Paper III

Correlation Analysis of Two-Dimensional Gel Electrophoretic Protein Patterns and Biological Parameters: p53 biosignatures reflect origin of cancer and stage of differentiation

Werner Van Belle[1§], Nina Ånensen[2], Ingvild Haaland[2], Øystein Bruserud[2,3] Bjørn Tore Gjertsen[3§]

[1]Bioinformatics Group, Norut IT, Research Park Tromsø

[2]Institute of Medicine, Hematology Section

University of Bergen, Bergen, Norway

[3]Department of Internal Medicine, Hematology Section

Haukeland University Hospital, Bergen, Norway

§Corresponding author correlation technique:

Werner Van Belle

Norut IT, Bioinformatics Group

Research park Tromsø

Postboks 6434, N9294 Tromsø

Email: werner.van.belle@itek.norut.no

Phone: +4777629404

Fax: +47776 29401

Corresponding author p53 biology:

Bjørn Tore Gjertsen

Institute of Medicine, Hematology Section

Haukeland University Hospital

University of Bergen, N-5021 Bergen, Norway

Email: bjorn.gjertsen@med.uib.no

Phone: +4755975000

Fax: +4755972950

Abstract

Two-dimensional gel electrophoresis (2DE) is a powerful technique to examine the post-translational modifications of complexly modulated proteins. We developed a novel correlation technique that determines the relationship between visualized two-dimensional gel electrophoresis images and biological data without requiring human interpretation on a gel per gel basis. After image alignment and normalization, the biological parameters and pixel values were replaced with their specific rank. The rank adjusted images and parameters were then used as input into a standard linear Pearson correlation and thereafter tested for significance and variance. The technique was first explored in a simulated dataset, thereafter applied on p53 2DE immunoblots from cancer cells known to have unique signaling networks. A p53 protein is targeted by a number of signaling networks, and is involved in growth regulation, cell death and differentiation. The p53 correlation analysis distinguished between the cancer forms acute lymphoblastic leukemia and acute myeloid leukemia. Further analysis of p53 and the differentiation stage of acute myeloid leukemia indicated that the level of differentiation in cancer might be reflected in the p53 protein pattern. This proposes that p53 protein isoforms can be read as a biosignature, and that the novel correlation method may be important for exploration of complex signal convergence on regulatory proteins in biological systems.

1 Introduction

Two-dimensional gel electrophoresis (2DE) has been a successful technique for identification and visualization of post-translational modifications [1] (reviewed in [2]), and is increasingly used to determine accessible parts of the proteome in human cells [3]. Only to a limited extent has 2DE been used to propose diagnosis or clinical classification in diseases [4]. The amount and complexity of data obtained from 2DE gel patterns have led to the development of analysis software for digitalized images [5, 6, 7], but human interpretation and validation of the data is usually necessary. Typically, one of the steps in 2DE analysis is the selection of spots followed by description of their position, volume and other parameters. Current methods for spot detection assume regular spot shapes [8] or model spots as bivariate Gaussian densities [9], and therefore cannot discriminate spot shapes and irregularity [10, 11]. We developed a method that omits the spot detection phase and does not require human interpretation on a gel-to-gel basis.

Given a set of gel images, the technique measures correlation between every pixel position and an external parameter. This makes it possible to study the protein distribution on 2DE gel as well as the actual relation to the external parameter. The method has been rigorously tested on a set of simulated 2DE gels with different levels of background, additional noise and outliers. Real-life evaluation of the technique was performed by testing the correlation analysis on p53 protein expression in cell samples from patients with well-characterized hematological malignancies.

The p53 protein regulates differentiation, growth and cell death through its complex post-translational modifications caused by multiple signaling networks directly or indirectly targeting the protein[12, 13]. The TP53 gene is frequently mutated in many cancers, and mutations in signaling pathways acting on p53 protein are found both in sporadic and hereditary cancers [14]. During differentiation, p53 undergoes modifications like phosphorylation and acetylation and is suggested to be involved in differentiation of acute myeloid leukemia [15, 16]. Different hematological malignancies, like acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) [17], are characterized by mutations or

varying expression of cell signalling encoding genes [18, 19]. We therefore hypothesized that a wide examination of the extensive post-translationally modified p53 protein could comprise information about type of cancer and the differentiation stage of the cancer.

A total of **39** unique acute myeloid leukemia (AML) and **8** unique acute lymphoblastic (ALL) patients were analyzed by 2DE and immunoblotting for visualization of the p53 protein pattern by an amino-terminal targeting antibody Bp53-12. The p53 patterns were correlated to cancer type (AML vs ALL) and the standardized French-America-British differentiation classification [20].

The overall results demonstrated that the presented correlation technique may aid the elucidation of complexly regulated proteins in biological systems.

2 Results

Overview of the study - The work-flow and the concept of the proposed correlation is indicated in Fig. 1. In order to illustrate how the correlation images ought to be interpreted, a simulated gel stack with defined spot characteristics in function of an external parameter t was created (Fig. 2). This simulation reassured a controlled environment in which the algorithmic behavior was observed. For simulation purposes t was chosen as the gel number in the simulated stack. In practice it could represent any biological parameter as illustrated later.

Spot positions, sizes and translations - The correlation analysis of the simulated gel stack (Fig. 2) present information about location, volume and translation of spots related to the external parameter t. The spots α and β in the gel stack (Fig. 2A) are altered in function of t. They are visualized in the correlation images (Fig 2B) at the same position, showing that the correlation image offers correct positional information. The correlation strength is presented in shades of green (positive correlation) and brown (negative correlation or anti-correlation). If spots do not relate to t, they might go unnoticed, as illustrated by the γ-spots. The two constant γ-spots were designed to be independent of the

gel sequence number t. This resulted in no visible correlation in Fig 2Bab. If the γ-spots were made to correlate with t due to the application of normalization, then the correlation image (Fig. 2Bc) visualized two spots of different sizes, reflecting the volume of the spots.

The δ-spots illustrated detection of spot shifts related to the external parameter. The original and destination positions will respectively correlate, then anti-correlate, resulting in a smear in the correlation image (Fig. 2B).

Reading of spot shape - All images in Fig. 2B showed the α-spot anti-correlate in the middle and correlate at its periphery. This is consistent with the creation of the gel-stack in which the amplitude of spot α lowers from 5.0 to 1.0 while the spots broadens from 10 to 100 pixels. Because the central spot widens, higher gel numbers will have relatively more signal in the periphery. This indicates that spots where diffusion-like alteration dominates will be detected based on difference in correlation between the inner and outer areas. Similar behavior can be observed in the shape changing β-spot. The initial vertical shape (low t-value) anti-correlates (it disappears) while the later horizontal shape (at higher t-values) correlates (it appears).

Significance visualized by masking the correlation image - In the simulated gel-stack, empty areas have a constant intensity. The raw correlation analysis indicates a strong correlation (Fig. 2Ba) or anti-correlation (Fig 2Bb) for those areas. Mathematically this is correct, but because of a lack in intensity variation (the area is constant) this correlation is without information. The putative empty information was filtered out using two masks, one mask to remove non significant correlations and a second mask to remove areas without variance (see material and methods, Step 4 for details). After masking the correlation image (Fig. 2B(abc)), only the areas with relevant spot modulations are indicated (Fig. 2B(a',b',c')).

Effect of different normalizations - Different background removal and scaling techniques on the simulated gel-stack were tested (Fig. 2), including background subtraction and background division. The original information that led to the creation of the gel-stack was

retrieved in all cases. The α, δ and β spot correlations were always visualized, indicating that the normalization technique used is of little importance for qualitative analysis. In the particular case of gel normalization obtained by division through the mean gel intensity, new information was found that did not directly originate from the creation of the simulation (Fig. 2Bc). Due to a t-dependent energy increase in spot α the mean intensity of the gel increases. As a result, the original constant γ-spots decreased in energy (division by a larger number leads to lower values). The γ-spots became t dependent and thus showed up in the correlation image. This unexpected behavior does not reduce the analytical power of the correlation method. Since normalization is performed on an individual gel basis it can always be repeated on any gel without taking into account previous gels. Quantitatively, normalization factors strongly influence correlation measures. If the technique is used as a quantitative method, then calibration spots ought to be used and exact understanding of machine specifications and camera properties should be known.

Background Correlations and the effect of noise in 2DE images - Adding noise to the images attenuated the appearance of background in the correlation images (Fig. 3Aa-c). The noised background was white in the correlation images, while the background was either correlating positive or negative in Fig. 2B. This indicates that small amounts of noise might enhance interpretation of the correlation analysis. Increasing noise up to 75% (of maximum image intensity) resulted in weaker correlations, but still important spots were identifiable (Fig. 3A a-c). Increasing the noise level in pace with increasing gel stack number (t-value) resulted in correct information about the negative correlation, but loss of information about the positive correlation (Fig 3Ad,d'). Such a situation could occur if an automatic camera acquired images at waning signal strength. We conclude that presence of uniform or Gaussian distributed noise in the dataset barely influences the analytical power of the presented correlation test.

Effect of randomization of the dataset - Two sets of random data were generated to be used as t-value. Instead of testing correlation towards the sequence number t, we now determined the effect of correlation of the images towards noise. The IDL function

'randomu' [21], generated the normally distributed pseudo random numbers. In the correlation images we always recognized the same shapes as with the original t-values. Areas that behave similarly had distinct pseudo-coloring, regardless of the external biological parameter set. This emphasizes the robustness of the algorithm to identify regions of interest. Fig 3B was made with the particular seeds 287 and 2874.

Outliers - A test with t-value outliers shows limited impact on the interpretation of the gels (Fig. 3B). We changed the t-values from {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14} to {0, 1, **15**, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}, resulting in a slight change in actual correlation magnitude, but the information content was surprisingly well preserved. Even with 13% outliers {0, 1, **15**, 3, 4, 5, 6, **4**, 8, 9, 10, 11, 12, 13, 14}, the original information is recovered. This is mainly due to the robust correlation which relies on ranking of the dataset instead of the numerical values (both the t-values and the image pixel values are ranked).

Application of correlation technique in cancer cells - We tested the algorithm on a set of p53 2DE immunoblots from acute leukemia. We first examined p53 2DE immunoblots in ALL versus AML (Fig. 4A). The impact of wrong diagnosis was examined by random swapping ALL and AML labels, resulting in lower correlation as expected (Fig. 4B). The strong positive correlation between ALL and p53-α and δ forms was examined in the normal and mature cell counterpart of ALL and AML, lymphocytes and granulocytes (Fig. 4C). Granulocytes hardly expressed α or δ p53 protein, while lymphocytes showed predominantly p53-δ forms, supporting the positive p53 correlation with lymphatic derived leukemia.

Correlation between p53 protein isoforms in acute myeloid leukemia (AML) and differentiation stage - Based on the observation that p53 is involved in leukemic cell differentiation [15, 22, 16, 23], and the fact that the standardized and widely used French-American-British (FAB) classification of AML is based on morphologically determined stage of myeloid maturation and direction of maturation [24, 20], we searched for correlations between FAB classification and the p53 2DE pattern. Using 73 gels we found striking correlations (Fig. 5). Image A is the masked correlation landscape, image B is the raw correlation image. The observations were that: a) The tail of the p53-α isoform

correlates negatively to the FAB classification (profile 4, region g and h). b) The p63 area correlates positively towards the FAB classification (profile 3, the i region). c) The p53-δ region has four positively correlating articulated spots (profile 1, a-d, r=0.2), d) The p53 sub-δ region has two negatively correlating spots (profile 2e,f). The combination of a positive correlation at the p53-δ region and a negative correlating sub-δ region indicates a spot shift from one area to another. e) Presence of the super-δ negative correlating region indicates that a change of spot shape also occurs. When the spots are larger and diffuse then the patient is classified as M0, M1 or M2. If the spots in the δ region are clear articulated and smaller, the patient are either M4 or M5. None of the above correlations are strong (r 0.25 using the stringent Spearman rank order correlation). Nonetheless they can all be observed in the 2DE images, which means that they can form an important tool in stratification of patients.

The presented correlation includes M3, a distinct subgroup of AML with signs of granulocytic differentiation, featuring the translocation t(15;17) and responsiveness to retinoic acid therapy [25]. FAB M3 is therefore a separate entity in the recent WHO classification [26]. The correlation of M0/1 versus M4/5 was weaker when M3 was removed (data not shown). This observation suggests that it is the maturation distance from M0/1 that is mirrored in the M3-including analysis (Fig. 3B); and it may also suggest that p53 patterns associated with M3 are unique, thereby contributing to a greater splitting of the patients into subgroups.

3 Discussion

In our opinion, the presented correlation technique offers unique advantages over standard 2D gel analysis techniques relying on spot detection methods, including (i) robust and general analysis; (ii) natural spot position comparison; (iii) fast interpretation; and (iv) no strict need for calibration.

The technique does not require perfect gels. Background noise and outliers do not influence the quality of the analysis. Since we rely on ranking of the data set, outliers (whether they

are in the gel images or in the external parameters) do not attribute any significant impact to the correlation image.

Spot detection and selection have become unnecessary. The analysis takes into account all information that does not easily fit into a bivariate Gaussian distribution. This implicates that spot shapes, tails and areas are all treated equally.

Because spots never occur at exactly the same position, a certain amount of jitter is allowed by spot detection methods before spots are considered different. Our analysis does not require such parameters and will automatically tune to the required resolution. When spots are small, then small translations will be considered relevant. When spots are broad, larger shifts are necessary to detect translation.

Standard procedures of selecting and quantifying spots take a long time due to parameter tuning. Since our technique does not require this, it is much faster. Also, the computational requirements are limited. Standard computers can easily calculate large correlation images in matters of minutes. Spot detection can afterwards be used on the final correlation image and can in some cases be a method for verification of the observed correlations.

We demonstrated that the correlation technique works surprisingly well without calibrated intensities. The use of mean background division and RMS scaling offers the same information quality as relying on exactly calibrated values.

A major drawback of the presented correlation technique is its requirement for properly aligned gels. Aligning many different gels is not an easy task without calibration spots. One often needs to rotate, translate and zoom gels appropriately. Another drawback is that the method does not allow comparison of s*pot volumes*. Instead of measuring protein concentrations related to spot volumes, this technique measures intensities on an individual pixel basis and provides information on the shape changes of spots.

Recently we demonstrated that signaling networks may be altered and potentiated in cancer cells suggesting a prognostic meaningful classification [27]. This includes altered p38

MAP-kinase signaling, known to phosphorylate p53. Signaling networks are emerging targets of new therapy [28], and the p53 biosignatures may be the result of integrated information about signaling networks in cancer, we hypothesize that these p53 signatures might be used for future therapy individualization. The correlation analysis of p53 in ALL versus AML suggests that p53 isoforms analyzed by two-dimensional immunoblots represent a biosignature that comprise information about the origin of cancer (Fig. 4). It is probable that the p53 biosignature is formed by the combinations of splice forms of p53 and by its various post-translational modifications [12, 29]. The p53 protein is tightly regulated by post-translational modifications [12], thereby involved in several positive and negative feedback networks [13]. This has ignited the hypothesis that p53 integrates information from all these signaling networks [30]. ALL and AML comprise different genetic abnormalities [25, 31], and analysis of growth factor receptor expression and global gene expression has pointed out that the expression of receptor tyrosine kinases and signaling modulators are different [32, 33]. Therefore, since the p53 protein is implied in various cancer related signaling networks, it is probable that a p53 isoform analysis using two-dimensional immunoblot provides a biosignature distinguishing ALL from AML.

Phosphorylation of p53 Ser315 is necessary for differentiation in mouse embryonic stem cells [34], and p53 is also able to direct differentiation in acute myeloid leukemia cell lines [23, 16]. The p53-deficient HL-60 cell line has potential for both monocytic and granulocytic differentiation, and introduction of wild type p53 direct differentiation in the granulocytic direction [22]. Recent reports indicate that the FAB classification, in particular the distinction between M1-2 and M4-5 in maturation level and direction of maturation, is associated with certain gene classes in unsupervised clustering of gene expression profiles [35, 36]. These observations suggest that the p53 biosignatures reflect the stage and direction of myeloid differentiation to certain extent. The p53 recognizing antibody Bp53-12 also detects p63 [37], and the correlation analysis supports that the p53 family-member p63 is increased at increased differentiation (Fig. 5A). This observation is interesting since p63 is more strongly related to embryonic development and differentiation than p53 [38].

Future development of the method should include 2DE gels with built in standards (pI, MW) and known protein amounts, together with consideration of hardware-parameters, grey value scales, noise types and noise floors. This may allow quantitative measurement of spot modulation.

Taking the correlation further than the pixel level should be possible by using canonical correlations and development of significance measures that take into account the similarity of neighboring pixels [39, 40].

Another worthwhile alley is prediction of external parameters based on a large training set. Using multivariate discriminant analysis [41], this might offer prospects for patient stratification. Finally, it should also be possible to insert clustering algorithms to pseudo color the final image or rely on image segmentation algorithms to classify areas automatically [42, 43]. However, even in its present provides our correlation technique a method to perform correlation analysis of complex biosignatures and signal network responses.

4 Material & Methods

The 2DE gel correlation technique relies on a large amount of 2DE gels of a biological system. Every gel needs to be described by an external numerical measure. E.g: life expectancy, differentiation stage of a cell sample, age of an organism, origin of a cancer cell sample, effect of cancer therapy, time, temperature, pressure, cell size, and so on. For every n gels (described as A_z in which z is the gel image number), there are n external parameters, described as T_z. Gels can further be annotated as $A_{x,y,z}$ in which (x, y) is the position on gel number z. $A_{x,y}$ is a vector containing the intensities of all gels:
$A_{x,y} = \left[\begin{array}{cccc} A_{x,y,1} & A_{x,y,2} & \dots & A_{x,y,n} \end{array} \right]$.

Step 1: Alignment and registration - The method requires proper direction and alignment of all gels. Presence of calibration spots facilitates this process, otherwise techniques such as Hough transformation [44, 45] for gel direction measurement and cross correlation [46] for multiple gel alignment can be used. Once the gels are aligned, further

basic warping and registration [47] techniques are useful to account for small shifts between the different gels. The aligned images are denoted $A_z^{'}$.

Step 2: Intensity Normalization - The second step normalizes the intensity values of the gels to allow for inter-gel pixel comparison. There is little known of the relation between pixel intensities and protein concentrations. Furthermore, pixel values can be relative or gamma corrected, depending on the hardware. The wide variety of possible pixel value interpretations leads us to embrace the use of relative grey values. The simulated gel stack showed that the choice of normalization technique barely influences the final correlation image.

Step 2a: Background intensity - The background floor of a 2DE gel refers to the brightness of empty gel areas. Different capture techniques produce different background floors. Background signal can be either added to all pixel values (additive background), or it can cumulate with a decaying signal (multiplicative background). As previously observed [48], most cameras introduce a mixture of additive and multiplicative backgrounds. Removal of additive noise can be done through subtracting the mean ($A_z^{"} := A_z^{'} - \overline{A_z^{'}}$) or median value ($A_z^{"} := A_z^{'} - median(A_z^{'})$). Removal of multiplicative noise can be done through $A_z^{"} := \frac{A_z^{'}}{\overline{A_z^{'}}} - 1$. We would emphasize that whatever normalization scheme is used in this step, it should be performed on an individual gel basis.

Step 2b: Scaling of gel intensity - After removal of the background floor, the dynamic range of the image is normalized through scaling of gel intensities. The presence of a calibration spot eases this process. If A' is the non-relative image and (x, y) is the calibration spot position, then the image $A^{"} := \frac{A^{'}}{A_{x,y}^{'}}$ defines the normalized image. Without calibration spot the total energy content (sum of all intensities or RMS value) forms a very reasonable scaling means: $A_z^{"} = \frac{A_z^{'}}{RMS(A_z^{'})}$

Step 3: Correlation Image - After alignment and normalization, the correlation analysis generates a new image visualizing the correlation measure between a specific position and an external parameter. The correlation image is composed of pixels, each testing one position

on the gel. The result of each test is a number between -1.0 (anti-correlation) and 1.0 (correlation), which, after appropriate scaling, defines the pixel color in the correlation image. The two vectors participating in the test are $A^{"}_{x,y}$ and B. The first vector contains the gel expression levels at position (x, y). Given 89 gel images, $A^{"}_{x,y}$ will contain 89 different expression values; one for each gel. The second vector B contains 89 external values associated with every gel. Repeating this correlation test for every pixel results in the correlation image C (Eq. 1)

$$C_{x,y} = \rho(A^{"}_{x,y}, T) \tag{1}$$

The correlation image can be visualized using different color schemes. In Fig. 1 green indicates positive correlations and brown negative correlations.

The preferred correlation is the robust Spearman rank order correlation (ρ-correlation)[21]. This non-parametric test allows us to ignore the specific distributions of gel intensity levels and external parameters. ρ-correlation requires a ranking of the two participating vectors and then relies on a standard linear Pearson correlation.

Step 4: Masking - Correlation does not necessarily imply a causal, significant, or useful relationship. To filter out some possibly useless relations, a number of masks limit the visible correlations. The first mask removes correlations that might be occurring by coincidence: some datasets easily correlate with any other dataset (significance). The second mask removes correlations that offer little useful information (E.g: a dataset containing all zero's).

Step 4a: Significance - To remove correlations that have a high probability of occurring, the significance test typically associated with the Spearman correlation test was used. In this context, it is defined as

$$S_{x,y} = 1 - C_{x,y}\sqrt{\frac{n-2}{1-C^2_{x,y}}} \tag{2}$$

If this number is close to 1 then there exists a low probability that some random data would happen to correlate with the given result set. Likewise, if this number is 0 then there exists a high probability that the correlation is coincidental.

Step 4b: Variance - The second mask avoids strong and significant correlations that have a low biological significance because the gel intensities do not change enough. It relies on the standard deviation [49] measured on the relative, non-ranked, gel intensities

$$D_{x,y} = \frac{\sqrt{\sum_{z=0}^{n-1}(\frac{A''_{x,y,z}}{A''_{x,y,*}} - 1)^2}}{N} \tag{3}$$

The standard variance (or RMS) of the mean divided gels will have a large value where there is a varying gel expression. At places where the gel expression is constant this value will be zero.

Step 4c: The masked correlation image - Multiplying the standard deviation mask (Eq. 3) with the significance mask (Eq. 2) gives a new mask that can be superimposed over the correlation image (Eq. 1).

$$R = C \times S \times D$$

The pixel values of R no longer relates to the correct correlation measure. Therefore, R forms an indicator, showing position of possible interest.

Simulation of a two-dimensional gel electrophoresis stack - The simulated gel-stack is based on the animation of different 2D gauss 'bumps', defined as

$$G(x,y) = a.exp(-\frac{(\frac{x-cx}{wx})^2 + (\frac{y-cy}{wy})^2}{2})$$

(cx, cy) is the center position, wx and wy are the width and height respectively. a is the amplitude of the curve. Based on this Gaussian 'bump' a gel-stack, containing 15 different

gels was constructed. Every gel contains: I) an out-fading spot (Fig 2, spot α) with a growing radius from 10 to 100 pixels and lowering amplitude from 5.0 to 1.0. II) An elliptical spot (Fig 2, spot β) which changes shape from being small and tall ($wx = 10$, $wy = 40$, $a = 5$) to broad and flat ($wx = 40$, $wy = 10$, $a = 5$). III) Two spots with minimal (1.0) and maximal (5.0) amplitudes (Fig 2, spots γ). IV) A moving spot (Fig 2, spot δ) from left to right. Every gel is annotated with its position in the stack.

Patients, leukemic cell separation and sample preparation - The study was approved by the local Ethics Committee and samples collected after informed consent. Cell separation, storage and culture of patient AML blasts were performed as previously described [50, 27]. ALL and AML blasts were isolated by density gradient separation with Lymphoprep (Nycomed Pharma AS, Oslo, Norway) and contained more than 90% malignant cells. Patient characteristics and number of gels are presented in Table 1. Normal granulocytes (97% neutrophile) and lymphocytes (peripheral blood mononuclear cells containing 10% monocytes and predominantly T lymphocytes) were separated by density gradient centrifugation combining Polymorphprep TM (Axis-Shield PoC AS, Oslo, Norway) and Lymphoprep following the manufacturers instructions. Preparation for two dimensional gel electrophoresis (2DE) and immunoblotting was performed as previously described [51, 52, 53]. Briefly, cells were washed in NaCl (9 mg/ml) and then lysed in 7% trichloroacetic acid. The precipitated protein was washed once in 5% trichloroacetic acid and three times in water saturated ether to remove salts. The protein pellet was resuspended in sample buffer for 2DE gel electrophoresis (7 M urea, 2 M thiourea, 100 mM dithiotreitol, 1.5% Ampholyte 3 - 10, 0.5% Ampholyte 5 - 6, 0.5% CHAPS). 2D was performed using 7 cm pH 3-10 (Zoom Strip, Invitrogen Corp., Carlsbad, CA, USA) isoelectric focusing gel strips, following the manufacturers' instructions. Electrophoresis was performed at 200 V for 60 minutes, after which the proteins were transferred to polyvinylidene fluoride membrane (Amersham Biosciences AB, Uppsala, Sweden) by standard electro-blotting. p53 protein was detected using primary Bp53-12 antibody (Santa Cruz Biotechnology, CA, USA) and secondary horse radish peroxidase conjugated mouse antibody (Jackson ImmunoResearch, West Grove, PA, USA) visualized using the Supersignal west Pico or Supersignal west Femto

system (Pierce Biotechnology, Inc., Rockford, IL, USA). Chemiluminescence imaging was performed using a Kodak Image Station 2000R (Eastman Kodak Company, Lake Avenue, Rochester, NY, USA) and were saved in TIFF format with the resolution of 300 dpi for correlation analysis.

5 Acknowledgements

The expert help of Nancy Gerits in preparing this manuscript and the expert technical assistance of Siv Lise Bedringaas and Stein-Erik Gullaksen is highly appreciated. This study was supported by The National Program for Research in Functional Genomics in Norway (FUGE grant no. 151859) at The Research Council of Norway, Innovation Norway and a Norut IT intramural grant.

References

[1] Farrell, P. H. O. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250, 4007–21 (1975).

[2] Gorg, A., Weiss, W., and Dunn, M. Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, 4, 3665–3685 (2004).

[3] Celis, J., Moreira, J., Cabezon, T., Gromob, P., Friis, R., et al. Identification of extracellular and intracellular signaling components of the mammary adipose tissue and its interstitial fluid in high risk breast cancer patients: toward dissecting the molecular circuitry of epithelial-adipocyte stromal cell interactions. *Mol Cell Proteomics*, 4, 492–522 (2005).

[4] Boyd, R., Adams, P., and et al., S. P. Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B. *Leukemia*, 17, 1605–1612 (2003).

[5] Garell, J. Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. *J. Biol. Chem.*, 254, 7961–7977 (1979).

[6] Curch, S. Advances in two-dimensional gel matching technology. *Biochem Soc Trans*, 32, 511–516 (2004).

[7] Blose, S. and Hamburger, S. Computer-analyzed high resolution two-dimensional gel electrophoresis: a new window for protein research. *Biotechniques*, 3, 232–236 (1985).

[8] Horgan, G. and Glasbey, C. A. Uses of digital image analysis in electrophoresis. *Electrophoresis*, 16, 298–305 (1995).

[9] Appel, R., Hochstrasser, D., Funk, M., Vargas, J., Pellegrini, C., et al. The melanie project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis*, 12, 722–735 (1991).

[10] Anindya, R., Kwan, R. L., Yaming, H., Marten, M., and Babu, R. Analyzing two-dimensional gel images. Technical report, Department of Mathematics and Statistics, University of Maryland (2003).

[11] Schlags, W., Walther, M., Masree, M., Kratzel, M., Noe, C. R., et al. Towards validating a method for two-dimensional electrophoresis/silver staining. *Electrophoresis*, 26, 2461–2469 (2005).

[12] Bode, A. and Dong, Z. Post-translational modifications of p53 in tumorigenesis. *Nat. Rev. Cancer*, 4, 793–805 (2004).

[13] Harris, S. and Levine, A. The p53 pathway: positive and negative feedback loops. *Oncogene*, 24, 2899–908 (2005).

[14] Lonning, P. Genes causing inherited cancer as beacons to identify the mechanisms of chemoresistance. *Trens Mol. Med.*, 10, 113–118 (2004).

[15] Shen, D., Real, F., DeLeo, A., Old, L., Marks, P., et al. Protein p53 and inducer-mediated erythroleukemia cell commitment to terminal cell division. *Proc Natl Aca Sci USA*, 80, 5919–22 (1983).

[16] Rizzo, M., Zepparoni, A., Cristofanelli, B., Scardigli, R., Crescenzi, M., et al. Wt-p53 action in human leukemia cell lines corresponding to different stages of differentiation. *Br J Cancer*, 77, 1429–1438 (1998).

[17] Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., et al. The importance of diagnostic cytogenetics on outcome in aml: analysis of 1612 patients entered into the mrc aml 10 trial. *The Medical Research Council Adult and Children's Leukaemia Working Parties; Blood*, 92, 2322–33 (1998).

[18] Gilliland, D., CT, C. J., and Felix, C. The molecular basis of leukemia. *Hematology (Am. Soc. Hematol. Educ. Program)*, pages 80–97 (2004).

[19] Chaladon, Y. and Schwaller, J. Targeting mutated protein tyrosine kinases and their signaling pathways in hematologic malignancies. *Haematologica*, 90, 949–68 (2005).

[20] Bennett, J., Catovsky, D., Daniel, M., Flandrin, G., Galton, D., et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol*, 33, 451–458 (1976).

[21] Veterling, W. T. and Flannery, B. P. *Numerical Recipes in C++*, chapter 10. Cambridge University Press, 2nd edition (2002).

[22] Soddu, S., Blandino, G., Citro, G., Scardigli, R., Piaggio, G., et al. Wild-type p53 gene expression induces granulocytic differentiation of hl-60 cells. *Blood*, 83, 2230–7 (1994).

[23] Tang, P. and Wang, F. Induction of IW32 erythroleukemia cell differentiation by p53 is dependent on protein tyrosine phosphatase. *Leukemia*, 14, 1292–1300 (2000).

[24] Bennett, J., Catovsky, D., Daniel, M., Flandrin, G., Galton, D., et al. Proposal for the recognition of minimally differentiated acute myeloid leukemia (Aml-Mo). *Br J Haematol*, 78, 325–329 (1991).

[25] Stone, R., O'Donnell, M., and Sekeres, M. Acute myeloid leukemia. hematology. *Am Soc Hematol Educ Program*, pages 98–117 (2004).

[26] Harris, N., Jaffe, E., Diebold, J., Flandrin, G., Muller-Hermelink, H., et al. 1999 world health organization classification of neoplastic diseases of the hematopietic and lymphoid tissues: report on the clinical advisory committee meeting, arlie house, virginia. *J. Clin Oncol*, 17, 3835–384 (1997).

[27] Irish, J., Hovland, R., Krutzik, P., Perez, O., Bruserud, Ø., et al. Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell*, 118, 217–228 (2004).

[28] Adjei, A. and Hidalgo, M. Intracellular signal transduction pathway proteins as targets for cancer therapy. *J Clin Oncol*, 10, 5386–403 (2005).

[29] Bourdon, J., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., et al. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev*, 19, 2122–37 (2005).

[30] Meek, D. Multisite phosphorylation and the integration of stress signals at p53. *Cell Signal*, 10, 159–66 (1998).

[31] Hoelzer, D., Gokbuget, N., Ottmann, O., Pui, C., Relling, M., et al. Acute lymphoblastic leukemia. *Hematology (Am Soc Hematol Educ Program)*, pages 162–92 (2002).

[32] Waele, M. D., Renmans, W., Gucht, K. V., Jochmans, K., Schots, R., et al. Growth factor receptor profile of cd34+ cells in aml and b-lineage all and in their normal bone marrow counterparts. *Eur J Haematol*, 66, 178–87 (2001).

[33] Sakhinia, E., Faranghpour, M., Yin, J. L., Brady, G., Hoyland, J., et al. Routine expression profiling of microarray gene signatures in acute leukaemia by real-time pcr of human bone marrow. *Br J Haematol*, 130, 233–48 (2005).

[34] Lin, T., Chao, C., Saito, S., Mazur, S., Murphy, M., et al. p53 induces differentiation of mouse embryonic stem cells by suppressing nanog expression. *Nat Cell Biol.*, 7, 165–171 (2005).

[35] Bullinger, L., Dohner, K., Beir, E., Frohling, S., Schlenk, R., et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med*, 16, 1605–16 (2004).

[36] Valk, P., Verhaak, R., Beijen, M., Erpelinck, C., van Waalwijk van Doorn-S. Khosrovani, B., et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*, 16, 1617–28 (2004).

[37] Ånensen, N., Oyan, A., Abrahamsen, J., Kalland, K., Øyestein Bruserud, et al. A distinct p53 protein isoform signature reflects the onset of induction chemotherapy for acute myeloid leukemia. *Submitted* (2005).

[38] Moll, U. and Slade, N. p63 and p73: roles in development and tumor formation. *Mol Cancer Res*, 2, 371–8 (2004).

[39] Branco, J., Croux, C., Filzmoser, P., and Oliviera, M. Robust canonical correlations: A comparative study. *Computational Statistics*, 20, 203–229 (2005).

[40] Dehon, C., Filzmoser, P., and Croux, C. Data analysis, classification, and related methods. pages 321–326 (2000).

[41] Croux, C. and Joossens, K. Influence of observations on the misclasification probability in quadratic discriminant analysis. *Journal of Multivariate Analysis* (2005).

[42] Stutz, J. and Cheeseman, P. *Maximum Entropy and Bayesian Methods, Cambridge 1994*, chapter AutoClass - a Bayesian Approach to Classification. Kluwer Acedemic Publishers, Dordrecht (1995).

[43] Cheeseman, P. and Stutz, J. *Advances in Knowledge Discovery and Data Mining*, chapter Bayesian Classification (AutoClass): Theory and Results. AAAI Press/MIT Press (1996).

[44] Gonzalez, R. C. and Woods, R. E. *Digital Image Processing*, chapter 7, pages 432–438. Prentice Hall, Upper Saddle River, New Jersey 07458, 2nd edition edition (2002). ISBN: 0201180758.

[45] Hough, P. Methods and means for recognizing complex patterns. *US Patent 3,069,654* (1962).

[46] Conradsen, K. and Pedersen, J. Analysis of two-dimensional electrophoresis gels. *Biometrics*, 48, 1273–1287 (1992).

[47] Wang, X. and Feng, D. D. Hybrid registration for two-dimensional gel protein images. In *Third Asia Pacific Bioinformatics Conference (APBC2005)* (2005).

[48] Belle, W. V., Sjøholt, G., Ånensen, N., Høgda, K.-A., and Gjertsen, B. T. A novel denoising technique that faciltates two-dimensional electrophoretic gel visualization, orientation, alignemtn and spot detection. *Submitted for publication* (2005).

[49] Kenny, J. and Keeping, E. *The Standard Deviation and Calculation of the Standard Deviation*, chapter 6.5–6.6, pages 77–80. Princeton NJ, 3th edition (1962).

[50] Øystein Bruserud, Hovland, R., Wergeland, L., Huang, T., and Gjertsen, B. T. Flt3-mediated signaling in human acute myelogenous leukemia (aml) blasts: a functional characterization of flt3-ligand effects in aml cell populations with and without genetic flt3 abnormalities. *Haematologica*, 88, 416–428 (2003).

[51] Ersvaer, E., Bertelsen, L., Espenes, L., Bredholt, T., Boe, S., et al. Characterization of ribosomal P autoantibodies in relation to cell destruction and autoimmune disease. *Scan J Immunol*, 60, 189–198 (2004).

[52] Gjertsen, B. T., Oyan, A., Marzolf, B., Hovland, R., Gausdal, G., et al. Analysis of acute myelogenous leukemia: preparation of samples for genomic and proteomic analysis. *J Hematother Stem Cell Res*, 11, 469–81 (2002).

[53] Gjertsen, B. T., Mellgren, G., Otten, A., Maronde, E., Genieser, H. G., et al. Novel (Rp)-camps analogs as tools for inhibition of camp-kinase in cell culture. basal camp-kinase activity modulates interleukin-1 beta action. *J Biol Chem*, 270, 20599–607 (1995).

[54] Research Systems Inc (RSI), C., Boulder. Idl, the interactive data language, v6.1.

6 Figures & Tables

Disease	Classification	Patients	Images
AML			
	M0/1	1	2
	M1	7	12
	M1/2	1	2
	M2	10	18
	M3	2	4
	M4	3	6
	M4/5	6	12
	M5	7	14
	M5a	2	3
	Sum		**73**
	Mean Age	**58.4 years**	
	Median Age	**61 years**	
	Age Range	**29-86 years**	
ALL		8	**16**
	Mean Age	**47.3 years**	
	Median Age	**41.5 years**	
	Age Range	**22-84 years**	

Table 2: Leukemia patient characteristics and corresponding 2DE imagesTotally 73 AML and 16 ALL images were included in the analysis. All ALL patients had B-cell disease and two patients comprised the bcr-abl fusion product.

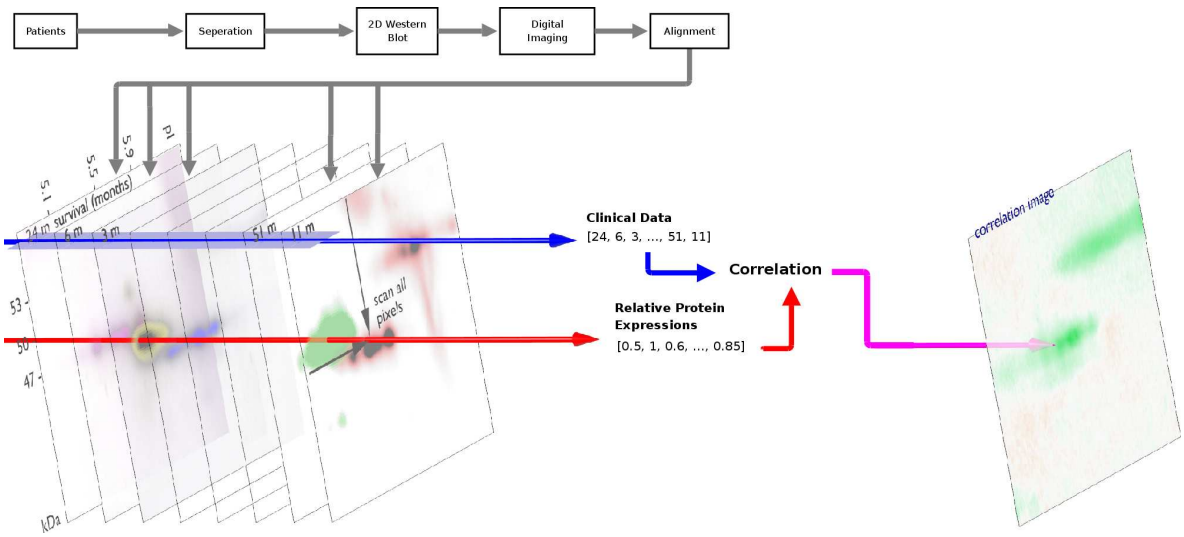

Figure 1: **2DE Gel Correlation** relies on an aligned, normalized stack of 2DE gel images and a numerical label associated with every gel. Pixel per pixel correlation creates a new image showing areas in the gel that relate to the external parameter. In comparison to standard gel analysis methods, spot detection is not necessary and therefor less bias is introduced into the analysis process. This technique also recognizes moving spots and spot shapes that change.

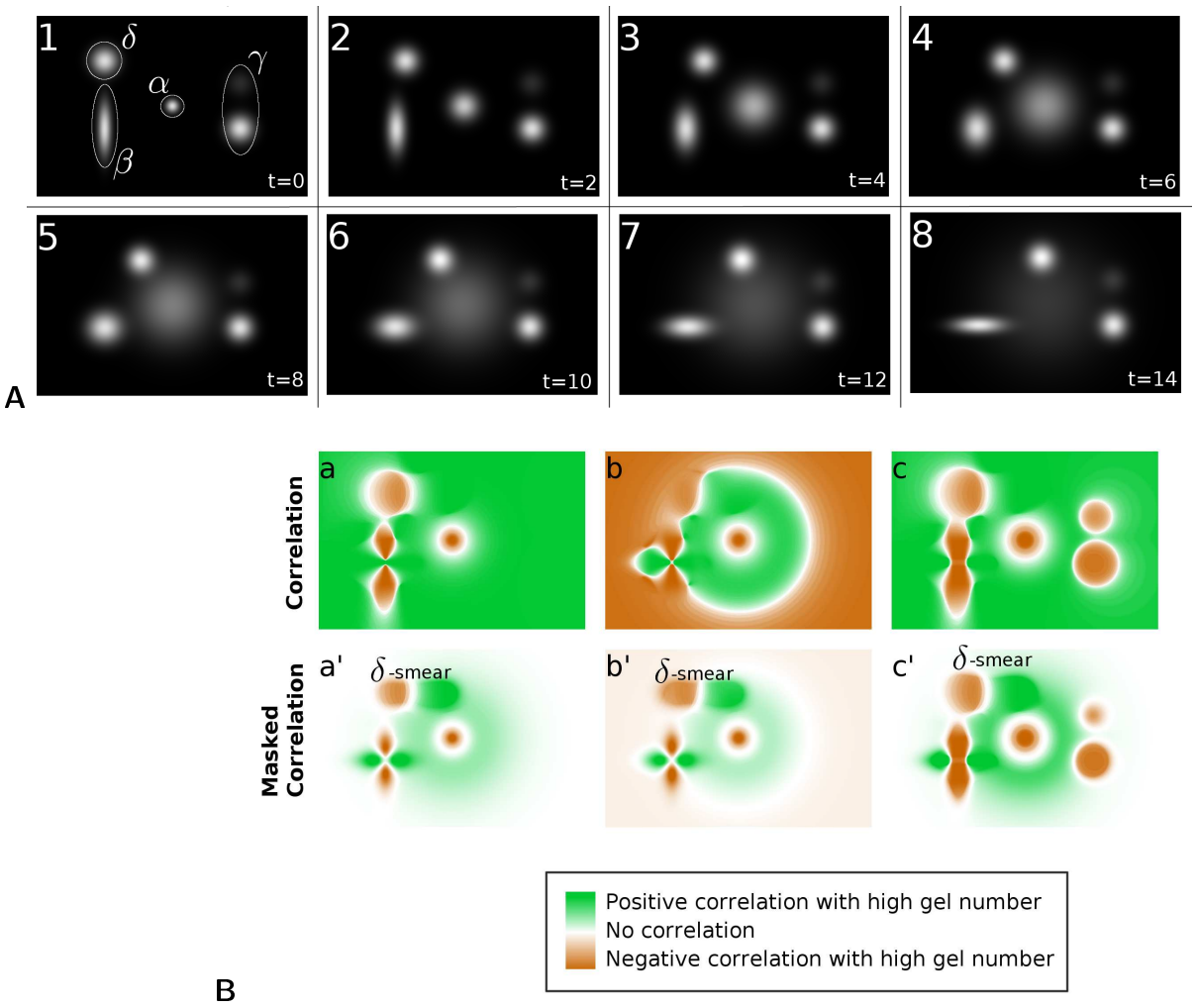

Figure 2: **Correlation of a simulated 2DE gel-stack and the effect of various background normalizations.** (A) 8 snapshots taken from a stack of 15 gels generated using Gaussian bumps, see material and methods for formula and details. Each gel contains simulated spots with particular characteristics: (α) an out-fading spot growing in size, (β) an elliptical spot changing shape from small and tall to broad and flat, (γ) two spots with constant amplitudes and (δ) a spot shifting from left to right. (B) Correlation between the gel-stack and the external parameter, in this case, the gels sequence number (t). Upper gels (a-c) visualizes the correlation, lower gels (a'-c') represent a masking that visualizes the areas of useful correlations. Correlation analysis was performed relying on different background removal methods. (a,a') without background removal; (b,b') with background subtraction, (c,c') using background division. See the result section for details.

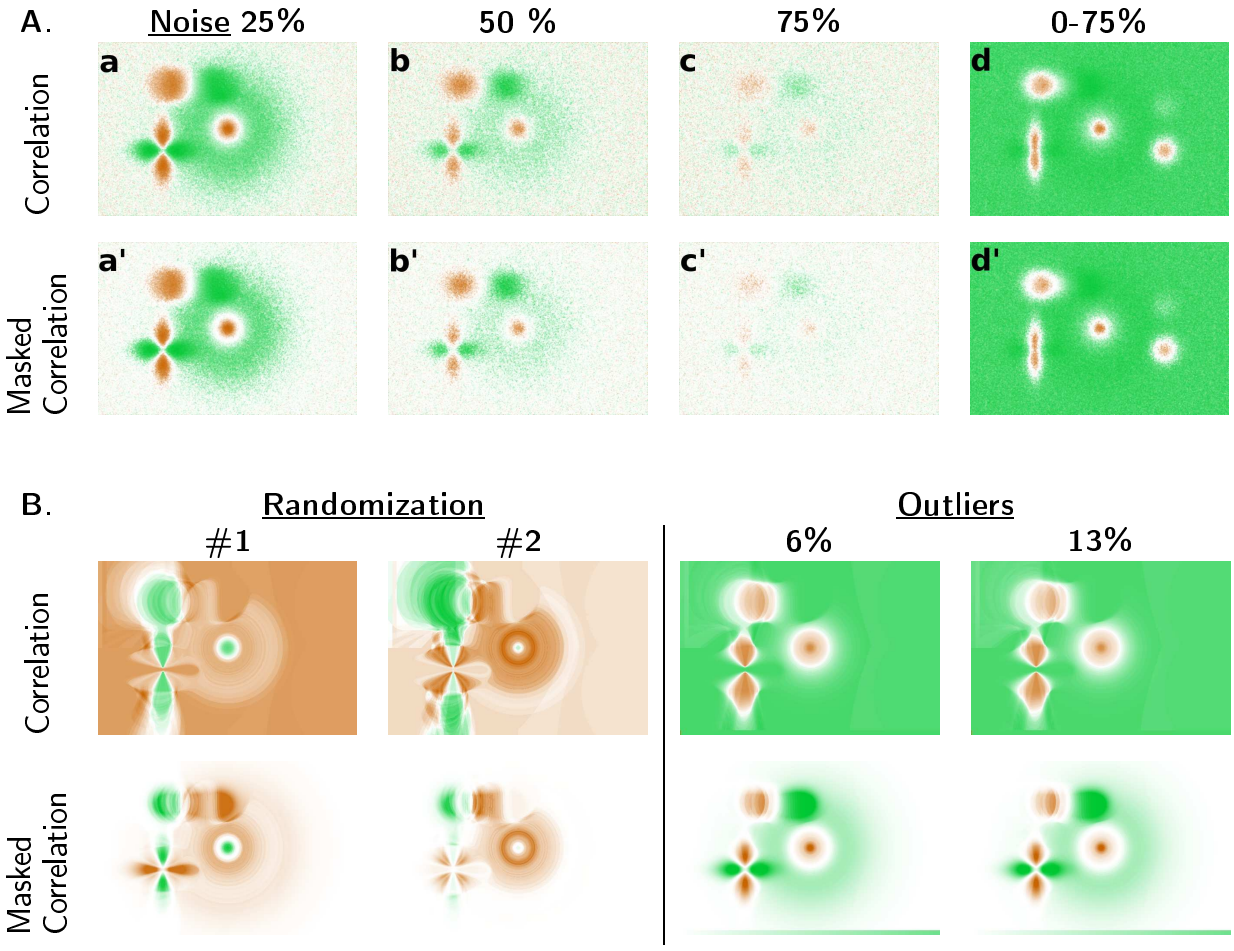

Figure 3: **The effect of noise, outliers and randomization in the correlation analysis.** (A) Correlation analysis with the addition of normal distributed noise to 25 % (a,a'), 50% (b,b'), 75% (c,c') and increasing noise from 0% to 75% in gels 0 to 14 (snapshot 1 to 8). The analysis was performed without normalization of the gels. (B) Two different randomizations were performed. #1(a,a') and #2(b,b') are the correlation images towards a fully randomized set of biological parameters. (c,c') and (d'd') show the correlation images when the biological parameter set is polluted by respectively 6% and 13% outliers.

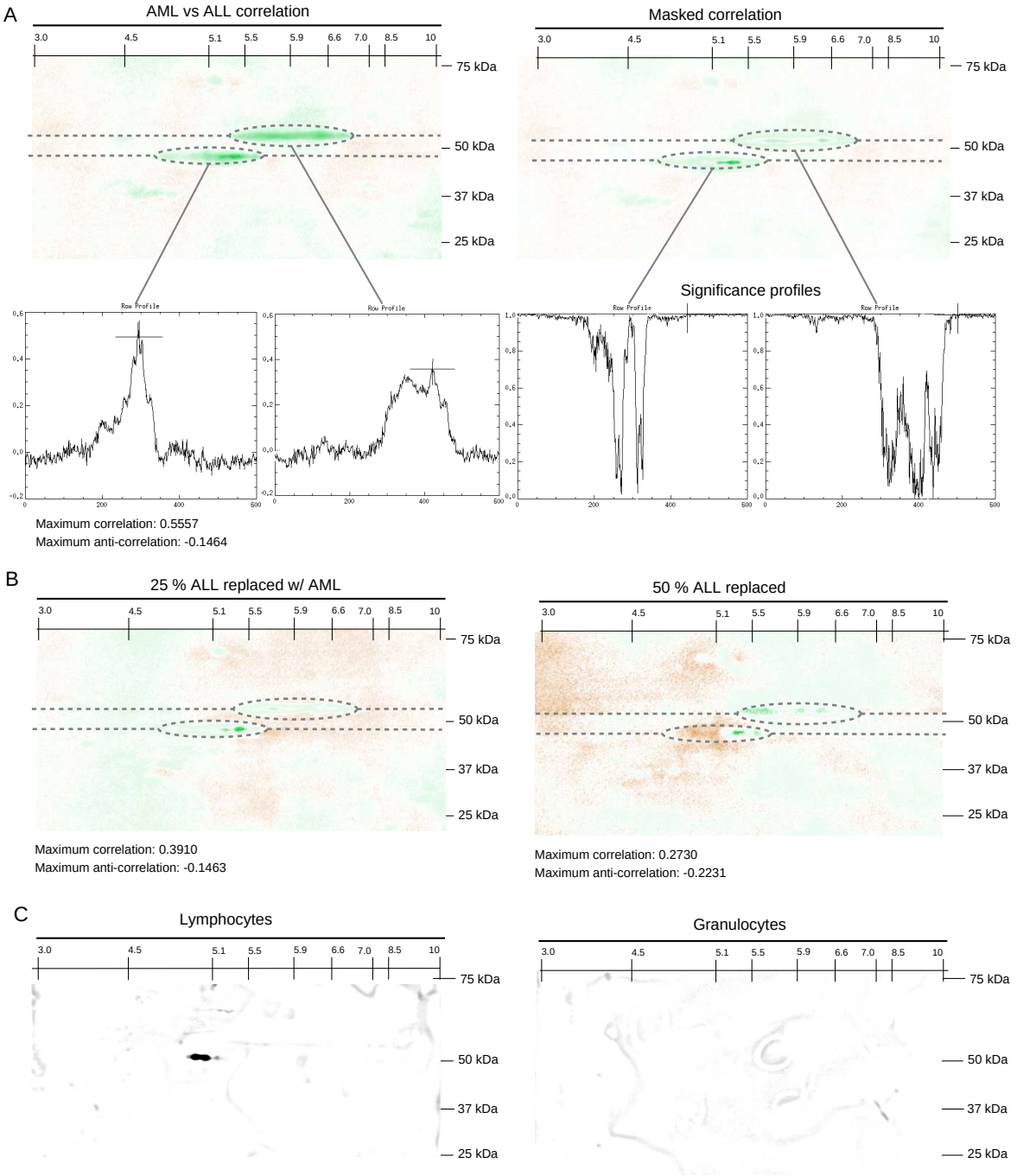

Figure 4: **Origin of cancer and p53 isoform distribution.** (A) 73 acute myeloid leukemia (AML) and 16 acute lymphoid leukemia (ALL) gel images were analyzed for correlation of p53 isoform distribution (left, correlation; right, masked correlation/significance of correlation). Green color indicates positive correlation for ALL (maximum positive correlation 0.5557), and brown color indicates negative correlation (maximum negative correlation 0.1464). Brown color indicates p53 protein forms predominantly found in AML. Inset boxes present the correlation (left) and significance (right) for the α and δ region, respectively. (B) Attenuated correlation when ALL images are replaced with images from AML. Replacement of 25% of the ALL cases with AML (a): Maximum positive correlation 0.3910, maximum negative correlation 0.1463. Replacement of 50% of ALL with AML (b): Maximum positive correlation 0.2730, maximum negative correlation 0.2232. (C) p53 expression in normal lymphocytes (left panel) and granulocytes (right panel). See Material and Methods for details on cell separation.

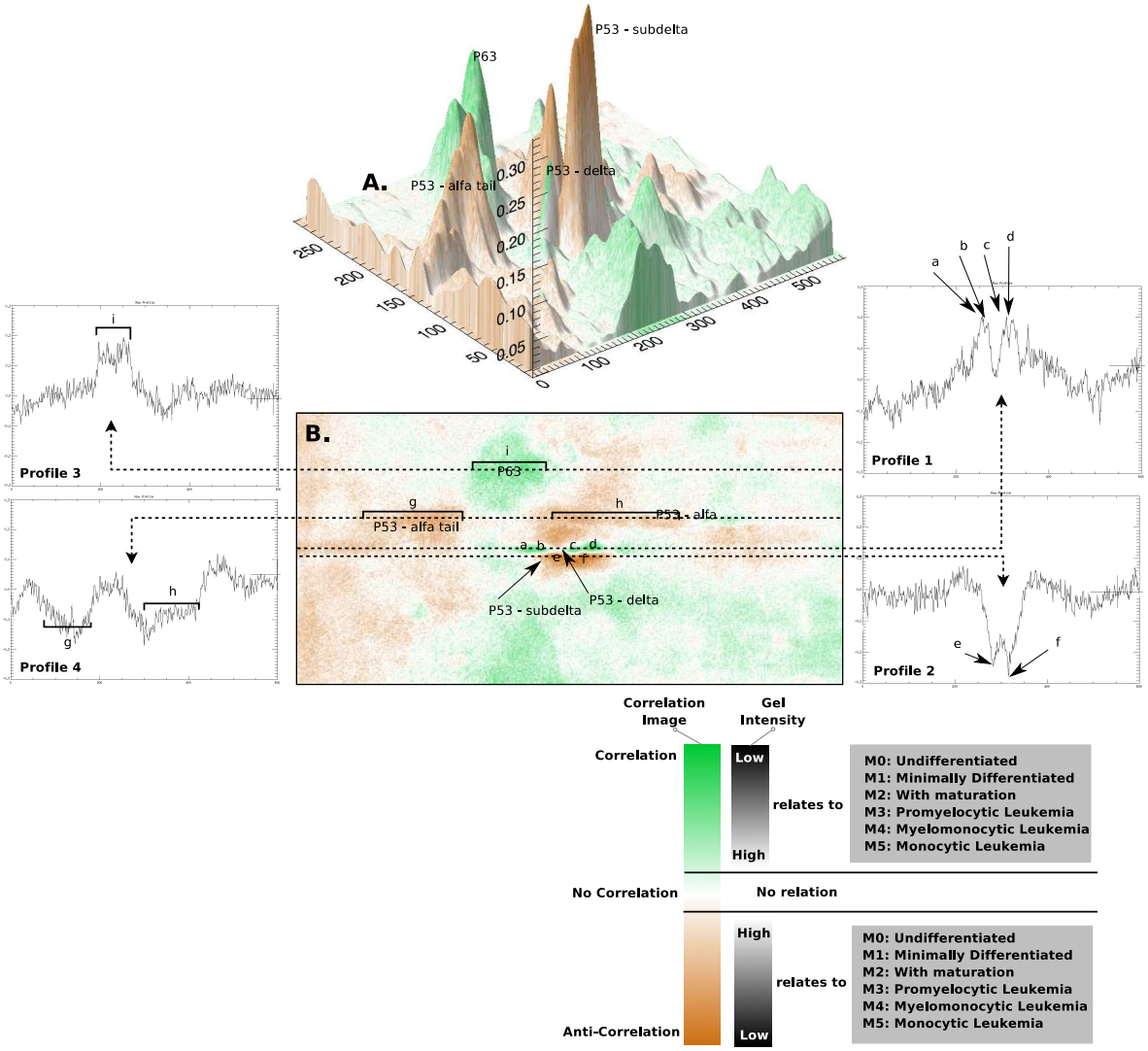

Figure 5: **Differentiation stages of cancer modulates p53 isoform distribution.** (A) Correlation landscape of p53 in 73 AML images related to differentiation direction and stage (FAB, French-American-British classification). Green indicates correlation with the more differentiated forms of AML. Brown indicates anti-correlation with the more mature forms of leukemia cells. The vertical axis sets out the absolute correlation value. (B) Correlation image demonstrating statistical significant alterations in p53. Profile 1 shows the p53-δ region containing four correlating spots (r = 0.2). Profile 2 shows the sub-δ region anti-correlating at positions e and f. Profile 3 is the p63 region (putative p53 isoform) correlating towards the more differentiated leukemia's. Profile 4, a p53 region anti-correlating with differentiated AML.

7 Supplementary Material

7.1 Correlation Implementation

The algorithm is implemented in IDLv6.1 [54] and takes two arguments. The first argument is the gel-stack, which is a three dimensional space. First dimension is the gel number, the second and third dimensions are the x and y-axis of the gels. The second argument is a vector describing the result of the different gels. The presented algorithm makes use of a mean scaling.

```
PRO correlate_images, all, result
  d = size(all,/dim)
  VX = d[1]
  VY = d[2]
  ; normalize the background
  for i = 0, d[0] - 1 do begin
    all[i,*,*] /= mean(all[i,*,*])
  endfor
  ; Rho correlation
  cor_pic = make_array(VX,VY,/double,value=0.0)
  f_pic = make_array(VX,VY,value=0.0)
  for x = 0, VX - 1 do begin
    for y = 0, VY - 1 do begin
      r = r_correlate(reform(all[*,x,y]),result)
      cor_pic[x,y]=r[0]
      f_pic[x,y]=1.0-r[1]
    endfor
  endfor
  ; we are interested in correlations with high variance on gel
  var_pic = make_array(VX,VY,/double,value=0.0)
  for x = 0, VX - 1 do begin
    for y = 0, VY - 1 do begin
      var_pic[x,y]=stddev(all[*,x,y])
    endfor
  endfor
  var_pic <= 1.0
  f_pic *= var_pic
  cor_pic <= 1.0
  cor_pic >= -1.0
  show_correlation, cor_pic, f_pic
end
```

7.2 Gauss Bumps

```
function gauss2d, sx, sy, cx, cy, wx, wy, a
  im = float(make_array(sx,sy,value=0.0))
  for x = 0, sx - 1 do begin
    for y = 0, sy - 1 do begin
      im[x,y]=float(((cx-x)/wx)^2 + ((cy-y)/wy)^2)
```

```
    endfor
  endfor
  im = -im/2
  im = exp(im)
  im *= a
  return, im
end
```

7.3 Simulated Gel Stack

```
function create_set, nr, sx, sy, wx1, wx2, wy1, wy2, a1, a2
  all = make_array(nr,sx,sy,value=0.0)
  for i = 0, nr - 1 do begin
    wx = wx1 + i*(wx2-wx1)/nr
    wy = wy1 + i*(wy2-wy1)/nr
    a = a1 + i*(a2-a1)/nr
    all[i,*,*] = gauss2d(sx, sy, sx/2, sy/2, wx, wy, a)
  endfor
  return, all
end
set1 = create_set(15, 600.0, 400.0, 10.0, 100.0, 10.0, 100.0, 5.0, 1.0)
set2 = create_set(15, 300.0, 300.0, 10.0, 40.0, 40.0, 10.0, 5.0, 5.0)
set3 = create_set(15, 300.0, 300.0, 20.0, 20.0, 20.0, 20.0, 5.0, 5.0)
set4 = create_set(15, 300.0, 300.0, 20.0, 20.0, 20.0, 20.0, 1.0, 1.0)
set1[*,0:299,0:299] += set2[*,*,*]
set1[*,300:599,0:299] += set3[*,*,*]
set1[*,300:599,100:399] += set4[*,*,*]
for i = 0, 14 do begin
  set1[i,0+i*10:299+i*10,200:399]+=set3[i,*,50:249]
endfor
for i = 0, 14 do begin
  set1[i,*,*]/=double(i)
endfor
set1 = relative(set1)
result1 = findgen(15)
correlate_images, set1, result1
end
```

7.4 Creating green / brown images

```
PRO show_correlation, cp, t_pic
  cor_pic = cp
  DDD = size(cp,/dim)
  VX = ddd[0]
  VY = ddd[1]
  ; the normal one
  shown = make_array(3,VX,VY,/double,value=255.0)
  multi = 1.0 / max(abs(cor_pic))
  multi = 1.0 / max(abs(cor_pic))
  shown[0,*,*] += (cor_pic[*,*] < 0) * multi * 55
  shown[1,*,*] += (cor_pic[*,*] < 0) * multi * 155
  shown[2,*,*] += (cor_pic[*,*] < 0) * multi * 255
  shown[0,*,*] -= (cor_pic[*,*] > 0) * multi * 255
  shown[1,*,*] -= (cor_pic[*,*] > 0) * multi * 55
```

```
  shown[2,*,*] -= (cor_pic[*,*] > 0) * multi * 205
  window, 1, title='Correlation', ret=2, xsize=vx, ysize=vy
  shown >= 0
  shown <= 255
  tvscl, shown, /true
  ; only the significant correlation
  wcor_pic = double(cor_pic) * double(t_pic)
  shown = make_array(3,VX,VY,/double,value=255.0)
  multi = 255.0 / max(abs(wcor_pic))
  shown[0,*,*] += (wcor_pic[*,*] < 0) * multi * 55
  shown[1,*,*] += (wcor_pic[*,*] < 0) * multi * 155
  shown[2,*,*] += (wcor_pic[*,*] < 0) * multi * 255
  shown[0,*,*] -= (wcor_pic[*,*] > 0) * multi * 255
  shown[1,*,*] -= (wcor_pic[*,*] > 0) * multi * 55
  shown[2,*,*] -= (wcor_pic[*,*] > 0) * multi * 205
  window, 3, title='Significant Correlations', ret=2, xsize=vx, ysize=vy
  tvscl, shown, /true
  shown = bytscl(shown)
  profiles, cor_pic
  ; significance
  window, 2, title='Significance', ret=2, xsize=vx, ysize=vy
  tvscl, t_pic
  profiles, t_pic
end
```