RESEARCH ARTICLE

Differences in embryo quality are associated with differences in oocyte composition: A proteomic study in inbred mice

Martin J. Pfeiffer[1,2], Leila Taher[2], Hannes Drexler[3], Yutaka Suzuki[4], Wojciech Makałowski[5], Caroline Schwarzer[1], Bingyuan Wang[1], Georg Fuellen[2] and Michele Boiani[1]

[1] Max Planck Institute for Molecular Biomedicine, Münster, Germany
[2] Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, Rostock, Germany
[3] MaxPlanck Institute for Molecular Biomedicine, Bioanalytical Mass Spectrometry Facility, Münster, Germany
[4] Laboratory of Functional Genomics, Department of Medical Genome Sciences, University of Tokyo, Japan
[5] Faculty of Medicine, Institute of Bioinformatics, University of Münster, Münster, Germany

Received: July 15, 2014
Revised: October 10, 2014
Accepted: October 29, 2014

Current models of early mouse development assign roles to stochastic processes and epigenetic regulation, which are considered to be as influential as the genetic differences that exist between strains of the species *Mus musculus*. The aim of this study was to test whether mouse oocytes vary from each other in the abundance of gene products that could influence, prime, or even predetermine developmental trajectories and features of derivative embryos. Using the paradigm of inbred mouse strains, we quantified 2010 protein groups (SILAC LC-MS/MS) and 15205 transcripts (RNA deep sequencing) present simultaneously in oocytes of four strains tested (129/Sv, C57Bl/6J, C3H/HeN, DBA/2J). Oocytes differed according to donor strain in the abundance of catalytic and regulatory proteins, as confirmed for a subset (bromodomain adjacent to zinc finger domain, 1B [BAZ1B], heme oxygenase 1 [HMOX1], estrogen related receptor, beta [ESRRB]) via immunofluorescence in situ. Given a Pearson's *r* correlation coefficient of 0.18–0.20, the abundance of oocytic proteins could not be predicted from that of cognate mRNAs. Our results document that a prerequisite to generate embryo diversity, namely the different abundances of maternal proteins in oocytes, can be studied in the model of inbred mouse strains. Thus, we highlight the importance of proteomic quantifications in modern embryology. All MS data have been deposited in the ProteomeXchange with identifier PXD001059 (http://proteomecentral.proteomexchange.org/dataset/PXD001059).

Keywords:
Cell biology / Deep RNA sequencing / Development / Mouse strains / Oocyte / SILAC

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

A cornerstone of research in developmental biology is the use of the mouse as a model to explore mechanisms and processes conserved in mammals. Current models of early mouse embryo development assign major roles to stochastic processes and epigenetic regulation, whether occurring in the natural environment of the oviduct or in a culture medium. For example, upon oocyte fertilization the division history of blastomeres influences their fating to the first tissues of the

Correspondence: Dr. Michele Boiani, Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, D-48149 Münster, Germany
E-mail: mboiani@mpi-muenster.mpg.de
Fax: +49-251 70365 399

Abbreviations: CV, coefficient of variation; **ICSI**, intracytoplasmic sperm injection; **PA**, parthenogenetic activation; **pEct**, primitive ectoderm; **pEnd**, primitive endoderm; **SCNT**, somatic cell nuclear transfer; **TE**, trophectoderm

blastocyst-stage embryo, namely primitive ectoderm (pEct), primitive endoderm (pEnd), and trophectoderm (TE) [1]. In the distinct but related field of embryo cloning, in which embryos are created using somatic cell nuclei instead of sperm nuclei, the role of stochastic processes and epigenetic regulation in determining developmental outcomes may be even more prominent [2, 3]. We envision that the presence of maternal gene products, namely RNAs and proteins, in oocytes could influence, prime, or even predetermine developmental trajectories and features of embryos.

Advances in "omics" research show that, in general, intermediate steps of the gene expression cascade, as well as post-translational protein modification and degradation, can affect the levels of functionally available proteins independently of transcription [4]. The protein composition of oocytes of different mouse inbred strains was studied in the 1990s with 2D gel protein electrophoresis; the studies showed that these oocytes can be distinguished by electrophoretic spot patterns, and that these proteins correlate with developmental differences. In particular, key to the developmental outcome are certain maternal proteins known as "egg modifiers" and maternal effect gene products [5–8]. Latham uncovered at least 17 proteins that exhibited significant, reproducible, quantitative differences in patterns of protein abundance between oocytes of C57Bl/6J and DBA/2J strains, whose support of embryo development after fertilization is different [5]. These differences in oocyte developmental potential, which have been in part mapped to specific loci [9], also are encountered when C57Bl/6J and DBA/2J oocytes are transplanted with somatic cell nuclei [3], suggesting that the different performances are not specific to oocyte–sperm interaction, but may be inherent to the recipient oocyte. Using 2D electrophoretic gel analysis, Richoux and colleagues revealed a 36.5 kDa peptide, back then named D14, which discriminated between developer and nondeveloper oocytes in the context of the DDK syndrome [10]. With the technology available in the 1990s, however, it was not possible to determine the identity of the proteins in question. Upon the advent of transcriptome profiling, for example, microarray and RNA-sequencing combined with preamplification methods in the early 2000s, the field shifted from 2D gel protein electrophoresis to the analysis of RNAs that encode those proteins. In this decade (2010s), improved sensitivity of proteomics technology based on LC-MS/MS made oocytes amenable to high-throughput quantitative proteomics analysis [11].

In this study, we raise the question of whether differences in mouse embryo quality can be explained, at least in part, by initial differences of oocyte composition, reaching beyond the level of mRNA analysis by adding the information layer of protein expression. A prerequisite to addressing this question is to expose the differences of oocyte composition, if they exist. Ideally, one would like to match the composition of a mouse oocyte with the developmental ability of the same oocyte. This direct match is precluded because the oocyte is almost invariably consumed in the assay that measures its composition. If this dual information cannot be obtained from the same oocyte, then it may be possible to use independent oocytes, provided they have minimal genetic variability within strains but still present defined genetic differences between strains. For these reasons we chose to work with inbred strains of mice, which have nearly all (>98%) loci in homozygosis owing to >20 rounds of brother–sister mating, while differing from strain to strain at defined genetic loci. Furthermore, inbred strains are well characterized (e.g., genome-sequenced [12]) and they present developmental differences that include, but are not limited to, duration of gestation, birth rate, and lifespan as reported in the Mouse Phenome Database [13].

After we verified the distinct abilities of oocytes from four inbred strains (129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J) to produce blastocysts, we analyzed oocyte proteomes using current quantitative proteomics technology based on the stable isotope labeling with aminoacids in cell culture (SILAC). Quantitative proteomic analysis of these oocytes revealed that they differ strain-wise in the abundance of catalytic and regulatory proteins (including epigenetic modifiers and maternal effect factors), as confirmed for a subset (bromodomain adjacent to zinc finger domain, 1B [BAZ1B], heme oxygenase 1 [HMOX1], estrogen related receptor, beta [ESRRB]) via immunofluorescence in situ. These differences in protein abundance could not be predicted from transcriptome analysis (Pearson's r correlation coefficient of 0.18–0.20). Thus, even before the genotype of the embryo is put together upon the union of oocyte and sperm, a set of maternal molecules present in the oocytes may predetermine how the genetic information is going to be used and how developmental decisions are going to be made. Our data highlight the importance of proteomic quantifications in modern embryology. Practically, different inbred strains of mice should not be used interchangeably when tackling questions about oogenesis and early development; and a variety of strains should, whenever possible, be studied to yield results of sufficient general validity.

2 Materials and methods

2.1 Study design

We studied oocytes from four inbred mouse strains (129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J) that support different developmental rates when triggered by the same developmental stimulus. We asked if the abundance of proteins in oocytes precedes and correlates with the observed differences in developmental rates.

2.2 Ethics approval for animal experiments

Mouse experiments were performed in accordance with the recommendations of the Federation of Laboratory Animal Science Associations (FELASA) and with the ethical permit issued by the Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) of the state of North Rhine-Westphalia, Germany (permit number: 87–51.04.2010.A160).

2.3 Oocyte collection

Metaphase II (MII) oocytes were collected after gonadotropin treatment and cervical dislocation, as described in [11], from mice of strains 129/Sv (Charles River), C57Bl/6J, C3H/HeN and DBA/2J (Janvier), aged 6–8 weeks. Mice were housed in groups of five in individually ventilated cages, and the four groups were always handled in parallel. For proteome analysis, we removed the zona pellucida of oocytes using acidic Tyrode's solution.

2.4 Embryo production

Embryos were produced by natural mating (CD1 sperm donor), intracytoplasmic sperm injection (ICSI, CD1 sperm donor), somatic cell nuclear transfer (SCNT, B6C3F1 cumulus cell donor), or parthenogenetic activation (PA) of MII oocytes. For each developmental stimulus the four groups were handled in parallel. Resultant embryos were cultured to blastocysts in alpha-MEM culture medium, as described in [14].

2.5 Confocal microscopy immunofluorescence of blastocysts to identify pEct, pEnd, and TE cells

All staining and imaging procedures to characterize the cell lineage composition of blastocysts have been performed as previously described [15].

2.6 Isotopic labeling, protein isolation, fractionation, MS, and protein identification/quantification

In compliance with the Minimum Information about a Proteomics Experiment (MIAPE) reporting guidelines, we used an established pipeline for the quantitative identification of oocyte proteins that has been described in detail before [11] with a few improvements. Our pipeline relies on the SILAC of a F9 carcinoma cell reference [11]. Estimates of protein abundance in the oocyte samples are made against the F9 reference and are given as heavy (F9)/light (oocyte) signal ratio (H/L). In brief, protein lysates from zona-free oocytes were mixed 1:1 (protein amount) with heavy F9 carcinoma cell lysate, acetone-precipitated, reduced and alkylated, and then digested with Endoproteinase Lys-C (3hr) and Trypsin (overnight). Following desalting on Empore 3M C18 discs, samples were offline fractionated by RP-HPLC at pH 10.2 (buffer A: 10 mM ammonium formiate, pH 10.2; buffer B: 10 mM ammonium formate, 90% ACN, pH 10.2; linear gradient from 0 to 35% B in 70 min; 35–70% B in 15min; 70% B for 10 min; Waters XBridge BEH C18 2.1 × 150 mm). Twenty pools were generated from each sample by concatenated fractionation, dried down in a SpeedVac, and subsequently analyzed individually by LC-MS/MS on a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Waltham, MA, USA), equipped with an Easy nano-LC system and a nanoelectrospray source (both from Proxeon, Odense, Denmark). The obtained raw data were processed by MaxQuant software (v 1.4.1.2). The MS proteomics data have been deposited to the ProteomeXchange Consortium [16] via the PRIDE partner repository with the dataset identifier PXD001059.

2.7 Deep RNA sequencing/next generation sequencing

Total RNA was extracted using RNAeasy (Qiagen). Using 200 μg of the total RNA, a sequencing library for Illumina deep sequencing was constructed using TruSeq RNA Seq kit (version 2) according to the manufacturer's instructions. Details of the procedures are also described elsewhere [17]. Sequencing was conducted on Illumina HiSeq 2500 platform by 36-base-single-end read sequencing. Approximately 20 million sequences were generated per sample and used for the analysis. The sequence data have been registered in the DNA Databank of Japan (DDBJ) under the accession number DRA002284.

2.8 Bioinformatics

Hundred base pairs long paired-end reads were first preprocessed to remove low-quality sequence on both the 3′ and 5′ ends using Trimmomatic v0.22, with the following parameters: LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50. Preprocessed reads were aligned to the mouse reference genome (Ensembl release 72, GRCh38) using TopHat v1.2.0.6 [18] with parameter -G and known splice junctions from Ensembl (ftp://ftp.ensembl.org/pub/release-72/gtf/mus_musculus/Mus_musculus.GRCm38.72.gtf.gz). Properly paired reads with a mapping quality score of at least 20 were extracted from the resulting BAM file using SAMtools [19] for further analyses. Read counts per gene were calculated for each sample with HTSeq [20] with default parameters. Genes with counts in the lowest 30% quantile were excluded from further analyses. Normalization and differential expression analysis was performed using the Bioconductor/R package DESeq [21].

2.9 Ranking of protein and mRNA abundance values

In order to perform functional enrichment analysis of the gene products that are most variably expressed in oocytes across the four strains of mice, we produced a ranking of these products' abundances based on statistical variance of the H/L ratios (proteomics) and molecule counts (RNA deep sequencing). As a measure of variance, we used the coefficient of variation (CV), defined as the ratio between SD and mean

of the H/L ratios or molecule counts. Analysis of CV values protects against detecting patterns in variability influenced by trends in absolute expression alone [22].

2.10 Confocal microscopy immunofluorescence of oocytes to verify most differently abundant proteins

The cell fixation, immunofluorescence, and imaging protocol was described previously [11]. The primary antibodies were anti-BAZ1B (Sigma, rabbit polyclonal Cat# W3516), anti-ESRRB (Santa Cruz, rabbit polyclonal Cat# sc-68879), anti-HMOX1 (Enzo Life Sciences, rabbit polyclonal Cat#ADI-SPA-895), and anti-GAPDH (where GAPDH is glyceraldehyde-3-phosphate dehydrogenase; EnCor Biotechnology, rabbit polyclonal Cat# RPCA-GAPDH). Appropriate Alexa Fluor-tagged secondary antibodies (Invitrogen, Life Technologies) were matched to the primaries (Alexa Fluor 647 Donkey Anti-Mouse, Cat# A31571; Alexa Fluor 647 Donkey Anti-Rabbit, Cat# A31573). Primary antibodies were applied in excess (5 μg/mL) to ensure that the limiting factor is the epitope not the antibody; secondary antibodies were applied at the standard concentration (1 μg/mL). The specimens were counterstained for DNA using YOPRO1 (Invitrogen, Life Technologies Cat# Y3603). The fluorescence signal intensity was quantified using Image J (U.S. National Institutes of Health, Bethesda, MD, USA).

2.11 GO analyses

Ontology analysis was conducted using DAVID [23] by computing overrepresentation of the top 50 most variant proteins and the top 100 most variant RNAs (according to the CV), respectively. Significance thresholds required a false discovery rate lower or equal to 0.05, and the entire known mouse transcriptome (proteome) was used as background.

2.12 Statistical data analysis

Statistical analyses were conducted using the statistical analysis package R.

3 Results and discussion

3.1 Oocytes from different inbred mice yield different blastocyst rates and qualities

The central question that we raised in this study was whether differences in embryo quality may be accounted for by initial differences in oocyte composition. Therefore, we first documented these differences in both embryos and oocytes. Using the paradigm of inbred mouse strains, we first showed that the oocytes ovulated by 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J mice (Fig. 1A and B) support different blastocyst

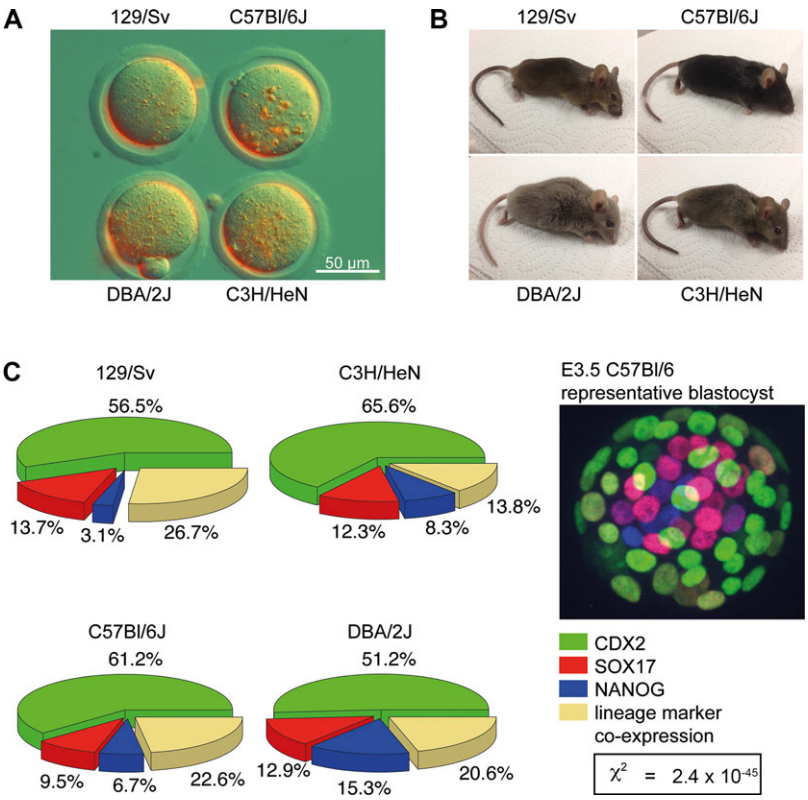

Figure 1. Oocyte and embryo features of the four inbred mouse strains 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J. (A) Appearance of MII oocytes obtained from the four strains. (B) Appearance of female mice from the four strains (anesthetized). (C) Pie chart representation and analysis of cell allocation in blastocysts retrieved from intercrosses of females of the four strains with CD1 males. Cdx2 (TE), Nanog (pEct), and Sox17 (pEnd) were used as markers to distinguish the different lineages. Coexpression of markers is indicated in yellow color. Depending on the mouse strain analyzed, the percentage of cells contributing to a certain lineage in the blastocyst differ (see chi-square test in the bottom-right of the figure).

Table 1. Developmental rates to the blastocyst stage of the four mouse strains C57Bl/6J, C3H/HeN, DBA/2J, and 129/Sv

	Fert (in vivo)		Fert (ICSI)		SCNT		Parthenogenesis	
	1-cell	blast (%)	1-cell	blast (%)	1-cell	blast (%)	1-cell	blast (%)
C57Bl/6J	27	26 (96)	51	0 (0)	31	10 (32)	182	6 (3)
C3H/HeN	39	35 (90)	55	29 (53)	34	3 (9)	61	44 (72)
DBA/2J	94	57 (61)	47	0 (0)	73	0 (0)	75	1 (1)
129/Sv	181	72 (40)	27	7 (26)	165	21 (13)	347	9 (3)
Chi-square test, p	4.46×10^{-12}		Test not possible		Test not possible		2.22×10^{-70}	

Development was triggered using four different techniques, namely natural mating with CD1 males (Fert [in vivo]), fertilization using ICSI of a CD1 sperm head (Fert [ICSI]), SCNT (with B6C3F1 cumulus cell nucleus), and parthenogenesis. For each group the starting number (1-cell) and number of obtained blastocysts (blast) are given. The percentage of developed embryos is given in brackets. Frequencies of blastocyst formation were analyzed by chi-square test (null observed frequencies preclude correct usage of this test).

rates and qualities, and then we looked for differences in the composition of the oocytes.

Following gonadotropin stimulation and induced ovulation, we subjected the MII oocytes to diverse developmental stimuli, namely in vivo fertilization (CD1 sperm), fertilization via ICSI (CD1 sperm), SCNT (B6C3F1 cumulus cell nuclei), and PA. In vivo fertilized oocytes were allowed to develop in vivo, while ICSI, parthenogenetic, and SCNT oocytes were cultured in the same medium. Blastocyst rates varied among the strains (χ^2, $p \leq 4.46 \times 10^{-12}$), although the ranking of the rates was not conserved (in vivo fertilization, C57Bl/6J > C3H/HeN > DBA/2J > 129/Sv; ICSI, C3H/HeN > 129/Sv > C57Bl/6J = DBA/2J; PA, C3H/HeN > C57Bl/6J = 129/Sv > DBA/2J; SCNT, C57Bl/6J > 129/Sv > C3H/HeN > DBA/2J; Table 1). Thus, it appears that the oocytes of the four inbred strains differ in their blastocyst potential, with C3H/HeN oocytes scoring among the best and DBA/2J oocytes scoring among the worst (Table 1). These observations suggest that the ability to develop in a specific environment or to recover from stress of manipulation depends on the composition of the oocyte.

As an indicator of blastocyst quality, we examined the allocation of blastomeres in the blastocysts derived from 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes fertilized in vivo using our established method [15]. The cell types of the blastocyst are the pEct or epiblast that provides pluripotent cells (the progenitors for the adult body); the pEnd or hypoblast that provides essential extraembryonic annexes, such as the yolk sac; and the TE that is essential for placenta formation. We analyzed the NANOG-positive (pEct), SOX17-positive (pEnd), and the CDX2-positive (TE) cells in the blastocysts, which had a total cell count of 40 ± 5 cells. Setting an interval (35–45 cells) is necessary to exclude the trivial effects of the number of cell cycles elapsed since fertilization. A representative picture of a stained blastocyst and a summary of the cell allocation distributions for the four different strains are given in Fig. 1C. Chi-square analysis revealed that the proportions of the three cell types in the blastocysts of the four strains were different overall (χ^2, $p = 2.47 \times 10^{-45}$). A caveat in our experimental design is that the in vivo fertilized oocytes of the four strains were allowed to develop in the genital tract of the own female (129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J) after mating to CD1 males. This means that, in addition to the oocytes, the genital tracts also were of different genotypes. However, this caveat does not bear on the comparison of the blastocyst rates obtained from the other three developmental stimuli (ICSI, SCNT, PA).

Taken together, these data document that the development of oocytes to blastocyst and the quality of these blastocysts are mouse strain-dependent, even if triggered to develop under conditions as natural as possible, and under the provision of equal gonadotropin doses for all four strains. While these differences of development were not surprising, we reasoned that the different molecular compositions of the oocytes across the four strains of mice may account, at least in part, for the developmental differences we observed. The mature mouse oocyte is transcriptionally quiescent [24] and relies on a stockpile of ready-made transcripts and proteins that allow it to jump-start and sustain development until the stage of oocyte-to-embryo transition. Whole proteome analysis of mouse oocytes is, however, severely impeded by the requirement of large numbers of oocytes as starting material. Although proteomic analysis of single oocytes is feasible in other animal taxa, for example, *Xenopus* [25], this is mainly because the biomass of a *Xenopus* oocyte is >1000 times larger than that of a mouse counterpart. For these reasons, we worked with pools of mouse oocytes, comparing the pools from 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J inbred strains using a "spike-in" reference added to each sample (see below).

3.2 Proteomes of oocytes from different inbred mouse strains and their relation to the transcriptomes

To shed light on the developmental diversity of 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes, we collected a total of 750 zona-denuded MII oocytes for each of the four strains and subjected them to our quantitative proteomics pipeline. Seven hundred fifty is the minimum number of oocytes required by our present-day proteomic pipeline to reach an analytical depth of 3000–4000 identified proteins, which is the size of the currently detectable mouse oocyte

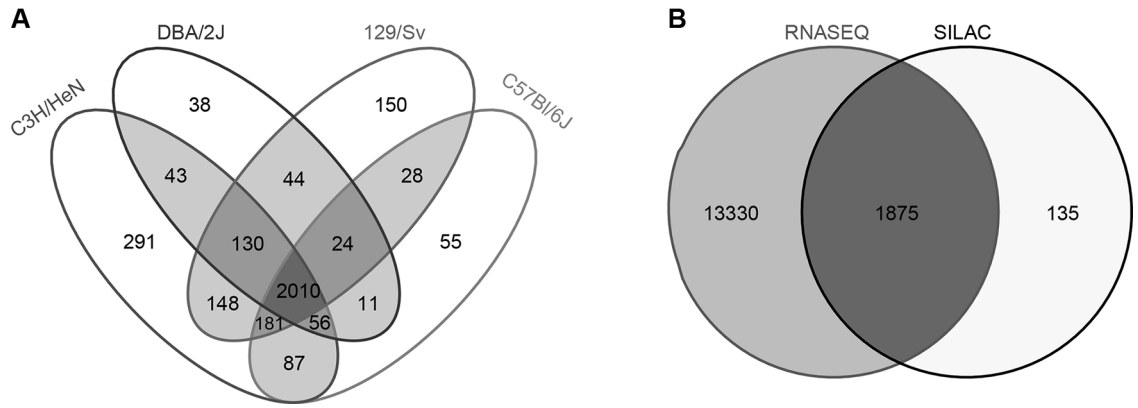

Figure 2. Venn diagrams showing (A) the intersections of proteins quantified (SILAC proteomics) in oocytes from four mouse strains and (B) the intersection of all RNAs quantified (RNAseq; 15205) and all proteins (2010) quantified simultaneously in oocytes from four mouse strains. The inbred mouse strains were 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J.

proteome. Our proteomic pipeline is based on the SILAC of a cell line reference. SILAC has significant advantages as compared to chemical labeling strategies, such as iTRAQ and TMT, since the labeling event occurs upstream in the sample preparation chain.

In our SILAC approach, cell lysates from zona-free oocytes were added with cell lysate from heavy (Lys8 and Arg10) F9 embryonal carcinoma cells in a 1:1 fashion (spike) prior to processing for LC-MS/MS. The labeling efficiency of F9 cells was 97.8%. We chose F9 cells as reference because they can easily be cultured feeder-free (feeder cells interfere with labeling efficiency by SILAC; Fig. 1 in [26]), because F9 cells have been used in proteomic studies before [27], and because F9 cells are a classic model for developmental pluripotency [28, 29]. Even if the F9 proteome is different from the proteome of oocytes, for a given protein the F9 amount is the same across the four groups of oocytes, hence the differences that we may observe between 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes are due to the oocytes. Using the F9 spike-in method, we previously were able to generate a quantitative catalogue of 2324 oocytic proteins from oocytes that were retrieved from B6C3F1 mice as old as 1 year [11]. Both in the previous and in the present study, peptides without counterparts (whether light or heavy) were excluded from our analysis, which therefore was based on proteins identified in both F9 cells (heavy peptides, H) and oocytes (light peptides, L). We were able to quantify a total 3043 protein groups based on two matching peptides, of which at least one was unique to the protein group. Considering ties that may occur because of isoforms/splice variants, mapping resulted in 3296 protein identities, of which 2715, 2452, 2946, and 2356 proteins were found in 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes, respectively. Furthermore, analysis was conducted when the H and L peptides were detected in all of the four groups of oocytes (129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J). Common to all four samples were 2010 proteins that provided the basis for further analysis (Fig. 2A). The primary H/L ratios for the proteins quantified in each strain are given in the output format of MaxQuant software in Supporting Information Table 1. It should be noted that some of the coming data analyses (e.g., Pearson's r correlation proteome transcriptome) and visual representations (e.g., Figs. 3–6) will be facilitated if the H/L ratios are inverted to L/H ratios, so as to have a direct instead of reciprocal ($1/x$) comparison between protein abundance and other physical quantities of oocytes.

Compared to our present results, previous proteomic studies based on label-free, that is, non-SILAC methods identified the proteins present in MII stage ICR or CD1 oocytes to a depth of 380 [30] and 625 [31] protein identities, respectively. Wang et al. raised the number to 2973 proteins in B6D2 oocytes and zygotes using a label-free method [32]. The largest coverage of the MII-stage oocyte proteome was achieved using oocytes from the B6C3F1 mouse strain and fell in the range of 3700 proteins, albeit without quantification [33]. Thus, both in the present and in previous studies, a substantial part of the mRNA-expressed genome seems to be missing from the protein call. Using deep-sequencing on a Illumina HiSeq 2500 platform after preamplification of RNA material collected from a total of 500 MII oocytes per group, we were able to assign the RNAs to 15 205 genes as expressed simultaneously in 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes. For 1875 of these RNAs, a protein counterpart was detected across all four strains as well (Fig. 2B). The expression values of RNAs as obtained by deep sequencing are given in Supporting Information Table 2. The small overlap, that is, large discrepancy between detected proteome and detected transcriptome (Fig. 2B, SILAC-RNAseq intersection) suggests that oocytes might not contain as many proteins as we would expect from studies of cell lines [34]. More likely, however, the proteomic method for minute specimens, such as oocytes, needs further improvement to be able to detect additional proteins that are presently missing from the call.

While we may not draw any conclusion based on proteins that we were unable to detect, we may comparatively analyze oocytes of different inbred strains based on proteins that are

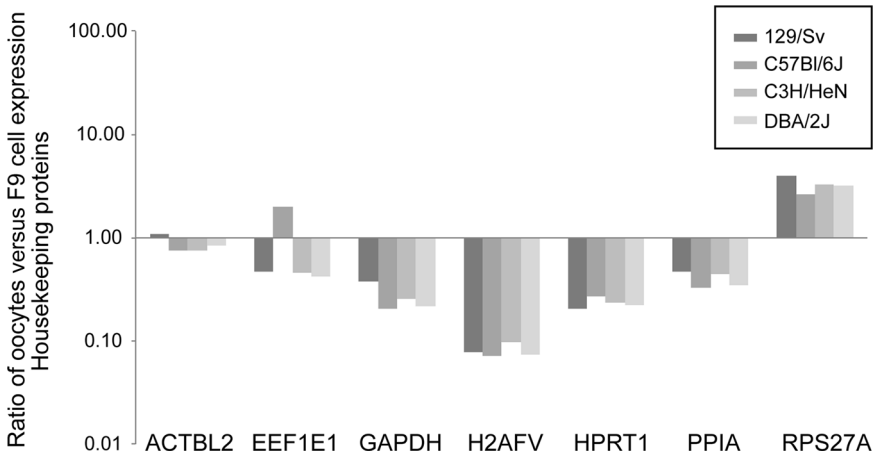

Figure 3. Graphical representation of the expression levels (L/H) of housekeeping proteins across the four mouse strains 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J. The gene names on the *x*-axis are given in alphabetical order. The *y*-axis of the chart is given in logarithmic scale.

detected simultaneously in all strains considered (and also in the F9 cells). The 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes share 2010 detected proteins; a cognate mRNA for 1875 of these was detected in the transcriptome. We calculated the Pearson's *r* correlation coefficient inside the set of 1875 pairs of protein and mRNA data (the SILAC-RNAseq intersection). L/H ratios for proteins and read counts for mRNAs are poorly correlated, with $r = 0.19$, 0.18, 0.20, and 0.18 for 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J oocytes, respectively. Although it is beyond the scope of this study to look for the reasons for the poor correlation, it may be noted that oocytes become transcriptionally silent shortly before ovulation [24], and they resume transcription after fertilization. So, correlation between mRNA and protein levels is necessarily lower in oocytes than in cells that are in a steady state of gene expression [35]. Given the low correlation coefficient, we focused subsequent analyses on the proteome that is closer to biological phenotype.

3.3 Overall abundance of oocyte proteins and interstrain variability

The central question that we address in this study is whether differences in embryo quality are associated, at least in part, with initial differences in oocyte composition. If this is the case, then we expect housekeeping proteins to be highly conserved in their abundance as opposed to other proteins with nonconstitutive roles. Gene expression data may be described by probability distributions; however, it is unlikely that there is a probability distribution for gene expression in general [36]. For this reason, we adopted a simple method of analysis that does not make any elaborate assumptions about protein abundances.

Since differences in protein abundance can be subtle when comparing oocytes with other oocytes (same cell type), we worked threshold-free (i.e., without setting an arbitrary threshold) using the CV to pursue the 50 most variable proteins among the 2010 proteins that are common to the four groups of oocytes (CV = SD/mean). The more variably a protein is expressed, the higher its CV. As shown in independent studies, basing the analysis on CV values protects against detecting patterns in variability influenced by trends in absolute expression alone [22]. As a safety measure, we excluded proteins with extreme H/L values from the CV analysis. To identify such extremes, Liao et al. created H/L mixtures of known ratio (1:1, 1:2, 1:5, 1:10, 1:25, 1:50, 1:100) using somatic cell lines, analyzed these mixtures with Orbitrap-based LS-MS/MS, and compared the measured ratios with the expected ratios [37]. The authors observed that the measured H/L ratios were as expected at ratios of 1:1, 1:2, 1:5, 1:10, and 1:25, started departing from the real value for ratios of 1:50, and were completely off for ratios of 1:100. Necessarily, the reciprocal ratios of 50:1 and 100:1 can be unreliable, too. Based on the results of Liao et al., we excluded 69 of 2010 oocytic proteins with H/L ratios lower than 1:25 (0.04) or higher than 25 from further analyses. Even though the existence of extreme H/L ratios is biologically plausible, they may not be measured reliably by the SILAC method.

We ranked proteins with H/L ratios satisfying $0.04 \leq H/L\ ratio \leq 25$ based on the CV of their H/L ratio across all four mouse strains, from high to low. As expected, proteins encoded by housekeeping genes exhibited a low variability across all four mouse strains, and thereby ranked in the bottom part of the CV list (typically, in the bottom 50%). We chose the housekeeping proteins based on a study of Mamo et al., who validated a robust set of housekeeping genes for studies of mouse oocytes and embryos [38]. A notable exception from the low variability was the eukaryotic translation elongation factor 1 epsilon 1 (Eef1e1), which presented a more than fourfold higher abundance in the mouse strain C57Bl/6J (Fig. 3; note the inverted H/L ratios in this figure).

We then examined the H/L ratios that were used to calculate the CVs to see how the H/L ratios of genes with high CV would compare to the ratios of genes with low CV. H/L ratios for housekeeping proteins ranged from 0.31 to 4.34

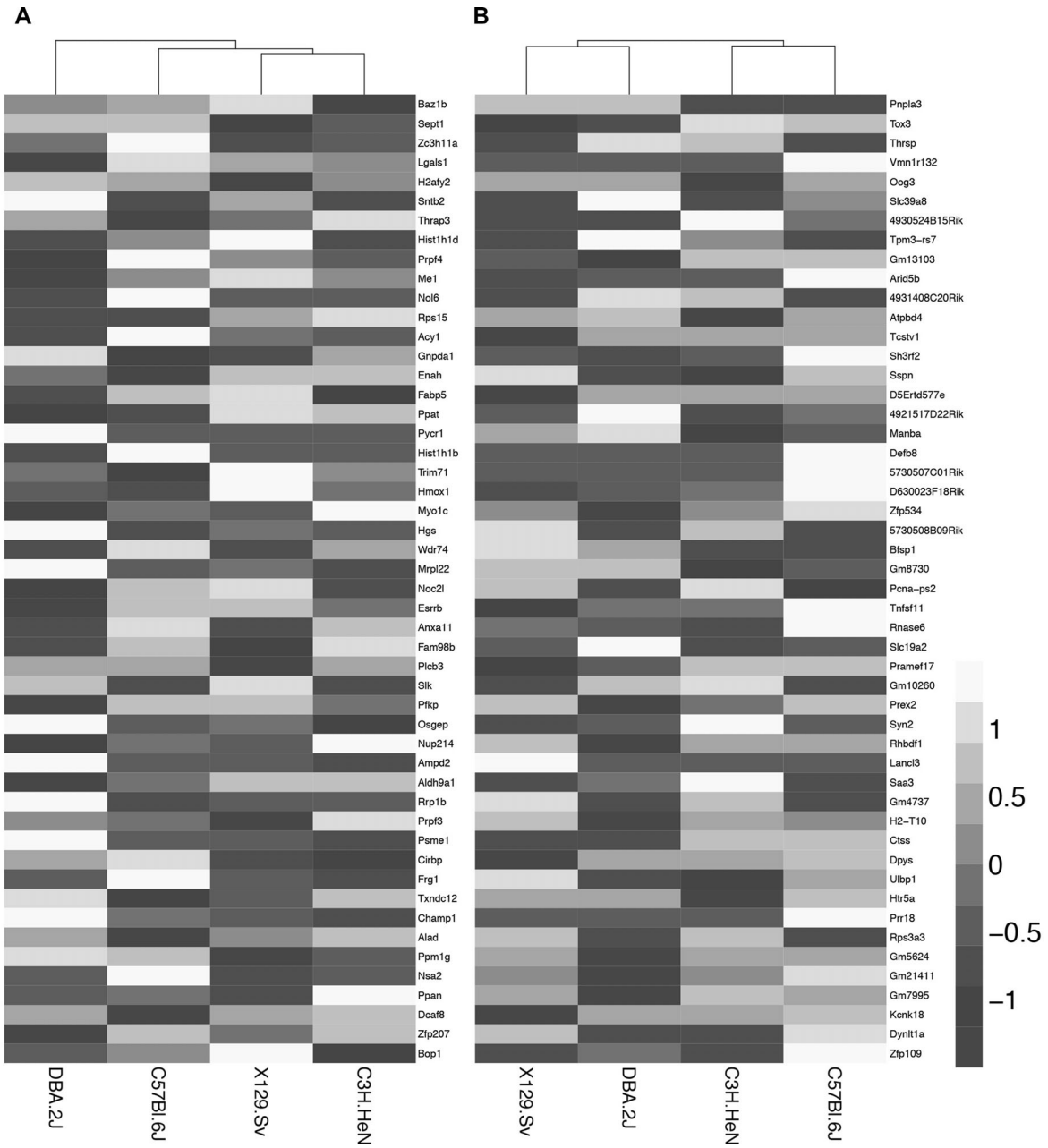

Figure 4. Heat map representation of the most variably expressed genes in oocytes across the four mouse strains 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J. Variability has been calculated using the CV, and the heat map entries are sorted from high to low CV. (A) Heat map of the 50 most variably expressed proteins (L/H). (B) Heat map of the 50 most variable expressed RNAs.

(H/L for Actlb2, 1.19; Ppia, 2.56; Hprt1, 4.34; Gapdh, 3.99; Ee1fe1, 1.80; Ubc, 0.31); the H/L ratios for the top 50 most variable proteins (Fig. 4A) ranged from 0.16 to 4.49 (mean 2.08). These H/L ratios are far from the possibly unreliable extremes that we discarded according to the observations of Liao et al. [37]. Thus, we observed substantial differences in the variability of protein abundances within comparable ranges of H/L ratios.

3.4 Reliability of SILAC LC-MS/MS abundance values

We validated our LC-MS/MS findings in situ by assessing the amount of protein present in oocytes using an independent method and additional 352 oocytes (106, 90, 67, and 89 oocytes from 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J mice, respectively). Just as one would validate RT-PCR,

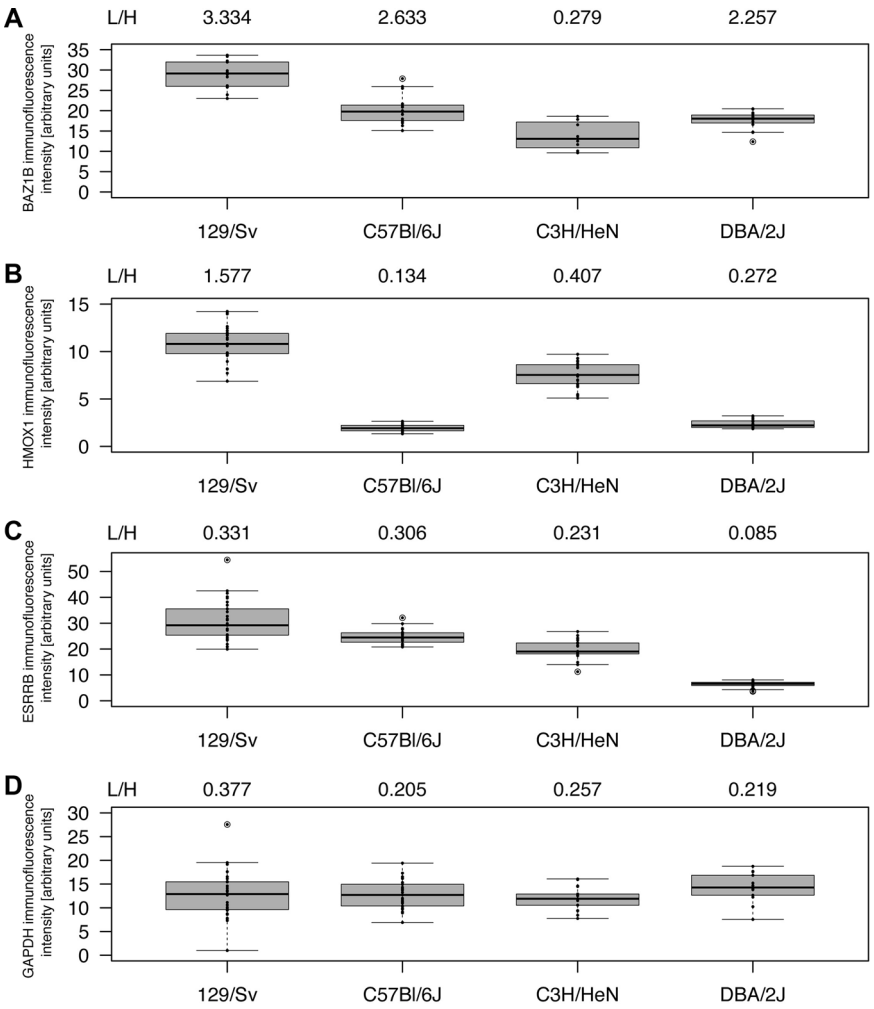

Figure 5. Box plots diagrams of the immunofluorescence signals of (A) BAZ1B, (B) HMOX1, (C) ESRRB, and (D) GAPDH compared with the L/H ratios of these proteins in the oocytes of the four inbred mouse strains 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J. The three horizontal lines of each box are the 25th, 50th, and 75th percentile of the values distribution, the whiskers indicate the minimum and maximum values. The *y*-axis of the diagrams provides the immunofluorescence intensity (background subtracted) of the oocytes. The L/H ratios of the proteins are shown above the box plots.

microarray, or RNA-seq data by riboprobe in situ hybridization, we chose to validate our protein data in situ by immunofluorescence followed by image quantitation.

We selected four proteins, three of which were picked from the top 50 proteins and the fourth from the housekeeping proteins. These proteins were BAZ1B, HMOX1, ESRRB (top-50), and GAPDH (housekeeping; see also Supporting Information Fig. 1). BAZ1B belongs to the *Modifiers of murine metastable epialleles* (*Mommes*), all of which show some degree of embryonic lethality when mutated in homozygosis [39]. HMOX1 is a marker of oocyte competence since a deficit of HMOX1 impairs ovulation as well as fertilization [40]. ESRRB is an embryonic transcription factor that can convert somatic cells to pluripotency [41]. GAPDH is a widely used housekeeping gene product [38].

We quantified the immunofluorescence signal of these proteins in the oocytes of the four strains (at least ten oocytes per strain), and then we compared it with their LC-MS/MS signals. To do so, we inverted the H/L ratios so that the oocyte value is the numerator and there is a direct instead of reciprocal ($1/x$) relationship to the fluorescence signal. Our established confocal microscopy pipeline [42] was used to measure and quantitate the signals of BAZ1B, ESRRB, HMOX1, and GAPDH above the background of the specimens that were not incubated with the primary antibody. Our data (Fig. 5) are in line with the proteomic analysis. The Pearson's correlation coefficients between the immunofluorescence signal and the L/H ratio approached the value of 1 (1 = perfect match) and were 0.873 (BAZ1B), 0.883 (HMOX1), and 0.989 (ESRRB). GAPDH was constant as expected, resulting in a correlation coefficient far from 1 (−0.311).

3.5 Functional enrichment analyses of the most differently expressed proteins in oocytes from different mouse strains

Next, we subjected the top 50 proteins of the CV-based rank (Fig. 4A) to a functional enrichment analysis (see Section 2). The most significantly enriched biological processes were related to chromatin organization, RNA processing, and ribosome biogenesis. In the GO domain "cellular component,"

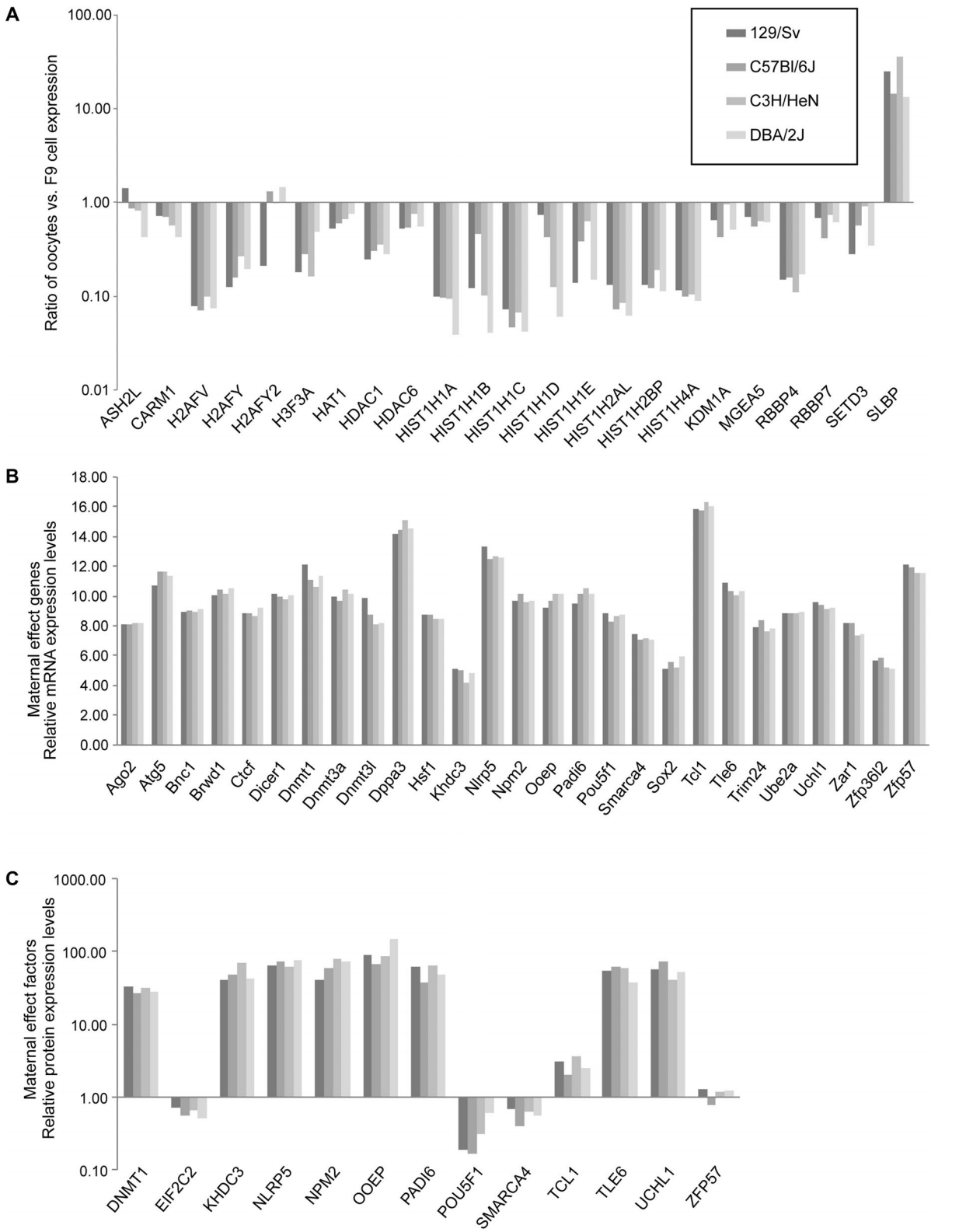

Figure 6. Expression levels of select genes of interest in oocytes of the four mouse strains 129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J. (A) Expression of histones and related genes on the protein level (L/H) across oocytes from all four strains. Expression of maternal effect genes on (B) the mRNA and (C) the protein level (L/H) across oocytes from all four strains. The gene names on the *x*-axis are given in alphabetical order. The *y*-axis of the charts (A) and (C) is given in logarithmic scale.

we observed enrichment in terms relating to chromatin, ribonucleoprotein complex, and nuclear and organelle lumen. As a functional enrichment analysis of the top 50 most variable RNAs (Fig. 4B) did not lead to any significant result, we extended the list to the top 100 most variable RNAs. These RNAs were enriched in biological processes related to signaling (transmembrane receptor protein tyrosine kinase and enzyme linked receptors), antigen processing and presentation, and, surprisingly for oocytes, male meiosis. Cellular component terms were enriched in membrane-related GOs (integral/intrinsic to plasma membrane) and in the terms lysosome, vacuole, and endoplasmatic reticulum. Therefore, the functional analyses of the most variant proteins and the most variant RNAs did not lead to the same results, as could be expected by the low correlation of the datasets exemplified by their Pearson's correlation coefficient described above.

Interestingly, the list of the 50 most variable proteins contains many entries linked to oocyte biology (Sept1, Rps15, Hmox1, and Gnpda1), embryo differentiation (Lgals1, Hgs, Esrrb, Wdr74, Baz1B), chromatin remodeling (Noc2l, Myo1c), and certain histones (H2afy2, Hist1h1d, and Hist1h1b). Further, the list includes proteins with a relation to RNA and especially RNA processing (Zc3h11a, Thrap3, Prpf4, Nol6, and Trim71) as well as proteins that play a role in general metabolism (Acy1, Fabp5, Ppat, Pycr1, and Mrpl22). A detailed description of the above-named proteins and their putative relation to embryology can be found in the discussion that is available as Supporting Information Material.

3.6 Histones and maternal effect genes in oocytes from different mouse strains

The fact that three histones (H2afy2, Hist1h1d, and Hist1h1b) were among the most variably expressed proteins prompted us to deepen the analysis of this family of proteins, which has recently contributed new members to the family of elusive factors that are in charge of nuclear reprogramming [43, 44]. Interestingly, other histones also presented variable expression levels across the four strains (Fig. 6A). The most variably expressed histones, H2afy2, Hist1h1d, and Hist1h1b, are of special interest in this context. H1 linker histones (Hist1h1d and Hist1h1b) are necessary for the condensation of nucleosome chains into higher order structures; embryos lacking multiple H1 variants die by midgestation [45]. H2afy2 is a macro-H2A variant that has been reported as enriched on the inactive X-chromosome in females [46] and for which a role in development has been proposed [47]. Further, macro-H2A variants have been shown to occupy pluripotency-related genes in fibroblasts (together with H2afy and H3K27me3) and to act as barriers upon reprogramming to pluripotency [48]. These functional observations suggest an impact of H2afy2 levels on embryonic development or lineage decision processes.

Last but not least, we specifically filtered our datasets for "maternal effect genes" that are crucial for embryonic genome activation and progression through the first cleavages and therefore may link to the observed developmental phenotypes described above. A total of 27 genes have been described in this category [7, 8] and the list is likely to grow in the future. Among them are transcription and chromatin remodeling factors, and modifiers of DNA methylation. All 27 factors were detected in the deep transcriptome of all four strains, although with seemingly similar expression levels (Fig. 6B). On the protein level, we were able to quantify 13 of the 27 described maternal effect genes in all four strains; like the mRNA levels, the majority of the protein levels were similar across strains (Fig. 6C). The notable exception was Oct4 (Pou5f1) that shows a fold difference of more than three between the highest (DBA/2J) and lowest expressing strain (C57Bl/6J). Oct4 is a transcription factor that is crucial for the maintenance of pluripotency and is expressed continuously in cells that comprise the germ line. It is interesting to note that the strains with lower Oct4 levels in their oocytes (129/Sv, C57Bl/6J; Fig. 6C) do not necessarily yield lower blastocyst rates after SCNT (Table 1), consistent with the dispensable role of maternal Oct4 in preimplantation mouse development [49].

4 Concluding remarks

In this study, we have shown that differences in oocyte composition precede differences in embryo quality when the sperm origin is the same and also when sperm is replaced using SCNT. We highlighted the different developmental phenotypes that are the result of different genomic backgrounds in mice (129/Sv, C57Bl/6J, C3H/HeN, and DBA/2J) and described strain-dependent differences in cell allocations at the blastocyst stage. Quantitative proteomic analysis of oocytes revealed that they differ strain-wise in the abundance of factors relevant to embryo biology, in a way that could not be predicted from transcriptome analysis. It is likely that additional proteins that we could not quantify (e.g., because they are not expressed in the reference F9 cells) are also differently expressed between the four mouse strains.

Our state-of-the-art quantitative proteomic catalogue can be of great value for future developmental studies in which one or more of the analyzed strains are used. The challenge for the near future will be to increase the coverage (at present, only part of the proteins expected to be there based on the cognate mRNA are indeed detected in the proteome), and to integrate the different layers of gene expression toward an understanding of the final result—the embryonic phenotype. Confirmation of proteomic results, possibly via independent methods, is also an issue. Deutsch et al. applied SRM to verify the abundance of selected proteins in lysates of small samples of bovine oocytes [50]. SRM is certainly a valid option, as is the in situ immunofluorescence approach we used here.

Studies to come will have to provide not only data on the oocyte itself, but also on the different stages of development. Then will it be possible to draw even stronger conclusions on the impact of the molecular makeup of the oocyte on cleavage

stage embryos and beyond. Such an understanding is also of paramount importance to the field of assisted reproductive technologies. Future efforts may allow us to generate such datasets and to finally integrate the different layers of gene expression (DNA regulation, RNA expression, translated protein) into a global and comprehensive description of the molecular processes that unfold during the first days of mammalian life. In the future, the ability to gather this type of information for mouse and human oocytes might enable predictions that would be of interest for human-assisted reproduction, for which the mouse is an experimental model.

The MS proteomics data in this paper have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [x]: dataset identifier PXD001059. We thank the Max Planck Institute for Molecular Biomedicine in Münster and its Director, Prof. Hans R. Schöler, for infrastructural support. We also thank Afroditi Nanou for help with the immunoconfocal microscopy analysis of the embryos, Dorota Buczek for help with the RNA extraction from oocytes, Terumi Horiuchi for help with the RNA sequencing and data handling. Support from the mouse housing facility and its animal caretakers is gratefully acknowledged. Amy Pavlak performed a language and grammar check on the final version of the manuscript. This study was supported by the priority program (Schwerpunktprogramm) no. 1356 of the Deutsche Forschungsgemeinschaft (DFG grants BO2540/3-2 and FU 583/2-2), by the European Framework Program 7 (FP7-People-2009-IRSES Project "EVOLGEN" No. 247633), and by the Max Planck Society.

The authors have declared no conflict of interest.

5 References

[1] Yamanaka, Y., Response: cell fate in the early mouse embryo—sorting out the influence of developmental history on lineage choice. *Reprod. Biomed. Online* 2011, *22*, 525–527; discussion 528.

[2] Wakayama, T., Yanagimachi, R., Mouse cloning with nucleus donor cells of different age and type. *Mol. Reprod. Dev.* 2001, *58*, 376–383.

[3] Gao, S., Czirr, E., Chung, Y. G., Han, Z., Latham, K. E., Genetic variation in oocyte phenotype revealed through parthenogenesis and cloning: correlation with differences in pronuclear epigenetic modification. *Biol. Reprod.* 2004, *70*, 1162–1170.

[4] Vogel, C., Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 2012, *13*, 227–232.

[5] Latham, K. E., Strain-specific differences in mouse oocytes and their contributions to epigenetic inheritance. *Development* 1994, *120*, 3419–3426.

[6] Engler, P., Haasch, D., Pinkert, C. A., Doglio, L. et al., A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell* 1991, *65*, 939–947.

[7] Li, L., Zheng, P., Dean, J., Maternal control of early mouse development. *Development* 2010, *137*, 859–870.

[8] Yurttas, P., Morency, E., Coonrod, S. A., Use of proteomics to identify highly abundant maternal factors that drive the egg-to-embryo transition. *Reproduction* 2010, *139*, 809–823.

[9] Latham, K. E., Sapienza, C., Localization of genes encoding egg modifiers of paternal genome function to mouse chromosomes one and two. *Development* 1998, *125*, 929–935.

[10] Richoux, V., Renard, J. P., Babinet, C., Synthesis and developmental regulation of an egg specific mouse protein translated from maternal mRNA. *Mol. Reprod. Dev.* 1991, *28*, 218–229.

[11] Schwarzer, C., Siatkowski, M., Pfeiffer, M. J., Baeumer, N. et al., Maternal age effect on mouse oocytes: new biological insight from proteomic analysis. *Reproduction* 2014, *148*, 55–72.

[12] Keane, T. M., Goodstadt, L., Danecek, P., White, M. A. et al., Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 2011, *477*, 289–294.

[13] Grubb, S. C., Bult, C. J., Bogue, M. A., Mouse phenome database. *Nucleic Acids Res.* 2014, *42*, D825–D834.

[14] Esteves, T. C., Balbach, S. T., Pfeiffer, M. J., Arauzo-Bravo, M. J. et al., Somatic cell nuclear reprogramming of mouse oocytes endures beyond reproductive decline. *Aging Cell* 2011, *10*, 80–95.

[15] Schwarzer, C., Esteves, T. C., Arauzo-Bravo, M. J., Le Gac, S. et al., ART culture conditions change the probability of mouse embryo gestation through defined cellular and molecular responses. *Hum. Reprod.* 2012, *27*, 2627–2640.

[16] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, *32*, 223–226.

[17] Kanematsu, S., Tanimoto, K., Suzuki, Y., Sugano, S., Screening for possible miRNA-mRNA associations in a colon cancer cell line. *Gene* 2014, *533*, 520–531.

[18] Trapnell, C., Pachter, L., Salzberg, S. L., TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, *25*, 1105–1111.

[19] Li, H., Handsaker, B., Wysoker, A., Fennell, T. et al., The sequence alignment/map format and SAMtools. *Bioinformatics* 2009, *25*, 2078–2079.

[20] Anders, S., Pyl, P. T., Huber, W., HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2014, DOI:10.1093/bioinformatics/btu638.

[21] Anders, S., Huber, W., Differential expression analysis for sequence count data. *Genome Biol.* 2010, *11*, R106.

[22] Mar, J. C., Matigian, N. A., Mackay-Sim, A., Mellick, G. D. et al., Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet.* 2011, *7*, e1002207.

[23] Huang da, W., Sherman, B. T., Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009, *4*, 44–57.

[24] Bouniol-Baly, C., Hamraoui, L., Guibert, J., Beaujean, N. et al., Differential transcriptional activity associated

with chromatin configuration in fully grown mouse germinal vesicle oocytes. *Biol. Reprod.* 1999, *60*, 580–587.

[25] Sun, L., Bertke, M. M., Champion, M. M., Zhu, G. et al., Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development. *Sci. Rep.* 2014, *4*, 4365.

[26] Graumann, J., Hubner, N. C., Kim, J. B., Ko, K. et al., Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* 2008, *7*, 672–683.

[27] Singhal, N., Graumann, J., Wu, G., Arauzo-Bravo, M. J. et al., Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* 2010, *141*, 943–955.

[28] Alonso, A., Breuer, B., Steuer, B., Fischer, J., The F9-EC cell line as a model for the analysis of differentiation. *Int. J. Dev. Biol.* 1991, *35*, 389–397.

[29] Zhang, X., Zhang, J., Wang, T., Esteban, M. A., Pei, D., Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J. Biol. Chem.* 2008, *283*, 35825–35833.

[30] Ma, M., Guo, X., Wang, F., Zhao, C. et al., Protein expression profile of the mouse metaphase-II oocyte. *J. Proteome Res.* 2008, *7*, 4821–4830.

[31] Zhang, P., Ni, X., Guo, Y., Guo, X. et al., Proteomic-based identification of maternal proteins in mature mouse oocytes. *BMC Genomics* 2009, *10*, 348.

[32] Wang, S., Kou, Z., Jing, Z., Zhang, Y. et al., Proteome of mouse oocytes at different developmental stages. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 17639–17644.

[33] Pfeiffer, M. J., Siatkowski, M., Paudel, Y., Balbach, S. T. et al., Proteomic analysis of mouse oocytes reveals 28 candidate factors of the "reprogrammome." *J. Proteome Res.* 2011, *10*, 2140–2153.

[34] Geiger, T., Wehner, A., Schaab, C., Cox, J., Mann, M., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* 2012, *11*, M111 014050.

[35] Schwanhausser, B., Busse, D., Li, N., Dittmar, G. et al., Global quantification of mammalian gene expression control. *Nature* 2011, *473*, 337–342.

[36] Thomas, R., de la Torre, L., Chang, X., Mehrotra, S., Validation and characterization of DNA microarray gene expression data distribution and associated moments. *BMC Bioinform.* 2010, *11*, 576.

[37] Liao, Z., Wan, Y., Thomas, S. N., Yang, A. J., IsoQuant: a software tool for stable isotope labeling by amino acids in cell culture-based mass spectrometry quantitation. *Anal. Chem.* 2012, *84*, 4535–4543.

[38] Mamo, S., Gal, A. B., Bodo, S., Dinnyes, A., Quantitative evaluation and selection of reference genes in mouse oocytes and embryos cultured in vivo and in vitro. *BMC Dev. Biol.* 2007, *7*, 14.

[39] Daxinger, L., Harten, S. K., Oey, H., Epp, T. et al., An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biol.* 2013, *14*, R96.

[40] Zenclussen, M. L., Jensen, F., Rebelo, S., El-Mousleh, T. et al., Heme oxygenase-1 expression in the ovary dictates a proper oocyte ovulation, fertilization, and corpora lutea maintenance. *Am. J. Reprod. Immunol.* 2012, *67*, 376–382.

[41] Feng, B., Jiang, J., Kraus, P., Ng, J. H. et al., Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat. Cell Biol.* 2009, *11*, 197–203.

[42] Cavaleri, F. M., Balbach, S. T., Gentile, L., Jauch, A. et al., Subsets of cloned mouse embryos and their non-random relationship to development and nuclear reprogramming. *Mech. Dev.* 2008, *125*, 153–166.

[43] Wen, D., Banaszynski, L. A., Liu, Y., Geng, F. et al., Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. *Proc. Natl. Acad. Sci. USA* 2014, *111*, 7325–7330.

[44] Shinagawa, T., Takagi, T., Tsukamoto, D., Tomaru, C. et al., Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. *Cell Stem Cell* 2014, *14*, 217–227.

[45] Fan, Y., Nikitina, T., Morin-Kensicki, E. M., Zhao, J. et al., H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol. Cell. Biol.* 2003, *23*, 4559–4572.

[46] Chadwick, B. P., Willard, H. F., Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum. Mol. Genet.* 2001, *10*, 1101–1113.

[47] Costanzi, C., Pehrson, J. R., MACROH2A2, a new member of the MARCOH2A core histone family. *J. Biol. Chem.* 2001, *276*, 21776–21784.

[48] Gaspar-Maia, A., Qadeer, Z. A., Hasson, D., Ratnakumar, K. et al., MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nat. Commun.* 2013, *4*, 1565.

[49] Wu, G., Han, D., Gong, Y., Sebastiano, V. et al., Establishment of totipotency does not depend on Oct4A. *Nat. Cell Biol.* 2013, *15*, 1089–1097.

[50] Deutsch, D. R., Frohlich, T., Otte, K. A., Beck, A. et al., Stage-specific proteome signatures in early bovine embryo development. *J. Proteome Res.* 2014, *13*, 4363–4376.

[51] Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A. et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, *41*, D1063–D1069.

