

Effects of HMGN variants on the cellular transcription profile

Mark Rochman¹, Leila Taher², Toshihiro Kurahashi¹, Srujana Cherukuri¹,
Vladimir N. Uversky^{3,4,5}, David Landsman², Ivan Ovcharenko² and Michael Bustin^{1,*}

¹Protein Section, Laboratory of Metabolism, Center for Cancer Research, National Cancer Institute,
²Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20894, ³Institute for Intrinsically Disordered
Protein Research, Center for Computational Biology and Bioinformatics, Department of Biochemistry and
Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, ⁴Department of Molecular
Medicine, University of South Florida, Tampa, FL 33612, USA and ⁵Institute for Biological Instrumentation,
Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

Received December 6, 2011; Accepted December 22, 2010

ABSTRACT

High mobility group N (HMGN) is a family of intrinsically disordered nuclear proteins that bind to nucleosomes, alters the structure of chromatin and affects transcription. A major unresolved question is the extent of functional specificity, or redundancy, between the various members of the HMGN protein family. Here, we analyze the transcriptional profile of cells in which the expression of various HMGN proteins has been either deleted or doubled. We find that both up- and downregulation of HMGN expression altered the cellular transcription profile. Most, but not all of the changes were variant specific, suggesting limited redundancy in transcriptional regulation. Analysis of point and swap HMGN mutants revealed that the transcriptional specificity is determined by a unique combination of a functional nucleosome-binding domain and C-terminal domain. Doubling the amount of HMGN had a significantly larger effect on the transcription profile than total deletion, suggesting that the intrinsically disordered structure of HMGN proteins plays an important role in their function. The results reveal an HMGN-variant-specific effect on the fidelity of the cellular transcription profile, indicating that functionally the various HMGN subtypes are not fully redundant.

INTRODUCTION

The dynamic architecture of the chromatin fiber plays a key role in regulating transcriptional processes necessary for proper cell function and mounting adequate responses to various internal and external biological signals. Architectural nucleosome-binding proteins such as the linker histone H1 protein family and the high mobility group (HMG) protein superfamily are known to continuously and reversibly bind to chromatin, transiently altering its structure and affecting the cellular transcription output (1,2). Although extensively studied, the cellular function and mechanism of action of these chromatin-binding architectural proteins are still not fully understood. A major question in this field is the extent of the functional specificity of the structural variants of histone H1 or of the various HMG families (3–6). Experiments with genetically altered mice lacking one or several H1 variants revealed that loss of one variant leads to increase synthesis of the remaining variants, suggesting functional redundancy between H1 variants (7,8). Yet, analysis of cells in which the levels of specific H1 variants have been altered suggests a certain degree of variant-specific effects on transcriptional output (9–11)

The HMG superfamily is composed of three families named HMGA, HMGB and high mobility group N (HMGN), each containing several protein members (3,4). It is known that HMG proteins affect transcription and modulate the cellular phenotype (12); however, the

*To whom correspondence should be addressed. Tel: +1 301 496 5234; Fax: +1 301 496 8419; Email: bustin@helix.nih.gov

Present address:

Srujana Cherukuri, University of Maryland School of Medicine, Department of Pediatrics, Center for Stem Cell and Regenerative Medicine 20 Penn Street, HSF II Room S103, Baltimore, MD 21204, USA.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Published by Oxford University Press 2011.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

transcriptional specificity of the various HMG variants has not yet been systematically studied. Here, we examine the role of the various HMGN variants in the regulation of the cellular transcription profile.

The HMGN family of chromatin architectural proteins consists of five members with a similar structure (13). All contain a bipartite nuclear localization signal (NLS), a highly conserved nucleosome-binding domain (NBD) and a negatively charged and highly disordered C-terminal domain. The HMGNs are the only nuclear proteins known to specifically recognize generic structural features of the 147-bp nucleosome core particle (CP), the building block of the chromatin fiber (3,4). HMGN binds to chromatin and CP without any known specificity for the sequence of the underlying DNA. In the nucleus, HMGNs are highly mobile moving among nucleosomes in a stop-and-go manner (2,14). The fraction of time that an HMGN resides on a nucleosome (stop period) is longer than the time it takes to 'hop' from one nucleosome to another; therefore, most of the time, most of the HMGNs are bound to chromatin. The amount of HMGN present in most nuclei is sufficient to bind only ~1% of the nucleosomes; however, the dynamic binding of HMGNs to chromatin ensures that potentially every nucleosome will temporarily interact with an HMGN molecule. Thus, potentially, HMGNs may affect the transcription of numerous genes.

HMGN variants share several functional properties, such as binding affinity to nucleosomes *in vitro* and *in vivo*, competition with linker histone H1 for the binding sites on nucleosomes, and effects on chromatin architecture. Likewise, both HMGN1 and HMGN2, the most abundant and ubiquitous members of this protein family, form multiple complexes with nuclear proteins (15). These findings, and the similarity of their domain structure, suggest that by enlarge, HMGN proteins could be functionally redundant. Yet, several studies indicate that HMGN proteins are not fully redundant. Both *in vivo* and *in vitro* studies indicate that the interaction of HMGN variants with CPs lead to the formation of complexes containing two molecules of a single type of variant; CPs containing two different HMGN variants are not formed under physiological conditions (16,17). In addition, while HMGN1 and HMGN2 seem to be ubiquitously expressed, HMGN3 and HMGN5 proteins show distinct developmental and tissue-specific expression (18–20). Most significantly, analysis of genetically altered mice and cells revealed variant-specific phenotypes and indication that the variants are not fully functionally redundant (12).

It has been repeatedly shown that interaction of HMGNs with chromatin affects transcription (21–24). However, the extent of specificity of HMGN variants in transcriptional regulation and the level of functional redundancy between them remain largely unknown, mainly because of the lack of systematic analysis of the effect of HMGNs on gene expression in a unified experimental system.

To gain insights into the extent of transcriptional specificity of the HMGN variants, we compared expression profiles of mouse embryonic fibroblasts (MEFs) in which

various HMGN variants were either knocked out or stably overexpressed, to double their cellular content. We found that loss of proteins affected the expression of a limited number of genes, while doubling the cellular levels of an HMGN variant affected the expression of hundreds of genes. While some of the genes were affected by more than one variant, the great majority of the genes were affected in a variant-specific manner.

Intrinsically disordered proteins are predicted to affect transcription even at low dosage overexpression because they form weak interactions with multiple partners (25,26). Thus, the significant transcriptional effects resulting from doubling the amount of HMGNs is in agreement with the highly intrinsically disordered structure of HMGN proteins and with their tendency to form multiple metastable protein complexes (15). We also found that specific variants affect the transcription profile in a cell-specific manner. Analysis of domain swap mutants suggests that the specificity of each HMGN is determined by a unique combination of a functional NBD and a C-terminal domain.

The results reveal an HMGN-variant-specific effect on the global transcription profile suggesting that these proteins fine tune the fidelity of the cellular transcription. We speculate that part of their specificity is due to their intrinsic highly disordered structure that enables each variant to form multiple types of complexes with nuclear components.

EXPERIMENTAL PROCEDURES

Isolation of MEFs and generation of stable cell lines

Mouse embryonic fibroblasts SV-40 transformed (MEFs) were purchased from ATCC.

MIN6 cell line was a gift from A.L. Notkins, NIDCR, NIH. Primary MEFs from variant-specific knock out mice were isolated from two embryos as described (27) and analyzed separately. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Calf Serum (FCS).

Retroviruses were produced in Phoenix helper cell line transfected with the pHAN vector bearing various HMGN proteins tagged with FLAG and HA at C-terminus. Stable cell lines were generated by retroviral infection in the presence of polybrene at the concentration of 5 µg/ml and subsequent selection with puromycin at the concentration of 1 µg/ml for 7 days. Cells were grown without antibiotics for 1 day prior to collecting samples for expression analysis and western blotting.

Antibodies and western blotting

All the antibodies used in the study were from our laboratory. Secondary Horseradish Peroxidase (HRP)-Conjugated antibodies were from Pierce.

Whole cell lysates were prepared in 2× Laemmli sample buffer (Bio-Rad) supplemented with protease inhibitors. Samples were separated on 15% pre-cast Criterion gels, transferred by semi-dry method to polyvinylidene difluoride (PVDF) membrane, blocked with non-fat milk in Phosphate Buffered Saline (PBS) and probed with

indicated antibodies. Chemiluminiscent detection by enhanced chemiluminescence (ECL) Detection Reagent (Amersham) was done according to the manufacturer's recommendations.

RNA preparation

RNA was prepared by TRIzol reagent according to the manufacturer's protocol. Subsequently, RNA was cleaned up by Qiagen RNeasy kit with on-column DNaseI treatment.

Gene arrays. Microarray expression analysis was performed using Affymetrix Mouse GeneChips 430 2 (430v2). Hybridization of biotin-labeled cRNA fragment to Mouse Genome 430 2.0 array, washing, staining with streptavidin-phycoerythrin (Molecular Probes), and signal amplification were performed according to the manufacturer's instructions at the Laboratory of Molecular Technology (LMT, Frederic, NCI).

Statistical analysis

We analyzed 51 array data sets ($n = 3$ for each particular experiment) to search for genes whose expression levels were significantly altered. All analyses were performed using R and BioConductor (28). R packages 'affy' (29) and 'simpleaffy' (30) and 'affyQCReport' (31) were employed to evaluate the quality of the arrays by means of images, histograms, box plots, degradation plots and scatter plots. Expression values were derived using the Robust Multichip Average protocol (32) with default settings. All analyses were done at the so-called sequence level, i.e. data from probes representing the same gene were combined. We did not apply any unspecific filter on the expression values.

Differentially expressed genes were identified using an empirical Bayes method implemented in the R package 'Limma' (33). *P*-values were corrected for multiple testing using a false discovery rate method (34). Genes for which the adjusted *P*-value was <0.001 (overexpression of different HMGN variants) or <0.05 (knock out of different HMGN variants) in at least one of the comparisons were considered differentially expressed. No fold-change cut-off was applied.

Mouse Genome 430 2.0 array has 45101 probe sets associated with approximately 20000 Mouse Genome Informatics (MGI) gene identifiers. Probe sets were mapped to MGI identifiers using information provided by the Jackson Laboratory (<http://www.informatics.jax.org/>).

Venn diagrams

Differentially expressed genes in all experiments were compared to controls and represented as Venn diagrams (R package 'Vennable') (35).

Functional analysis

Functional analysis of microarray data was based on overrepresentation of GO terms (36). *P*-values were corrected for multiple comparisons using Bonferroni's method.

Bioinformatics structural analysis

Composition profiling. Analysis of amino acid composition of HMGN proteins was performed using Composition Profiler online service (<http://www.cprofiler.org>) (37) with default settings. The following reference protein sets were used: DisProt 3.4 (38), PDB Select 25 (39) and mouse HMGN proteins. The set DisProt 3.4 comprises consensus sequences of experimentally determined disordered regions; PDB Select 25 contains PDB structures with $<25\%$ sequence identity, biased toward the composition of proteins amenable to crystallization studies. Amino acids are arranged in the order of increase of their disorder propensity, according to the scale by Radivojac *et al.* (40).

Intrinsic disorder prediction. Per-residue predictions of intrinsic disorder in HMGN proteins were performed using a PONDR[®] VLXT predictor, access to which was provided by Molecular Kinetics, Inc. (<http://www.pondr.com>). PONDR[®] (Predictor Of Natural Disordered Regions) is a set of neural network predictors of disordered regions on the basis of local amino acid composition, flexibility, hydrophathy, coordination number and other factors. These predictors classify each residue within a sequence as either ordered or disordered. PONDR[®] VL-XT integrates three feed forward neural networks: the Viciously characterized Long, version 1 (VL1) predictor (41), which predicts non-terminal residues, and the X-ray characterized N- and C-terminal predictors (XT) (42), which predicts terminal residues. Output for the VL1 predictor starts and ends 11 amino acids from the termini. The XT predictor output provides predictions up to 14 amino acids from their respective ends. A simple average is taken for the overlapping predictions; a sliding window of nine amino acids is used to smooth the prediction values along the length of the sequence. Unsmoothed prediction values from the XT predictors are used for the first and last four sequence positions.

RESULTS

Structural characterization of HMGN variants

Examination of the structure of genes coding for the various members of the HMGN family suggests that they originated from a common ancestor. All the genes contain relatively long 5'- and 3'-untranslated regions, six exons and the boundaries of the first four exons are highly conserved (Figure 1A). The gene coding for HMGN5 evolved recently because it is found only in mammals. All the proteins encoded by the genes contain a positively charged, highly conserved, NBD (Figure 1A) that serves as their main chromatin-binding site. Embedded in the NBD is the sequence RRSARLSA (K,M)P that has been shown to be the core sequence that specifically anchors HMGN proteins to the 147-bp nucleosome CP, the building block of the chromatin fiber (43). A NLS that is localized at the N-terminal part of the proteins is also highly conserved in all HMGN variants. The C-terminal region of the proteins, encoded by exons

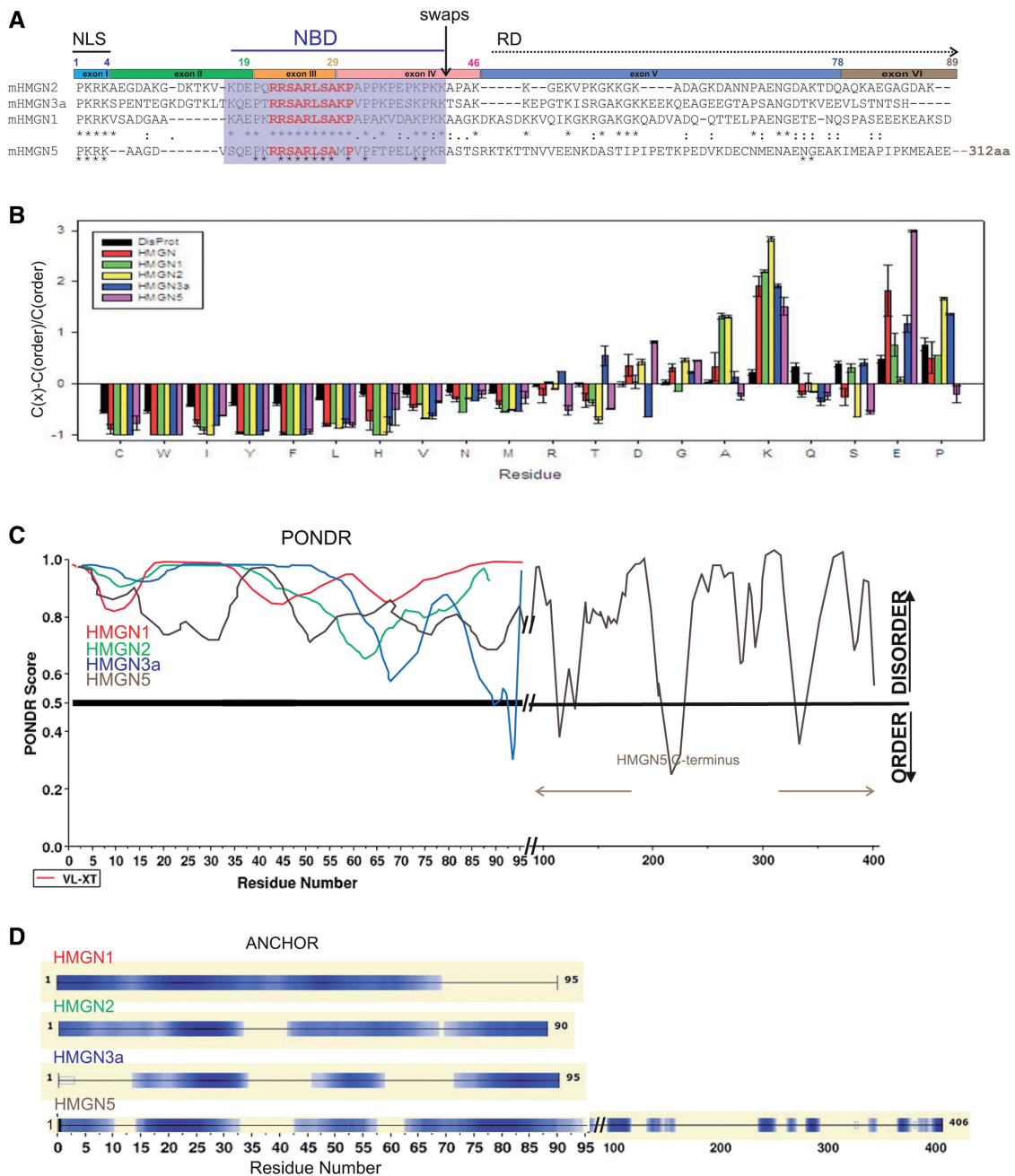


Figure 1. HMGN proteins are intrinsically disordered. (A) Multiple sequence alignment of mouse HMGN1, HMGN2, HMGN3a and HMGN5 proteins by ClustalW. Only the first 94 amino acids of HMGN5 are aligned. The positively charged NBD, the hallmark of HMGN proteins, is shaded by a blue square. The core sequence of NBD that is conserved in all HMGN proteins is labeled in red. The exon structure of the HMGN genes is color-coded over the sequences; numbers over the exons correspond to the last amino acid encoded by the exons of the *Hmgn2* gene because HMGN2 is the most evolutionarily conserved HMGN variant. Asterisks indicate identical amino acid, colon indicates conserved substitutions and dot indicates semi-conserved substitutions. The alignment of HMGN5 is separate from that of HMGN1-3. NLS, nuclear localization signal; RD, regulatory domain. Solid arrow indicates the position of the swap tail mutants (Figure 4). (B) Relative amino acid composition of various HMGN proteins in comparison with ordered proteins. Bars are calculated as $C(x) - C(\text{order})/C(\text{order})$, where $C(x)$ is the content of a given residue in HMGN and $C(\text{order})$ is its content in ordered proteins from Protein Data Bank (<http://www.pdb.org/pdb/home/home.do>). Negative bars correspond to residues underrepresented in HMGN, whereas positive bars correspond to residues overrepresented in HMGN. Data for typical intrinsically disordered proteins are shown for comparison (DisProt, <http://www.disprot.org>, black bars). Sets of bars correspond to mean values for all HMGNs (HMGN) as well as for individual HMGNs (HMGN1, HMGN2, HMGN3a and HMGN5). The graph demonstrates that potentially HMGNs are more disordered than the averaged disordered proteins. (C) PONDR VL-XT disorder prediction for mouse HMGNs. In PONDR plots, segments with scores >0.5 correspond to the disordered regions, whereas those <0.5 correspond to the ordered regions/binding sites. Note that disorder distribution in NBD (residues 18–42) is conserved for HMGN1, HMGN2 and HMGN3a. HMGN5 shows much less disorder conservation. (D) Predicting potential binding sites by ANCHOR algorithm. Potential binding sites are indicated by blue boxes.

5 and 6, differs significantly among the HMGN variants. The HMGN5 C-terminal domain is especially long and contains several repeats of a negatively charged sequence motif (18). The alignment shown in Figure 1A illustrates the major similarities and differences between the mouse HMGN variants. Mouse HMGN1, HMGN2 and HMGN3a are similar in size, ranging between 89 and 95 amino acids, and are more similar to each other than to HMGN5, which is 406 amino acid long. The alignment does not contain the splice variant HMGN3b, which lacks the 21 C-terminal residues of HMGN3a, nor the HMGN4 variant, which has not yet been investigated in detail.

Analysis of the amino acid composition of the HMGN proteins in comparison to ordered proteins listed in the Protein Data Bank (<http://www.pdb.org/pdb/home/home.do>) reveals that potentially all HMGNs are highly disordered proteins (Figure 1B). In fact, HMGNs are expected to be significantly more disordered than an 'average' disordered protein, because they are much more depleted in major order-promoting residues (compare the colored bars with negative values for various HMGNs with the black bars for Intrinsically Disordered Proteins at the left side of the plot) and are significantly enriched in major disorder-promoting residues (right side of the plot). Interestingly, the HMGN variants are different from each other and show significant variability in amino acid compositions, as exemplified by the large variations in R, T, D, G, A, S, E and P.

In agreement with intrinsic disorder prediction of HMGN proteins based on amino acid composition, PONDR analysis (41,44) predicts a high degree of structural disorder in all HMGN variants, with a few short regions with increased order propensity (Figure 1C, dips in the graph). These relatively ordered regions often correspond to potential binding sites that fold upon interaction with binding partners (45–47). Notably, all HMGNs contain several regions that according to the ANCHOR algorithm (48) are predicted to serve as binding sites for other interacting proteins (Figure 1D), an observation fully compatible with our previous findings that both HMGN1 and HMGN2 can be found in numerous metastable multiprotein complexes (15). The number and localization of the predicted protein binding sites although highly similar are not identical between HMGN variants.

In summary, although all HMGNs share several physical properties and are nuclear proteins that bind to nucleosome CPs through a highly conserved domain, each variant has a distinct structure and has several sites for interacting with other proteins. These characteristics raise the possibility of HMGN-variant-specific effects on the cellular transcription profile.

Transcriptional impact of HMGN proteins

To investigate the transcriptional specificity of HMGN variants we first analyzed the transcriptional profile of primary MEFs isolated from *Hmgn1*^{-/-}, *Hmgn3*^{-/-} and *Hmgn5*^{-/-} mice using mouse 430.2 Affymetrix expression arrays. *Hmgn2*^{-/-} are not available because these mice are

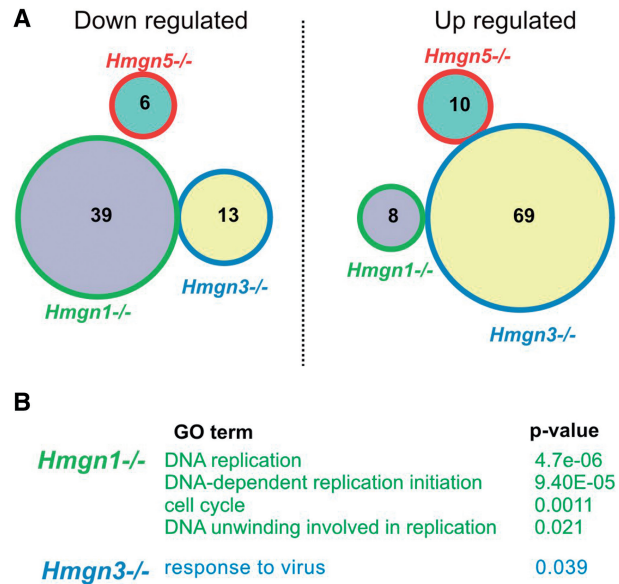


Figure 2. Effects of HMGNs knock out on transcription in primary MEFs. (A) Venn diagrams of down- and upregulated genes in primary MEFs. (B) GO analysis of affected genes ($P < 0.05$).

embryonic lethal (M.B. unpublished data). The results reveal variant-specific changes in gene expression profile; no overlap was observed for the genes affected by the knockout of different HMGN variants (Figure 2A). The changes involved both up- and downregulation of transcript levels, a finding that is fully compatible with the notion that HMGNs enhance transcriptional fidelity by affecting chromatin structure and optimizing the fidelity of transcription. Even though transcription of a relatively small number of genes was affected, Gene Ontology (GO) analysis revealed significant enrichment in a few non-overlapping pathways for *Hmgn1*^{-/-}, *Hmgn3*^{-/-} MEFs (Figure 2B). These results are in agreement with our previous observations that the phenotypes of *Hmgn1*^{-/-} and *Hmgn3*^{-/-} mice are distinct but not severe (19,49).

Because HMGN proteins are intrinsically disordered proteins (Figure 1) and because dosage changes in such proteins may lead to large changes in transcription (25), we reasoned that a mild increase in the cellular levels of HMGN variants, in a uniform system, may give a more sensitive indication of the potential transcriptional specificity of the HMGN variants. To compare the effect of the overexpression of HMGN variants on transcription in a uniform system, we used retroviral infection to generate MEFs cell lines stably expressing specific HMGN variants tagged with FLAG and HA at their C-terminus. Vectors expressing HMGN1, HMGN2, HMGN3a, HMGN5 and the HMGN5-S17,21E double-point mutant, which does not bind to chromatin (18), were generated and efficiently expressed in MEFs. Following infection, cells were subjected to the selection procedure and all the cells that passed the selection were analyzed as a pool.

Western blot analysis of the infected cells revealed that the level of expression of each exogenous protein

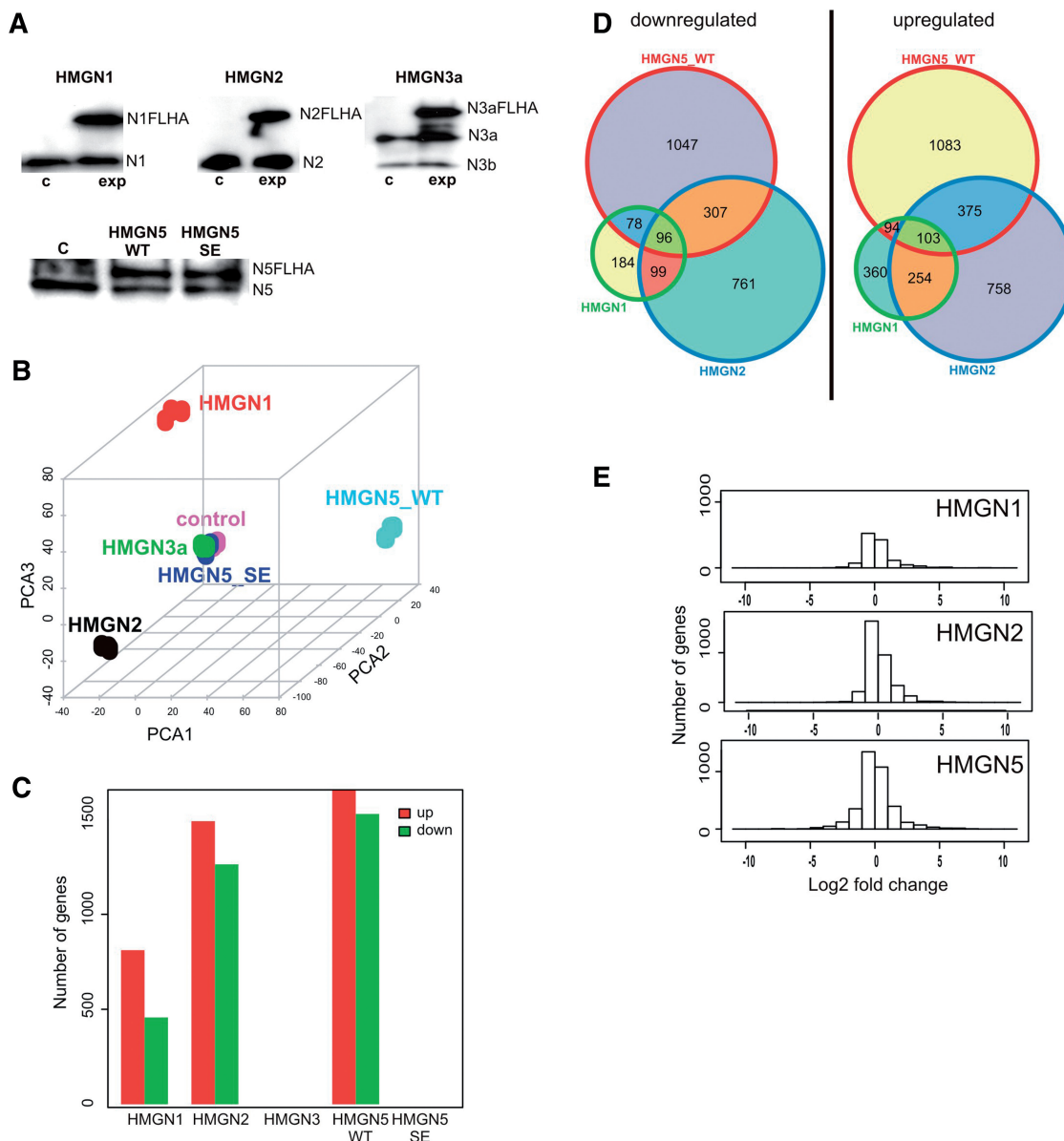


Figure 3. Effects of elevated expression of HMGNs on transcription in MEFs. (A) Western blot analysis of stably infected MEFs. Shown are western analysis of MEFs stably expressing FLAG and HA tagged HMGN variants. Endogenous and exogenous HMGN proteins are indicated. C, control infection with empty virus; exp, experimental infection with indicated protein. Note comparable amounts of exogenous and endogenous proteins for all cell lines. (B) PCA of gene expression profiles in infected MEFs. Each sample was analyzed in triplicate. Stably expressed proteins are indicated. Each dot corresponds to individual pool of indicated HMGN variant. (C) The graph represents the number of genes changed following stable expression of an HMGN protein, compared with the control empty vector expression. Note the negligible effect of HMGN3a and HMGN5S17,21E on transcription. (D) Venn diagrams of down- and upregulated genes in infected MEFs. (E) The plot represents fold change in transcription for all affected genes following HMGN1, HMGN2 and HMGN5 overexpression. Note that most of the genes are affected up to 2-fold.

was comparable to the level of its endogenous counterpart (Figure 3A). Thus, stably infected MEFs express ~2-fold higher levels of a specific HMGN variant. The HMGN5-S17,21E protein contains mutations in two serine residues in the NBD which abolish its binding to nucleosomes (18) and thus served as a control for transcriptional effects due to chromatin binding. As an additional control, we infected MEFs with virus carrying an empty vector. For each variant, we analyzed the transcription profile of three independently infected pools of MEFs using mouse 430.2 Affymetrix expression arrays.

We compared the transcription profile of MEFs over-expressing specific HMGN variants to the control cell lines transfected with empty vectors or with the HMGN5-S17,21E double-point mutant.

Class comparison between cell lines indicated that overexpression of HMGNs altered the expression level of 5203 genes. Three-dimensional clustering of these transcripts, based on principal component analysis (PCA), revealed that the various cell lines formed four distinct expression clusters (Figure 3B). Three of the clusters were formed by the cell overexpressing either HMGN1,

or HMGN2, or HMGN5. The fourth cluster was formed by cell lines overexpressing either the HMGN5-S17,21E double-point mutant, cell transfected with an empty vector, or by the cells overexpressing the HMGN3a variant. These results demonstrate that in MEFs, each HMGN variant has specific effects on transcription.

The cells differed not only in the specificity of genes affected but also in the number of genes affected. Doubling the levels of HMGN1, HMGN2 and HMGN5 affected the expression of 1268, 2753 and 3183 genes, whereas HMGN3a and HMGN5-S17,21E caused no significant changes in gene expression. For HMGN1, HMGN2 and HMGN5 proteins, the proportion of up- and down regulated genes was roughly equal, indicating that the proteins did not preferentially activate or inhibit transcription (Figure 3C).

More detailed comparison of the genes affected by each of the HMGN variants revealed that while each protein either up- or downregulated the expression of a unique set of genes, a fraction of the genes was affected by more than one HMGN protein, suggesting partial redundancy in transcriptional regulation (Figure 3D). Thus, of the 457 genes that were downregulated by overexpressing HMGN1, 40% were uniquely affected, 17% were also downregulated by HMGN5, 22% were also downregulated by HMGN2 and 21% were downregulated by all three HMGNs. Of the 811 genes that were upregulated in HMGN1-overexpressing cells, 44% were uniquely affected, 12% were also affected by HMGN5, 31% were also affected by HMGN2 and 13% were upregulated by all the HMGNs. For HMGN2, a total of 1263 genes were downregulated, of these 60% were specifically downregulated only by HMGN2, and 24% were also downregulated by HMGN5. Likewise, ~50% of the 1490 genes were specifically upregulated by HMGN2 and ~70% of the genes up- or downregulated by HMGN5 were specifically affected by HMGN5.

Only a small proportion of the 5203 genes whose transcription changed by overexpression of the HMGNs was commonly affected by all the three HMGN proteins. In all, 96 genes (2%) were downregulated and 103 (2%) were upregulated. The most extensive overlap was observed between HMGN1 and HMGN2 proteins; 44% of the genes upregulated by HMGN1 were also upregulated by HMGN2. We also found a large number of genes regulated by both HMGN5 and HMGN2; 478 and 403 genes were up- and downregulated by both proteins, respectively. While the total number of the genes whose transcription levels changes was statistically significant was relatively large, the transcription levels of most of the genes changed ~2-fold (Figure 3E). These findings are in agreement with previous studies indicating that while HMGN proteins affect the expression of many genes, the changes in transcription levels are relatively small (19,27,50).

Next, we performed a functional analysis on the sets of genes exclusively regulated by each individual protein, as well as on the sets of genes regulated by combinations of several proteins (Table 1) for significantly overrepresented GO terms. The results indicate that overexpression of individual HMGN variants affected gene from different

categories. Whereas HMGN1 affected genes involved in cell division and mitosis, HMGN2 regulated genes involved in regulation of transcription, development and chromatin binding. Notably, genes involved in cell cycle regulation were also preferentially affected in *Hmgn1*^{-/-} MEFs (Figure 2B). HMGN5-induced transcriptional changes were mainly associated with metabolic processes, protein, metal ion and transcription factor binding. We note, however, that the GO analysis suggests a certain degree of redundancy among the HMGN variants. For instance, GO term 'response to virus' (GO:0009615) was enriched for genes commonly regulated by HMGN1 and HMGN2 proteins. In addition, several biosynthetic processes, such as sterol biosynthesis (GO:0016126), cholesterol (GO:0006695) and lipid biosynthesis (GO:0008610), and others were enriched for the genes regulated by all three HMGN variants. In fact, of 199 genes regulated by all HMGNs, 52 genes are involved in various biosynthetic processes.

Taken together, the gene expression profiles and the GO analyses reveal a surprising degree of specificity in the effects of the various HMGN variants on the cellular transcription profile. The functional redundancy among the variants is lower than what would be expected from a set of proteins with structural similarities, which bind to nucleosomes with similar affinities, have highly similar NBDs and use an identical sequence motif to bind specifically to nucleosome CPs.

Transcriptional effects of HMGN swap mutants

Because the N-terminal half of the HMGNs are highly similar, while their C-terminal domains are clearly distinct (Figure 1A), we assumed that the functional specificity of the proteins resides in their C-terminal region. To test this assumption, we generated retroviral vectors expressing tail swap mutants in which C-terminal region of either HMGN2 or HMGN3a protein was fused to N-terminal part of HMGN1 protein, immediately after the conserved NBD (see Figure 1A for exact location of the regions swapped). We named these mutants as N1-N2 swap and N1-N3 swap. The correct expression of the swap mutants in MEFs infected with the retroviral vectors was verified by western blot analysis (Figure 4A) using an antibody elicited against the conserved NBD of the HMGN protein family, which recognizes all the HMGN variants (51).

The transcription profile of the MEFs expressing the swap mutants was determined by mouse 430.2 Affymetrix expression arrays and compared with that of cells infected with vectors expressing the native proteins. Three-dimensional clustering of the results using PCA (Figure 4B) revealed that the effect of the swap mutants on transcription was distinct from that of their 'source' proteins. Thus, while HMGN3a did not affect transcription (Figure 3) in MEFs, the N1-N3 swap mutant significantly affected the transcription of 1522 genes, most of which were distinct from the genes affected by HMGN1 (Figure 4C, 3-4). Further comparison of the genes affected by the swap mutants with the genes affected by either HMGN1 or HMGN2 (Figure 4C) supported the

Table 1. GO analysis of gene expression in MEFs with elevated HMGN levels

Gene set	No. of genes*	GO category	Description	Ontology	P-value**
HMGN1	544	GO:0051301	Cell division	BP	4.2e-05
		GO:0007067	Mitosis	BP	0.00066
HMGN2	1519	GO:0008270	Zinc ion binding	MF	0.038
		GO:0045449	Regulation of transcription	BP	6.6e-06
		GO:0005515	Protein binding	MF	1.6e-05
		GO:0000166	Nucleotide binding	MF	0.00024
		GO:0007507	Heart development	BP	0.0009
		GO:0006350	Transcription	BP	0.0042
		GO:0003682	Chromatin binding	MF	0.0049
HMGN5	2130	GO:0030324	Lung development	BP	0.035
		GO:0005515	Protein binding	MF	1.8e-09
		GO:0016740	Transferase activity	MF	1.3e-06
		GO:0046872	Metal ion binding	MF	9.1e-05
		GO:0005524	ATP binding	MF	0.0035
		GO:0008152	Metabolic process	BP	0.0046
		GO:0003700	Sequence-specific DNA-binding transcription factor activity	MF	0.0074
		GO:0045449	Regulation of transcription	BP	0.012
		GO:0006974	Response to DNA damage stimulus	BP	0.013
		GO:0000166	Nucleotide binding	MF	0.015
		GO:0016310	Phosphorylation	BP	0.036
		GO:0019838	Growth factor binding	MF	0.048
		GO:0042802	Identical protein binding	MF	0.049
HMGN1/HMGN2	353	GO:0009615	Response to virus	BP	3.2e-06
HMGN2/HMGN5	682	GO:0007155	Cell adhesion	BP	6.6e-05
		GO:0008201	Heparin binding	MF	0.0028
HMGN1/HMGN2/HMGN5	199	GO:0005515	Protein binding	MF	0.017
		GO:0016126	Sterol biosynthetic process	BP	5.3e-20
		GO:0006695	Cholesterol biosynthetic process	BP	6.9e-18
		GO:0006694	Steroid biosynthetic process	BP	9.1e-17
		GO:0008610	Lipid biosynthetic process	BP	3e-15
		GO:0008299	Isoprenoid biosynthetic process	BP	2.9e-09
		GO:0006418	tRNA aminoacylation for protein translation	BP	0.00035
		GO:0004812	aminoacyl-tRNA ligase activity	MF	0.0029
		GO:0003824	Catalytic activity	MF	0.0035
		GO:0016874	Ligase activity	MF	0.0037
		GO:0019287	Isopentenyl diphosphate biosynthetic process, mevalonate pathway	BP	0.011
		GO:0055114	Oxidation-reduction process	BP	0.027
		GO:0016491	Oxidoreductase activity	MF	0.037

BP, biological process; MF, molecular function.

* sets are exclusive.

** significance threshold is 0.05.

notion that transcriptional changes induced by the tail swap mutants differed from those observed for HMGN1, HMGN2 or HMGN3a proteins. The swap mutants specifically downregulated 621 genes (Figure 4C 1,3) and upregulated 402 genes (Figure 4C 2,4). Interestingly, the two swap mutants had very similar effects on the cellular transcription profile (Figure 4D). Of the 1252 and the 1177 genes, respectively, downregulated by the N1–N2 and the N1–N3 swap mutants, close to 80% genes overlapped. Likewise, most of the genes that were up-regulated by the N1–N2 swap mutant were also upregulated by the N1–N3 swap protein (Figure 4D). The similarity in the genes regulated by the swap mutants points out to the importance of their shared NBD region in determining the effect on the transcription profile. Yet, the effects were clearly distinct from HMGN1, the donor of their shared NBD, an indication that ultimately, the structure of the entire protein, the combination of individual N- and C-terminal domains, determines the functional specificity of the HMGN variants.

Cell type-specific effects on transcription

Surprisingly, our analysis revealed that overexpression of HMGN3a had no effect on transcription profile of the transfected MEFs. We previously reported that in MEFs the protein levels of HMGN3 are lower than those of HMGN1 and HMGN2 that are robustly expressed in most cells. However, HMGN3a is highly expressed in MIN6 cell, a mouse pancreatic cell line that secretes insulin (19). In these cells, small interfering RNA-mediated downregulation of HMGN3, but not that of HMGN1 or HMGN2, affects the transcription of genes involved in insulin secretion, suggesting that HMGN3a may affect transcription in a cell type-specific manner. To test this possibility, we first re-examined the relative amount of HMGN3 protein in MIN6 and MEF cells. Western blot analyses revealed that indeed the HMGN3 protein levels in MIN6 were significantly higher than in MEFs (Figure 5A). Next, we infected MIN6 cells with retroviral vectors expressing either HMGN1 or

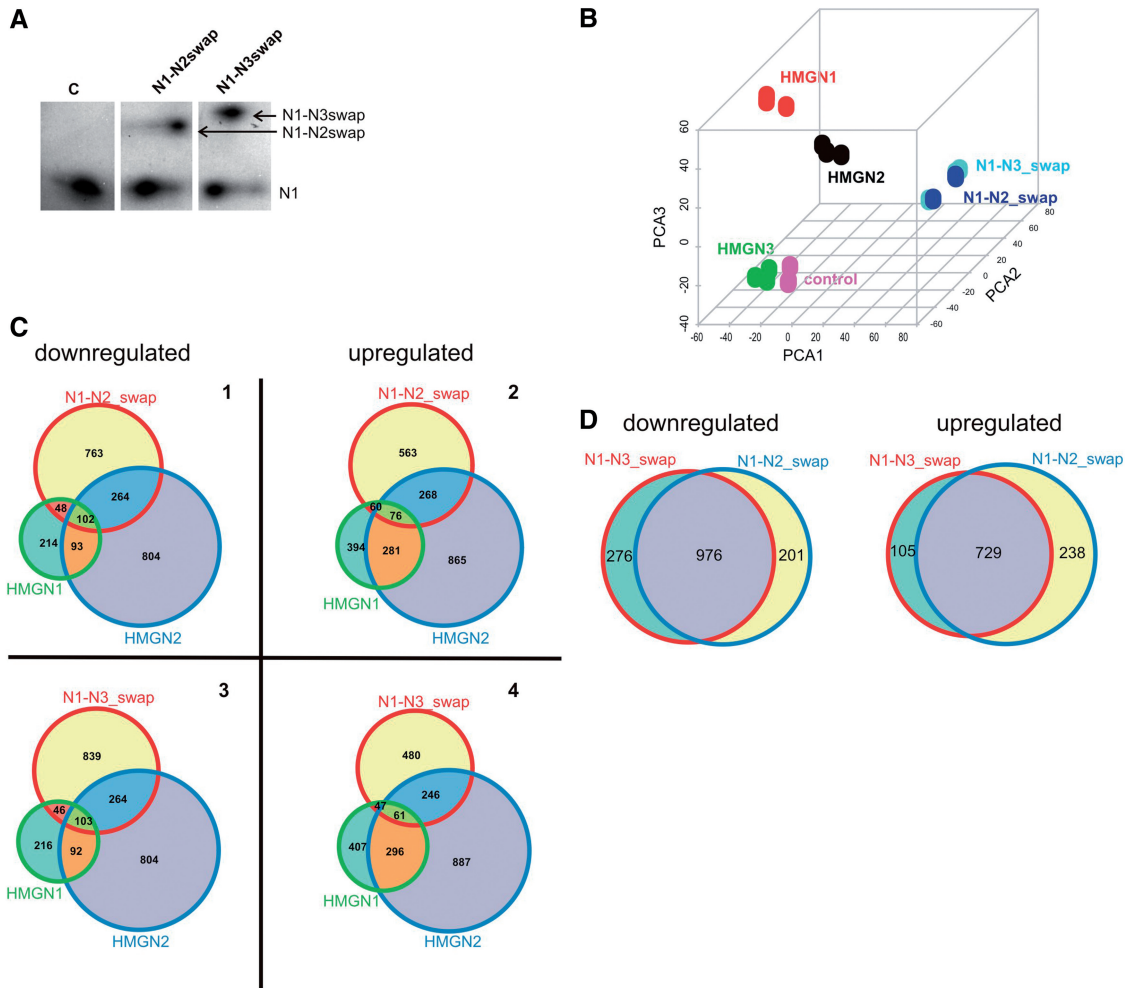


Figure 4. Comparison of the effect of HMGN tail swap mutants on transcription in MEFs. (A) Western blot analysis of MEFs stably expressed swap mutants using an antibody that recognizes the conserved NBD. N1, endogenous HMGN1; c, control expression of empty vector. (B) PCA of gene expression profiles in MEFs. Each sample was analyzed in triplicate. Stably expressed proteins are indicated. Each dot corresponds to individual pool of indicated HMGN variant. (C) Venn diagrams of down- and up-regulated genes in stable cell lines comparing N1-N2 swap (1,2) or N1-N3 swap (3,4) with HMGN1 and HMGN2 proteins. (D) Venn diagrams of down- and up-regulated genes in cells expressing tail swap mutant proteins.

HMGN3a proteins and verified protein expression by western blot analysis (Figure 5B and C). Transcriptional array analysis revealed that in MIN6 cells, HMGN3a significantly changed the expression of 1429 genes; of these 471 were up-regulated and 958 genes were down-regulated (Figure 5D). In contrast, overexpression of HMGN1 in MIN6 cell line had no significant effect on the transcription profile. Because HMGN1 had significant effects on the transcription profile of MEFs (Figure 3), these results suggest cell type-specific transcription effects of HMGN variants.

DISCUSSION

The major goal of this study is to examine whether the various members of the HMGN protein family can affect the cellular transcription profile in an HMGN-variant-specific manner. Although previous studies indicated that the binding of HMGN protein to chromatin

alters the cellular transcription profile, the degree to which these changes are HMGN-variant specific has not yet been investigated.

The dynamic nature of HMGN binding to chromatin and the lack of any DNA sequence specificity in their chromatin interactions, taken together with the conservation of their nuclear-binding domain and similarities in their overall organization and physical properties, raised the possibility that the individual HMGN variants would be functionally redundant and have similar effects on the cellular transcription profile. Conversely, the widespread expression of HMGN1 and HMGN2, but not HMGN3 and HMGN5, in most tissues and the sequence specificity of their C-terminal domains suggest that potentially the proteins may have variant-specific effects on the transcription profile. Indeed, *in vitro* experiments revealed variant-specific effects on histone modifications, and experiments with genetically altered mice also suggest that the HMGN variants are not fully functionally redundant.

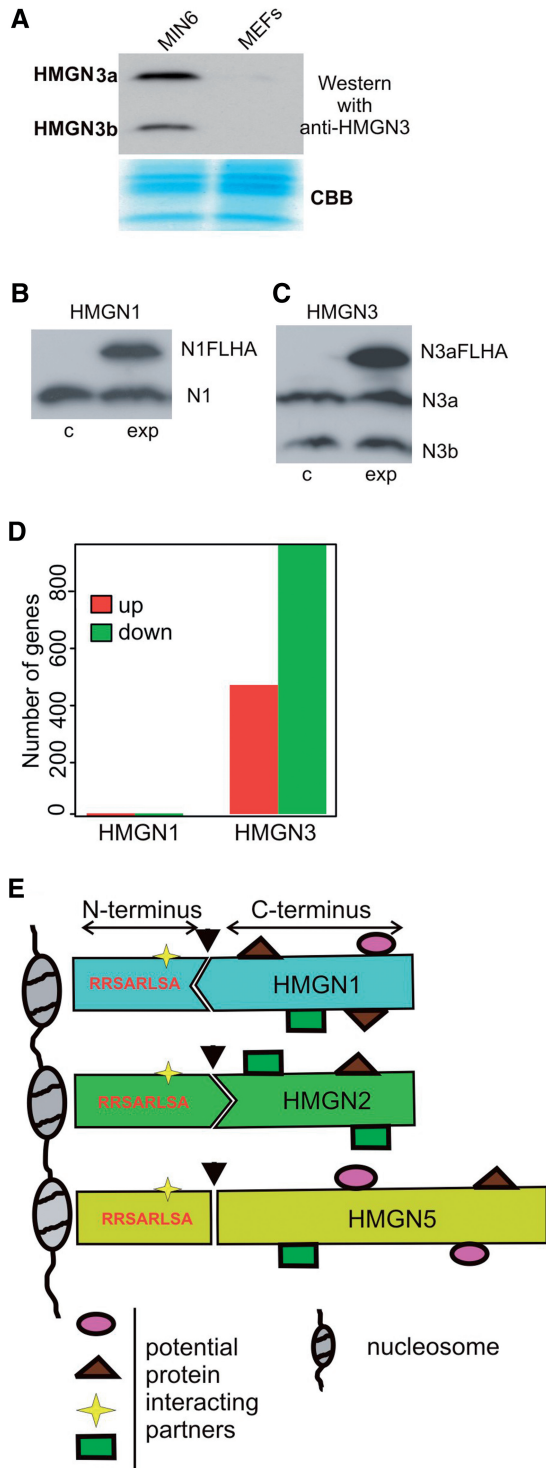


Figure 5. Effect of HMGN1 and HMGN3a on transcription in MIN6 cells. (A) Comparison of the protein levels of HMGN3 and HMGN1 in MIN6 and MEFs by western blot. CBB, Coomassie Blue staining. Western blot analysis of MIN6 cell lines stably expressing HMGN1 (B) or HMGN3a (C) proteins. Endogenous and exogenous FLAG and HA tagged (FLHA) proteins are indicated. c, control infection with empty virus; exp, experimental infection with indicated protein. The graph (D) represents the number of genes changed following overexpression of HMGN1 and HMGN3a proteins compared with the control empty vector expression. (E) Model for structural specificity of individual HMGN proteins. HMGN proteins consist of a conserved N-terminal region, which contains the NBD and the conserved

Our experiments suggest that each HMGN variant can affect the expression of numerous genes, especially when overexpressed, and by enlarge in an HMGN-specific manner. The amplitude of transcriptional changes was moderate; for most of the affected genes being in the limits of 2-fold difference. Importantly, nearly equal amount of genes were either up- or downregulated by each HMGN, suggesting that HMGNs are neither transcriptional activators nor repressors. The GO analyses indicated that multiple cellular processes were affected by individual HMGNs or by combinations of several HMGNs, suggesting that the HMGNs are general modulators of the cellular transcriptional fidelity.

Two molecular mechanisms whereby HMGN affect the transcription profile could be envisioned. One possibility is that by binding to nucleosomes, HMGN induce structural changes that alter the ability of transcriptional regulators, either positive or negative, to interact with their chromatin targets. A second possibility is that the HMGN interact with specific regulators and affect their chromatin interactions. Both possibilities suggest that the ability of HNGN variants to bind to chromatin is a major effect on the transcriptional output. Indeed, our previous experiments (50), and our present analyses of the HMGN5-S19,23E mutant, indicate that HMGNs affect transcription by binding to nucleosomes.

While nucleosome binding seems to be an absolute requirement for any noticeable effects on transcription, the variant-specific effects on the transcription profile suggest that additional properties of these proteins play a role in determining their biological specificity. Because the C-terminal domain of HMGNs is highly variable in sequence between individual HMGN proteins, we tested the possibility that the specific transcriptional effects of HMGNs reside in this domain and expressed several HMGN swap mutants in MEF cells. Surprisingly, the transcriptional outcome following the expression of these swap mutants with a common NBD from HMGN1, and a C-terminal domain from HMGN1, HMGN2 or HMGN3 was different from either one of their 'source' proteins. Thus, the variant-specific effects of HMGNs on transcription are the consequence of coordinate effects of the various structural domains of each variant. In other words, neither the NBD nor the C-terminal domain alone defines the transcriptional effect of each HMGN protein, but rather the entire structure of the protein defines its specific role in transcription (Figure 5E).

In considering the molecular mechanisms leading to HMGN-variant-specific effects on transcription, we

octapeptide, RRSARLSA and a C-terminal region with a more variable sequence. The N- and C-terminal regions of each HMGN variant fit to give the specific property of each variant. Arrow marks the hypothetical connection between N- and C-regions; the geometry indicates unique combination of regions in each HMGN protein. Both the N- and the C-terminal regions interact with various protein partners. Some partners are shared between all HMGNs, whereas others are specific to individual proteins. Combinations of different interacting proteins will define the properties of each HMGN protein and its ability to affect chromatin architecture and transcription.

note that early structural studies indicated that HMGNs have little ordered structure (52), and our computational analysis (Figure 1) reveal that HMGNs are among the most intrinsically disordered proteins known. Intrinsically disordered proteins can interact with multiple protein partners with relatively low affinity and acquire more ordered structures (45–47,53–58). It has been recently reported that the harmful effect of elevated cellular levels of many proteins is correlated with the degree of their disorderness (25). At the same time, cells with decreased amount of these proteins function robustly and do not demonstrate significant changes in cellular functions. Our observation that knock out of HMGNs has significantly smaller effect on transcription supports this theory and strongly argues that disordered structure of HMGNs is one of the major functional properties of these proteins.

Variations in structure of HMGN proteins due to interaction with different protein partners can modulate the effects of HMGN variants on local nucleosome structure, global chromatin architecture and transcription (Figure 5E). Indeed, both HMGN1 and HMGN2 have been shown to form multiple metastable macromolecular complexes (15), and specific protein partners have been identified for several HMGN variants. Thus, HMGN3 interacts specifically with the thyroid hormone receptor (59) and with the transcription factor PDX1 (19), HMGN1 forms a complex with ERalpha and SRF (15), and HMGN2 was shown to interact with PITX2 (60). Our observation of cell-specific effects of HMGN3a protein on transcription in the pancreatic derived MIN6 cell line, but not in MEFs (19), supports the idea of existence of specific protein partners for individual HMGN proteins.

In conclusion, our results reveal both specific and redundant roles of HMGN variants in the global regulation of gene expression. Each HMGN preferentially affects a unique set of genes with little or no specificity for defined cellular processes. Thus, changes in the expression of an HMGN may disrupt the fidelity of the cellular transcription and render the organism more susceptible to further damage. Indeed, experiments with genetically altered mice and with cells derived from these mice indicate that loss of HMGN1 leads to an impaired DNA damage repair response and increased tumorigenicity (27,49,61). Likewise, loss of HMGN3, which is highly expressed in beta cells of the pancreatic islets, affects insulin secretion leading to a mild diabetic phenotype (19). The transcriptional specificity of the HMGN variants is similar to that of the H1 variants. It seems that the dynamic interaction of HMGN, H1 and other structural proteins with chromatin is part of the mechanism that ultimately fine tunes the transcription profile to optimize cellular function.

ACKNOWLEDGEMENT

We thank Dr Nina Bubunenko, LMT, Frederick for help with the Affymetrix arrays.

FUNDING

Intramural program of the CCR, NCI, NIH and Research Program of the National Library of Medicine, NIH. Funding for open access charge: Intramural, CCR, NCI, NIH.

Conflict of interest statement. None declared.

REFERENCES

- Bustin,M., Catez,F. and Lim,J.H. (2005) The dynamics of histone H1 function in chromatin. *Mol. Cell*, **17**, 617–620.
- Catez,F. and Hock,R. (2010) Binding and interplay of HMG proteins on chromatin: lessons from live cell imaging. *Biochim. Biophys. Acta-Gene Regul. Mech.*, **1799**, 15–27.
- Bianchi,M.E. and Agresti,A. (2005) HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.*, **15**, 496–506.
- Bustin,M. (1999) Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell. Biol.*, **19**, 5237–5246.
- Woodcock,C.L., Skoultchi,A.I. and Fan,Y. (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.*, **14**, 17–25.
- Khochbin,S. (2001) Histone H1 diversity: bridging regulatory signals to linker histone function. *Gene*, **271**, 1–12.
- Fan,Y. and Skoultchi,A.I. (2004) Genetic analysis of H1 linker histone subtypes and their functions in mice. *Methods Enzymol.*, **377**, 85–107.
- Fan,Y., Sirotkin,A., Russell,R.G., Ayala,J. and Skoultchi,A.I. (2001) Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. *Mol. Cell. Biol.*, **21**, 7933–7943.
- Alami,R., Fan,Y., Pack,S., Sonbuchner,T.M., Besse,A., Lin,Q., Grealley,J.M., Skoultchi,A.I. and Bouhassira,E.E. (2003) Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc. Natl Acad. Sci. USA*, **100**, 5920–5925.
- Bhan,S., May,W., Warren,S.L. and Sittman,D.B. (2008) Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation. *Gene*, **414**, 10–18.
- Sancho,M., Diani,E., Beato,M. and Jordan,A. (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet.*, **4**, e1000227.
- Hock,R., Furusawa,T., Ueda,T. and Bustin,M. (2007) HMG chromosomal proteins in development and disease. *Trends Cell Biol.*, **17**, 72–79.
- Postnikov,Y. and Bustin,M. Regulation of chromatin structure and function by HMGN proteins. *Biochim. Biophys. Acta*, **1799**, 62–68.
- Catez,F., Brown,D.T., Misteli,T. and Bustin,M. (2002) Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep.*, **3**, 760–766.
- Lim,J.H., Bustin,M., Ogryzko,V.V. and Postnikov,Y.V. (2002) Metastable macromolecular complexes containing high mobility group nucleosome-binding chromosomal proteins in HeLa nuclei. *J. Biol. Chem.*, **277**, 20774–20782.
- Postnikov,Y.V., Trieschmann,L., Rickers,A. and Bustin,M. (1995) Homodimers of chromosomal proteins HMG-14 and HMG-17 in nucleosome cores. *J. Mol. Biol.*, **252**, 423–432.
- Postnikov,Y.V., Herrera,J.E., Hock,R., Scheer,U. and Bustin,M. (1997) Clusters of nucleosomes containing chromosomal protein HMG-17 in chromatin. *J. Mol. Biol.*, **274**, 454–465.
- Shirakawa,H., Landsman,D., Postnikov,Y.V. and Bustin,M. (2000) NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression. *J. Biol. Chem.*, **275**, 6368–6374.
- Ueda,T., Furusawa,T., Kurahashi,T., Tessarollo,L. and Bustin,M. (2009) The nucleosome binding protein HMGN3 modulates the transcription profile of pancreatic beta cells and affects insulin secretion. *Mol. Cell. Biol.*, **29**, 5264–5276.

20. West,K.L., Ito,Y., Birger,Y., Postnikov,Y., Shirakawa,H. and Bustin,M. (2001) HMG3a and HMG3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins. *J. Biol. Chem.*, **276**, 25959–25969.
21. Ding,H.F., Rimsky,S., Batson,S.C., Bustin,M. and Hansen,U. (1994) Stimulation of RNA polymerase II elongation by chromosomal protein HMG-14. *Science*, **265**, 796–799.
22. Paranjape,S.M., Krumm,A. and Kadonaga,J.T. (1995) HMG17 is a chromatin-specific transcriptional coactivator that increases the efficiency of transcription initiation. *Genes Dev.*, **9**, 1978–1991.
23. Bustin,M. (2001) Chromatin unfolding and activation by HMGN(*) chromosomal proteins. *Trends Biochem. Sci.*, **26**, 431–437.
24. Rochman,M., Malicet,C. and Bustin,M. HMG5/NSBP1: a new member of the HMGN protein family that affects chromatin structure and function. *Biochim. Biophys. Acta*, **1799**, 86–92.
25. Vavouri,T., Semple,J.I., Garcia-Verdugo,R. and Lehner,B. (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*, **138**, 198–208.
26. Hansen,J.C., Lu,X., Ross,E.D. and Woody,R.W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.*, **281**, 1853–1856.
27. Birger,Y., West,K.L., Postnikov,Y.V., Lim,J.H., Furusawa,T., Wagner,J.P., Laufer,C.S., Kraemer,K.H. and Bustin,M. (2003) Chromosomal protein HMG1 enhances the rate of DNA repair in chromatin. *EMBO J.*, **22**, 1665–1675.
28. Gentleman,R., Carey,V., Bates,D., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
29. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
30. Wilson,C. and Miller,C. (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, **21**, 3683–3685.
31. Parman,C., Halling,C. and Gentleman,R. affyQCReport: QC Report Generation for affyBatch objects. R package version 1.18.0.
32. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
33. Smyth,G.K. (2005) Limma: linear models for microarray data. In: Gentleman,R., Carey,V., Dudoit,S., Irizarry,R. and Huber,W. (eds), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer, New York, pp. 397–420.
34. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B (Methodological)*, **57**, 289–300.
35. Swinton,J. (2009) Vennerable: Venn and Euler area-proportional diagrams. R package version 2.0/r75.
36. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
37. Vacic,V., Uversky,V.N., Dunker,A.K. and Lonardi,S. (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*, **8**, 211.
38. Sickmeier,M., Hamilton,J.A., LeGall,T., Vacic,V., Cortese,M.S., Tantos,A., Szabo,B., Tompa,P., Chen,J., Uversky,V.N. *et al.* (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.*, **35**, D786–D793.
39. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
40. Radivojac,P., Iakoucheva,L.M., Oldfield,C.J., Obradovic,Z., Uversky,V.N. and Dunker,A.K. (2007) Intrinsic disorder and functional proteomics. *Biophys. J.*, **92**, 1439–1456.
41. Romero,P., Obradovic,Z., Li,X., Garner,E.C., Brown,C.J. and Dunker,A.K. (2001) Sequence complexity of disordered protein. *Proteins*, **42**, 38–48.
42. Li,X., Romero,P., Rani,M., Dunker,A.K. and Obradovic,Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.*, **10**, 30–40.
43. Ueda,T., Catez,F., Gerlitz,G. and Bustin,M. (2008) Delineation of the protein module that anchors HMGN proteins to nucleosomes in the chromatin of living cells. *Mol. Cell. Biol.*, **28**, 2872–2883.
44. Romero,P., Obradovic,Z. and Dunker,A.K. (2004) Natively disordered proteins: functions and predictions. *Appl Bioinformatics*, **3**, 105–113.
45. Garner,E., Romero,P., Dunker,A.K., Brown,C. and Obradovic,Z. (1999) Predicting binding regions within disordered proteins. *Genome Inform. Ser. Workshop Genome Inform.*, **10**, 41–50.
46. Oldfield,C.J., Cheng,Y., Cortese,M.S., Romero,P., Uversky,V.N. and Dunker,A.K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, **44**, 12454–12470.
47. Cheng,Y., Oldfield,C.J., Meng,J., Romero,P., Uversky,V.N. and Dunker,A.K. (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry*, **46**, 13468–13477.
48. Dosztanyi,Z., Meszaros,B. and Simon,I. (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, **25**, 2745–2746.
49. Birger,Y., Catez,F., Furusawa,T., Lim,J.H., Prymakowska-Bosak,M., West,K.L., Postnikov,Y.V., Haines,D.C. and Bustin,M. (2005) Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMG1. *Cancer Res.*, **65**, 6711–6718.
50. Rochman,M., Postnikov,Y., Correll,S., Malicet,C., Wincovitch,S., Karpova,T.S., McNally,J.G., Wu,X., Bubunencko,N.A., Grigoryev,S. *et al.* (2009) The interaction of NSBP1/HMG5 with nucleosomes in euchromatin counteracts linker histone-mediated chromatin compaction and modulates transcription. *Mol. Cell*, **35**, 642–656.
51. Bustin,M., Crippa,M.P. and Pash,J.M. (1990) Immunochemical analysis of the exposure of high mobility group protein 14 and 17 surfaces in chromatin. *J. Biol. Chem.*, **265**, 20077–20080.
52. Bradbury,E.M. (1983) Conformations and flexibilities of histones and high mobility group (HMG) proteins in chromatin structure and function. *Ciba Found. Symp.*, **93**, 246–270.
53. Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **27**, 527–533.
54. Dunker,A.K., Silman,L., Uversky,V.N. and Sussman,J.L. (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.*, **18**, 756–764.
55. Oldfield,C.J., Meng,J., Yang,J.Y., Yang,M.Q., Uversky,V.N. and Dunker,A.K. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*, **9(Suppl. 1)**, S1.
56. Tompa,P. and Fuxreiter,M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.*, **33**, 2–8.
57. Uversky,V.N. and Dunker,A.K. Understanding protein non-folding. *Biochim. Biophys. Acta*, **1804**, 1231–1264.
58. Lu,X., Hamkalo,B., Parseghian,M.H. and Hansen,J.C. (2009) Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder. *Biochemistry*, **48**, 164–172.
59. Lee,J.W., Choi,H.S., Gyuris,J., Brent,R. and Moore,D.D. (1995) Two classes of proteins dependent on either the presence or absence of thyroid hormone for interaction with the thyroid hormone receptor. *Mol. Endocrinol.*, **9**, 243–254.
60. Amen,M., Espinoza,H.M., Cox,C., Liang,X., Wang,J., Link,T.M., Brennan,R.G., Martin,J.F. and Amendt,B.A. (2008) Chromatin-associated HMG-17 is a major regulator of homeodomain transcription factor activity modulated by Wnt/beta-catenin signaling. *Nucleic Acids Res.*, **36**, 462–476.
61. Kim,Y.C., Gerlitz,G., Furusawa,T., Catez,F., Nussenzweig,A., Oh,K.S., Kraemer,K.H., Shiloh,Y. and Bustin,M. (2009) Activation of ATM depends on chromatin interactions occurring before induction of DNA damage. *Nat. Cell Biol.*, **11**, 92–96.