Dna Research, 2017, 0(0), 1–15
doi: 10.1093/dnares/dsw057
Full Paper

OXFORD

Full Paper

An integrated transcriptomics-guided genome-wide promoter analysis and next-generation proteomics approach to mine factor(s) regulating cellular differentiation

Kamal Mandal[1], Samuel L. Bader[2], Pankaj Kumar[3], Dipankar Malakar[4], David S. Campbell[2], Bhola Shankar Pradhan[1], Rajesh K. Sarkar[1], Neerja Wadhwa[1], Souvik Sensharma[1], Vaibhav Jain[5], Robert L. Moritz[2], and Subeer S. Majumdar[1,6]*

[1]Cellular Endocrinology Laboratory, National Institute of Immunology, New Delhi, India, [2]Institute for Systems Biology, Seattle, WA 98109, USA, [3]G.N.R. Knowledge Centre for Genome Informatics, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India, [4]Sciex, Gurgaon, Haryana, India, [5]Next-Generation Sequencing Facility, National Institute of Immunology, New Delhi, India, and [6]National Institute of Animal Biotechnology, Miyapur, Hyderabad, India,

*To whom correspondence should be addressed. Tel. 26717121; 26717145. Fax: 91-11-26742125/91-11-26742626. Email: subeer@nii.ac.in

Edited by Dr. Toshihiko Shiroishi

Received 11 July 2016; Editorial decision 22 November 2016; Accepted 24 November 2016

Abstract

Differential next-generation-omics approaches aid in the visualization of biological processes and pave the way for divulging important events and/or interactions leading to a functional output at cellular or systems level. To this end, we undertook an integrated Nextgen transcriptomics and proteomics approach to divulge differential gene expression of infant and pubertal rat Sertoli cells (Sc).Unlike, pubertal Sc, infant Sc are immature and fail to support spermatogenesis. We found exclusive association of 14 and 19 transcription factor binding sites to infantile and pubertal states of Sc, respectively, using differential transcriptomics-guided genome-wide computational analysis of relevant promoters employing 220 Positional Weight Matrices from the TRANSFAC database. Proteomic SWATH-MS analysis provided extensive quantification of nuclear and cytoplasmic protein fractions revealing 1,670 proteins differentially located between the nucleus and cytoplasm of infant Sc and 890 proteins differentially located within those of pubertal Sc. Based on our multi-omics approach, the transcription factor YY1 was identified as one of the lead candidates regulating differentiation of Sc.YY1 was found to have abundant binding sites on promoters of genes upregulated during puberty. To determine its significance, we generated transgenic rats with Sc specific knockdown of YY1 that led to compromised spermatogenesis.

Key words: cellular differentiation, multi-omics, sertoli cell, SWATH-MS, TRANSFAC

1. Introduction

Tissue differentiation is an obligatory phenomenon in higher organisms. There exists a cohort of cellular and molecular mechanisms to fulfil this requisite of biological system. The proteome complement of a cell type defines its differentiation/transformation status (e.g. normal vs malignant status) and functionality. However, the proteome profile is primarily determined by the transcriptomic complement of a cell that is governed by the nucleus.[1,2] The coordinated modulation and dynamical expression of a complement of transcription factors (TF) direct lineage-specific developments of each cell type.[3] Such spatio-temporal regulation of differential gene expression occurs due to interaction of *trans* acting factors (along with their interacting partners) with genomic components, mainly promoters and enhancers.[4] These are guided by the nuclear proteome of each cell dictated by the extracellular and intracellular signals and cumulatively govern transcriptional activity of chromatin carried out in the nucleus.

We chose testicular Sertoli cells (Sc) to visualize dynamical gene/proteome regulation upon maturation because this cell undergoes maturation after birth. Infant Sc from 5-day-old rats are immature and incapable to support germ cell (Gc) development, whereas by 12 days of age, Sc are well differentiated and regulate Gc development successfully leading to robust onset of the process of spermatogenesis.[5] Although levels of circulating FSH and testosterone (T), which regulate spermatogenesis via Sc, are similar in infant (5 days) and pubertal (12 days) rats,[6–8] robust Gc differentiation is noticed only in 12-day-old testis. This suggests that there are endogenous cellular switches or mechanisms leading to maturation of Sc during puberty, which are missing during infancy. Use of modern analysis tools to explore the transcriptomic and proteomic complement of such maturation of cells, associated with its phenotype, may help us in resolving unexplained mechanisms identified in such cellular transitions.

Recent technological advancements provide new powerful tools to compare comprehensive transcriptome or proteome profiles obtained from multiple biological samples. Significantly large amounts of information, including new regulatory pathways and hub molecules, have been identified using high-throughput omics-based data.[9,10]

We report the key TF binding sites (TFBS) and associated nuclear proteins present in infant and pubertal Sc of rats. For this, genome-wide computational analysis of the promoters was performed. Promoters of differentially expressed transcripts were determined using data from microarray analysis of infant and pubertal Sc. TFBS prediction of the promoters was done using TRANSFAC database. Since activity of these TFBS is governed by nuclear proteome of the cell type, quantitative nuclear proteome profiling was performed. Since activity and functions of a protein can depend exclusively on post-translational modifications (PTMs), discovery proteomics approach was performed to identify PTMs in the cytoplasmic and nuclear proteome of these two age groups of Sc. Comprehensive quantitative analyses of proteome in nuclear and cytoplasmic fractions of Sc were performed using data-independent acquisition (DIA) proteomics (SWATH-MS) analysis.[11,12]

We established the role of YY1, a lead candidate obtained in this study, in Sc differentiation using a transgenic knockdown model of rat. To our knowledge, this is the first report of cataloguing the potential TFBS that are active and required for onset of differentiation, post birth. Such biological validation of omics-based output generated substantial support for undertaking such multi-omics approach-based studies. This is the first report of comprehensive proteome quantification (using SWATH-MS) in mammalian system at organelle level.

2. Materials and methods

2.1. Promoter sequence retrieval and TF binding analysis

The ±2-kb promoter sequences centred at annotated transcription start site (TSS) for all differentially expressed (up and down) and control-set gene (un-expressed genes of same cardinality) promoters were retrieved from UCSC-Galaxy (https://usegalaxy.org/). All these promoter sequences [differentially expressed (up and down] and control-set gene promoter sequences) were analysed for the presence of different TFBSs using MATCH™ program available with TRANSFAC® professional 12.1. The total occurrence of individual TFBS on each differentially expressed gene-set promoter was considered as observed frequency. Similarly, the total occurrence of an individual TFBS from control-set promoters divided by total promoters in control-set gave the expected frequency of individual TFBS (Fig. 1A). The discrepancy between the observed frequency and expected frequency for individual TFBS on differentially expressed (up and down) gene-set promoters was evaluated by determining the statistically variable chi-square (χ^2). All significantly enriched TFBS ($P < 0.05$) from both up-regulated and down-regulated gene-set promoters were further interrogated for positional binding distribution analysis (with respect to TSS).

2.2. TFBS interactome analysis

Interactome analysis was performed using STRING database in order to determine the proteins associated with a set of TFBS. A set of TFBS was queried against the STRING database to extract all their interacting genes and proteins. Only high confidence (score > 0.7) hits were allowed to appear with no >50 interacting partners.

2.3. Sertoli cell culture

Rattus norvegicus (Wistar rats) for this study was procured from the small animal facility of National Institute of Immunology, New Delhi. Animals were maintained and bred in accordance to the guidelines provided by CPCSEA, Government of India. All the experiments involving animals in this study were approved by Institutional Animal Ethics Committee.

Infant and Pubertal rat Sc were isolated and cultured as described previously by us.[5] Briefly, testes were surgically removed from the animal and were decapsulated. Hanks Balanced Salt Solution was used for all the tissue processing of the testis. The decapsulated testis was then chopped and enzymatically digested with collagenase (Sigma-Aldrich) and centrifuged down to eliminate Leydig cells and other interstitial cells. Pancreatin (Sigma-Aldrich) was used for digestion to disrupt the seminiferous tubules and separate the SC clusters (*groups of SC with attached germ cells*) from the peritubular cells and germ cells upon centrifugation. These Sc clusters were then plated with DMEM-F12 medium with 1% serum foetal bovine serum (FBS). The serum media was then replaced by 1% growth factor DMEM-F12 media. Germ cells were removed by hypotonic shock 48 hours after plating. Pulsatile hormone treatment of FSH and T was given 72 hours post plating as described previously.[5] The cells were harvested for further processing to obtain protein.

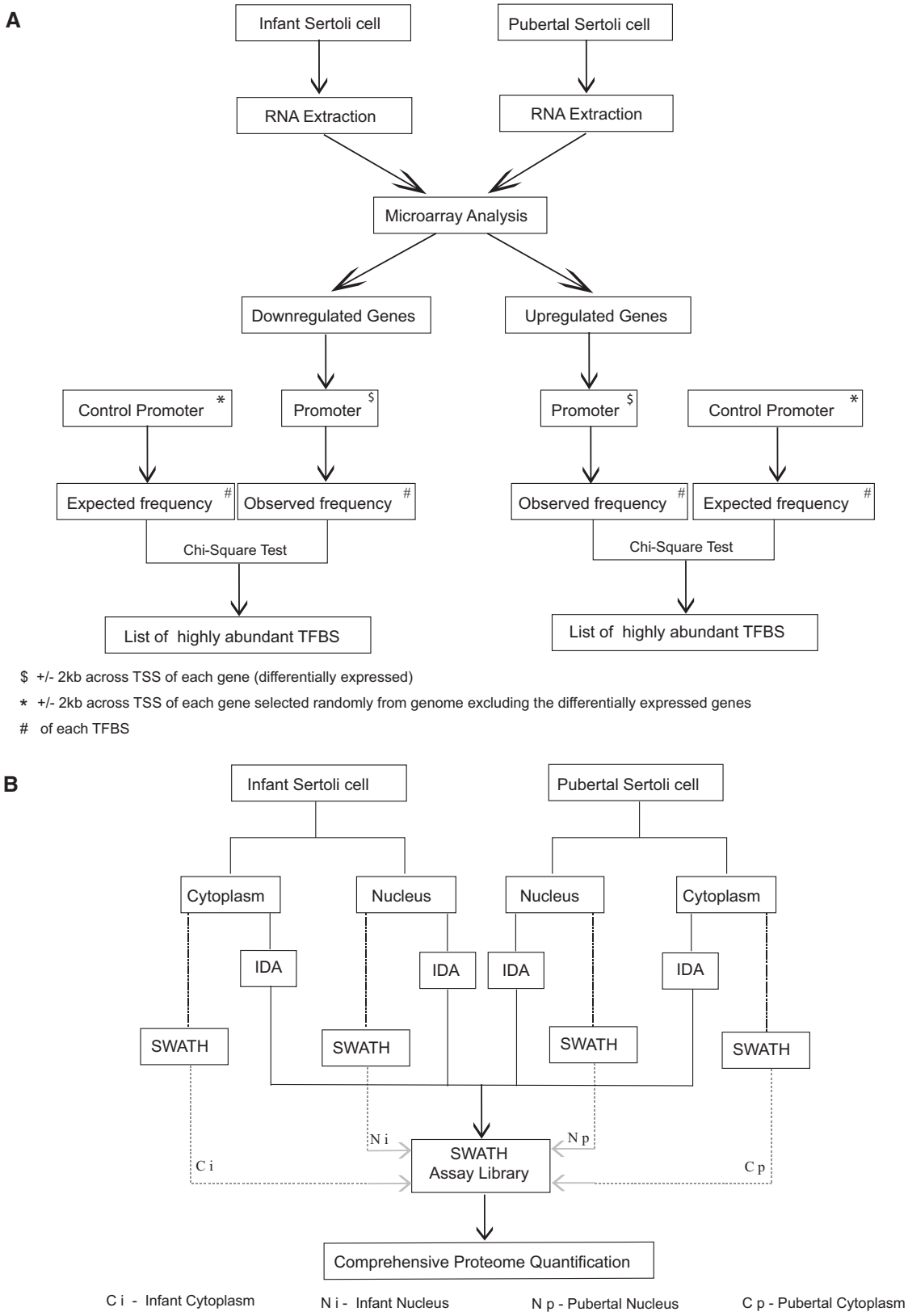

Figure 1. Workflow used for (A) TFBS analysis using TRANSFAC database. (B) Comprehensive proteome quantification using SWATH analysis.

2.4. Sample preparation for mass spectrometry

Nuclear and cytoplasmic protein fractions were isolated from cultured Sc using NE-PER™ (Thermo Scientific, USA) Nuclear and Cytoplasmic Extraction Reagents as per the manufacturer's instructions. Protein fractions were diluted in 10 mM ammonium bicarbonate. The diluted samples were then reduced with 25 mM dithiothreitol (Sigma-Aldrich) at 60°C for 30 min followed by alkylation with 55 mM iodoacetamide (Sigma-Aldrich) at room temperature for 20 min. The samples were then incubated with porcine sequencing grade modified trypsin (Promega V111) for overnight at 37°C in 1:10 ratio (trypsin to protein). The trypsin-digested samples were dried up using speed vac. The peptides obtained were reconstituted in 0.1% formic acid solution for mass spectrometry analysis.

2.5. Nano LC–MS/MS analysis

Chromatographic separation was performed on a NanoLC system (NanoLC-2D Ultra, Eksigent). Peptides were trapped on a NanoLC pre-column (ChromXP C18-CL-3 μm, I.D. 0.35 × 0.5 mm Eksigent) and then eluted onto an analytical column (C18-CL-120, I.D. 0.075 × 150 mm, Eksigent). A 1,000-ng sample was loaded, trapped and desalted at 3 μl/min for 37 min with 100% mobile phase A [2% acetonitrile (Sigma-Aldrich) in 0.1% formic acid (Sigma-Aldrich)]. Peptides were separated at a flowrate of 250 nl/min using a stepwise gradient of buffer B [98% acetonitrile (Sigma-Aldrich) in 0.1% formic acid (Sigma-Aldrich)] from 5% to 10% B in the first 10 min and from 10% to 40% in the following 60 min, from 40% to 50% in the next 10 min. The column was washed after the gradient with 90% B for 10 min before re-equilibrating to the initial chromatographic conditions for 10 min.

2.6. Mass spectrometric analysis

Data acquisition was performed with a TripleTOF 5600 System (SCIEX, Concord, ON, CA) fitted with a Nanospray III source (SCIEX, Concord, ON, CA) and a pulled quartz tip as the emitter (New Objectives, Woburn, MA). The electrospray source was operated with ion spray voltage of 2.2 kV, a curtain gas of 25 PSI, a nebulizer gas of 20 PSI, and an interface heater temperature of 125 °C. The MS was operated with a resolution greater than or equal to 30,000 full width at half maximum for TOF MS scans with mass range 350–1250 *m/z*. For shotgun proteomics, information-dependent acquisition (IDA) experiment, survey scans were acquired in 250 Mrs. Product ion scans with mass range 200–1800 *m/z* were collected in high-sensitivity mode for the 20 most intense precursor with an intensity higher than 150 counts per second and with a charge-state between 2+ and 5+. The cycle time was kept constant at 1.7 s. Four time bins were summed for each scan at a pulser frequency value of 11 kHz through monitoring of the 40 GHz multichannel TDC detector with four-anode/channel detection. Rolling collision energy (CE) with the default equations and a spread of 5 V was used for collision-induced dissociation. Selected precursors were excluded for 5 seconds after one occurrence.

2.7. SWATH-MS analysis

For data-independent acquisition, a SWATH-MS acquisition method was generated using the 'create SWATH' mode. The mass range from 350 to 1250 Da was covered with 36 selection windows with an isolation width of 26 Da (25 Da of optimal ion transmission efficiency 1 Da for the window overlap).[11] The CE for each window was determined using the same equation as for the IDA experiments based on the 2+ charge precursor with a spread of 5 eV. The total duty cycle was of 3.0 seconds (total 2.95 seconds for stepping through the 36 isolation windows − 0.05 seconds for the optional survey scan). The MS/MS acquisition was performed in high-sensitivity mode corresponding to the mass resolution of about 15,000, which also enables to extract fragment ions with 10–50 ppm accuracy.

2.8. SWATH-assay library generation

The Trans-Proteomic Pipeline was used for data analysis.[13,14] All the .wiff files obtained in IDA experiments were converted to .mzML format using msconvert from ProteoWizard version 3.0.4624 selecting vendor-specific peak picking algorithm.[15] All the .mzML files were then independently searched against rat protein sequence file using X!Tandem[16] and Comet[17] search. The rat protein sequence contained all the sequence available in UniProt database (February 2015) supplemented with the common laboratory contaminant proteins and decoy sequences generated by randomizing the rat protein sequences using TPP v4.8.0 PHILAE, Build 201412031424-6764 (Linux). Carbamidomethylation of cysteine (+57.021464) was set as a fixed modification. Oxidation of methionine and acetylation of lysine (+14.0165, +28.0313) was considered as variable modifications. A maximum of two missed cleavage was allowed. Precursor and fragment ion mass tolerance were set to 75 and 30 ppm, respectively. The .pep.xml files obtained from each search were statistically evaluated using PeptideProphet followed by combining the results using iProphet.

A consensus spectral library was built using SpectraST based on the peptide spectrum matches with a $P > 0.65$, which corresponds to a decoy estimated error rate of 0.41% on peptide level and SWATH-MS assays were generated from this spectral library essentially as reported previously.[18] In short, we selected precursor ions in the mass range of 400–1,250 Da, with at least 4 and at the most 100 fragment ions with a charge state between 1+ and 4+ that fall in the mass range of 350–2,000, but excluding the precursor mass window.

To normalize the retention time of the peptide, the retention time of the peptides of all the runs were adjusted to the run with maximum number of identifications by linear regression. The normalized retention time of each peptide was then determined, by taking the median of all the retention time determined for the individual runs. Finally, the rat SWATH-assay library was updated with the normalized retention times.

2.9. SWATH-MS quantitation

The SWATH-MS files were interrogated with SWATH-assay library (Fig. 1B) using MS/MS^{ALL} with SWATH-MS Acquisition MicroApp version 2.0 and PeakView version 2.2 software (SCIEX, Concord, ON, CA). Maximum six peptides were allowed for quantification of a single protein with a maximum of six transitions per peptide. The transition was extracted in a retention time window of 5 min (±2.5 min) and with a mass tolerance of 75 ppm. Peptides with a minimum confidence level of 95% were used. False discovery rate (FDR) threshold was kept below 5%. The filtered proteins and associated peptides were exported to MarkerView version 1.2.1 software (Sciex, Concord, ON, CA) for relative quantitative analysis. The median peak ratio method was used for the normalization of the quantitative protein data for all samples of different groups. Statistical analysis of the relative quantitation was performed by *t*-test analysis. The protein levels with a $P < 0.05$ were considered significant. For

all SWATH-MS quantitation in this study, three biological replicates were used. For each biological replicate, there were three technical replicates.

2.10. Gene ontology analysis

Gene ontology (GO) analysis was performed using GORILLA (Gene Ontology Enrichment Analysis and Visualization Tool).[19] The proteins found to be in high abundance in nuclear or cytoplasm (fold change >2, $P < 0.05$) using SWATH-MS analysis were considered for performing GO analysis.

2.11. PTM analysis

PTM search was performed using ProteinPilot software 5.0 (Sciex, USA) employing Paragon Algorithm.[20] All the three IDA files obtained from the different biological replicates of each sample were processed together. UniProt database in FASTA format was used as sequence library of *Rattus norvegicus*. All biological and *in vivo* modifications available with ProteinPilot were checked with global FDR <1%.

2.12. Plasmid constructs design

The hairpin (TCAAGAG) and seed sequence of Yy1 (CGACGGTTGTAATAAGAAGTTTGCTCAGT)- and Luciferase (GGATTTCAGTCGATGTACACGTTCGTCAC)-specific shRNA were obtained from OriGene (USA). The forward and reverse DNA strand of shRNA was procured from Sigma-Aldrich (USA). The sequence of both the strands was designed such that upon annealing it generates a staggered ends compatible for ligation into *Eco*RI (5′) and *Sal*I (3′) digested vector. shRNA strands were annealed in NEB buffer-2 using thermal cycler with initial incubation at 95 °C for 5 min. This mixture was then incubated at every temperature at the interval of 2 °C for 1 second till it reached 85 °C. The temperature of the mixture was thereafter decreased by interval of 0.1 °C every cycle with incubation time of 1 second at every temperature till it reaches 25 °C followed by quick chill down to 4 °C. The shRNA annealing was confirmed in a 4% agarose gel. The annealed oligo was then ligated in the vector digested with *Eco*RI and *Sal*I. Positive colonies were screened by digesting the plasmid with *Age*I (the restriction enzyme site for *Age*I was added after the shRNA seed sequence and is absent in the vector backbone).

2.13. Generation of transgenic rat

Transgenic rats with Sc-specific knockdown were generated using the method of testicular electroporation as described previously by us.[21] Briefly, the maxi preparation of the YY1 knockdown plasmid (GenElute^{TM} HP Plasmid Maxiprep Kit; Sigma Aldrich) was linearized using the restriction enzyme *Stu*1 followed by ethanol precipitation. The linearized DNA was then injected into the testes of 38-day-old Wistar rats to generate a fore-founder animal. This fore-founder was then co-habitated with wild-type female rat 60 days after electroporation. The pups born were screened for the presence of transgene using slot blot analysis.

2.14. Genomic DNA isolation and slot blot analysis

Genomic DNA was isolated from the tail snip of the pups using phenol:chloroform isolation method.[22] The isolated genomic DNA was quantitated using Nanodrop 2000c spectrophotometer (Thermo Scientific, USA). For slot blot analysis, 2 µg of g-DNA was blotted onto positively charged nylon-66 membrane (M\DI Membrane Technologies, Ambala Cantt, India) and probed with α-P^{32}-labelled DNA probe specific to the YY1 knockdown construct. The membrane was then analysed using phosphorimager Typhoon 9400 (GE Healthcare) to detect the transgene positive genomic DNA samples (Supplementary Figure S2).

2.15. RNA extraction and real-time PCR analysis

Whole testicular RNA was extracted using TRI reagent (Sigma Aldrich) as per the manufacturer's instructions. All the real-time PCR analysis was done as previously described by us.[23] Extracted RNA was quantified using NANO drop spectrophotometer (Thermo Scientific, USA). One microgram of RNA was subjected to DNase treatment followed by single-strand c-DNA synthesis using M-MLV reverse transcriptase (Promega, USA) as per the manufacturer's protocol. The c-DNA was then used for real-time PCR analysis using Mesa- green master mix kit (Eurogentech, Belgium) in StepOnePlus Real Time PCR Systems (Applied Biosystems, USA). Gene expression was normalized using cyclophilin as a reference and expression value calculated as $2^{-\Delta\Delta ct}$. Real-time PCR primers used in this study are listed in Supplementary Table S12).

2.16. RNA sequencing analysis

The quality of the RNA was assessed using Bioanalyzer followed by library preparation. These samples were then sequenced on Illumina HiSeq 2500 platform with 2×100 Paired End cycles. RNA Seq reads were aligned/mapped to the rat reference genome (UCSC rn4) using TopHat2 (version 2.0.9).[24] Expression quantification in terms of Fragments Per Kilobase of transcript per Million mapped reads for each of the transcripts were determined using Cufflinks (version 2.2.1)[25] and differentially expressed genes at $P < 0.05$ were identified using Cuffdiff (version 2.2.1).[26] RNA sequencing experiments were performed in duplicates (biological replicates).

2.17. Protein extraction and immunoblot analysis

Total testicular protein was extracted using RIPA lysis buffer (G-Biosciences, St. Louis, USA) with 1X protease inhibitor cocktail (AMRESCO, USA). Protein concentration was determined using BCA assay (G-Biosciences, St. Louis, USA). The primary antibodies used were anti-YY1 alpha (Abcam, Cat. No. ab12132), anti-β-actin (Cell Signalling Technologies, Cat. No.4967L). Secondary antibodies used were anti-rabbit-horseradish peroxidase (HRP) conjugated (Epitomics, Cat. No.3053-1). Luminol and hydrogen peroxide were used as the substrate of HRP for developing the blot.

2.18. Histological analysis

Briefly, the rat testes were surgically removed and fixed in 4% paraformalydehyde followed by paraffin block preparation. Tissue sectioning were done using microtome machine (2040AUTOCUT, Reichert- Jung), followed by hematoxylin-eosin staining. The testes sections were observed under bright field microscope (Eclipse TiNikon).

The tubule diameters were measured in testis of rats by randomly selecting 10 tubules in each section of testis from three rats. Mean diameter of 10 tubules were used to determine average value.

2.19. Immunostaining

Sc were cultured and fixed in 2% paraformaldehyde. The cells were permeabilized using 0.1% triton X-100 for 2 min followed by blocking with 3% BSA for 30 min. Primary antibody (Abcam, Cat. no.

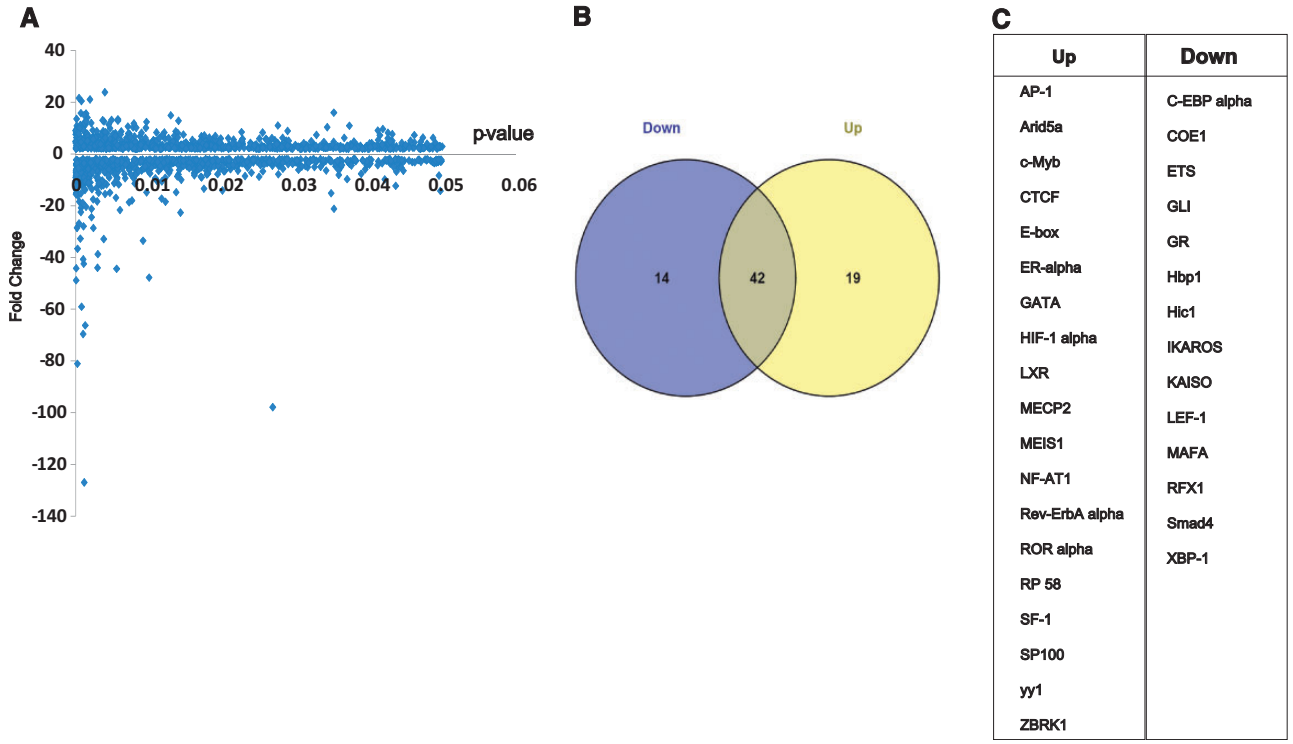

Figure 2. Transcriptomic analysis and TFBS analysis using TRANSFAC database. (A) Volcanic map representing the genes differentially expressed in infant and pubertal Sc of rat as obtained from microarray analysis. Positive y-axis represents the genes up regulated in pubertal Sc as compared to infant Sc. Similarly, negative y-axis represents the genes down regulated in pubertal Sc as compared to infant Sc. $P < 0.05$ as represented by the x-axis. (B) Pie chart showing the TFBS found to be in high abundance over the promoters of up regulated and down regulated genes of pubertal Sc with respect to infant Sc (comparison was performed with the promoters of control genes displaying no change in expression pattern using χ^2 test, $P < 0.05$). (C) List of TFBS found exclusively in high abundance in each case (up and down with reference to pie chart in B). See Supplementary Tables S1 and S2.

ab12132) was used at a dilution of 1:100 (overnight incubation, 4 °C). Secondary antibody alexafluor 488- goat anti-rabbit (Life Technologies, Cat. no. A-11008) was used at a dilution of 1:500 (45 min incubation, RT), Nuclei were stained using Hoechst 3342. The coverslips were mounted on glass slides using prolong gold Antifade reagent (Invitrogen). Cell imaging was done using Ri-2 Epi-flouresscence microscope (Eclipse TiNikon).

Testes were fixed in 4% paraformaldehyde followed by paraffin block preparation and tissue sectioning as described above (Histological analysis). The tissue sections were permeabilized in 0.5% triton X-100 for 20 min followed by blocking with 2% normal goat serum for 30 min. Primary antibody (Abcam, Cat. no. ab12132) was used at a dilution of 1:100 (overnight incubation, 4 °C). Secondary antibody alexafluor 488-goat anti-rabbit (Life Technologies, Cat. no. A-11008) was used at a dilution of 1:500 (4-h incubation, RT), Nuclei were stained using Hoechst 3342. Coverslips were mounted over the tissue sections using prolong gold Antifade reagent (Invitrogen). Tissue imaging was performed using Ri-2 Epi-flourescence microscope (Eclipse Ti Nikon).

2.20. Sperm count analysis

The caput epididymis was surgically removed and minced in 5-ml PBS solution in a petriplate. Ten microliters of the solution containing sperm was loaded into a Neubauer Haemocytometer, and the spermatozoa were counted to determine the sperm count for each animal.

3. Results and discussion

3.1. Unique sets of TFBS govern gene expression programme in infant and pubertal Sc

Genes expressed during puberty allow onset of spermatogenesis, which is restricted during infancy in majority of species.[5,8] In order to define TF-mediated differential gene regulation in infant and pubertal Sc, we used gene expression data (DNA microarray) from NCBI-GEO (GSE48795) previously submitted by our lab. We identified 735 genes predominantly down-regulated and 663 genes predominantly up-regulated ($P < 0.05$ and fold change > 2) in pubertal Sc as compared to infant Sc (Fig. 2A and Supplementary Table S1). Next, we asked which TFBSs were predominant within the promoters of these differentially expressed genes. In order to address this, we searched for 220 TFBS (vertebrate non-redundant Positional Weight Matrices from TRANSFAC database) within ±2-kb promoter sequences of 735 up-regulated, 663 down-regulated and their respective control set promoters (see Material and methods) using the MATCH^{TM} utility.[27] The discrepancy between the frequency of individual TFBS on up and down-regulated gene-set promoters (observed occurrence) and that of control sets (expected by chance) was evaluated independently (Fig. 1A) by determining the statistical variable chi-square (χ^2).[28] Considering a significance level of $P < 0.$

05, we identified 61 TFBSs on up-regulated and 56 TFBSs on down-regulated gene-set promoters (Supplementary Table S2). Out of these, 19 and 14 TFBSs were found exclusive on up-regulated and down-regulated gene-set promoters, respectively (Fig. 2B). Interestingly, we observed that binding sites for NF-AT1, C-Myb, ER-alpha, SP-100, YY1 and HIF-1alpha were uniquely enriched on up-regulated gene promoters (Figs 2C and 3C and Supplementary Table S5) while C-EBP alpha, GR, LEF-1, MAFA, RFX1 and Smad4 were uniquely enriched on down-regulated gene promoters (Figs 2C and 3C and Supplementary Table S5). The binding enrichment of each TFBS on individual gene promoters (both up and down) has been represented by a heat map (Fig. 3A and Supplementary Table S3). Furthermore, clustering analysis suggested groups of genes with exclusively high binding enrichment corresponding to a particular TFBS (Fig. 3A).

Higher occurrence of 3 of the 19 TFBS, namely, GATA box, SF-1 and ER- alpha, in the promoter region of upregulated genes of pubertal Sc was found compared to infant Sc. Significance of TFs corresponding to these three TFBS have been shown previously in the testis displaying active spermatogenesis,[29–32] strengthening the legitimacy of our TFBS analysis. Interestingly, the circadian rhythm-related TFBS, namely, E-box, ROR alpha and Rev-Erba alpha, were abundantly present in up-promoter-set. Association of Sc maturation with photoperiod has already been reported before.[33–35]

3.2. Pattern of TFBS distribution along TSS

We analysed the positional distribution of each TFBS with reference to annotated TSS of up- and down-regulated gene-set promoters. We observed that in case of down-regulated gene sets, majority of TFBS, namely, CEBP-alpha, COE1, GR, Kaiso, Lef-1, MAFA, RFX1, Smad4, were uniformly distributed across ± 2-kb region flanking TSS (Fig. 3B and Supplementary Table S4). However, in the case of up-regulated gene sets, the majority of TFBS, namely, AP-1, Arid5a, LXR, MECP2, Rev-ErbA alpha, ROR alpha, RP58, SF-1, SP100 and ZBRK1, had specific positional location across the ± 2-kb region centred to TSS, unlike uniform distribution observed in the case of down-regulated gene sets (Fig. 3B). This observation is in agreement with the previous report that suggests that TFBS associated with development are uniformly distributed across the TSS.[36] The positional abundance of a particular TFBS with respect to TSS was calculated in terms of its cumulative abundance in all the genes of its corresponding gene set (up and down-gene-set).

In pubertal Sc, restricted positional abundance was observed for many TFBSs. This is in agreement with a previous report suggesting that for tissue-specific expression of genes (potentially regulating cellular status of differentiation), there is a relation between positional binding of TF with reference to TSS.[37,38] We found that promoters of up-regulated-gene-set of pubertal Sc, a terminally differentiated cell, show strict positional restriction in the occurrence of TFBS with reference to TSS (Fig. 3B). Heat map analysis revealed positional abundance of many of the TFBS suggesting their symmetrical presence on both sides of TSS, in promoters of up-regulated genes of pubertal Sc. Although there is no experimental evidence in support of this observation, close proximity binding of some Zinc finger protein flanking TSS (± 200 bases) has been previously reported using TRANSFAC.[28]

3.3. Quantitative proteomics of cytoplasmic and nuclear proteins in infant and pubertal Sc

The transcriptional activity of TFBS is governed by the nuclear proteome profile. TF(s) that are found in cytoplasm but not in the nucleus may not participate in gene expression. Interactome analysis using String database[39] analysis suggested 75 genes to be associated with the TFBS in case of downregulated-gene-set and 87 genes in case of upregulated-gene-set (using only high confidence hits, score > 0.7) of pubertal Sc as compared to infant Sc (Fig. 4A and B and Supplementary Table S6). We applied state-of-the-art SWATH-MS analysis to quantitatively profile the proteome of the nuclear and the cytoplasmic fraction of infant and pubertal Sc. To enable the targeted extraction of peptide assays from SWATH-MS data, we generated a SWATH-assay library from biological triplicates of each fraction and time point used in this study and combined all assays into one comprehensive library (see Materials and methods for details).The final SWATH assay library consists of 21,163 assays (an assay is a set of all the fragment ions corresponding to a specific peptide ion) for 19,776 peptides mapping to 4,532 proteins (Fig. 1B).

Comprehensive relative quantification was performed using SWATH-MS analysis (Fig. 5A–D and Supplementary Tables S7 and S8). Upon comparing the nuclear and cytoplasmic proteome of 5-day-old Sc, 3,164 proteins were quantified (Fig. 5A and Supplementary Tables S7 and S8), of which 1,670 proteins were differentially present in nucleus and cytoplasm fractions ($P < 0.05$). In case of 12-day-old Sc, 3,769 proteins could be quantified (Fig. 5C, Supplementary Tables S7 and S8) out of which 890 proteins were found to be differentially present in nucleus and cytoplasm fractions ($P < 0.05$). In the nuclear proteome fraction of infant and pubertal Sc, 3,027 proteins could be quantified (Fig. 5B, Supplementary Tables S7 and S8), out of which 612 were found to be differentially ($P < 0.05$) localized in their nuclei. GO analysis was done for the nuclear (nuclear/cytoplasmic, fold change > 2, $P < 0.05$) and cytoplasmic (cytoplasmic/nuclear, fold change > 2, $P < 0.05$) proteins in both cases (infant and pubertal Sc). Our GO analysis showed high enrichment of chromatin and other DNA-binding proteins in nuclear fraction and structural proteins associated with cytoskeleton were found in cytoplasmic fraction as expected. This indicates a successful nuclear and cytoplasmic fractionation of Sc (Supplementary Figure S1).

Merging the IDA data of all the samples (nuclear and cytoplasmic of infant and pubertal Sc) and generation of a comprehensive SWATH-assay library, ensured the availability of identification spectra (high-quality spectra matching to a specific peptide) of all those proteins that were present in high abundance even in one of the samples. The availability of identification spectra for such a wide range of proteins in the assay library ensured its quantification by SWATH-MS analysis in all the samples even when it was present in low abundance in some samples. Thus, such comprehensive proteome quantification workflow enabled us to conduct quantitative analysis across all proteome fractions at an unprecedented quantitative depth, fulfilling our objective of quantifying most of the differentially expressed proteins among the nuclear and cytoplasmic fractions of infant and pubertal Sc. This is the first report of quantification of such a wide range of proteins in Sc utilizing next generation proteomic SWATH-MS analysis.

GATA 4 and GATA zinc finger protein (Gatad2a) was found to be in high abundance in nucleus of pubertal Sc (Supplementary Tables S7 and S8). GATA box also appeared in our TFBS

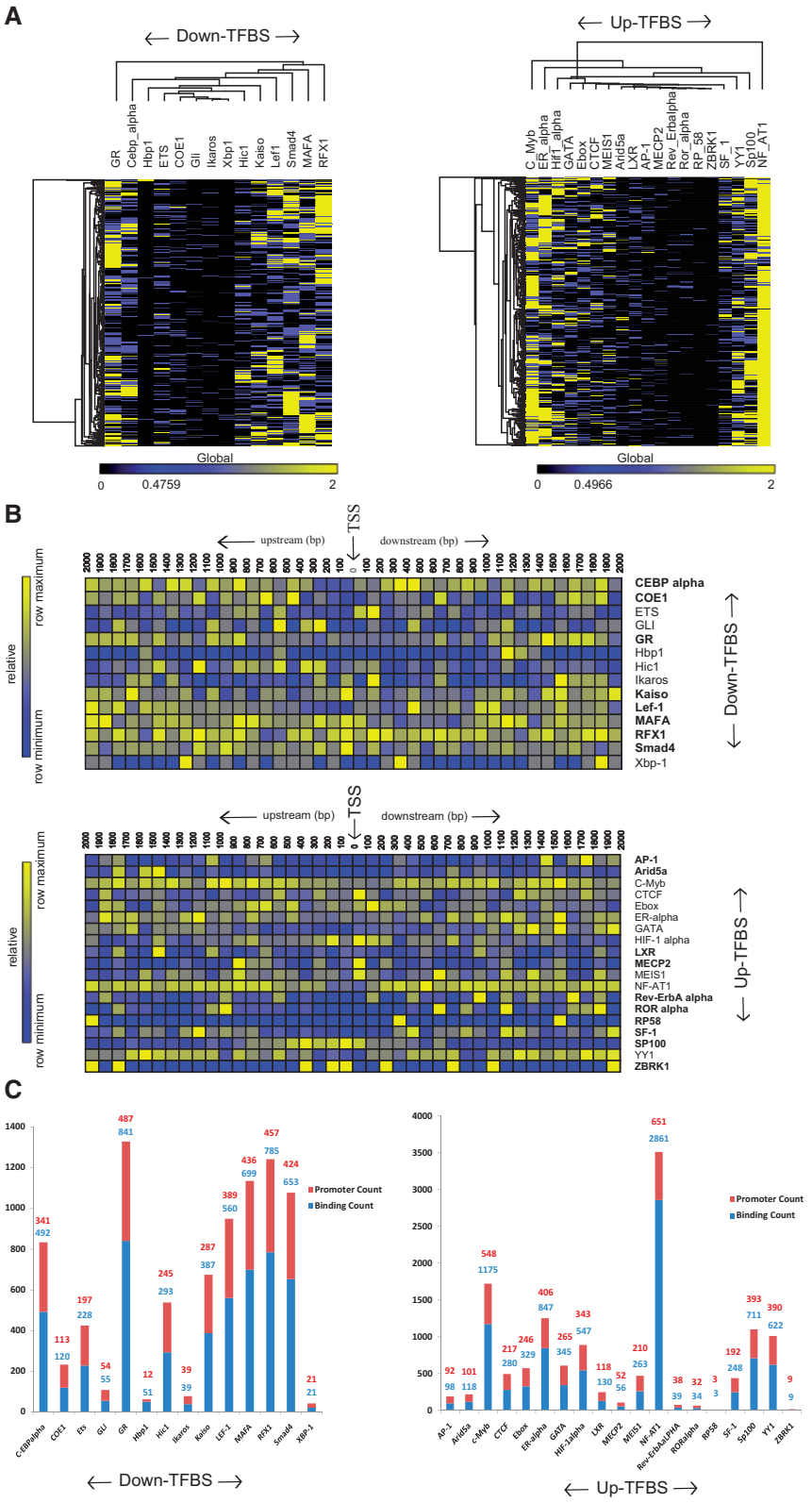

Figure 3. TRANSFAC analysis (A) Heat map showing abundance of each TFBS on the promoters of each gene. (B) Heat map showing positional abundance of each TFBS with reference to TSS cumulatively along whole promoter set. (C) Bar graph showing promoter count and binding against their corresponding promoter set (Binding count represents total number of a particular TFBS located within up and down-regulated gene-set promoter. Promoter count is the number of promoters found in a gene-set promoter (up or down), in which a particular TFBS is present at least once). See Supplementary Tables S3–S5.

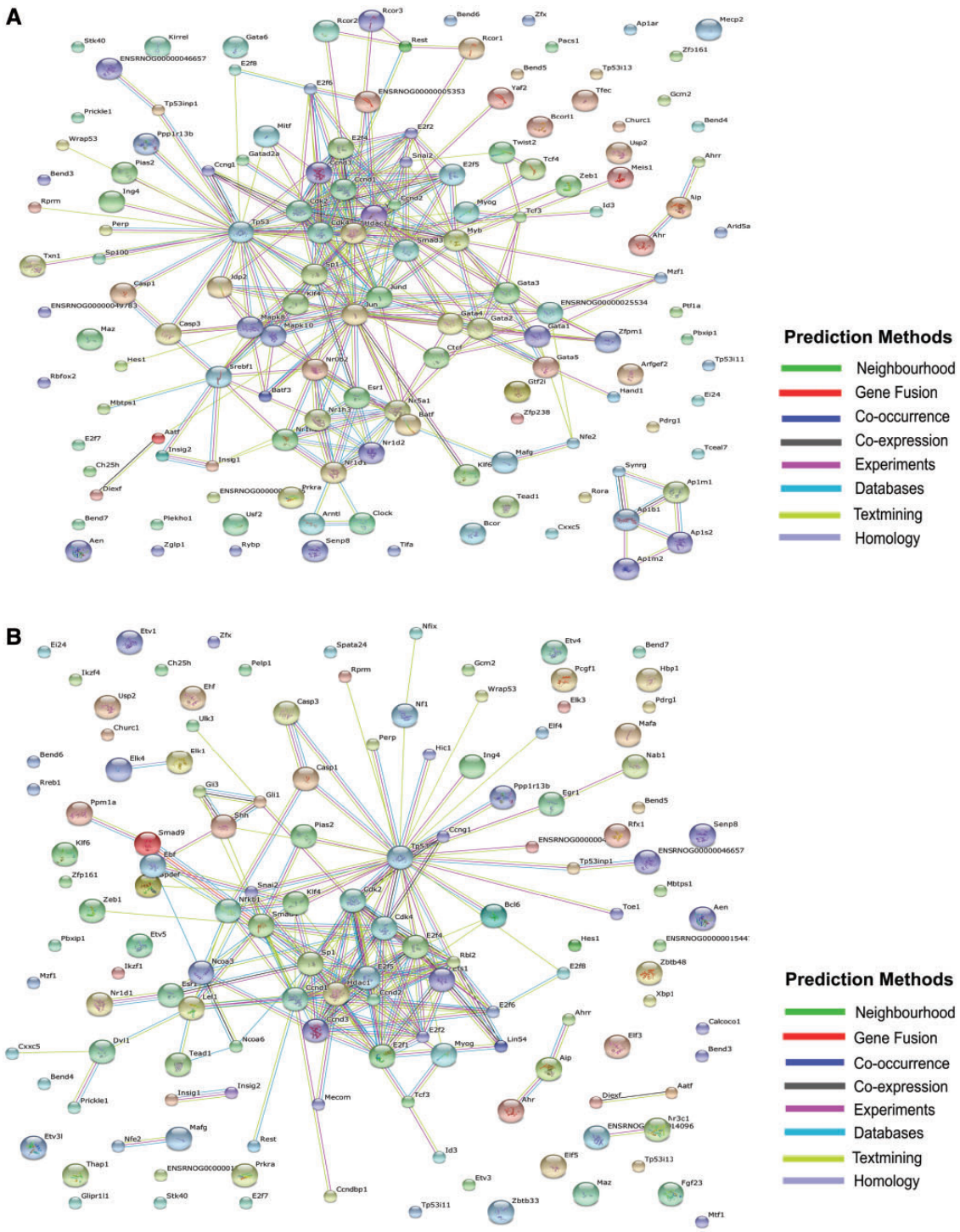

Figure 4. Depiction of interacting partners (build using STRING database) associated with the TFBS obtained from TRANSFAC analysis. (A) Up network and (B) down network. See Supplementary Table S6.

analysis, whose binding sites were found to be in high abundance in the promoters of upregulated genes of pubertal Sc. Only few of the proteins found in our TFBS interactome analysis could be quantified using our SWATH-MS analysis. This is mainly because of low abundance of TFs in general, whose inclusion in a SWATH assay library demands their detection using a very specialized strategy (acquisition of transition list using peptides specific to a particular transcription factor obtained from *in vitro* synthesis) of mass spectrometry.[40] However, this can be performed when the complete SWATH-assay library of rat covering all expressed ORF's becomes publicly available in SWATHAtlas (www.swathatlas.org) (5 December 2016, date last accessed).

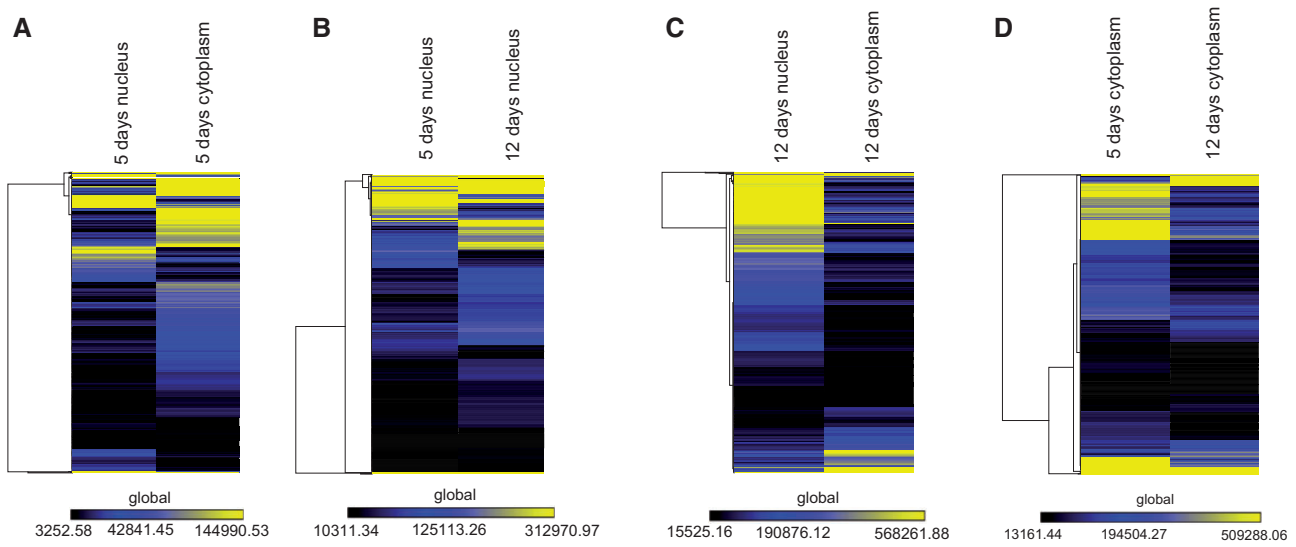

Figure 5. Heat maps depicting comprehensive proteome quantification profile comparison of nuclear and cytoplasmic extracts of infant (5 days old) and pubertal (12 days old) Sc. (FDR < 5%, peptide confidence > 95%, $P < 0.05$). (A) Five-day nucleus vs 5-day cytoplasm. (B) Five-day nucleus vs 12-day nucleus. (C) Twelve-day nucleus vs 12-day cytoplasm. (D) Five-day cytoplasm vs 12-days cytoplasm. See Supplementary Table S7.

3.4. Differential set of unique PTM signatures in nuclear and cytoplasmic proteome of infant and pubertal Sc

Since activity of proteins (TFs, signalling components etc) depends heavily on the PTMs, we interrogated our shotgun proteomic data of nuclear and cytoplasmic fractions, which revealed numerous PTMs in each proteome fraction. Most of the PTMs obtained were exclusive to their nuclear or cytoplasmic presence in a specific age group, though some overlap was there amongst them (Fig. 6A and B and Supplementary Tables S9). This was expected since both the stages of Sc are characteristically distinct from each other. The exclusive appearance of the PTMs in their respective proteome fraction suggested the uniqueness of each proteome. All PTMs obtained in different proteome fractions (nuclear and cytoplasmic of infant and pubertal Sc) were categorized and represented in terms of the absence or presence of a particular PTM in a particular proteome. All the PTMs identified across all fractions were analysed for categorical representation (Fig. 6C and Supplementary Table S10).

Our PTM analysis suggested histone-1.5, a linker histone, to be geranylgeranylated in the nucleus only during infancy. However, its nuclear enrichment with respect to cytoplasm was found to be approximately same in infant and pubertal Sc. Similarly, Rho-GDI alpha was found to be succinylated in the nucleus only in infant Sc, having roughly same nuclear enrichment in infant and pubertal Sc. All these observation suggested our analysis to be a powerful resource for further studies on postnatal maturation of a cell type.

3.5. *In vivo* validation using transgenic knockdown rat model

The TRANSFAC and SWATH-MS analysis of this study provided a very powerful resource based on which several biological follow-up studies can be undertaken. In order to authenticate it, we intended to validate the role of one of the lead candidates obtained from this study. YY1 was one such lead candidate found in this study. Our TRANSFAC analysis showed high abundance of the binding sites of YY1 on the promoters of the genes upregulated in pubertal Sc. The SWATH-MS analysis also suggested YY1 to be localized in the nucleus of pubertal Sc. Immunostaining studies of cultured pubertal Sc further established the nuclear localization of YY1 (Fig. 7A). Immunohistochemistry studies of pubertal testis section was also performed in order to confirm the nuclear localization of YY1 in pubertal Sc at *in vivo* conditions (Supplementary Figure S3).These observations suggested a strong basis for *in vivo* validation of TF YY1 in Sc differentiation. To this end, we have generated a transgenic knockdown rat model of YY1 for *in vivo* validation of its role in Sc differentiation. In order to do so, we generated a transgenic rat that expressed shRNA against YY1 under Pem promoter. The expression of Pem promoter is restricted only in Sc from puberty onwards.[41] Thus, YY1 knockdown rat model (YY1-KD) was supposed to have low abundance of YY1 in pubertal and adult Sc. The control animals were also generated employing similar strategy, expressing shRNA targeted to Luciferase (Luc-KD).

The level of YY1 was found to be low in the YY1 knockdown rats (Fig. 7B). As expected, the level of Anti-Mullerian Hormone (*Amh*), a marker of immature Sc[42] was found to be significantly high in the YY1 knockdown rats as compared to the age-matched control animals (Fig. 7C). Histological studies revealed reduction in seminiferous tubule size of knockdown rats by 35% (Fig. 7D and E). On account of the presence of relatively less matured Sc in YY1 knockdown rats, their sperm count declined by 60% as compared to the control animals (Fig. 7E). The level of *Gdnf* also declined significantly ($P < 0.05$) in the YY1 knockdown rats as compared to the controls (Fig. 7C). It was previously reported that *Gdnf* is required for self renewal of Spermatogonial stem cells.[43] Low level of *Gdnf* in the YY1 knockdown rats explained its compromised ability to support quantitatively normal spermatogenesis leading to reduced sperm count.

In an attempt to determine the comprehensive alteration in the transcriptome profile on account of the knockdown of YY1, we

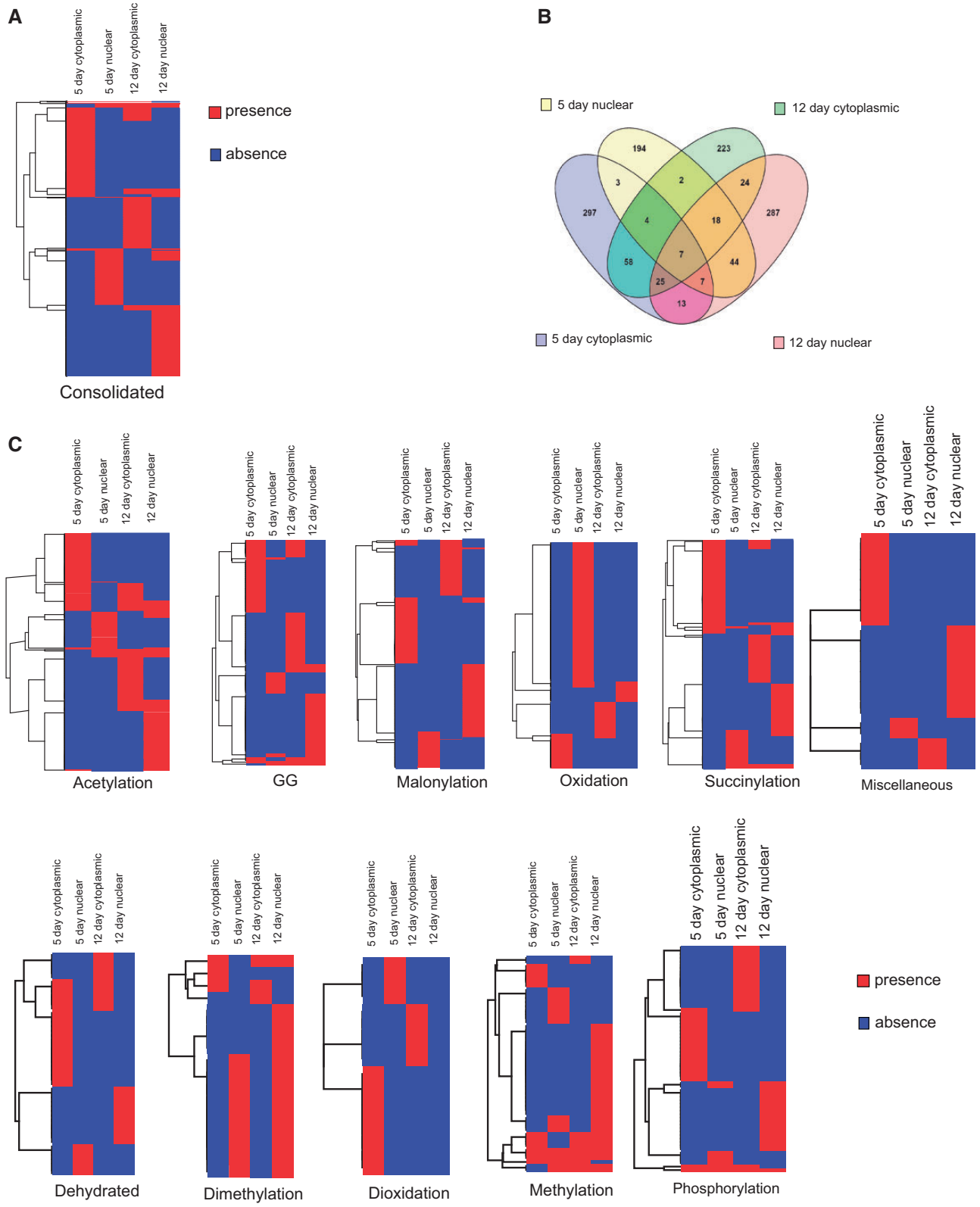

Figure 6. Depiction of comprehensive PTM analysis. Each modification on a single protein has been recognized as a unique modification for representation (red colour represents presence and blue colour represents absence of each modification). (A) Heat map depicting comprehensive PTM analysis of each proteome. (B) Pie chart showing number of PTMs identified in each sample set. (C) Categorical representation of different PTMs. See Supplementary Tables S9 and S10.

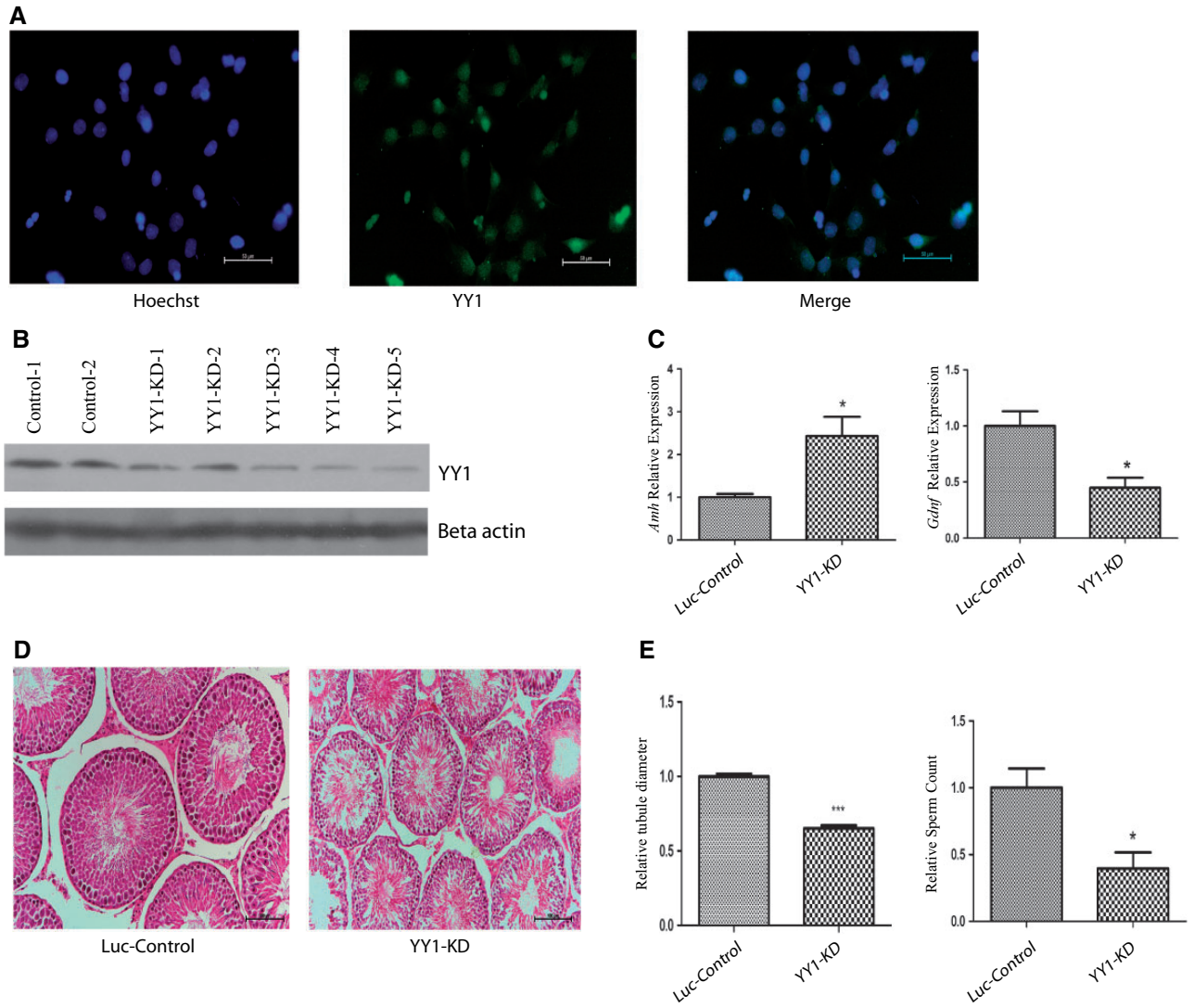

Figure 7. *In vivo* validation of the role of YY1 in Sc differentiation (A) Nuclear localization of YY1 in cultured pubertal Sc. (B) Western blot showing knockdown of YY1 in knockdown (YY1-KD) animals compared to age matched wild type (control) animals. (C) Real-time PCR analysis showing mRNA levels of Amh and Gdnf. (D) Histology of YY1-KD and control animals (Luc-KD) showing reduced seminiferous tubule diameter. (E) Quantitative analysis of seminiferous tubule diameter and sperm count of YY1-KD animals compared to Luc-KD animals. The animals used for real-time PCR, western blot, sperm count and histological analysis were >8 weeks old. Error bars were represented as ±SEM, $n = 3$. Statistical significance was determined using Student's *t*-test. $^{*}P < 0.05$, $^{***}P < 0.001$.

performed transcriptome sequencing analysis (RNA-seq) of testicular RNA. It was found that 191 genes were differentially expressed significantly ($P < 0.05$) in YY1 knockdown rats as compared to the age-matched control animals (Fig. 8A). Interestingly, we observed that the promoters of 64 genes out of these 191 differentially expressed genes harbouring at least one YY1 binding sites within close proximity (±2 kb) of their TSS (Fig. 8B). GO analysis of these 191 differentially expressed genes using GeneCodis[44] suggested that these genes are known to be involved in important cellular processes (Fig. 8C). This indicated the importance of YY1 in maintenance of normal functional status of pubertal Sc. In our transcriptomic analysis, genes like Npy4r had several fold higher expression in YY1 knockdown rat as compared to control animals (Fig. 8A) which was also validated by real time PCR (Supplementary Figure S4). Npy4r is known to play an important role in obesity.[45] Its elevated level in pubertal Sc might be interesting for future studies. On the other hand, Krt72, a keratin protein was found down-regulated in YY1 knockdown rats (Fig. 8A), which was also confirmed by real time PCR (Supplementary Figure S4). Future studies on functional link between Krt72 and YY1 might be interesting.

All these observations suggested that YY1 plays a crucial role in Sc differentiation and therefore attainment of functional maturity. This study was conducted using knockdown approach and thus ended up in partial loss of function of Sc in knockdown rats in terms of sperm count. Sc specific as well as puberty specific knockout of YY1 (using Pem promoter and lox-cre system) may result in complete loss of function of Sc. However, this *in vivo* validation of the role of YY1 in Sc differentiation using knockdown approach provided basis for further studies using other lead candidates obtained in this study.

The next-generation omics methodologies coupled together with computational integration of the TRANSFAC database used in this work exemplify the ability to rapidly divulge molecular bases of cellular differentiation. The computational prediction of differential

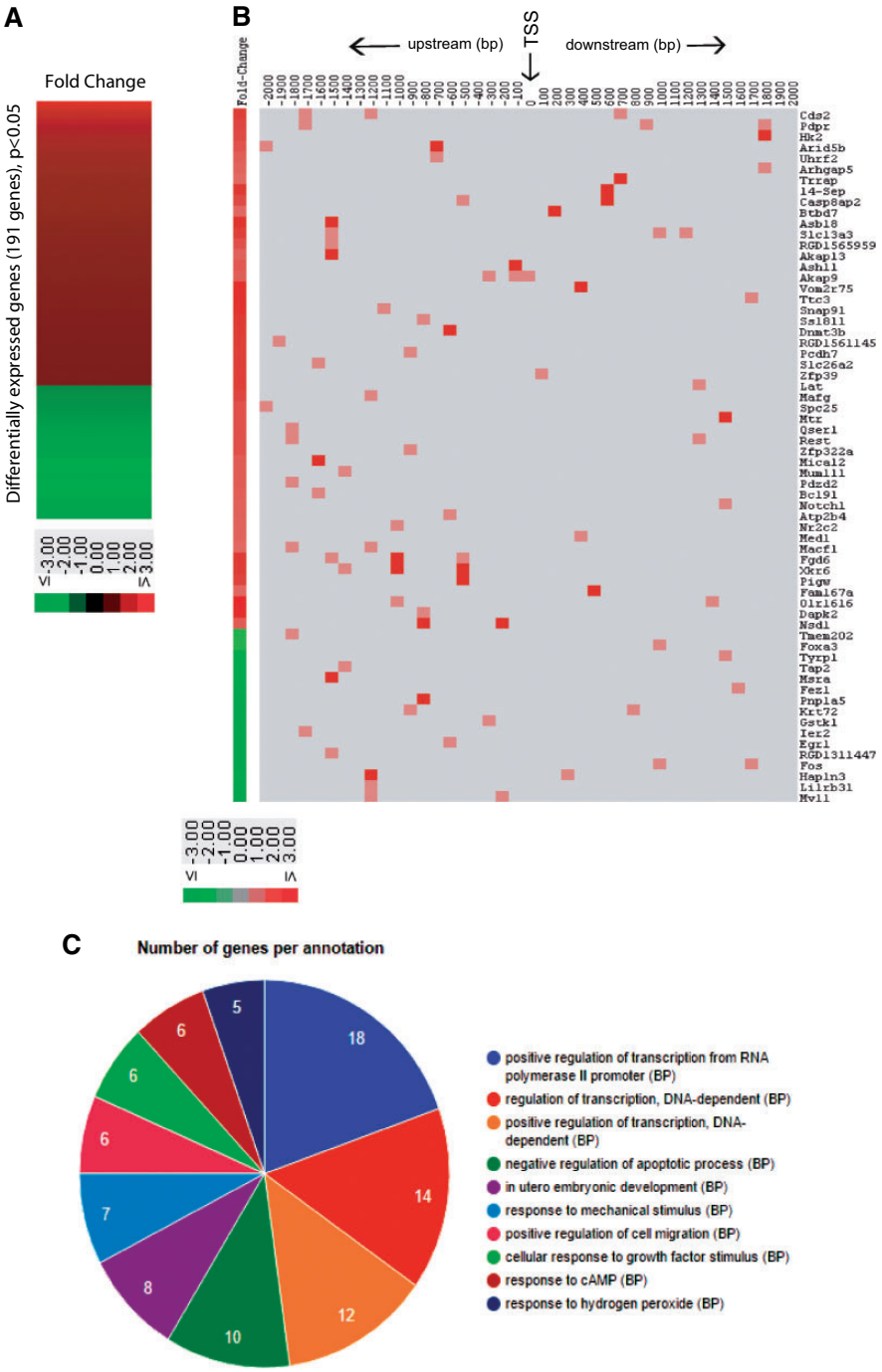

Figure 8. Transcriptome analysis of YY1 knockdown animals. (A) Heat map showing fold change of each differentially expressed gene in testis of YY1 knockdown animal as compared to the age matched Luc-KD control animals. (B) Representation of the YY1 binding sites over the promoters (±2 kb across TSS) of the differentially expressed genes (those having at least one YY1 binding site) along with their fold change in YY1-KD animals. (C) GO analysis of the differentially expressed genes in YY1-KD animals. See Supplementary Table S11.

abundance of TFBS on the promoters of the differentially regulated genes followed by the differential enrichment of nuclear proteins under two or more different context, using quantitative proteomics approach, would provide more precise information about functional regulation of different cellular states. Since functional aspects of proteins are well coordinated by the PTMs, the comprehensive PTM search in nuclear and cytoplasmic fraction of two cell type may be useful in deciphering the mechanism behind cellular programming from one state to another. This approach can be used to determine differentially acting sets of TFBS and nuclear proteins in any two altered states of a cell type.

This study evaluated differential abundance of nuclear proteins and differential abundance of TFBS on the promoters of differentially expressed genes under the two different cellular contexts. Elucidation

of such differentially acting sets of nuclear proteins and TFBS can provide substantial information to predict the differentially acting sets of signalling pathways of the two concerned cellular states. The knowledge about the differentially activated signalling pathways can also be correlated to the cohort of external stimulus acting on the cell.[46–48] Altogether, the outcome of the whole work can be used to visualize the overall picture about the differentially acting molecular concert under two physiologically different states. To our knowledge, this is the first report with an unbiased approach to comprehensively interrogate differentially acting nuclear proteins and TFBS under the two or more cellular context using a multi-omic approach.

Although at this point, robust correlative studies of differential TFs, TFBS and nucleus as well as cytoplasm-specific proteomes along with their pattern of PTM is not possible, availability of more advanced species-specific bioinformatics tools in future may allow one to do so. We believe that use of such multi-omics strategies in other system showing differential regulation of any function may pinpoint meaningful candidates which are crucial to such functions after reduction of the noise which is generated while using only single-omics approach.

3.6. Availability

All the mass spectrometry data was deposited in ProteomXchange with accession number PXD003050 (IDA data) and PXD003056 (SWATH-MS data). The Rat SWATH-assay library was deposited in the online resource SWATHAtlas with dataset identifier PASS00778 (www.swathatlas.org). The transcriptome sequencing data was deposited to NCBI (SRA) with BioProject ID PRJNA349782.

Acknowledgements

We thank Dr. Shantanu Sengupta (IGIB, New Delhi) and Dr.TrayambakBasak (IGIB, New Delhi) for their generous help and Ayushi Jain (NII, New Delhi) for real-time PCR primer standardizations. We also thank Dr. Shantanu Chowdhury (IGIB, New Delhi), Dr.Debasisa Mohanty(NII, New Delhi), Ms Neetu Saini (NII, New Delhi) and Dr. Arnab Mukhopadhyay (NII, New Delhi) for helping in informatics related work.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

The work was supported by National Institute of Immunology, New Delhi and Department of Biotechnology, Government of India (grant no. BT/HRD/35/01/01/2010, BT/PR10805/AAQ/1/576/2013). This work was also supported in part with US federal funds from the American Recovery and Reinvestment Act (ARRA) funds through National Institutes of Health, from the National Human Genome Research Institute grant No. RC2 HG005805 (to RLM), and the National Institute of General Medical Sciences under grant No. R01 GM087221, 2P50 GM076547/Center for Systems Biology and S10 RR027584 (to RLM).

References

1. Jackson, D. A., and Cook, P. R. 1985, Transcription occurs at a nucleoskeleton. *EMBO J.*, **4**, 919–25.
2. Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. 1996, Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J. Cell Sci.*, **109 (Pt 6)**, 1427–36.
3. Orkin, S. H. 2000, Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.*, **1**, 57–64.
4. Simicevic, J., and Deplancke, B. 2010, DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. *Mol. Biosyst.*, **6**, 462–8.
5. Bhattacharya, I., Pradhan, B. S., Sarda, K., Gautam, M., Basu, S., and Majumdar, S. S. 2012, A switch in Sertoli cell responsiveness to FSH may be responsible for robust onset of germ cell differentiation during prepubartal testicular maturation in rats. *AJP Endocrinol. Metab.*, **303**, E886–98.
6. Abel, M. H., Wootton, a N., Wilkins, V., Huhtaniemi, I., Knight, P. G., and Charlton, H. M. 2000, The effect of a null mutation in the follicle-stimulating hormone receptor gene on mouse reproduction. *Endocrinology*, **141**, 1795–803.
7. Chang, C., Chen, Y.-T., Yeh, S.-D., et al. 2004, Infertility with defective spermatogenesis and hypotestosteronemia in male mice lacking the androgen receptor in Sertoli cells. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 6876–81.
8. Sharpe, R. M., McKinnell, C., Kivlin, C., and Fisher, J. S. 2003, Proliferation and functional maturation of Sertoli cells, and their relevance to disorders of testis function in adulthood. *Reproduction*, **125**, 769–84.
9. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., and Mardis, E. R. 2013, XThe next-generation sequencing revolution and its impact on genomics. *Cell*, **155**, 27–38.
10. Lavigne, J. P., Espinal, P., Dunyach-Remy, C., Messad, N., Pantel, A., and Sotto, A. 2013, Mass spectrometry: A revolution in clinical microbiology? *Clin. Chem. Lab. Med.*, **51**, 257–70.
11. Gillet, L. C., Navarro, P., Tate, S., et al. 2012, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics*, **11**, O111.016717.
12. Selevsek, N., Chang, C.-Y., Gillet, L. C., et al. 2015, Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. *Mol. Cell. Proteomics*, **14**, 739–49.
13. Deutsch, E. W., Mendoza, L., Shteynberg, D., et al. 2010, A guided tour of the trans-proteomic pipeline. *Proteomics*, **10**, 1150–9.
14. Deutsch, E. W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., and Moritz, R. L. 2015, Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clin. Appl.*, **9**, 745–54.
15. Chambers, M. C., Maclean, B., Burke, R., et al. 2012, A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, **30**, 918–20.
16. Craig, R., and Beavis, R. C. 2004, TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, **20**, 1466–7.
17. Eng, J. K., Jahan, T. a., and Hoopmann, M. R. 2013, Comet: An open-source MS/MS sequence database search tool. *Proteomics*, **13**, 22–4.
18. Schubert, O. T., Gillet, L. C., Collins, B. C., et al. 2015, building high quality assay libraries for targeted analysis of SWATH MS data. *Nat. Protoc.*, **10**, 426–41.
19. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. 2009, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48.
20. Shilov, I. V, Seymour, S. L., Patel, A. A., et al. 2007, The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics*, **6**, 1638–55.
21. Pradhan, B. S., and Majumdar, S. S. 2016, An efficient method for generation of transgenic rats avoiding embryo manipulation. *Mol. Ther. Acids*, **5**, e293.
22. Sambrook, J., and Russell, D. W. 2006, Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc.*, **2006**.
23. Sarkar, H., Arya, S., Rai, U., and Majumdar, S. S. 2016, A Study of Differential Expression of Testicular Genes in Various Reproductive Phases of Hemidactylus flaviviridis (Wall Lizard) to Derive Their

Association with Onset of Spermatogenesis and Its Relevance to Mammals. *PLoS One*, **11**, e0151150.
24. Kim, D., Pertea, G., Trapnell, C., et al. 2013, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
25. Trapnell, C., Williams, B. A., Pertea, G., et al. 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–5.
26. Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. 2012, Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.*, **31**, 46–53.
27. Kel, a. E., Gößling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. V., and Wingender, E. 2003, MATCH^{TM}: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–9.
28. Kumar, P., Yadav, V. K., Baral, A., Kumar, P., Saha, D., and Chowdhury, S. 2011, Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. *Nucleic Acids Res.*, **39**, 8005–16.
29. Wakabayashi, J., Yomogida, K., Nakajima, O., et al. 2003, GATA-1 testis activation region is essential for Sertoli cell-specific expression of GATA-1 gene in transgenic mouse. *Genes to Cells*, **8**, 619–30.
30. Kato, T., Esaki, M., Matsuzawa, A., and Ikeda, Y. 2012, NR5A1 is required for functional maturation of Sertoli cells during postnatal development. *Reproduction*, **143**, 663–72.
31. Royer, C., Lucas, T. F. G., Lazari, M. F. M., and Porto, C. S. 2012, 17Beta-Estradiol Signaling and Regulation of Proliferation and Apoptosis of Rat Sertoli Cells. *Biol. Reprod.*, **86**, 108–108.
32. Lucas, T. F., Pimenta, M. T., Pisolato, R., Lazari, M. F., and Porto, C. S. 2011, 17beta-estradiol signaling and regulation of Sertoli cell function. *Spermatogenesis*, **1**, 318–24.
33. Newton, S. C., Welsh, M. J., and Bartke, a. 1992, Effects of age, photoperiod and follicle-stimulating hormone on lactate production by cultured Sertoli cells from prepubertal Siberian hamsters (Phodopus sungorus). *J. Reprod. Fertil.*, **95**, 87–95.
34. van Haaster, L. H., van Eerdenburg, F. J., and de Rooij, D. G. 1993, Effect of prenatal and postnatal photoperiod on spermatogenic development in the Djungarian hamster (*Phodopus sungorus sungorus*). *J. Reprod. Fertil.*, **97**, 223–32.
35. Seco-Rovira, V., Beltrán-Frutos, E., Ferrer, C., Sáez, F. J., Madrid, J. F., and Pastor, L. M. 2014, The death of sertoli cells and the capacity to phagocytize elongated spermatids during testicular regression due to short photoperiod in Syrian hamster (*Mesocricetus auratus*). *Biol. Reprod.*, **90**, 107.
36. Koudritsky, M., and Domany, E. 2008, Positional distribution of human transcription factor binding sites. *Nucleic Acids Res.*, **36**, 6795–805.
37. Tharakaraman, K., Bodenreider, O., Landsman, D., Spouge, J. L., and Marĩno-Ramírez, L. 2008, The biological function of some human transcription factor binding motifs varies with position relative to the transcription start site. *Nucleic Acids Res.*, **36**, 2777–86.
38. Bellora, N., Farré, D., and Albà, M. M. 2007, Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters. *BMC Genomics*, **8**, 459.
39. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. 2003, STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.*, **31**, 258–61.
40. Stergachis, A. B., MacLean, B., Lee, K., Stamatoyannopoulos, J. A., and MacCoss, M. J. 2011, Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat. Methods*, **8**, 1041–3.
41. Rao, M. K., Wayne, C. M., Meistrich, M. L., and Wilkinson, M. F. 2003, Pem homeobox gene promoter sequences that direct transcription in a Sertoli cell-specific, stage-specific, and androgen-dependent manner in the testis in vivo. *Mol. Endocrinol.*, **17**, 223–33.
42. Donahoe, P. K., Ito, Y., Marfatia, S., and Hendren, W. H. 1976, The Production of mullerian inhibiting substance by the fetal, neonatal and adult rat. *Biol. Reprod.*, **15**, 329–34.
43. Oatley, J. M., and Brinster, R. L. 2008, Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.*, **24**, 263–86.
44. Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. M., and Pascual-Montano, A. 2007, GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.*, **8**, R3.
45. Aerts, E., Beckers, S., Zegers, D., et al. 2016, CNV analysis and mutation screening indicate an important role for the NPY4R gene in human obesity. *Obesity (Silver Spring).*, **24**, 970–6.
46. Bohmann, D. 1990, Transcription factor phosphorylation: a link between signal transduction and the regulation of gene expression. *Cancer Cells*, **2**, 337–44.
47. Hao, N., Budnik, B. A., Gunawardena, J., and O'Shea, E. K. 2013, Tunable signal processing through modular control of transcription factor translocation. *Science*, **339**, 460–4.
48. Chang, F., Steelman, L. S., Lee, J. T., et al. 2003, Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, U.K*, **17**, 1263–93.