

# Quantitative signature for architectural organization of regulatory factors using intranuclear informatics

Daniel W. Young<sup>1</sup>, Sayyed K. Zaidi<sup>1</sup>, Paul S. Furcinitti<sup>2</sup>, Amjad Javed<sup>1</sup>, Andre J. van Wijnen<sup>1</sup>, Janet L. Stein<sup>1</sup>, Jane B. Lian<sup>1</sup> and Gary S. Stein<sup>1,\*</sup>

<sup>1</sup>Department of Cell Biology and Cancer Center and <sup>2</sup>Program in Molecular Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655-0106, USA

\*Author for correspondence (e-mail: gary.stein@umassmed.edu)

Accepted 18 March 2004

Journal of Cell Science 117, 4889-4896 Published by The Company of Biologists 2004  
doi:10.1242/jcs.01229

## Summary

**Regulatory machinery for replication and gene expression is punctately organized in supramolecular complexes that are compartmentalized in nuclear microenvironments. Quantitative approaches are required to understand the assembly of regulatory machinery within the context of nuclear architecture and to provide a mechanistic link with biological control. We have developed ‘intranuclear informatics’ to quantify functionally relevant parameters of spatially organized nuclear domains. Using this informatics strategy we have characterized post-mitotic re-establishment of focal subnuclear organization of Runx (AML/Cbfa) transcription factors in progeny cells.**

**By analyzing point mutations that abrogate fidelity of Runx intranuclear targeting, we establish molecular determinants for the spatial order of Runx domains. Our novel approach provides evidence that architectural organization of Runx factors may be fundamental to their tissue-specific regulatory function.**

Supplementary material available online at <http://jcs.biologists.org/cgi/content/full/117/21/4889/DC1>

Key words: Nuclear architecture, Runx (AML/Cbfa), Informatics, Quantitative, Transcription

## Introduction

The architectural organization of nucleic acids and cognate factors in subnuclear microenvironments is linked to gene regulation, replication and repair (Stein et al., 2000a; Stein et al., 2000b; Lemon and Tjian, 2000; Dundr and Misteli, 2001; Iborra and Cook, 2002; Spector, 2003; Stein et al., 2003). Spatio-temporal changes in this subnuclear organization accompany cell cycle progression and cell differentiation (Ma et al., 1998; Francastel et al., 2000). Perturbations in subnuclear organization have been functionally related with compromised gene expression that accompanies the onset and progression of disease (Dyck et al., 1994; Karpuj et al., 1999; McNeil et al., 1999). Traditionally, biological control of gene expression has been experimentally addressed by the identification and characterization of promoter elements and cognate regulatory and co-regulatory proteins, as well as by mechanistically defining the dynamics of chromatin structure and nucleosome organization. It is becoming increasingly evident that regulatory parameters of gene expression are operative within a higher-order subnuclear organization of nucleic acids and regulatory proteins. Observations made by epifluorescence and confocal microscopy have provided the initial insight into assembly of nuclear microenvironments that support the combinatorial compartmentalization of regulatory factors and chromosomal domains (Cook, 1999; Stein et al., 2000a; Stein et al., 2000b; Stein et al., 2003; Spector, 2003). Quantitative strategies are necessary to mechanistically associate the subnuclear organization of regulatory factors with biological control.

Here we describe a novel approach, intranuclear informatics,

to examine the subnuclear organization of regulatory factor domains from digital microscopic images. Intranuclear informatics utilizes parameters with biologically relevant variability to characterize subnuclear organization. We have developed an image-processing algorithm to acquire and evaluate these parameters of subnuclear organization. The result is a multivariable data-set that can be used for exploratory analysis techniques and for quantitatively testing specific biological hypotheses.

Runx transcription factors provide a paradigm for compartmentalization of gene expression and nuclear matrix association of regulatory proteins (Lian and Stein, 2003). A conserved intranuclear targeting signal (NMTS) within the C-terminus directs Runx factors to matrix associated subnuclear sites that support transcriptional control in the interphase nucleus (Choi et al., 2001; Zaidi et al., 2001; Zeng et al., 1997). By the application of intranuclear informatics we elucidate that Runx regulatory proteins exhibit an interphase architectural signature that is restored following mitosis. Furthermore, our analysis of NMTS mutant proteins provides evidence that architectural association of Runx factors may be fundamental to their tissue-specific gene regulatory functions. Thus, intranuclear informatics quantitatively bridges the spatial organization of protein domains with regulatory determinants of biological control.

## Materials and Methods

Cell culture and transfections

ROS 17/2.8 osteosarcoma cells were maintained in F12 with penicillin

and streptomycin, 2 mM L-glutamine and 5% fetal bovine serum (FBS). HeLa cells were maintained in DMEM with PS, 2 mM L-glutamine and 10% FBS. Exponentially growing HeLa cells were transfected with 500 ng of either HA-tagged wild-type Runx2, an HA-tagged C-terminal deletion, or one of the four HA-tagged NMTS point mutants for 24 hours with Superfectamine (Invitrogen, San Diego, CA).

### Immunofluorescence

HeLa and Ros cells were grown on gelatin-coated coverslips (BD Biosciences, Lexington, KY). Cells were processed for in situ immunofluorescence as described (Javed et al., 2000). In brief, cells were rinsed twice with ice-cold PBS and fixed in 3.7% formaldehyde in PBS for 10 minutes on ice. After rinsing once with PBS, the cells were permeabilized in 0.1% Triton X-100 in PBS, and rinsed twice with PBSA [0.5% bovine serum albumin (BSA) in PBS] followed by antibody staining. Antibodies and their dilutions used are as follows: rabbit polyclonal antibodies against Runx2 (1:200; Oncogene, Carlsbad, CA) and rabbit polyclonal antibodies against HA-epitope (1:500, Santa Cruz Biotechnology, Santa Cruz, CA). The secondary antibodies used were either anti-rabbit or -mouse Alexa568 or Alexa488 (1:800, Molecular Probes, Eugene, OR).

### Image acquisition and restoration

Immunostaining of cell preparations was recorded using a CCD camera attached to an epifluorescence Zeiss Axioplan 2 microscope (Zeiss, Thornwood, NY). For Runx2 interphase/telophase studies single image planes were acquired and deconvoluted using the Metamorph Imaging Software (Universal Imaging, Downingtown, PA). For NMTS mutation experiments Z-series image stacks were acquired at 0.25  $\mu\text{m}$  intervals with 67 nm/pixel (xy). Restoration of images was carried out by 3D deconvolution using a measured point-spread function as described previously (Carrington et al., 1995).

### Image processing

We have developed an image processing algorithm that automatically performs image segmentation, feature extraction and parameter computation. Our algorithm requires the input of any number of image pairs and a text-file that lists the name of the images to be analyzed. For each pair of images, one is the digital micrograph and a second is the nuclear mask image. The nuclear mask, which is generated using Metamorph imaging software or Adobe Photoshop (Adobe Systems, San Jose, CA), is utilized to eliminate intensity data that is located outside the nucleus and restrict analysis to intensity data within the nucleus. For mitosis studies we analyzed a single image plane per cell. For NMTS mutation studies we analyzed z-section images from deconvoluted Z-series stacks. Image segmentation is carried out using a threshold technique, where the selected threshold is the intensity value that maximizes the number of detectable nuclear domains. The image analysis is implemented using the MATLAB<sup>®</sup> image processing and statistics toolboxes (The Mathworks, Natick, MA) and Metamorph Imaging Software (Universal Imaging).

### Image feature extraction

Our algorithm extracts the total number of domains within the nucleus, the size of each domain, the location of each domain-centroid in image pixel coordinates, the nuclear cross-sectional area and the nuclear cross-sectional perimeter from the segmented and mask images. From these measurements we determine the following statistics for both domain size and nearest-neighbor distances: mean, median, variance, standard deviation, index of dispersion, coefficient of variation, skewness and kurtosis. The index of dispersion and coefficient of variation are mean normalization measures of variation

and standard deviation, respectively. Skewness reflects the degree of asymmetry in the distribution with positive values indicating right skewness and negative values indicating left skewness. Kurtosis is a measure of the peakedness of the distribution: positive values indicate a tall peak and negative values indicate a flat peak (or plateau) (Norman, 2000). To assess the spatial domain randomness we measured Euclidean nearest-neighbor distances (NN distance) between domain centroids. The mean and variance of the Euclidean nearest-neighbor distances between domains is compared with a Poisson point-process of an equivalent density (i.e. domains per unit nuclear area) (Clark and Evans, 1954). Expected nearest-neighbor distance parameters are corrected for edge effects (Donnelly, 1978; Sinclair, 1985). The ratio of observed (Ro) to expected (Re) mean nearest-neighbor distances is referred to as the Clark and Evans statistic (Ro/Re<1, clustered; Ro/Re=1, random; Ro/Re>1, ordered) (Clark and Evans, 1954). We also examined the radial position of domains within the nucleus. This statistic is determined by measuring the mean distance from each domain centroid to the nuclear centroid (mean domain radius) and the mean distance from the nuclear centroid position to the each perimeter pixel (mean perimeter radius); for a circle this would be the radius. The ratio of the two values is the mean relative domain radius. Values between 0 and 0.5 reflect a tendency for domains to be positioned in the nuclear interior and values between 0.5 and 1 reflect a tendency for domains to be positioned toward the nuclear periphery.

### Statistical analyses

For mitosis studies ANOVA tests were conducted on subnuclear organization data to determine the significance of observed differences in each parameter. Asterisks indicate parameters with differences that are considered to be statistically significant on a 0.05 level. *P*-values were adjusted to account for the false-discovery rate; asterisks are indicative of this adjustment. Analysis was performed using the general linear model (GLM) procedures in SAS/STAT (Sas Institute, Cary, NC). These statistical tests were conducted to compare each telophase (T<sub>1</sub> and T<sub>2</sub>) nuclei as well as to compare telophase nuclei with interphase (I) nuclei. We analyzed 60 nuclei for Runx2; 20 for each nucleus (see supplementary material for the complete dataset). For NMTS studies, statistical tests were conducted to compare wild-type Runx2 with each of the five mutants. In total, 330 Z-sections were analyzed, 55 for each protein from two independent experiments (see supplementary material for the complete dataset). Five Z-sections were analyzed per cell to account for within cell variability. Thus, the effect of NMTS mutation was assessed using a repeated measure ANOVA at a 0.05 level.

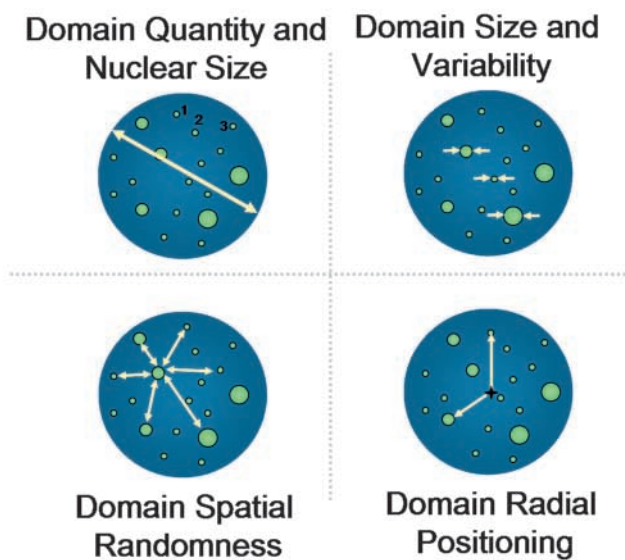
Factor analysis was performed on parameters of subnuclear organization for each of the wild-type Runx and the five mutant proteins using the data obtained from 330 nuclear images. This analysis represents the observed subnuclear organization parameters in terms of a smaller number of uncorrelated 'factors' (or groups of parameters) that account for most of the information contained in the complete data set (Norman, 2000). Factors are extracted using principal component analysis and rotated using the varimax method. Factor scores were computed for each image and represent the sum of the standardized subnuclear organization parameters multiplied by their respective factor loadings. Factor loading refers to the correlation of each subnuclear organization parameter with a particular factor. Factor loadings greater than 0.65 were considered to be significant. This analysis was carried out using the factor procedure in SAS/STAT.

Hierarchical cluster analysis was performed on mean subnuclear organization parameters from wild-type Runx and the five mutant proteins using the data from 330 nuclear images. Cluster analysis was performed using the Euclidean distance metric with complete linkage. Clusters were displayed using a dendrogram. Cluster analysis was carried out using the cluster procedure in SAS/STAT.

## Results

### Intranuclear informatics: a signature of nuclear architecture for regulatory proteins

We have developed intranuclear informatics to characterize spatially organized protein domains within the nucleus in terms of parameters with inherent biological variability. The conceptual framework for quantifying nuclear organization is outlined in Fig. 1 and briefly described here. Alterations in size and number of protein domains with respect to physiological conditions, cell cycle stage, and/or cellular differentiation have been observed (Ma et al., 1998; Stenoien et al., 2001; Nielsen et al., 2002; Zaidi et al., 2003). Intranuclear informatics exploits this variability in domain size and number, to elucidate changes between different biological conditions (Fig. 1). Another prominent feature of nuclear organization is the non-random localization of chromosome territories and protein domains (Noordmans et al., 1998; Cremer and Cremer, 2001; Shiels et al., 2001; Kozubek et al., 2002; Tanabe et al., 2002). Our approach employs first-order nearest-neighbor statistics, commonly used in ecological studies (Clark and Evans, 1954; Sinclair, 1985), to characterize the spatial randomness of



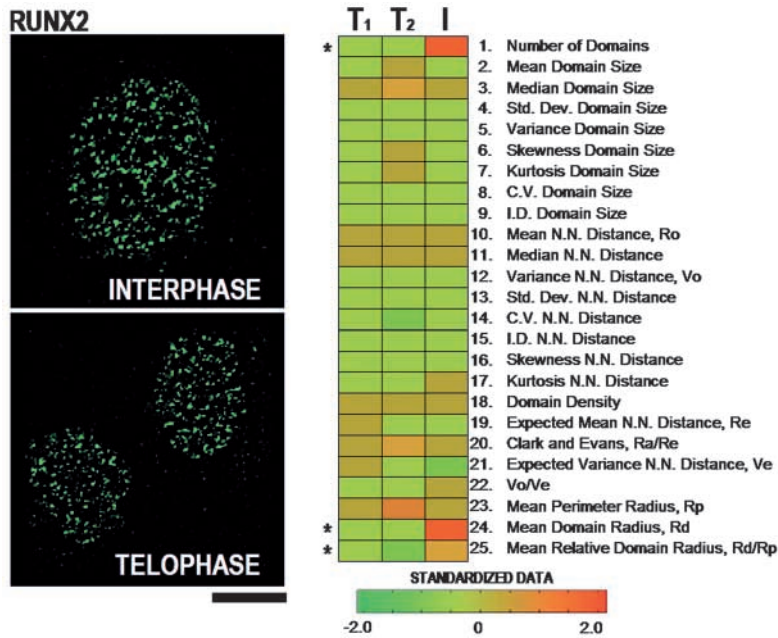
**Fig. 1.** Conceptual framework for the quantitation of subnuclear organization by intranuclear informatics. Four main groups of parameters, selected on the basis of inherent biological variability, are examined. Parameters that describe domain quantity and nuclear size comprise group 1 (upper left panel). Group 1 includes: number of domains and domain density. Parameters that describe domain size and variability comprise group 2 (upper right panel). Group 2 includes: domain size mean, median, standard deviation, variance, skewness, kurtosis, coefficient of variation, and index of dispersion. Parameters that describe the domain spatial randomness, which is based on domain nearest-neighbor distances, comprise group 3 (lower left panel). Group 3 includes: domain nearest-neighbor mean, median, standard deviation, variance, skewness, kurtosis, coefficient of variation, index of dispersion, domain density, nearest-neighbor distance mean and variance expected for a random distribution, ratios between actual and expected mean and variance, and the standard error in the nearest-neighbor distances. Parameters that characterize the radial position of domains comprise group 4 (lower right panel). Group 4 includes: mean perimeter radius, mean domain radius, and mean relative domain radius.

nuclear microenvironments (Fig. 1). Finally, the radial position of regulatory machinery for replication and transcription is functionally interrelated to the location of chromosomal territories as well as chromatin structure (Ma et al., 1998; Cook, 1999; Cremer and Cremer, 2001; Tumber and Belmont, 2001; Kozubek et al., 2002). Intranuclear informatics establishes the placement of regulatory foci within the context of nuclear morphology (Fig. 1). Based on these biological observations, our approach describes and defines intranuclear organization utilizing 25 parameters, evaluated from digital fluorescence microscopic images. We have developed an image-processing and statistical algorithm to acquire measurements and compute the parameters from any number of images. The resulting data are then analyzed to quantitatively address specific biological questions using statistical approaches such as factor analysis and multivariate clustering techniques. In summary, intranuclear informatics incorporates the principal features of intranuclear organization to provide a vehicle for quantitatively defining nuclear structure-function interrelationships.

Intranuclear informatics reveals that the post-mitotic restoration of Runx subnuclear domain organization is functionally conserved

The hematopoietic and osteogenic Runx transcription factors are involved in tissue-specific gene expression and support cell differentiation (Tracey and Speck, 2000; Komori, 2002; Lutterbach and Hiebert, 2000; Lian and Stein, 2003). In the interphase nucleus Runx proteins are associated with the nuclear matrix and are organized into punctate domains (Zaidi et al., 2001; Zeng et al., 1997). These nuclear microenvironments spatially coincide with sites of active transcription and colocalize with several coregulatory proteins (Thomas et al., 2001; Javed et al., 2000; Lian and Stein, 2003; Harrington et al., 2002; Zaidi et al., 2002; Westendorf et al., 2002; Kundu et al., 2002; Zaidi et al., 2004). These observations suggest a direct link between the activity of Runx proteins and their spatiotemporal organization within the nucleus. We have recently demonstrated that Runx1 and Runx2 protein domains persist during mitosis, and undergo spatial and temporal reorganization resulting in equal partitioning into progeny nuclei (Zaidi et al., 2003). These mitotic alterations reflect natural perturbations in both nuclear structure and function and serve as a biological template for understanding Runx domain organization. Together, the dynamic distribution of Runx proteins provides a model for quantitative and comparative analysis of the subnuclear organization of regulatory proteins.

Here we have applied intranuclear informatics to understand the spatial organization of endogenous Runx2 domains in the interphase nucleus as well as following mitosis. Immunofluorescence microscopy confirms that the protein is distributed in punctate subnuclear domains (Fig. 2). We analyzed and compared 25 parameters of subnuclear organization in interphase nuclei and in both telophase nuclei. Our quantitative results show that most parameters are similar in interphase and telophase nuclei for Runx2. As expected telophase nuclei are significantly smaller than interphase. There are the same number of domains in progeny telophase nuclei but higher numbers in interphase nuclei. This



**Fig. 2.** Post-mitotic restoration of the spatially ordered Runx subnuclear organization is functionally conserved. ROS 17/2.8 osteosarcoma cells (top panel) were subjected to in situ immunofluorescence microscopy for endogenous Runx2. Runx2 is distributed at punctate subnuclear domains throughout the interphase and telophase nucleus (lower panel). Subnuclear organization parameters were computed from deconvoluted images for Runx2 for interphase nuclei (I), and both progeny telophase nuclei, denoted at random as telophase nucleus 1 (T<sub>1</sub>) or telophase nucleus 2 (T<sub>2</sub>). A color map has been applied to the standardized data assigning red to higher values and green to lower values (see supplementary material). Each increment of one reflects one (row) standard deviation (inner left and right panels). ANOVA was performed to assess the significance of observed differences between T<sub>1</sub>, T<sub>2</sub>, and I nuclei. Asterisks indicate statistically significant differences based on a 0.05 level with correction for false discovery rate. Bonferroni's multiple comparison tests were used to determine which nuclei differed significantly at a 0.05 level. In each case significant differences were observed between telophase (T<sub>1</sub>, T<sub>2</sub>) and interphase nuclei (I), but differences were not observed between telophase nuclei. Overall mean Clark and Evans statistics (Ro/Re) were 1.4 for Runx2, indicating a non-random organization with spatial order. Bar, 10  $\mu$ m.

observation is consistent with the mitotic partitioning of Runx proteins (Zaidi et al., 2003) (Fig. 2). We further find that Runx2 domains exhibit a non-random organization with spatial order. We conclude that the post-mitotic restoration of Runx subnuclear organization is functionally conserved in progeny cells.

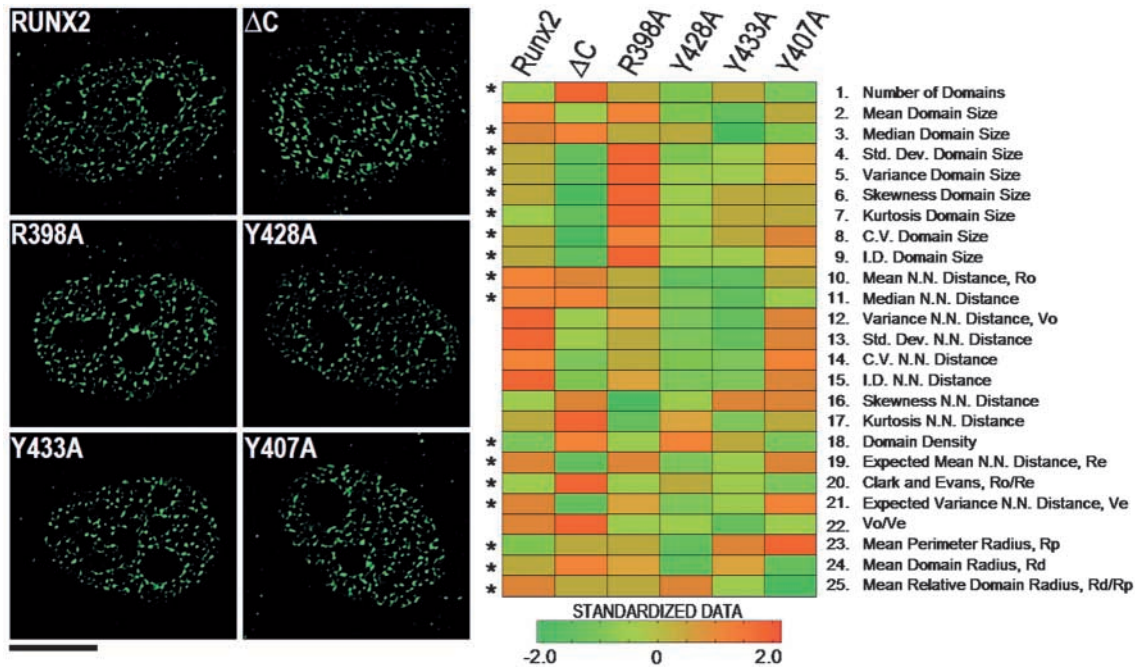
#### Intranuclear informatics establishes molecular determinants for the spatial domain organization of Runx transcription factors

A viable candidate for elucidating the underlying requirements for Runx domain organization is the nuclear matrix targeting signal (NMTS). The NMTS is a conserved and unique Runx protein motif that is necessary and sufficient for directing the protein to matrix-associated intranuclear sites (Zaidi et al., 2001; Zeng et al., 1997). Biochemical, cellular and in vivo genetic approaches have established the requirement of the NMTS and associated functions in Runx control of cell differentiation and tissue-specific development (Choi et al., 2001; Yergeau et al., 1997). Importantly, mutations in Runx proteins that alter subnuclear targeting are associated with skeletal disease and leukemia (McNeil et al., 1999; Choi et al., 2001; Barseguian et al., 2002; Zhang et al., 2000).

Our experimental strategy combines mutagenesis, microscopy, and intranuclear informatics to understand the contribution of the NMTS to Runx domain organization. We examined wild-type Runx2, a C-terminal deletion (Runx2- $\Delta$ C) that lacks the NMTS, as well as four NMTS point mutations, using immunofluorescence microscopy. These mutants exhibit varying degrees of compromised intranuclear targeting and selective alterations in physical and functional protein-protein interactions (Zaidi et al., 2002; Zaidi et al., 2004), (our unpublished observations). Our intranuclear informatics analysis was performed on deconvoluted images ( $n=330$ ) from nuclei of cells expressing these proteins. All of the Runx proteins localize to punctate domains within the nucleus (Fig.

3). Initial evaluation of subnuclear organization data reveals that there are significant differences in 17 of 25 parameters, as identified by ANOVA. By contrast, parameters that reflect variation in nearest-neighbor distances are not significantly altered by the mutations. These results are schematically demonstrated by a color representation of standardized data (Fig. 3). Collectively, our observations indicate that there are indeed alterations in the spatial domain organization of Runx proteins as a consequence of mutations in the NMTS.

Our analysis reveals that R398A and Y407A mutants share most of their properties with the wild-type protein; except for a reduction in the domain radial positioning for the Y407A mutant protein, and an increase in the domain size variability for the R398A mutant protein. Significant alterations in subnuclear organization were observed for the Y428A, Y433A and  $\Delta$ C mutations. It has been proposed that nuclear microenvironments represent the steady state local accumulation of proteins resulting from dynamic molecular interactions providing threshold concentrations of regulatory factors for combinatorial control (Stein et al., 2000a; Stein et al., 2000b; Misteli, 2001; Stein et al., 2003). Consistent with this concept, the Y428A mutant, which has a significant reduction in the mean and variability in domain size (Fig. 3), functionally abrogates interactions between Smad and Runx, thereby blocking integration of TGF $\beta$  signals at matrix-associated subnuclear sites (Zaidi et al., 2002). Furthermore, impairment of Src/YAP signaling by the Y433A mutation correlates with a decrease in average domain size (Fig. 3). Similarly, deletion of the C-terminus of Runx2 abolishes interactions with several known co-regulators and results in reduced domain size and variability (Hanai et al., 1999; Javed et al., 2000; Westendorf et al., 2002; Thomas et al., 2001) (Fig. 3). Hence, the reduction in Runx domain size is likely to be a consequence of abrogated and/or altered protein-protein interactions. We further find that the NMTS may contribute to the spatial distribution of domains within the nucleus. This is demonstrated by a reduced variability in domain nearest-



**Fig. 3.** Mutation of NMTS alters the interphase Runx subnuclear organization. Point mutations within the Runx2 NMTS were generated using PCR-mediated mutagenesis. Deconvoluted images were analyzed of whole cells (HeLa) expressing either HA-tagged wild-type Runx2, an HA-tagged C-terminal deletion or one of the four HA-tagged NMTS point mutants. As shown, each of these mutants and wild-type Runx exhibits a punctate subnuclear distribution (top panel). Standardized mean subnuclear organization data for the indicated proteins are shown (right panel). A color map has been applied to the standardized values assigning red to higher values and green to lower values (see supplementary material). Using a repeated-measure analysis of variance (ANOVA) we detect significant differences at a 0.05 level in 17 of 25 parameters measured, as indicated by asterisks. Bar, 10  $\mu\text{m}$ .

neighbor distances for Y428A, Y433A and the  $\Delta\text{C}$  protein. In addition, the Runx2- $\Delta\text{C}$  protein, which has abrogated subnuclear targeting, exhibits a significant increase in mean domain nearest-neighbor distances as well as in domain density. We conclude that NMTS mediated intranuclear targeting is a functional determinant for the characteristic spatially ordered distribution of Runx domains.

#### Intranuclear informatics selectively discriminates between the subnuclear organization of wild-type and mutant Runx protein domains

To identify on a broader level the biological features of subnuclear organization that are predominantly influenced by the NMTS mutations we used factor analysis, a multivariate analytical tool for grouping related parameters ('factors'). Three factors that capture a large proportion of the biological variability and are readily interpretable describe the domain size (factor A), the domain packing (factor B) and the domain spatial randomness (factor C) (Table 1). We evaluated the subnuclear distribution of each protein by calculating 'factor scores' and generated star-plots to compare wild-type Runx2 with each of the NMTS mutants (see Fig. 4 legend and Materials and Methods for details). Differences are evident in all three factors. Our analysis of these changes reveals that NMTS mutations have selective effects on Runx subnuclear organization (Fig. 4). Based upon the observed differences we can categorize the proteins into two groups. One group contains wild-type Runx2, R398A and Y407A, which exhibit

similar spatial randomness and domain packing. The second group contains the remaining mutants with similar effects on domain packing, but selective effects on size and spatial randomness. Whereas Y428A and Y433A mutants display similar changes in spatial randomness, domain size alterations are common between the Y428A mutant and the Runx2- $\Delta\text{C}$  protein. Of all the mutants, the Runx- $\Delta\text{C}$  protein has the most prominent effect on the three factors collectively. Notably, this mutant protein exhibits compromised subnuclear targeting, fails to promote osteoblast differentiation, and has been linked to the human disease cleidocranial dysplasia (CCD) (Choi et al., 2001; Zhang et al., 2000). Taken together, our analysis selectively distinguishes between wild-type Runx2 and NMTS mutant proteins based upon the three factors of subnuclear organization.

#### Intranuclear informatics quantitatively bridges the spatial organization of protein domains with regulatory determinants of biological control

We have demonstrated that mutations in the NMTS have selective and specific effects on the architectural signature of Runx proteins (Figs 3 and 4). Consequently, it is important to comprehensively assimilate all the data to establish the overall degree of domain organizational similarity between wild-type and mutant proteins. Here we utilized hierarchical cluster analysis to group each protein on the basis of the 25 parameters that describe and define their subnuclear organization (Fig. 5). The dissimilarity between the subnuclear organization of wild-

**Table 1. Factor analysis of subnuclear organization**

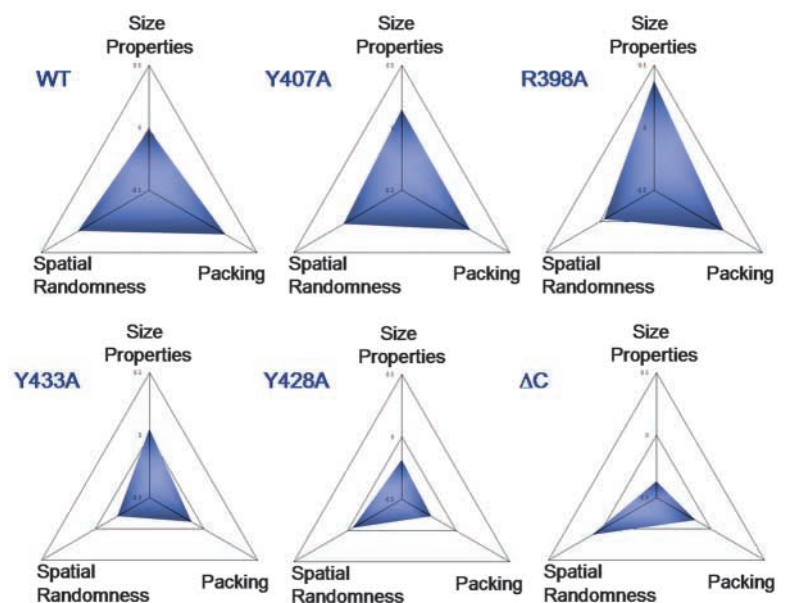
Factor	Common domain characteristic	Constituent parameters	Loadings
A	Size properties	Mean domain size	0.66
		Standard deviation in domain size	0.94
		Variance in domain size	0.93
		Skewness in domain size	0.90
		Kurtosis in domain size	0.90
		Coefficient of variation in domain size	0.95
		Index of dispersion in domain size	0.97
B	Packing	Mean domain nearest neighbor distance (observed), Ro	0.94
		Median domain nearest neighbor distance	0.91
		Domain density	-0.91
		Mean domain nearest neighbor distance (expected), Re	0.90
		Variance in domain nearest neighbor distance (expected), Ve	0.88
C	Spatial randomness	Variance in domain nearest neighbor distance (observed), Vo	0.87
		Standard deviation in domain nearest neighbor distance	0.85
		Coefficient of variation in domain nearest neighbor distance	0.95
		Index of dispersion in domain nearest neighbor distance	0.92
		Vo/Ve	0.96

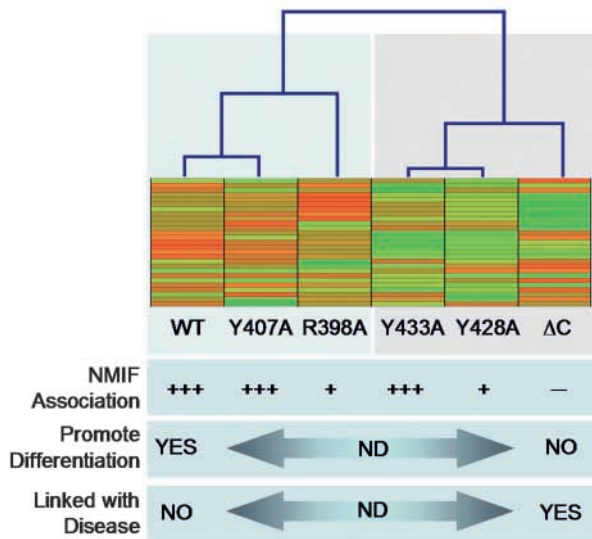
The covariance structure between the 25 subnuclear organization parameters measured on wild-type and mutant images ( $n=330$ ) indicates that there is a large degree of correlation between parameters (data not shown). Factor analysis was carried out to represent the observed subnuclear organization parameters in terms of a smaller number of uncorrelated variables. The strategy reduces the 25 parameters to a subset of three factors that retain the information (variability) contained within the entire data set. In multivariate analysis, 'factors' reflect groups of correlated parameters that are related to a common property of subnuclear organization. Each of the factors has a biological interpretation based upon the grouped subnuclear organization parameters. We restricted our analysis to the first three factors (referred to as A, B and C), as they reflect meaningful aspects of the domain spatial. Factor A, which represents 'domain size properties', accounts for approximately 30% of the information describing the subnuclear organization of the wild-type Runx and mutant images (i.e. 30% of the variation). The domain size properties are highly correlated with parameters that describe the variability in domain size and to a lesser degree the mean domain size. Factor B, which reflects 'domain packing', describes 23% of the variation and is directly correlated with parameters that characterize the mean nearest neighbor distances and inversely related to domain density. 'Packing' indicates that this factor relates the number of domains to the distances between domains. Factor C, which reflects 'domain spatial randomness', describes 10% of the variation and is directly correlated with parameters that described the variability in domain nearest neighbor distance.

type and the functionally compromised Runx2- $\Delta$ C mutant is evident by the presence of two distinct clusters. This mutation removes the entire C-terminus including the subnuclear targeting signal and associated functions; homozygosity for the Runx2- $\Delta$ C allele results in embryonic lethality (Choi et al., 2001). We find that the subnuclear organization of the Y433A

and Y428A mutants is similar to that of the Runx2- $\Delta$ C mutant. This observation is in agreement with our factor analysis and is consistent with the evidence that these mutant proteins are functionally compromised and are incompetent for integrating physiological signals, which include BMP/TGF $\beta$  and Src/YAP signaling (Zaidi et al., 2001; Zaidi et al., 2004). To provide

**Fig. 4.** Discrimination between wild-type Runx2 and NMTS mutants on the basis of domain size, packing, and spatial randomness. To understand the subnuclear organization of the wild-type Runx protein and the five mutants, we analyzed factors scores, which reflect the sum of standardized subnuclear organization parameters multiplied by respective factor loadings. Factor scores assign a value to each of the unobservable factors (factor A: domain size properties, factor B: domain packing, and factor C: domain spatial randomness). Using the data acquired from the 330 nuclear image sections, we computed factor scores for wild-type and each of the mutants and analyzed star-plots of these scores on three axes (see supplementary material). The center of the star-plot has a value of  $-0.5$ , the end of each axis has a value of  $0.5$ , and the mid-point on each axis is zero; these values are in standardized units. The three mean factor scores for each protein define the points of a filled triangle that has been drawn to illustrate the similarities and differences among each of the proteins. Based upon the shape of each of the filled triangles, we can discriminate two groups of domain organizations: one comprised of the wild-type Runx2 protein along with the Y407A and R398A mutants and a second group containing Y433A, Y428A, and the functionally compromised Runx2- $\Delta$ C mutant. Differences in the shape of the triangles highlight the selective alterations in subnuclear organization as a consequence of NMTS mutations.





**Fig. 5.** The subnuclear organization of Runx domains is linked with subnuclear targeting, biological function and disease. In order to determine the extent to which the subnuclear organization of each mutant differs from wild-type we performed hierarchical cluster analysis using the Euclidean distance matrix and complete linkage. Cluster organization is illustrated using a dendrogram. Subnuclear organization data is presented in a compressed form with a color map as described in Fig. 3. As shown, there are two main clusters: one including wild-type and one including the Runx2-ΔC protein, which does not contain the NMTS. We find a clear parallel between this cluster analysis and our factor analysis, particularly with respect to the clustering of Runx2-ΔC with Y433A and Y428A. This parallel lends strength to the observed clusters. Shown at the bottom is a symbolic representation of the extent to which each protein associates with the nuclear matrix as determined by biochemical fractionation and western blot analysis, i.e. ranging from '+++' (associated) for wild-type to '-' (no association for Runx2-ΔC) (Zaidi et al., 2001; Choi et al., 2001) (our unpublished observations). We find a correlation between subnuclear organization and nuclear matrix association. The schematic below indicates whether a protein will promote differentiation or is involved in disease (i.e. cleidocranial dysplasia) [yes, no, or not determined (ND)].

further insight into the Runx nuclear structure-function relationships, we directly compare the hierarchical cluster arrangement and the intranuclear targeting competency of each protein, as well as their contribution to development and disease. This analysis reveals a link between Runx subnuclear domain organization and biological function. We conclude that the architectural organization of Runx transcription factors within the nucleus is fundamental to their tissue specific regulatory function.

## Conclusion

Knowledge of the biochemical and genetic components of gene regulation, replication and repair far exceeds our understanding of the integration of these processes within the context of nuclear architecture (Stein et al., 2000a; Stein et al., 2000b; Lemon and Tjian, 2000; Dundr and Misteli, 2001; Iborra and Cook, 2002; Spector, 2003; Stein et al., 2003). Here we have established a bioinformatics approach to describe and define

organization of protein domains within the nucleus. Intranuclear informatics provides the quantitative platform to capture the relevant parameters of subnuclear organization and relate them to the fundamental requirements for biological control. Using this approach, we have demonstrated that the focal subnuclear organization of Runx proteins is conserved in progeny cells. Our strategy has enabled us to discriminate between functional and non-functional Runx proteins based upon their domain organization within the nucleus. Furthermore, we have identified an architectural signature of Runx transcription factors that is coupled with fidelity of intranuclear targeting. In a broader context, intranuclear informatics can be applied to analyze subtle alterations in any spatially organized nuclear microenvironments under normal and pathological conditions.

We thank Brian Wiener of Worcester Polytechnic Institute and Jitesh Pratap of our laboratory for helpful comments in preparing this manuscript. We thank Elizabeth Bronstein and Judy Rask for editorial assistance. Studies reported were in part supported by grants from the National Institutes of Health (PO1CA82834, PO1AR48818, AR39588), as well as grant 5P30 DK32520 from the National Institute of Diabetes and Digestive and Kidney Diseases. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

## References

- Barseguian, K., Lutterbach, B., Hiebert, S. W., Nickerson, J., Lian, J. B., Stein, J. L., van Wijnen, A. J. and Stein, G. S. (2002). Multiple subnuclear targeting signals of the leukemia-related AML1/ETO and ETO repressor proteins. *Proc. Natl. Acad. Sci. USA* **99**, 15434-15439.
- Carrington, W. A., Lynch, R. M., Moore, E. D., Isenberg, G., Fogarty, K. E. and Fay, F. S. (1995). Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. *Science* **268**, 1483-1487.
- Choi, J.-Y., Pratap, J., Javed, A., Zaidi, S. K., Xing, L., Balint, E., Dalamangas, S., Boyce, B., van Wijnen, A. J., Lian, J. B. et al. (2001). Subnuclear targeting of Runx/Cbfa/AML factors is essential for tissue-specific differentiation during embryonic development. *Proc. Natl. Acad. Sci. USA* **98**, 8650-8655.
- Clark, P. J. and Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**, 445-453.
- Cook, P. R. (1999). The organization of replication and transcription. *Science* **284**, 1790-1795.
- Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301.
- Donnelly, K. P. (1978). Simulations to determine the variance and edge effect of total nearest-neighbor distance. In *Simulation studies in archaeology* (ed. I. Hodder), pp. 91-95. London, UK: Cambridge University Press.
- Dundr, M. and Misteli, T. (2001). Functional architecture in the cell nucleus. *Biochem. J.* **356**, 297-310.
- Dyck, J. A., Maul, G. G., Miller, W. H., Chen, J. D., Kakizuka, A. and Evans, R. M. (1994). A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. *Cell* **76**, 333-343.
- Francstel, C., Schubeler, D., Martin, D. I. and Groudine, M. (2000). Nuclear compartmentalization and gene activity. *Nat. Rev. Mol. Cell Biol.* **1**, 137-143.
- Hanai, J., Chen, L. F., Kanno, T., Ohtani-Fujita, N., Kim, W. Y., Guo, W.-H., Imamura, T., Ishidou, Y., Fukuchi, M., Shi, M. J. et al. (1999). Interaction and functional cooperation of PEBP2/CBF with smads. Synergistic induction of the immunoglobulin germline  $\alpha$  promoter. *J. Biol. Chem.* **274**, 31577-31582.
- Harrington, K. S., Javed, A., Drissi, H., McNeil, S., Lian, J. B., Stein, J. L., van Wijnen, A. J., Wang, Y.-L. and Stein, G. S. (2002). Transcription factors RUNX1/AML1 and RUNX2/Cbfa1 dynamically associate with stationary subnuclear domains. *J. Cell Sci.* **115**, 4167-4176.
- Iborra, F. J. and Cook, P. R. (2002). The interdependence of nuclear structure and function. *Curr. Opin. Cell Biol.* **14**, 780-785.

- Javed, A., Guo, B., Hiebert, S., Choi, J.-Y., Green, J., Zhao, S.-C., Osborne, M. A., Stifani, S., Stein, J. L., Lian, J. B. et al. (2000). Groucho/TLE/R-Esp proteins associate with the nuclear matrix and repress RUNX (CBF $\alpha$ /AML/PEBP2 $\alpha$ ) dependent activation of tissue-specific gene transcription. *J. Cell Sci.* **113**, 2221-2231.
- Karpuj, M. V., Garren, H., Slunt, H., Price, D. L., Gusella, J., Becher, M. W. and Steinman, L. (1999). Transglutaminase aggregates huntingtin into nonamyloidogenic polymers, and its enzymatic activity increases in Huntington's disease brain nuclei. *Proc. Natl. Acad. Sci. USA* **96**, 7388-7393.
- Komori, T. (2002). Runx2, a multifunctional transcription factor in skeletal development. *J. Cell Biochem.* **87**, 1-8.
- Kozubek, S., Lukasova, E., Jirsova, P., Koutna, I., Kozubek, M., Ganova, A., Bartova, E., Falk, M. and Pasekova, R. (2002). 3D Structure of the human genome, order in randomness. *Chromosoma* **111**, 321-331.
- Kundu, M., Javed, A., Jeon, J. P., Horner, A., Shum, L., Eckhaus, M., Muenke, M., Lian, J. B., Yang, Y., Nuckolls, G. H. et al. (2002). Cbfbeta interacts with Runx2 and has a critical role in bone development. *Nat. Genet.* **32**, 639-644.
- Lemon, B. and Tjian, R. (2000). Orchestrated response, a symphony of transcription factors for gene control. *Genes Dev.* **14**, 2551-2569.
- Lian, J. B. and Stein, G. S. (2003). The temporal and spatial subnuclear organization of skeletal gene regulatory machinery, integrating multiple levels of transcriptional control. *Calcif. Tissue Int.* **72**, 631-637.
- Lutterbach, B. and Hiebert, S. W. (2000). Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene* **245**, 223-235.
- Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P. C., Meng, C. and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. *J. Cell Biol.* **143**, 1415-1425.
- McNeil, S., Zeng, C., Harrington, K. S., Hiebert, S., Lian, J. B., Stein, J. L., van Wijnen, A. J. and Stein, G. S. (1999). The t(8;21) chromosomal translocation in acute myelogenous leukemia modifies intranuclear targeting of the AML1/CBF $\alpha$ 2 transcription factor. *Proc. Natl. Acad. Sci. USA* **96**, 14882-14887.
- Misteli, T. (2001). Protein dynamics, implications for nuclear architecture and gene expression. *Science* **291**, 843-847.
- Nielsen, J. A., Hudson, L. D. and Armstrong, R. C. (2002). Nuclear organization in differentiating oligodendrocytes. *J. Cell Sci.* **115**, 4071-4079.
- Noordmans, H. J., van der Kraan, K., van Driel, R. and Smeulders, A. W. (1998). Randomness of spatial distributions of two proteins in the cell nucleus involved in mRNA synthesis and their relationship. *Cytometry* **33**, 297-309.
- Norman, G. R. (2000). *Biostatistics, The Bare Essentials*. Hamilton, OH: Decker.
- Shiels, C., Islam, S. A., Vatcheva, R., Sasieni, P., Sternberg, M. J., Freemont, P. S. and Sheer, D. (2001). PML bodies associate specifically with the MHC gene cluster in interphase nuclei. *J. Cell Sci.* **114**, 3705-3716.
- Sinclair, D. F. (1985). On tests of spatial randomness using mean nearest neighbor distance. *Ecology* **66**, 1084-1085.
- Spector, D. L. (2003). The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573-608.
- Stein, G. S., Montecino, M., van Wijnen, A. J., Stein, J. L. and Lian, J. B. (2000a). Nuclear structure – gene expression interrelationships, implications for aberrant gene expression in cancer. *Cancer Res.* **60**, 2067-2076.
- Stein, G. S., van Wijnen, A. J., Stein, J. L., Lian, J. B., Montecino, M., Choi, J.-Y., Zaidi, K. and Javed, A. (2000b). Intranuclear trafficking of transcription factors, implications for biological control. *J. Cell Sci.* **113**, 2527-2533.
- Stein, G. S., Zaidi, S. K., Braastad, C. D., Montecino, M., van Wijnen, A. J., Choi, J.-Y., Stein, J. L., Lian, J. B. and Javed, A. (2003). Functional architecture of the nucleus, organizing the regulatory machinery for gene expression, replication and repair. *Trends Cell Biol.* **13**, 584-592.
- Stenoien, D. L., Nye, A. C., Mancini, M. G., Patel, K., Dutertre, M., O'Malley, B. W., Smith, C. L., Belmont, A. S. and Mancini, M. A. (2001). Ligand-mediated assembly and real-time cellular dynamics of estrogen receptor alpha-coactivator complexes in living cells. *Mol. Cell. Biol.* **21**, 4404-4412.
- Tanabe, H., Habermann, F. A., Solovei, I., Cremer, M. and Cremer, T. (2002). Non-random radial arrangements of interphase chromosome territories, evolutionary considerations and functional implications. *Mutat. Res.* **504**, 37-45.
- Thomas, D. M., Carty, S. A., Piscopo, D. M., Lee, J. S., Wang, W. F., Forrester, W. C. and Hinds, P. W. (2001). The retinoblastoma protein acts as a transcriptional coactivator required for osteogenic differentiation. *Mol. Cell* **8**, 303-316.
- Tracey, W. D. and Speck, N. A. (2000). Potential roles for RUNX1 and its orthologs in determining hematopoietic cell fate. *Semin. Cell Dev. Biol.* **11**, 337-342.
- Tumbar, T. and Belmont, A. S. (2001). Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. *Nat. Cell Biol.* **3**, 134-139.
- Westendorf, J. J., Zaidi, S. K., Cascino, J. E., Kahler, R., van Wijnen, A. J., Lian, J. B., Yoshida, M., Stein, G. S. and Li, X. (2002). Runx2 (Cbf1, AML-3) interacts with histone deacetylase 6 and represses the p21(CIP1/WAF1) promoter. *Mol. Cell. Biol.* **22**, 7982-7992.
- Yergeau, D. A., Hetherington, C. J., Wang, Q., Zhang, P., Sharpe, A. H., Binder, M., Marin-Padilla, M., Tenen, D. G., Speck, N. A. and Zhang, D. E. (1997). Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. *Nat. Genet.* **15**, 303-306.
- Zaidi, S. K., Javed, A., Choi, J.-Y., van Wijnen, A. J., Stein, J. L., Lian, J. B. and Stein, G. S. (2001). A specific targeting signal directs Runx2/Cbfa1 to subnuclear domains and contributes to transactivation of the osteocalcin gene. *J. Cell Sci.* **114**, 3093-3102.
- Zaidi, S. K., Sullivan, A. J., van Wijnen, A. J., Stein, J. L., Stein, G. S. and Lian, J. B. (2002). Integration of Runx and Smad regulatory signals at transcriptionally active subnuclear sites. *Proc. Natl. Acad. Sci. USA* **99**, 8048-8053.
- Zaidi, S. K., Young, D. W., Pockwinse, S. H., Javed, A., Lian, J. B., Stein, J. L., van Wijnen, A. J. and Stein, G. S. (2003). Mitotic partitioning and selective reorganization of tissue specific transcription factors in progeny cells. *Proc. Natl. Acad. Sci. USA* **100**, 14852-14857.
- Zaidi, S. K., Sullivan, A. J., Medina, R., Ito, Y., van Wijnen, A. J., Stein, J. L., Lian, J. B. and Stein, G. S. (2004). Tyrosine phosphorylation controls Runx2-mediated subnuclear targeting of YAP to repress transcription. *EMBO J.* **23**, 790-799.
- Zeng, C., van Wijnen, A. J., Stein, J. L., Meyers, S., Sun, W., Shopland, L., Lawrence, J. B., Penman, S., Lian, J. B., Stein, G. S. et al. (1997). Identification of a nuclear matrix targeting signal in the leukemia and bone-related AML/CBF $\alpha$  transcription factors. *Proc. Natl. Acad. Sci. USA* **94**, 6746-6751.
- Zhang, Y. W., Yasui, N., Kakazu, N., Abe, T., Takada, K., Imai, S., Sato, M., Nomura, S., Ochi, T., Okuzumi, S. et al. (2000). PEBP2 $\alpha$ A/CBFA1 mutations in Japanese cleidocranial dysplasia patients. *Gene* **244**, 21-28.