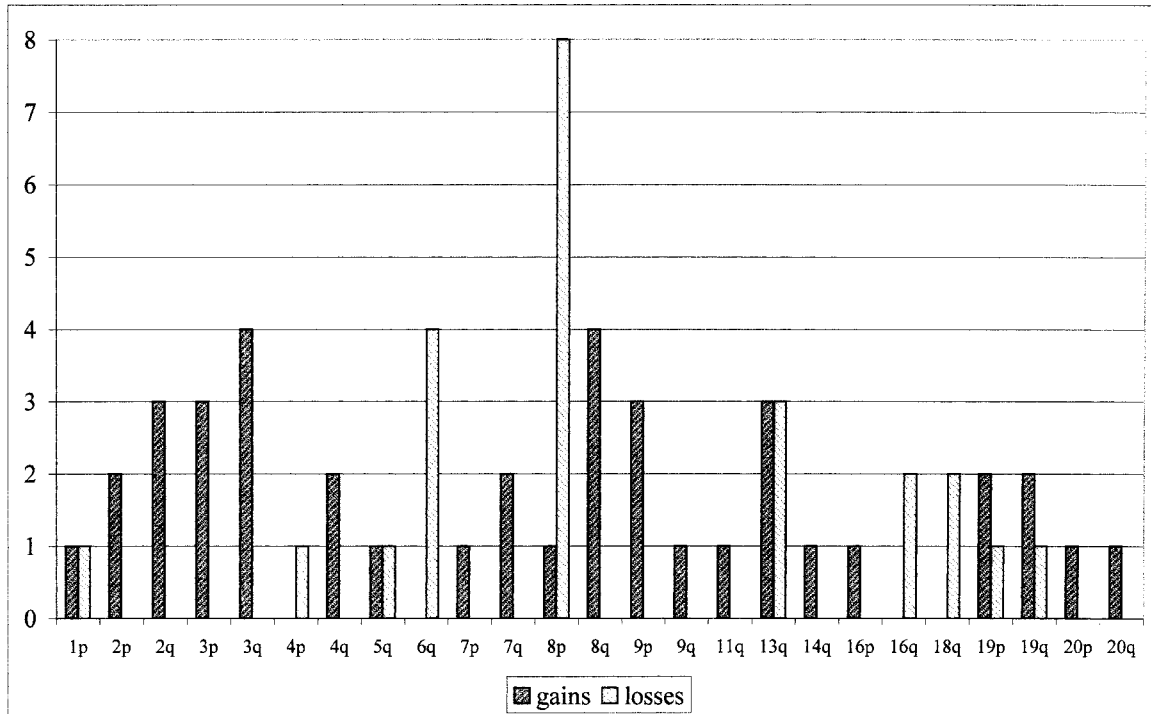
(b)

Fig. 1. (a) CGH-data of group I. Summary of all gains and losses of DNA sequences observed in 40 prostate carcinomas by CGH – without follow-up. (b) CGH-data of group II. Summary of all gains and losses of DNA sequences observed in 20 prostate carcinomas by CGH – with follow-up.
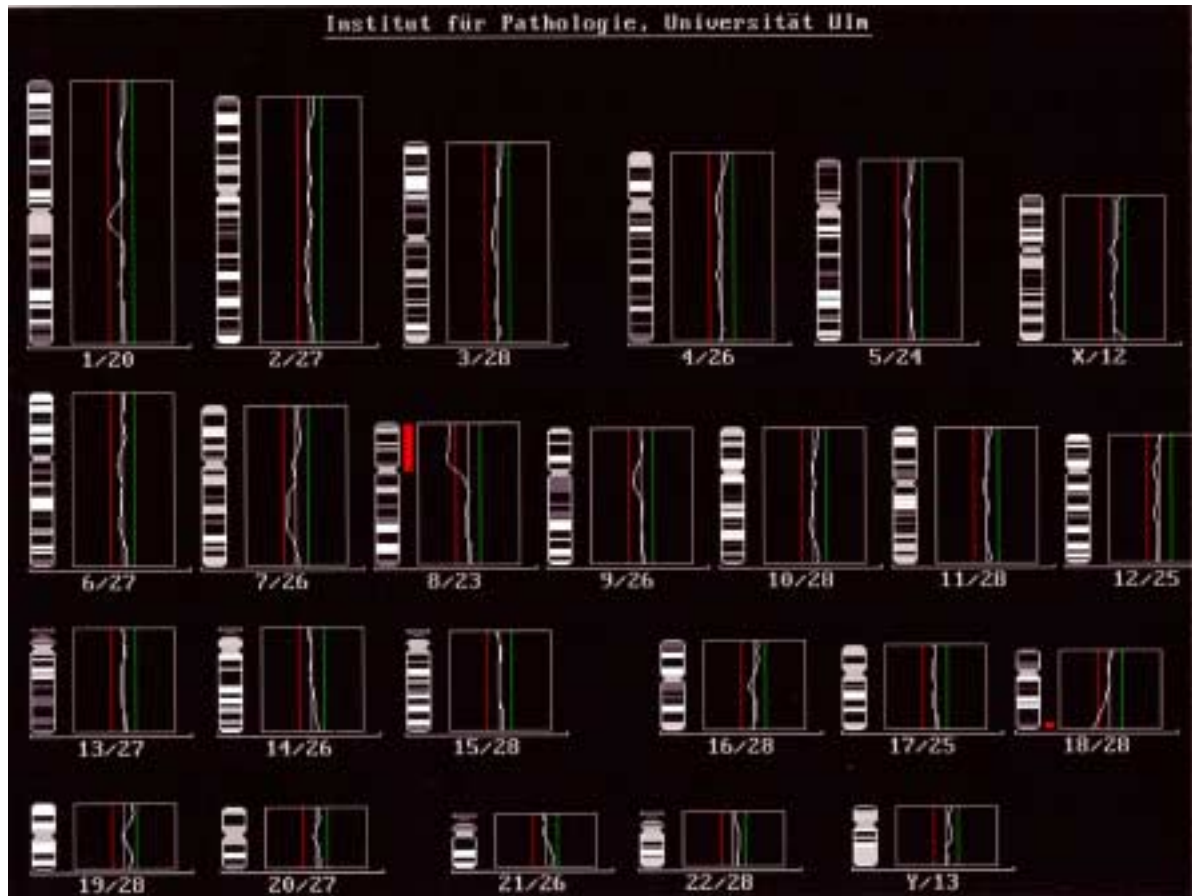
Fig. 2. Example of the CGH analysis. Ratios of the fluorescence intensities generated by tumour and control DNA were obtained from 15 metaphase cells, and the average ratio profile was plotted along each single chromosome. The three vertical lines next to each ideogram indicate the balanced state of the chromosomal copy number (central line, ratio value 1.0) as well as the diagnosis thresholds (red line, ratio value 0.8; green line, ratio value 1.25) indicating chromosomal losses or gains, respectively. Red bars on the left of the chromosomes indicate DNA loss, and green bars on the right indicate DNA gains (here not shown). This figure can be viewed on http://www.esacp.org/acp/2001/23-1/mattfeldt.htm.

loss on 6q, 8p and 13q, which are the most frequent losses. The Gleason score and WHO grade of cluster 3 are the lowest, it has the fewest losses and gains in general. The clusters 1 and 2 are very similar to each other in many aspects, however there are more gains in cluster 1 than in cluster 2. Whether this has a biological implication, cannot be determined from this group because there is no follow-up. Anyway, it is known that the Gleason-score is a very strong indicator of prognosis, but the Gleason-score is not increased in cluster 1 compared to cluster 2 here. We come to the results of group II (20 cases, Table 3).

Also in this group, Genecluster has created two small clusters and one large cluster. Just as above, the large cluster (3) has the most 'relatively benign' cases: lowest proportion of progression (25%), lowest Gleason-score, lowest WHO grade, smallest number

of losses, the number of gains is also relatively small. Note there is no 8p-loss in this cluster. In contrast, both clusters 1 and 2 have 8p-loss in all cases, and nearly no losses of 6q and 13q. In these clusters, progression is 75–100%. Table 4 shows the results of group III (pooled data, 60 cases).

In this table, the proportions of cases with progression is only given for the cases from group II, because only for this group follow-up is known. Also in the pooled group, we found a similar distribution of cases in clusters as in groups I and II. A large group of 40 cases constituted a cluster of cases of relatively low malignancy: low Gleason-scores and WHO grades, low numbers of losses and gains, no 8p-loss. The outcome of the 8 cases in cluster 2 and the 12 cases in cluster 3, in contrast, was relapse in a high proportion (80%). All cases in cluster 3 and half of

Group	Without progression	Progression	U	P
Median number of losses	0	2.0	19.0	0.019
Median number of gains	0	1.5	23.0	0.041
Median Gleason score	6.5	8.5	14.0	0.007
Median WHO grade	2.0	2.75	13.5	0.006

Group II consists of 20 cases for whom a follow-up is available. The left column gives median values for the 10 cases without, the right column for the 10 cases with progression. The number of gains and losses as well as the tumour grades are strongly elevated in the group with postoperative tumour progression. The nonparametric U-test was used for comparisons to be free from assumptions on distributions of the variables which might be not fulfilled. U = test statistic of the U-test, p = probability of getting a result as extreme or more extreme than the one observed, assuming that there is no difference between the groups.

Table 2
Analysis of CGH-data of group I using Genecluster

Cluster	1	2	3
Cases	6	9	25
Cases with 6q-loss	0/6	2/9	0/25
Cases with 8p-loss	3/6	5/9	0/25
Cases with 13q-loss	3/6	5/9	0/25
Mean Gleason score	5.8	6.3	5.3
Mean WHO grade	1.9	2.1	1.8
Mean number of losses	2.2	2.2	0
Mean number of gains	3.5	0.2	0.2

The CGH data of group I (40 cases without follow-up) have been subjected to a cluster analysis by Genecluster. The three resulting clusters are indicated as the three columns with 6, 9, and 25 cases. Cluster 3 has nearly no aberrations. Clusters 1 and 2 are rich in 8p-losses and 13q-losses. These clusters have also higher Gleason-scores and WHO grades than cluster 3. The only larger difference between them is the very high number of gains in cluster 1.

the cases in cluster 2 have an 8p-loss. Gleason-score, WHO grade and number of losses are similar, however cluster 3 has a higher number of gains.

4. Discussion

In the present paper, we have applied a self-organizing map for the first time to data from comparative genomic hybridization. The small datasets used here (20–60 cases) are typical for CGH, because evaluation of a single case is rather laborious. In order to make SOMs for CGH popular, we presented the rather easy

Table 3
Analysis of CGH-data of group II using Genecluster

Cluster	1	2	3
Cases	4	4	12
Cases with 6q-loss	0/4	1/4	3/12
Cases with 8p-loss	4/4	4/4	0/12
Cases with 13q-loss	0/4	0/4	3/12
Progression	3/4	4/4	3/12
Mean Gleason score	7.8	8	6.7
Mean WHO grade	2.6	2.6	2.2
Mean number of losses	2.2	2.0	0.8
Mean number of gains	6.8	0.8	0.9

This table shows the result of a cluster analysis by Genecluster in group II (20 cases with follow-up). The two small clusters 1 and 2 have 100% 8p losses, whereas cluster 3 has no 8p-loss. On the other hand, there was 75–100% progression in clusters 1 and 2 and only 25% progression in cluster 3. The small clusters again differ in the total number of gains (see Table 2).

Table 4
Analysis of CGH-data of group III using Genecluster

Cluster	1	2	3
Cases	40	8	12
Cases with 6q-loss	2/40	4/8	2/12
Cases with 8p-loss	0/40	4/8	12/12
Cases with 13q-loss	9/40	1/8	1/12
Progression	2/10	3/4	5/6
Mean Gleason score	5.6	7.4	7.0
Mean WHO grade	1.8	2.4	2.4
Mean number of losses	0.4	2.2	2.1
Mean number of gains	0.4	0.6	3.8

Group III is the pooled set of cases from Groups I and II. Progression is therefore indicated only for 20 cases. The large cluster 1 has only 20% progression, no 8p-loss but an appreciable number of 13q-losses. Gleason score and WHO grade are low. In contrast, the smaller clusters 2 and 3 have 50–100% 8p-losses and 75–83% progression, with high Gleason score and WHO grade. Only cluster 3 has a high number of gains.

to use program 'Genecluster', equipped with a graphical interface of Windows-type and preset parameters (which can be changed if desired, nevertheless).

From our point of view, the results as manifested by the clusters are reasonable. We have set the network to produce three clusters, because the data sets are small and we wanted to get clusters with three degrees of malignancy. In fact, the numerical outcome was comparable: the SOM regularly produced one larger cluster and two relatively small clusters (group I: 25, 6 and 9 cases; group II: 12, 4 and 4 cases; group III: 40, 8 and 12 cases). Closer examination showed that the large cluster included the majority of the cases with a low grade

of malignancy in all groups: they had the lowest Gleason scores and WHO grades in general; where known, they had the smallest probability of progression; they had the lowest number of gains and losses altogether. The two small clusters had cases with much worse properties. In general, these tended to higher WHO-grades and Gleason-scores, more gains and losses, and in group II progression was drastically more frequent. The main differences between the small clusters consisted mainly in the fact that one of these clusters had a higher number of gains – this could be shown reproducibly in all groups I–III. So we may summarize that our SOM managed, from the CGH information alone, to discern clusters in all groups which fit well to the biological characteristics of the tumours – strong evidence that in prostate cancer, genetic information as obtained by CGH is tightly coupled to biological aggressiveness.

Subsequently we analyzed the genetic differences between the clusters more thoroughly. As our research program is focused on early prostatic cancer and pre-neoplasic lesions, we have concentrated on losses of genetic material because these seem to form the earliest alterations of the genome here. Among these, losses at 8p, 6q and 13q were most frequent. In fact, loss of 8p seems to be associated with clusters with poor prognosis, whereas loss of 6q and 13q are not clearly associated with poor prognosis. This is easiest to observe in the small clusters of group II and III, where all or nearly all cases have a loss at 8p. From these data it seems that losses at 8p, 6q and 13q are all frequent in pT2N0 prostatic cancer, but the prognostic importance of a loss at 8p is larger than a loss at 6q or 13q. In general, loss of 8p in prostate cancer is well-known [1,4,6,7]. The eminent role of an 8p-loss in early carcinogenesis has been documented also for the urinary bladder [12]. It is tempting to speculate which changes at the level of the genes occur in the aforementioned losses. Potential tumor suppressor genes at 6q are: CCNC (cyclin C) and IGFR2 (insulin-like growth factor II) [5], at 8p: MSR1 (macrophage scavenger receptor I) [4] and N33 (Putative prostate cancer tumour suppressor) [18], and at 13q: RB1 (retinoblastoma 1) [25] and BRCA 2 (breast cancer 2) [30], but the relative role of these genes in prostate cancer has not been fully elucidated.

5. Conclusions

A cluster analysis of CGH data can be performed successfully with self-organizing maps. The results in-dicate that losses at chromosome arms 6q, 8p and 13q are all frequent in pT2N0 prostate cancer, but a loss at 8p seems to have the highest prognostic importance. The results indicate that a simple rule based on these preliminary findings of losses could improve prognosis.

References

[1] J.C. Alers, J. Rochat, P.-J. Krijtenburg, W.C.J. Hop, R. Kranse, C. Rosenberg, H.J. Tanke, F.H. Schröder and H. van Dekken, Identification of genetic markers for prostatic cancer progression, *Lab. Invest.* **80** (2000), 931–942.

[2] M. Bentz, A. Plesch, L. Bullinger, S. Stilgenbauer, G. Ott, H.K. Müller-Hermelink, M. Baudis, T.F. Barth, P. Möller, P. Lichter and H. Döhner, Positive mantle cell lymphomas exhibit complex karyotypes and share similarities with B-cell chronic lymphocytic leukemia, *Genes Chromos. Cancer* **27** (2000), 285–294.

[3] J.C. Bezdek and N.R. Pal, Two soft relatives of learning vector quantization, *Neural Networks* **8** (1995), 729–743.

[4] G.S. Bova, B.S. Carter, M.J.G. Bussemakers, M. Emi, Y. Fujiwara, N. Kyprianou, C.C. Jacobs, J.C. Robinson, J.I. Epstein, P.C. Walsh and W.B. Isaacs, Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer, *Cancer Res.* **53** (1993), 3869–3873.

[5] K.A. Cooney, J.C. Wetzel, C.M. Consolino and K.J. Wojno, Identification and characterization of proximal 6q deletions in prostate cancer, *Cancer Res.* **56** (1996), 4150–4153.

[6] M.R. Emmert-Buck, C.D. Vocke, R.O. Pozzatti, P.H. Duray, S.B. Jennings, C.D. Florence, Z. Zhuang, D.G. Bostwick, L.A. Liotta and W.M. Linehan, Allelic loss on chromomosome 8p12–21 in microdissected prostatic intraepithelial neoplasia, *Cancer Res.* **55** (1995), 2959–2962.

[7] W. Fu, L. Bubendorf, N. Willi, H. Moch, M.J. Mihatsch, G. Sauter and T.C. Gasser, Genetic changes in clinically organ-confined prostate cancer by comparative genomic hybridization, *Urol.* **56** (2000), 880–885.

[8] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, 1988.

[9] S. Joos, U.S.R. Bergerheim, Y. Pan, H. Matsuyama, M. Mentz, S. Du Manoir and P. Lichter, Mapping of chromosomal gains and losses in prostate cancer by comparative genomic hybridization, *Genes Chromos. Cancer* **14** (1995), 267–276.

[10] A. Kallioniemi, O.P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman and D. Pinkel, Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258** (1992), 818–821.

[11] O.P. Kallioniemi, A. Kallioniemi, A. Piper, J. Isola, F.M. Waldman, J.W. Gray and D. Pinkel, Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors, *Genes Chromos. Cancer* **10** (1994), 231–243.

[12] M.A. Knowles, M.E. Shaw and A.J. Proctor, Deletion mapping of chromosome 8 in cancers of the urinary bladder using restriction fragment length polymorphisms and microsatellite polymorphisms, *Oncogene* **8** (1993), 1357–1364.

[13] T. Kohonen, Self-organized formation of topologically correct feature maps, *Biological Cybernetics* **43** (1982), 59–69.

[14] T. Kohonen, *Self-Organization and Associative Memory*, 3rd edn, Springer-Verlag, Berlin, 1989.

[15] T. Kohonen, *Self-Organizing Maps*, 2nd edn, Springer, Heidelberg, 1997.

[16] T. Kohonen, J. Hynninen, J. Kangas and J. Laaksonen, SOM_PAK: The self-organizing map program package, Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland, 1996.

[17] J. Mao and A.K. Jain, A self-organizing network for Hyperellipsoidal Clustering (HEC), *IEEE Transactions on Neural Networks* **7** (1996), 16–29.

[18] D. Mac Grogan, A. Levy, G.S. Bova, W.B. Isaacs and R. Bookstein, Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22, *Genomics* **35** (1996), 55–65.

[19] P. Mangiameli, S.K. Chen and D. West, A comparison of SOM neural network and hierarchical clustering, *Eur. J. Operat. Res.* **93** (1996), 402–417.

[20] T. Mattfeldt, H.A. Kestler, R. Hautmann and H.W. Gottfried, Prediction of prostatic cancer progression after radical prostatectomy using artificial neural networks: a feasibility study, *BJU International* **84** (1999), 316–323.

[21] T. Mattfeldt, H.A. Kestler, R. Hautmann and H.-W. Gottfried, Prediction of postoperative prostatic cancer stage on the basis of systematic biopsies using two types of artificial neural networks, *Eur. Urol.* **39** (2001), 520–527.

[22] B.J.T. Morgan and A.P.G. Ray, Non-uniqueness and inversions in cluster analysis, *Appl. Statist.* **44** (1996), 117–134.

[23] F.K. Mostofi, I.A. Sesterhenn and L.H. Sobin, eds, Histological typing of prostate tumours, in: *International histological classification of tumours*, Vol. 22. World Health Organization, Geneva, 1980.

[24] N. Nupponen and T. Visakorpi, Molecular cytogenetics of prostate cancer, *Microsc. Res. Techn.* **51** (2000), 456–463.

[25] S.M.A. Phillips, C.M. Barton, S.J. Lee, D.G. Morton, D.M. Wallace, N.R. Lemoine and J.P. Neoptolemos, Loss of the retinoblastoma susceptibility gene (RB1) is a frequent and early event in prostatic tumorigenesis, *Br. J. Cancer* **70** (1994), 1252–1257.

[26] H. Ritter, K. Schulten and T. Martinetz, *Neuronale Netze*, Addison-Wesley, Bonn, 1990.

[27] H.P. Sattler, V. Rohde, H. Bonkhoff, T. Zwergel and B. Wullich, Comparative genomic hybridization reveals DNA copy number gains to frequently occur in human prostate cancer, *Prostate* **39** (1999), 79–86.

[28] F.H. Schroeder, P. Hermanek and L. Denis, The TNM classification of prostate carcinoma, *Prostate* (Suppl.) **4** (1992), 129–139.

[29] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander and T.R. Golub, Interpreting patterns of gene expression with self-organizing maps: Methods and application to hemopoietic differentiation, *Proc. Natl. Acad. Sci.* **96** (1999), 2907–2912.

[30] T. Wooster, S.L. Neuhausen, J. Mangion, Y. Quirk, D. Ford, N. Collins, K. Nguyen, S. Seal, T. Tran, D. Averill, P. Fields, G. Marshall, S. Narod, G.M. Lenoir, H. Lynch, J. Feunteun, O. Devilee, C.J. Cornelisse, F.H. Menko, P.A. Daly, W. Ormiston, R. McManus, C. Pye, C.M. Lewis, L.A. Cannon-Albright, J. Peto, B.A.J. Ponder, M.H. Skolnick, D.F. Easton, D.E. Goldgar and M.R. Stratton, Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13, *Science* **265** (1994), 2088–2090.