OPINION

Sperm RNA code programmes the metabolic health of offspring

Yunfang Zhang, Junchao Shi, Minoo Rassoulzadegan, Francesca Tuorto and Qi Chen

Abstract | Mammalian sperm RNA is increasingly recognized as an additional source of paternal hereditary information beyond DNA. Environmental inputs, including an unhealthy diet, mental stresses and toxin exposure, can reshape the sperm RNA signature and induce offspring phenotypes that relate to paternal environmental stressors. Our understanding of the categories of sperm RNAs (such as tRNA-derived small RNAs, microRNAs, ribosomal RNA-derived small RNAs and long non-coding RNAs) and associated RNA modifications is expanding and has begun to reveal the functional diversity and information capacity of these molecules. However, the coding mechanism endowed by sperm RNA structures and by RNA interactions with DNA and other epigenetic factors remains unknown. How sperm RNA-encoded information is decoded in early embryos to control offspring phenotypes also remains unclear. Complete deciphering of the 'sperm RNA code' with regard to metabolic control could move the field towards translational applications and precision medicine, and this may lead to prevention of intergenerational transmission of obesity and type 2 diabetes mellitus susceptibility.

Perhaps one of the most undisputed facts in biology is that heritable traits are coded in DNA sequences. However, DNA sequence-based inheritance alone is not sufficient to explain heritable diseases (see BOX 1 for definitions of common terms) that cannot be pinpointed to mutations and/or variations in the DNA sequence, and this accounts for a number of diseases associated with modern lifestyles (including obesity and type 2 diabetes mellitus). It is especially difficult to explain the phenomenon that certain life experiences can induce phenotypic alterations in immediate offspring and sometimes across multiple generations in mammals[1–3] (BOX 2). This 'missing heritability' problem[4] has stimulated a wave of research that aims to explore the epigenetic mechanisms that could be responsible for intergenerational and transgenerational transmission of certain disease susceptibilities or phenotypes in a non-DNA sequence-based manner and to address how such epigenetic information is altered by environmental factors and further transmitted via the germline between generations, as extensively reviewed in recent papers[5–9]. This so-called intergenerational epigenetic inheritance or transgenerational epigenetic inheritance is a tantalizing research direction that has the potential to revolutionize our understanding of the aetiology of many human diseases that originate from environmental factors. In addition to well-known human population-based studies indicating that maternal malnutrition and/or over-nutrition during periconception and the postnatal family environment will lead to metabolic disease in the offspring (such as the Dutch famine effect)[10–12], emerging research in animal models from the last decade has further demonstrated that intergenerational and transgenerational transmission of metabolic disorders can originate from a variety of both maternal and paternal environmental factors, including diet, chemical toxicants, endocrine disruptors and mental stressors[1–3]. These findings might generate widespread medical and social concerns for human health, as similar environmental factors could damage the metabolic health of offspring.

The processes by which environmental information is coded and transmitted inter-generationally via the germline remains unclear. This lack of clarity is partly due to the extensive epigenetic reprogramming process in mammals, which erases most well-known epigenetic modifications such as DNA methylation[13,14] in the preimplantation embryo and during germline development. Other possibilities, such as histone retention in the process of histone-to-protamine replacement and the role of histone modifications[15–17], might contribute to storage of mammalian epigenetic memory from environmental input.

Many correlative studies in mammals (including humans) have suggested that various epigenetic factors, such as DNA methylation, histone modification and RNA, in germ cells (both sperm and egg) could contribute to intergenerational inheritance of environment-induced phenotypes. Moreover, causal evidence has been established in the past 5 years from injection of sperm RNAs (total RNA or a subset of sperm RNAs) from males exposed to various environmental stimuli (such as mental stressors or an unhealthy diet) into healthy zygotes. A number of studies from multiple groups have shown that the injection of sperm RNAs induces offspring phenotypes that fully or partially recapitulate the paternal environmental input, including behaviour changes, obesity and altered glucose metabolism[18–22]. The unique signature of mammalian sperm RNAs is developmentally and spatially organized and is controlled by genetic and environmental factors[8,22]. The RNA sequences are decorated by various RNA modifications that have been suggested to form a 'sperm RNA code' to programme specific offspring phenotypes during embryonic development[22]. This process is probably achieved by RNA sequence-specific and/or structure-specific interactions with other RNAs or epigenetic, transcriptional and translational mechanisms in the sperm and early embryo, and these processes should be considered as a direction for future research.

In this Perspectives, we focus on mammalian sperm RNA as an emerging

hereditary information carrier that regulates the metabolic health of offspring and discuss the evidence and logic behind the concept of the sperm RNA code with regard to its composition, regulation and information capacity. We also discuss how the RNA code in sperm might be decoded in the early embryo and transformed into other molecular signals to influence embryonic development and offspring phenotypes. In the face of the current obesogenic environment and the pandemic of diabetes mellitus, a comprehensive understanding of the sperm RNA code and its functional specificity could set the stage for translational applications in precision medicine that aim to prevent these metabolic disorders.

The making of the sperm RNA code

Before it is possible to establish the concept of the 'sperm RNA code', and to fully decipher its function in sensing the paternal environment and the programming of offspring phenotypes, several levels of evidence are needed. First, we need evidence that shows that the composition and structure of sperm RNAs are regulated by environmental or genetic signals in response to paternal exposure. Second, we need evidence that supports the mediation of offspring phenotypes either by the compositional signature or the stoichiometry of sperm RNA subpopulations. Third, we need to identify the functional diversity and specificity of sperm RNAs, which might be based on the compositional, stoichiometric and/or amplification effects of sperm RNA, with defined working mechanisms. So far, the body of evidence is not complete, but emerging data have begun to support the overarching hypothesis by demonstrating the role of the sperm RNA code in programming offspring phenotypes based on the signature of sperm RNA expression and RNA modification profiles.

Diverse types of sperm RNA. Although sperm RNAs were once considered simple remnant products of spermatogenesis, emerging data from multiple groups provide compelling evidence that they are in fact dynamically regulated during development and are spatially compartmentalized at the mature stage[23–25] (BOX 3).

According to previous RNA-sequencing data, the small RNA population in mature mouse sperm is dominated by tRNA-derived small RNAs (tsRNAs)[18,24,26], and a smaller population of microRNAs (miRNAs) and Piwi-interacting RNAs, as well as an appreciable amount of ribosomal RNA (rRNA)-derived small RNAs (rsRNAs) that were only systematically discovered following the advent of improved bioinformatic pipelines in the past 2 years[22,27,28]. Mature sperm also contains various populations of large RNAs, including mRNAs, long non-coding RNAs and circular RNAs, and the list of sperm RNA subtypes could continue to grow with improved sequencing methods and analytical power.

The composition of the sperm RNA population could differ slightly in other mammals[29,30], but among the known mammalian species, the composition of RNA in mature sperm are all distinct from that in somatic cells or spermatogenic cells found in the testes, which suggests that RNA in mature sperm has unique functions. The potential interactions between different RNA species in sperm could synergistically generate the level of complexity needed to achieve functional diversity and specificity in regulating complex biological processes[31], such as optimizing early embryonic development[32,33] or transmitting paternal traits to offspring[18–21]. We discuss these aspects in more detail later in the article.

RNA modifications increase complexity. In addition to the sequence diversity of sperm RNAs, various RNA modifications in sperm RNAs were identified in 2016, and identification of these modifications

Box 1 | **Glossary terms**

- Heritable diseases: diseases inherited from one generation to another through genetic or epigenetic regulation.
- Missing heritability: the phenomenon that single genetic variations cannot account for much of the heritability of diseases, behaviours and other phenotypes.
- Intergenerational or transgenerational epigenetic inheritance: the transmission of heritable information across two (F0 to F1) or more than two generations (F0 to F2 and beyond) without alteration of DNA sequence.
- Disease susceptibilities: conditions of the body that tend to make the individual more than usually susceptible to certain diseases.
- Precision medicine: a personalized medicine tailoring of medical treatment to the individual characteristics of each patient.
- Spermatogenesis: the biogenesis of haploid spermatozoa from germ cells in the seminiferous tubules of the testes.
- Epididymis: a curved structure at the back of the testicle where sperm maturation takes place that can be divided into three main sections: the proximal (caput), middle (corpus) and distal (cauda) parts.
- Extracellular vesicles: cell-derived membrane-enclosed vesicles that contain proteins, lipids and nucleic acids.
- Queuosine: a hypermodified 7-deazaguanosine nucleoside located in the anticodon wobble position of certain tRNAs, which cannot be synthesized de novo in eukaryotes.
- RNA-editing events: the insertion, deletion and base substitution of nucleotides within the edited RNA molecule.
- RNA G-quadruplexes: RNA helical structures formed by Hoogsteen hydrogen-bonded guanine tetrads that involved by one, two or four RNA strands.
- RNA-dependent RNA polymerase: an enzyme that catalyses the replication of RNA from the RNA template.
- Histone marks: the post-translational modification of histone proteins, including methylation, phosphorylation, acetylation, ubiquitylation and sumoylation.
- Haploinsufficiency: a diploid organism that has lost one copy of a gene and is left with a single functional copy of that gene.
- Transposon elements: DNA sequences that can change their locations within a genome, including retrotransposon and DNA transposon.
- Parthenogenetic embryo: an embryo that developed from an oocyte without fertilization.
- Ribosome heterogeneity: ribosomal subpopulations that preferentially translate distinct sub-pools of mRNAs to exhibit a functional specificity.
- Symmetry breaking: the transition process from symmetry towards asymmetry by small fluctuations occurring in the system.
- Third-generation RNA sequencing: characterized as single-molecule sequencing, which is able to determine both the sequence and modifications of the RNA molecules and avoid amplification biases.
- Machine learning: the process by which a computer learns from data and adjusts to new inputs to recognize patterns and to accomplish specific tasks.
- Watson–Crick base-pairing interactions: the most basic nucleotide recognition principle of the base pair, where guanine (G) is bound to cytosine (C), and adenine (A) is bound to thymine (T) or uracil (U) by hydrogen bonds that permit the formation of double-stranded helices in DNA and RNA.

has revealed new dimensions of complexity regarding RNA structural and functional diversity[18,22]. High-throughput quantitative approaches powered by liquid chromatography–tandem mass spectrometry (LC-MS/MS) have revealed that different sperm RNA fractions contain distinct profiles of RNA modifications. For example, the 30–40 nucleotide RNA fraction, which is dominated by tsRNAs and rsRNAs, contains markedly more RNA modifications than the 15–25 nucleotide RNA fraction, which is enriched in miRNAs[18,22]. The differences in the level of RNA modification could reflect the fact that tsRNAs and rsRNAs are derived from tRNA and rRNA precursors, respectively, which are known to harbour a plethora of RNA modifications, many of whose biological functions are just beginning to be revealed[34,35].

RNA modifications have been reported to increase overall sperm RNA stability and thus might prolong their function after entry into the oocyte after fertilization[18]. This effect could be related to the influence of RNA structures mediated by specific RNA modifications. A 2018 study revealed that adding or removing 5-methylcytosine (m^5C) at a specific site on tsRNA can alter its secondary structure, changing the resistance of tsRNA to RNase digestion. This change in RNA structure is also correlated with altered biological functions that differentially affect cell transcriptomic profiles[22]. In another study, pseudouridine (Ψ, an isomer of the nucleoside uridine (U)) on tsRNAs was found to contribute to the binding potential between tsRNAs and specific proteins in the translational initiation complex, and thus to affect global translational efficiency[36]. Some RNA modifications, such as *N*1-methyladenosine (m^1A), can block Watson–Crick base-pairing interactions and generate unexpected secondary RNA structures during folding[37].

Importantly, m^5C, m^1A and Ψ have been found to be enriched in the sperm tsRNA fraction and to be present in other fractions of sperm RNAs[18,22]. The site-specific distribution of these modifications on sperm RNAs is expected to not only alter RNA structures and RNA–RNA interactions, but also RNA binding potential to DNA and proteins, which could in turn increase the information capacity (for example, in regulating metabolism) of sperm RNAs beyond primary sequences. Considering that one tsRNA can carry multiple site-specific RNA modifications, it is notable that to date no study has isolated a single tsRNA or other small RNA (either from sperm or other tissue) to thoroughly study its whole RNA modification profile. In addition, different micro-species of a single tsRNA with different modification levels could exist[31]. This newly appreciated complexity of sperm RNA modification suggests that our understanding of the nature of sperm RNA remains preliminary; however, this vast complexity might serve as the very source of regulatory information that generates the specificity of the sperm RNA code.

Dynamic environmental and genetic regulation. Evidence suggests that both RNA profiles and RNA modifications in mammalian sperm are sensitive to the paternal environment. A wide range of environmental factors, including paternal diet, mental stress, chemical exposure, exercise and alcoholism, can change the composition of sperm RNAs[8,38,39], including miRNAs, tsRNAs and rsRNAs. In addition, paternal high-fat diet (HFD) increases RNA modifications such as m^5C, m^1A and *N*2-methylguanosine (m^2G) in the 30–40 nucleotide mouse sperm RNA fraction, which is enriched in tsRNAs and rsRNAs[18,22].

HFD in mice also elevates the expression of a tRNA methyltransferase, *Dnmt2*, in the caput epididymis[22], which is the region where sperm undergo maturation and can acquire their small RNA content[25,26]. Deletion of *Dnmt2* in mice abolishes the ability of sperm RNA to transfer HFD-induced paternal metabolic disorders to offspring. This change could be due in part to the influence of RNA modification (m^5C or m^2G) in the 30–40 nucleotide sperm RNA fraction under an HFD and to alterations in sperm small RNA expression profiles[22].

In another mouse model, it was found that genetic deletion of *Kit* at one allele results in accumulation in the sperm of abnormal transcripts from the *Kit* locus, an effect that is associated with the murine 'white-tail phenotype', which is an indicator of *Kit* gene disruption[40]. The injection of heterozygotic sperm RNAs that contain the abnormal *Kit* RNA fraction into zygotes results in wild-type offspring that express the white-tail phenotype in mice[40], whereas injection of sperm RNAs from DNMT2-negative *Kit* heterozygotes does not induce such a phenotype[41].

Moreover, in an insect (red flour beetles) model of transgenerational immune priming, in which paternal exposure to a non-lethal dose of bacteria protects the offspring against a potentially lethal dose of the same pathogen, it was found that paternal lack of DNMT2 attenuates the transgenerational immune priming effect and increases offspring susceptibility to bacterial infection[42]. This converging evidence suggests that DNMT2 might be critical in establishing sperm RNA-mediated offspring phenotypes in multiple models across species, probably due to its evolutionary conservation[43].

Mechanistically, the function of DNMT2 has now been linked with tsRNA and rsRNA biogenesis in sperm. DNMT2 is known to catalyse m^5C at position C38 of a few types of tRNA, increasing their stability. The loss of C38 m^5C due to *Dnmt2* deletion makes the tRNAs unstable and inclined to be cleaved at the anticodon region, generating more tsRNAs[43,44]. This effect has been found to be highly conserved in flies[45], as well as in testes and mature sperm in mouse models with *Dnmt2* knockout[45,46]. *Dnmt2* deletion alters sperm tsRNA composition, and this might explain the disruption of sperm RNA-mediated intergenerational transmission of phenotypes in different models. Interestingly, deletion of DNMT2

Box 2 | Inheritance of complex traits from ancestral environmental stressors

The coding of complex traits, such as metabolic disorders, stress behaviour, drug addiction and thermal adaptation, is believed to go beyond simple protein-coding DNA sequences. Instead, coding is proposed to involve multiple layers of codes such as DNA methylation, histone marks and RNAs that synergistically dictate the expression of series of genes with spatiotemporal precision. While the inheritance of these complex traits might have a genetic basis[4], certain traits have also been found to be modified by parental environmental input and can be intergenerationally, or transgenerationally (more rare in mammals than in yeast, flies and worms), inherited by offspring in a process that most likely involves epigenetic mechanisms. In addition to effects such as adverse metabolic health induced by both unhealthy parental diets[1–3] and toxicants (such as arsenic[95]), the inheritance of some environmentally induced traits can be very specific, such as sensitivity to particular odours[96]. A recent study in 2018 also showed that cold exposure of males before mating results in enhanced metabolism in male offspring that protects them from diet-induced obesity[97], suggesting an adaptive nature. In another study, inheritance of cocaine-seeking behaviour in rats was elicited by voluntary paternal cocaine administration, but not by high paternal intake of cocaine itself[98]. These results suggest very specific intergenerational transmission of brain-elicited changes beyond simple drug exposure, which may involve (heritable) information flow from somatic to germ cells.

in mice also downregulates sperm rsRNA biogenesis[22], most probably independently of its RNA methylation function, as deletion of *Dnmt2* does not cause hypomethylation of m^5C on rRNAs[47]. Of note, however, the underlying mechanism by which DNMT2 regulates rsRNAs, remains unknown and further investigation is warranted.

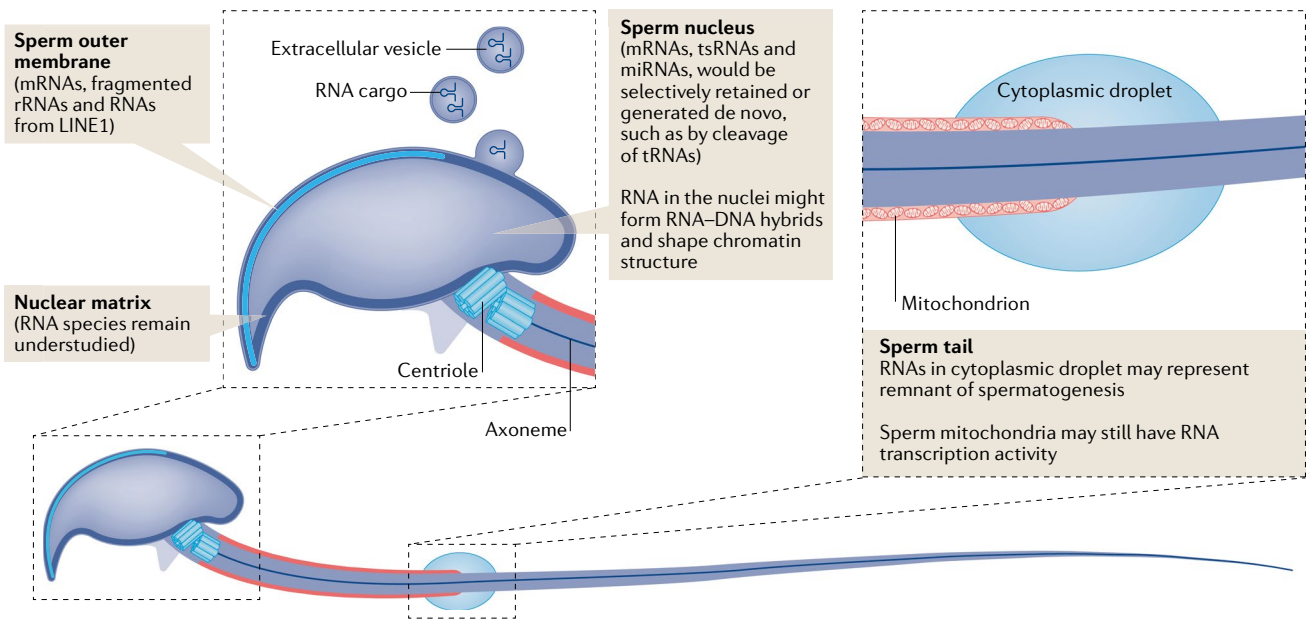
Interestingly, queuine, which is a nutritional component derived from food or microbiome, is required as a substrate for tRNA queuosine modification, and the levels of queuosine-modified tRNA (Q-tRNA) promote DNMT2-mediated m^5C on tRNA[48,49]. Indeed, nutritional deprivation of queuine in germ-free mice and human cancer cells results in abnormal tRNA-Asp methylation (m^5C) and protein biogenesis due to altered translational speed[48]. In addition, queuine depletion has been shown to increase the expression level of tsRNA upon cellular stress[50], which might also relate to m^5C deficiency-induced tsRNA accumulation. The increase in the expression level of tsRNAs might further control translational machinery in a multitude of ways[51]. In addition, the mouse microbiome can strongly affect host *N*6-methyladenosine (m^6A) mRNA modification (m^6A is an abundant modification in mRNA that regulates multiple aspects of mRNA metabolism[34]) in a tissue-specific manner, by altering the expression of m^6A writer and eraser proteins[52]. These intriguing lines of evidence suggest the existence of a complex interplay that links food intake and the gut microbiome to RNA modifications. These data suggest that this interplay might also be involved in regulating the composition of sperm tsRNA, although further evidence to confirm this hypothesis is awaited.

In addition to DNMT2, another enzyme related to RNA-editing events, APOBEC1, has been reported to contribute to transgenerational inheritance of susceptibility to testicular germ-cell tumours in mice[53]. The contribution of *Apobec1* to transgenerational inheritance might be related to altered sperm RNA profiles and RNA structures. The list of enzymes that might affect the sperm RNA signature is expected to expand in the coming years, especially with regard to enzymes associated with RNA modifications (such as NSUN2 and PUS7) and tRNA and/or rRNA metabolism[34]. Understanding how these enzymes are regulated by various types of paternal environmental exposure could hold the key to further deciphering the environmental–genetic interplay that shapes the sperm RNA code.

Functional specificity for embryonic development and offspring phenotypes. If the combination of sperm RNA and RNA modifications represents a code, then there must be biological functions that it codes for. While some controversy still exists regarding the function of some individual

Box 3 | **Developmental origin and compartmentalization of sperm RNAs**

Mature sperm are uniquely structured with a condensed nucleus covered by minimal cytoplasm and a long mitochondria-containing tail. The compartmentalization of various RNA species in mature sperm is closely associated with their developmental origins and biogenesis (see the figure). RNAs in sperm could be selectively retained from spermatogenesis. It has been shown that most large RNAs (>200 nucleotides) are enriched in, or attached with the sperm outer membrane, including mRNAs, fragmented ribosomal RNAs (rRNAs) and RNAs derived from repeat sequences such as LINEs[23]. Removal of the outer membrane with detergent results in an approximately two-thirds reduction in total sperm RNA[23]. On the other hand, some RNA species, such as tRNA-derived small RNAs (tsRNAs) and microRNAs (miRNAs), can be deeply imbedded in the sperm nucleus, since RNA sequencing of detergent-treated sperm heads has shown the clear presence of tsRNAs and miRNAs[24,99]. In addition, certain RNAs can hybridize with nuclear DNA by forming highly structured RNA–DNA complexes, and they could be more difficult to isolate during RNA extraction. These RNAs might be involved in chromatin repackaging during spermatogenesis[100] or in protamine–histone exchange at fertilization. The sperm tail also contains tsRNAs, miRNAs and relatively more Piwi-interacting (piRNAs) than the sperm head[25]. Extra piRNAs in the tail might be derived from the residual cytoplasmic droplet as remnant of spermatogenesis. While mature sperm are generally believed to be transcriptionally silent, mitochondria could retain certain transcriptional activities[101]; some RNAs such as tsRNAs and rRNA-derived small RNAs, could be generated by de novo cleavage from precursor tRNAs or rRNAs and to fine-tune the sperm RNA reservoir[22]. In addition, sperm might also acquire RNA from extracellular vesicles from the male or female reproductive tract via membrane fusion[25,102], which would support Charles Darwin's Pangenesis hypothesis (1868) as a way to enable information flow from somatic to germ cells[103] and to transfer environmental information (such as changes to diet[26]) to germ cells.



sperm RNAs[54,55], experiments involving enzymatic removal of sperm RNAs[32] and genetic disruption of the sperm small RNA biogenesis pathway[33] have demonstrated that overall disruption of sperm RNA integrity can cause abnormal early embryonic development.

Sperm have been found to show dynamic changes in the composition of small RNAs with tsRNA fractions increasing during the transition from the testis to the caput (proximal) epididymis and to the cauda (distal) epididymis[24,25]. A recent study in mice using intracytoplasmic sperm injection (ICSI) of sperm extracted from different segments of the epididymis (caput versus cauda) showed that immature caput sperm cannot generate viable pups owing to abnormal embryonic development. The authors showed that embryo lethality could be rescued by supplementation with 18–26 nucleotide RNAs (mostly miRNAs with some tsRNAs and rsRNAs) extracted from cauda epididymal extracellular vesicles[56], which suggests a stage-specific function of RNAs during sperm maturation. Another study, however, has shown that ICSI of caput sperm can generate viable pups[57]. The discrepancies in these findings might be the result of different experimental procedures, and further clarification is awaited.

In addition to the overall effect of sperm RNAs on embryonic development, sperm RNAs can carry environmental information and transmit certain phenotypes from paternal environmental exposure to offspring. For example, zygotic injection of sperm total RNA extracted from male mice exposed to HFD[18] or a high-fat, high-sugar diet (HFSD)[20] induces offspring phenotypes that partially or fully mimic the paternal metabolic phenotype. In the HFD exposure mouse model, further examination of the functional sperm RNA fraction through separate injections of different sizes of sperm RNAs (15–25 nucleotides, 30–40 nucleotides and >40 nucleotides) revealed that only injection of the 30–40 nucleotide RNA fraction could efficiently mimic injection of total sperm RNA; that is, to transmit impaired glucose tolerance but not insulin resistance[18], while in the HFSD model, injection of total sperm RNA induced both impaired glucose tolerance and insulin resistance[20]. In addition to the potential discrepancies in experimental procedures, the different metabolic phenotypes induced by sperm RNA injection might represent differential dietary effects on sperm RNA composition, as HFSD is known to induce more severe metabolic disorder than HFD[58]. Additionally, in a recent 2019 study using a maternal HFD (MHFD) in mice, it was found that zygotic injection of tsRNA-enriched sperm RNA fraction (30–34 nucleotides) but not other RNA fractions (40–45 or 70–79 nucleotides) extracted from the F1 male (the immediate offspring from the MHFD mother), induced enhanced hedonic behaviours (such as overconsumption of palatable food and alcohol preference) and a metabolic phenotype (obesity and glucose metabolism) in the resultant F2 generation[59]. Together, these findings suggest functional specificity of the sperm RNA code, in which different environmental inputs can determine different outcomes, and that environmental information can be encoded in a specific fraction of sperm RNAs, especially in tsRNAs (FIG. 1).

Another line of evidence in mice has shown that zygotic injection of the sperm total RNA extracted from mentally stressed fathers (unpredictable maternal separation combined with maternal stress) can induce behavioural and metabolic alterations reminiscent of the fathers in the F1/F2 mice offspring[21]. However, compared with total RNA injection, further separation and injection of fractions of either long (>200 nucleotides) or short (<200 nucleotides) sperm RNAs in this stress model induced only sub-phenotypes. More specifically, injection of long RNAs affected food intake, the glucose response to insulin and risk-taking tendencies, while the small RNA fraction affected body weight and behavioural despair during the forced swim test[19]. Importantly, unlike intact RNAs, fragmented long RNAs (fragmented by ultrasonication) do not induce embryonic transcriptome changes after zygotic injection[19], which suggests that the specific

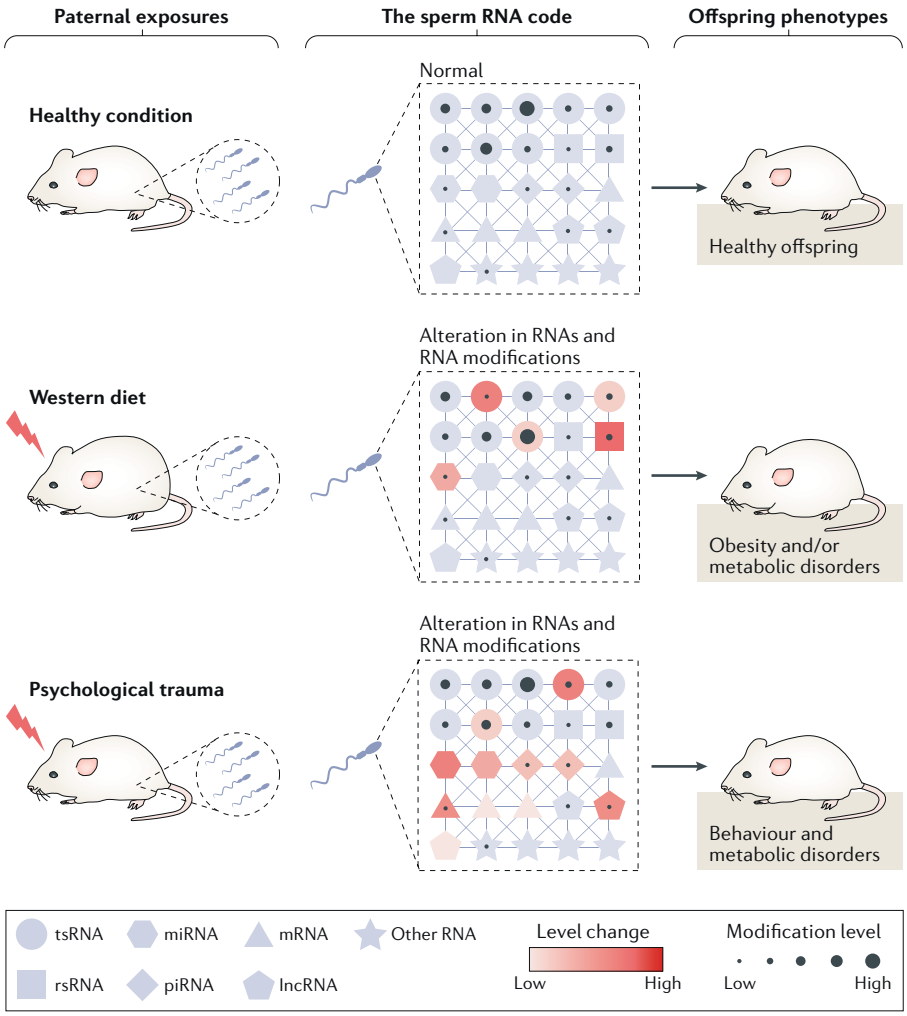


Fig. 1 | **Information capacity and functional specificity of the sperm RNA code.** Schematic representation of the sperm RNA code (an interconnected combination of RNA expression and RNA modification profile) that reflects paternal environmental exposure and mediates the transmission of specific phenotypes to the offspring. Different paternal information can be encoded in specific subsets of sperm RNA fractions. lncRNA, long non-coding RNA; miRNA, microRNA; piRNA, Piwi-interacting RNA; rsRNA, ribosomal RNA-derived small RNA; tsRNA, tRNA-derived small RNA.

sequence signature is key to inducing the offspring phenotype. These data further demonstrate the functional specificity of different fractions of sperm RNAs and support the existence of the sperm RNA code; however, the critical regulator of sperm RNA signature under each specific environmental condition remains to be discovered.

Decoding the sperm RNA code

Interplay with DNA methylation, histone marks and transposon elements. If the sperm RNA code exists and can encode environmental information, how can this information be decoded during embryonic development to programme an offspring phenotype? This has been an open question in the field of mammalian epigenetic inheritance, as clear molecular mechanisms remain poorly understood in mammals in contrast to our understanding of the mechanisms in other model organisms[5–9].

Small RNA-mediated DNA methylation in plants[7] and small RNA-mediated histone modification in worms[60] and yeast[61] have been shown to effectively enable transgenerational gene regulation. In these models, the exertion and maintenance of the effects of RNAs involve either amplification of initial small RNAs via RNA-dependent RNA polymerase (RdRP) in plants and worms[7,60] or a positive feedback loop between small RNAs and histone marks as implicated in yeast[61] — the positive feedback in yeast might also involve the function of RdRP[62]. In fly models in which the RdRP system is not found, the paternal diet can transmit obesity and metabolic disorder intergenerationally by altering the chromatin state[63]. Interestingly, in these fly models, disruption of an H3K9 methyltransferase, *SetDB1*, prevents heterozygotic male flies from transmitting paternal diet-mediated obesity in both mutant and wild-type offspring, suggesting the involvement of *trans*-acting molecules such as RNAs (carried by the wild-type sperm)[63]. Similarly, in a mouse model, paternal haploinsufficiency of *Setdb1* can trigger a *trans*-acting signal that affects the penetrance of a well-established coat-colour phenotype[64], which is regulated by the DNA methylation level of intracisternal A particle retrotransposon-controlled agouti viable yellow (*Avy*)[65].

These converging data are thought-provoking, as it has been demonstrated that the deletion of *SetDB1* in mouse embryonic stem cells triggers elevated expression of tsRNAs, which play an active role in silencing transposon elements[66]. Additionally, deletion of *Dnmt2* or *Nsun2* in flies has been found to affect the accumulation of transposon elements[67], which could also be associated with the enhanced biogenesis of tsRNAs. The engagement of tsRNAs to regulate transposon elements in cells where retrotransposon activation takes place (such as plant pollen grains and mammalian pre-implantation embryos) may involve various mechanisms and has been discussed in other papers[68–70]. All these studies suggest highly integrated synergism among sperm RNAs, histone marks, DNA methylation and transposon elements in transmitting specific paternal phenotypes. A related question that needs answering is whether the deletion of *Setdb1* can prevent intergenerational transmission of the paternal HFD effect in mice, as has been found in the fly. In addition, whether this process might involve sperm tsRNAs is intriguing, as the answer may set the stage to establish a general principle by which tsRNAs and histone marks interact with each other to intergenerationally relay environment-induced traits in multiple species.

In mammals, in which the RdRP system has not been found (with the rare exception that a RdRP gene has been found hidden in the genome of certain bat species encoded by endogenous bornavirus-like L elements[71]), it remains unknown whether any sperm RNAs can be directly amplified or initiate RNA–histone feedback in the early embryo to generate prolonged effects. Interestingly, small RNA sequencing of stage-dependent mouse preimplantation embryos has revealed a surge of tsRNAs at the eight-cell stage. These eight-cell stage embryos have a tsRNA composition similar to that of sperm tsRNAs[72]. We are still unsure whether these data suggest self-induction of the initial sperm tsRNA input, and if they do, we need to find the trigger and investigate the biological meaning of such a signal.

It has also been shown that the distribution pattern of H3K4me3, a histone mark, in sperm is initially removed in zygotes, but is re-established in both paternal and maternal chromosomes at the late two-cell embryo stage[73,74]. This finding suggests that paternal information in the form of sperm H3K4me3 is recorded and inherited. It would be very interesting to investigate whether the reconstruction of the sperm H3K4me3 pattern in the embryo is related to sperm RNAs, and whether these histone marks are the signal triggering tsRNA expression during later embryonic stages (FIG. 2a). In the human embryo, the mechanism of re-establishment of nucleosomal chromatin domains may involve different histone marks but also needs the information input from the sperm[75], with a possible contribution from sperm RNAs as well. It is a tantalizing question as to how the sperm RNA code can interact with the histone code throughout embryo development; part of this question can be tested through experiments using parthenogenetic embryos.

Regulating ribosomal machinery and translational specificity? In addition to potential interactions with DNA methylation, histone marks and transposable elements, sperm RNAs may amplify their effects in other ways, if they are not themselves amplified. tsRNAs have been found to interact with translational machinery and ribosome biogenesis, for example binding with translational initiation factors (facilitated by RNA modifications and RNA structure) and Ago-dependent or Ago-independent regulation of ribosomal protein mRNAs with sequence specificity[51]. In addition, exciting discoveries during the past 5 years have revealed that alterations in ribosomal protein composition endow the ribosomal machinery with different selectivity for translating sub-pools of mRNAs (referred to as ribosome heterogeneity), including those that control metabolism and development[76,77]. These converging findings have led to the hypothesis that modified sperm tsRNAs might reprogram the embryonic metabolic state by targeting specialized ribosomes, thus biasing the trajectory of embryonic development[76]. In other words, paternal environmental information in the form of the sperm RNA code might be transformed into an embryonic 'ribosome code' that generates translational specificity to define the metabolic phenotype of the offspring (FIG. 2b).

Cascade effects of metabolic state and the role of metabolites. Zygotic injection of sperm RNAs extracted from HFD-treated fathers can induce altered transcription of metabolic gene pathways in both early embryos and offspring tissues in mice[18], but how this altered metabolic transcriptional cascade penetrates the whole developmental process remains unknown. One explanation could be that once a biased metabolic state is triggered in the early embryo, it might alter metabolites or mitochondrial function to further trigger a

chain reaction that continuously influences the metabolic state. The chain reaction could possibly act via epigenetic feedback to the nucleus, as many central metabolites, such as *S*-adenosyl methionine and acetyl coenzyme A, are known substrates to be incorporated into chemical modifications in DNA and histones[78] (FIG. 2c). This 'butterfly effect' scenario might be particularly true in the mammalian early embryonic system, as initial small biases between blastomeres have been shown to be used to facilitate symmetry breaking through crucial transcriptional and epigenetic factors that are differentially expressed between

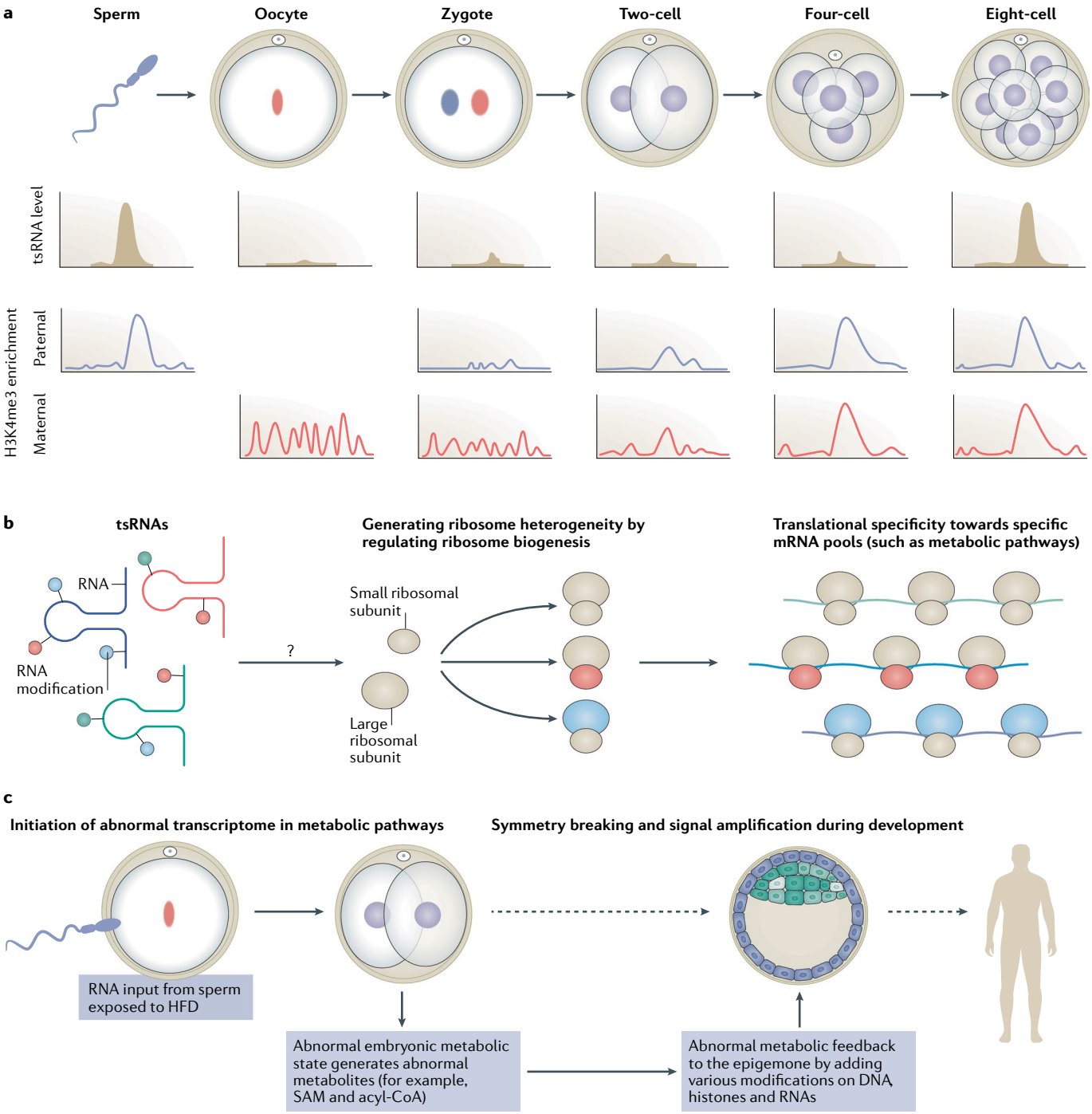


Fig. 2 | **Potential mechanisms in transformation of the sperm RNA code during embryo development. a** | In mouse early embryo, the tRNA-derived small RNA (tsRNA) expression level increases from the four-cell to eight-cell transition, and the tsRNA composition is similar to that of mature sperm[72]. At certain genomic loci, the H3K4me3 pattern in sperm is initially removed in zygotes but re-established in both paternal and maternal chromosomes from the late two-cell embryo stage[73,74]. It would be interesting to explore the causal relationship between tsRNAs and the H3K4me3 pattern in sperm and embryo. **b** | The hypothesis of tsRNA-mediated ribosome heterogeneity that generates biased translational specificity towards different pools of mRNA subpopulations. **c** | A hypothetical scenario to explain how the initial abnormal metabolic transcriptome induced by sperm RNA can be maintained throughout embryo development. This could be triggered by a self-amplifying loop of abnormal metabolic transcriptome, metabolites and epigenome during embryo development. Acyl-CoA, acetyl coenzyme A; HFD, high-fat diet; SAM, S-adenosyl methionine.

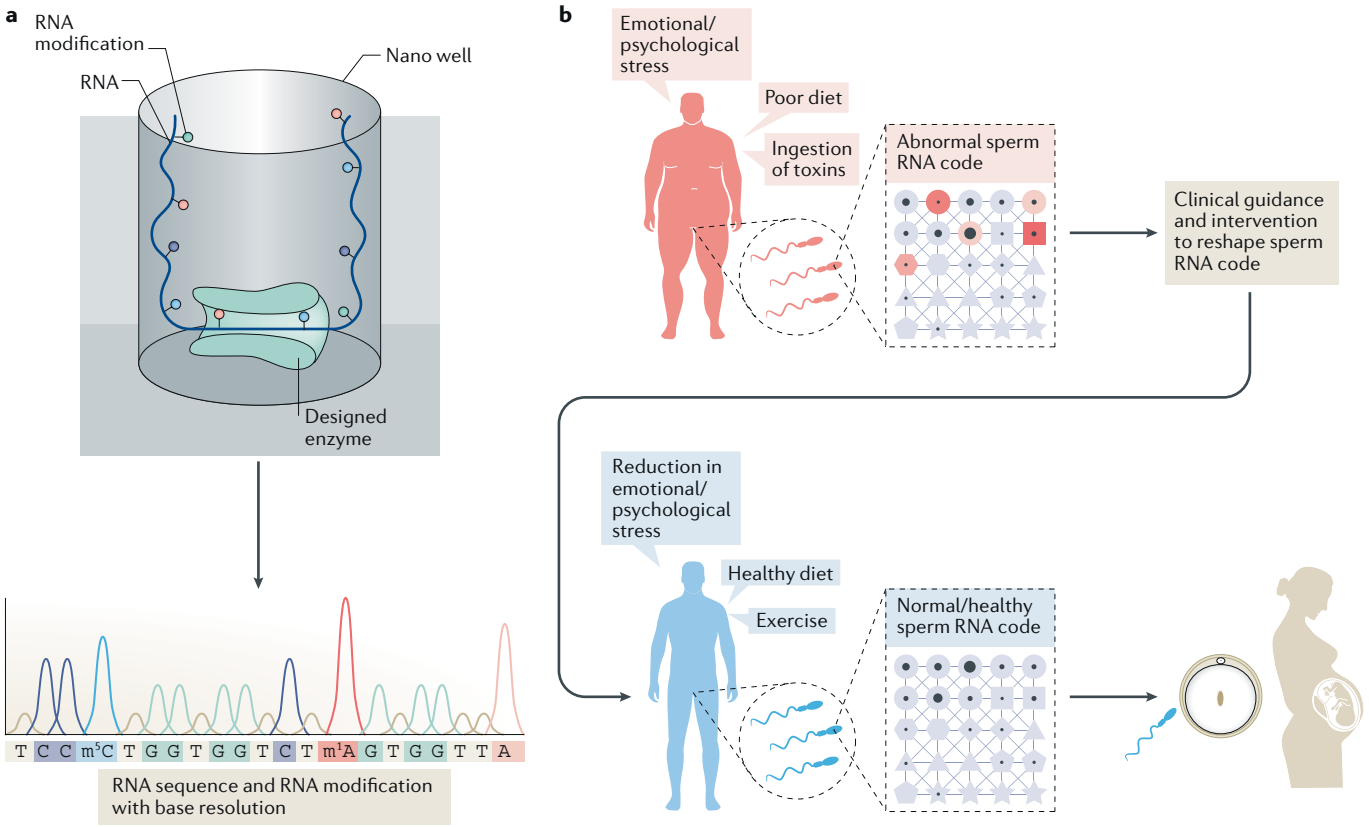


Fig. 3 | **Future applications and precision medicine based on high resolution of sperm RNA code. a** | Future application of third-generation RNA-sequencing platform to simultaneously obtain full RNA sequences and meanwhile pinpoint different RNA modifications at base resolution. **b** | Future use of sperm RNA code information as clinical guidance for well-timed pregnancy and to reduce the susceptibility to metabolic disorder in the offspring.

blastomeres[79,80]. This amplification system might be similarly 'hijacked' under a specific paternal environment to amplify the initial regulatory networks that programme an abnormal metabolic state throughout embryonic development to affect offspring metabolic health (FIG. 2c).

Application and precision medicine

Aside from understanding the nature of the sperm RNA code with regard to programming offspring phenotypes to satisfy scientific interest, perhaps an equally important mission is to control offspring health by harnessing this code. Interestingly, it seems that the metabolic state in animals is particularly vulnerable to a variety of environmental inputs. For example, in addition to an unhealthy diet, parental exposure to a range of endocrine-disrupting chemicals (EDCs) and other toxicants such as arsenite can cause metabolic disorder in the parental generation, and such phenotypes can be transmitted to the offspring[2]. Under these circumstances, it is conceivable that certain crucial regulatory elements (such as RNAs and histone marks) in the sperm, oocyte or embryo can be disrupted by different causes but similarly lead to metabolic disorders.

Previous studies have indeed found that paternal exposure to EDCs (vinclozolin or DDT) can cause changes in sperm RNA (for example, miRNAs and tsRNAs) and that some changes can persist in the sperm of offspring[81,82]; however, whether the alterations in sperm RNA are the cause of offspring metabolic disorders remains unknown due to a lack of functional studies. It would be interesting to compare sperm RNA profiles and RNA modifications under different paternal environmental exposure scenarios that similarly cause offspring metabolic diseases in a well-controlled laboratory setting. Such an experiment could help reveal the essential sperm RNA signature that causes the altered metabolic outcome, thus generating fundamental knowledge for future therapeutic interventions. For example, the sperm RNA signature could be used as a clinical marker of offspring metabolic disease susceptibility. In addition, as human sperm RNA profiles are known to be efficiently altered by improved body condition[1,83–85], the signature could be monitored before a planned pregnancy and clinical guidance (such as improvements to lifestyle) could also be introduced if necessary to reshape the sperm RNA code towards a healthy condition (FIG. 3). Moreover, synthetic modified RNAs might be designed and used to counteract adverse conditions or improve metabolic performance in livestock.

Caution, however, should be exercised when applying this RNA-based information, as the current RNA sequencing protocol generates biased results. This bias is due to the fact that some RNA modifications interfere with library construction processes such as adaptor ligation and reverse transcription[44]. Furthermore, RNA secondary structures, such as RNA G-quadruplexes[86] and RNA–RNA interactions, might further impede acquisition of a holistic picture of the sperm RNA profile. New sequencing methods are expected to overcome these difficulties. In particular, future development of third-generation RNA sequencing platforms (such as PacBio or Oxford Nanopore) could provide a unique opportunity to simultaneously obtain both RNA sequences and RNA modification information at base resolution[87]. Such advances are expected to allow resolution of

the sperm RNA code with unprecedented precision and will facilitate translational applications using this quantitative information (FIG. 3).

Conclusions

The research community is beginning to accept the existence of additional biological codes beyond the protein-coding DNA sequences. The discovery of additional codes for biological traits, such as histone code and RNA code, is exciting but also introduces tremendous challenges. The traditional DNA code is linear and encodes how DNA is transcribed into mRNA and translated into protein. Additional codes, however, such as the RNA code discussed in this Perspectives, regulate the on-and-off switches of DNA code in space and time to generate gene products (that is RNA and protein) in an exact order, which involves information in 3D and 4D that is required to organize biological processes with precision[88,89].

As we have discussed in this article, emerging evidence has begun to show the versatility and specificity of the sperm RNA code in programming offspring phenotypes, such as metabolic performance. The complexity of this RNA code, however, is obvious. It contains multiple layers of information, including its sequence, modifications and secondary structure, not to mention its interactions with environmental, genetic and epigenetic factors. The possible permutations, and thus the information capacity endowed by this complexity, could be astronomical[31]. However, given the fast-evolving capacity of technology to simultaneously analyse multiple layers of information in single cells, such as DNA, transcriptome, epigenome and chromatin structure information[90,91], and to decipher interactions among different levels of information, including RNA–DNA and RNA–chromatin interactions[92,93], along with the equally fast-evolving power of machine learning[94], it is highly likely that we will be able to crack these new biological codes at the most fundamental level and to understand the nature of inheritance with unprecedented insight.

*Yunfang Zhang[1,2,6], Junchao Shi[2,3,6], Minoo Rassoulzadegan[4], Francesca Tuorto[5] and Qi Chen[2,3]**

[1]*Medical Center of Hematology, The Xinqiao Hospital of Army Medical University, Chongqing, China.*

[2]*Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV, USA.*

[3]*Division of Biomedical Sciences, School of Medicine, University of California, Riverside, Riverside, CA, USA.*

[4]*Université de Nice Sophia Antipolis, Inserm U1091, CNRS U7277, Nice, France.*

[5]*Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany.*

[6]*These authors contributed equally: Yunfang Zhang, Junchao Shi*

**e-mail: qi.chen@medsch.ucr.edu*

https://doi.org/10.1038/s41574-019-0226-2

Published online 24 June 2019

1. Barres, R. & Zierath, J. R. The role of diet and exercise in the transgenerational epigenetic landscape of T2DM. *Nat. Rev. Endocrinol.* **12**, 441–451 (2016).
2. Sales, V. M., Ferguson-Smith, A. C. & Patti, M. E. Epigenetic mechanisms of transmission of metabolic disease across generations. *Cell Metab.* **25**, 559–571 (2017).
3. Fleming, T. P. et al. Origins of lifetime health around the time of conception: causes and consequences. *Lancet* **391**, 1842–1852 (2018).
4. Eichler, E. E. et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
5. Skvortsova, K., Iovino, N. & Bogdanovic, O. Functions and mechanisms of epigenetic inheritance in animals. *Nat. Rev. Mol. Cell Biol.* **19**, 774–790 (2018).
6. Miska, E. A. & Ferguson-Smith, A. C. Transgenerational inheritance: models and mechanisms of non-DNA sequence-based inheritance. *Science* **354**, 59–63 (2016).
7. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95–109 (2014).
8. Chen, Q., Yan, W. & Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat. Rev. Genet.* **17**, 733–743 (2016).
9. Perez, M. F. & Lehner, B. Intergenerational and transgenerational epigenetic inheritance in animals. *Nat. Cell Biol.* **21**, 143–151 (2019).
10. Bouret, S., Levin, B. E. & Ozanne, S. E. Gene-environment interactions controlling energy and glucose homeostasis and the developmental origins of obesity. *Physiol. Rev.* **95**, 47–82 (2015).
11. Oestreich, A. K. & Moley, K. H. Developmental and transmittable origins of obesity-associated health disorders. *Trends Genet.* **33**, 399–407 (2017).
12. Godfrey, K. M. et al. Influence of maternal obesity on the long-term health of offspring. *Lancet Diabetes Endocrinol.* **5**, 53–64 (2017).
13. Radford, E. J. et al. In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).
14. Kazachenka, A. et al. Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. *Cell* **175**, 1259–1271 (2018).
15. Hammoud, S. S. et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478 (2009).
16. Brykczynska, U. et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* **17**, 679–687 (2010).
17. Siklenka, K. et al. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).
18. Chen, Q. et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397–400 (2016).
19. Gapp, K. et al. Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. *Mol. Psychiatry*. https://doi.org/10.1038/s41380-018-0271-6 (2018).
20. Grandjean, V. et al. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci. Rep.* **5**, 18193 (2015).
21. Gapp, K. et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat. Neurosci.* **17**, 667–669 (2014).
22. Zhang, Y. et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat. Cell Biol.* **20**, 535–540 (2018).
23. Johnson, G. D. et al. Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic Acids Res.* **43**, 6847–6859 (2015).
24. Peng, H. et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res.* **22**, 1609–1612 (2012).
25. Sharma, U. et al. Small RNAs are trafficked from the epididymis to developing mammalian sperm. *Dev. Cell* **46**, 481–494 (2018).
26. Sharma, U. et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391–396 (2016).
27. Chu, C. et al. A sequence of 28S rRNA-derived small RNAs is enriched in mature sperm and various somatic tissues and possibly associates with inflammation. *J. Mol. Cell Biol.* **9**, 256–259 (2017).
28. Shi, J. et al. SPORTS1.0: a tool for annotating and profiling non-coding RNAs optimized for rRNA- and tRNA-derived small RNAs. *Genomics Proteomics Bioinformatics* **16**, 144–151 (2018).
29. Schuster, A. et al. SpermBase: a database for sperm-borne RNA contents. *Biol. Reprod.* **95**, 99 (2016).
30. Hua, M. et al. Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discov.* **5**, 20 (2019).
31. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
32. Guo, L. et al. Sperm-carried RNAs play critical roles in mouse embryonic development. *Oncotarget* **8**, 67394–67405 (2017).
33. Yuan, S. et al. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development* **143**, 635–647 (2016).
34. Frye, M. et al. RNA modifications modulate gene expression during development. *Science* **361**, 1346–1349 (2018).
35. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404 (2018).
36. Guzzi, N. et al. Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* **173**, 1204–1216 (2018).
37. Safra, M. et al. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251–255 (2017).
38. Murashov, A. K. et al. Paternal long-term exercise programs offspring for low energy expenditure and increased risk for obesity in mice. *FASEB J.* **30**, 775–784 (2016).
39. Rompala, G. R. et al. Heavy chronic intermittent ethanol exposure alters small noncoding RNAs in mouse sperm and epididymosomes. *Front. Genet.* **9**, 32 (2018).
40. Rassoulzadegan, M. et al. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
41. Kiani, J. et al. RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. *PLoS Genet.* **9**, e1003498 (2013).
42. Schulz, N. K. E., Diddens-de Buhr, M. F. & Kurtz, J. Paternal knockdown of Dnmt2 increases offspring susceptibility to bacterial infection. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/422063v1 (2018).
43. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2018).
44. Zhang, X. et al. Small RNA modifications: integral to function and disease. *Trends Mol. Med.* **22**, 1025–1034 (2016).
45. Schaefer, M. et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24**, 1590–1595 (2010).
46. Tuorto, F. et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).
47. Legrand, C. et al. Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* **27**, 1589–1596 (2017).
48. Tuorto, F. et al. Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.* **37**, e99777 (2018).
49. Muller, M. et al. Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine. *Nucleic Acids Res.* **43**, 10952–10962 (2015).
50. Wang, X. et al. Queuosine modification protects cognate tRNAs against ribonuclease cleavage. *RNA* **24**, 1305–1313 (2018).
51. Shi, J. et al. tsRNAs: the Swiss army knife for translational regulation. *Trends Biochem. Sci.* **44**, 185–189 (2018).
52. Wang, X. et al. Transcriptome-wide reprogramming of N(6)-methyladenosine modification by the mouse microbiome. *Cell Res.* **29**, 167–170 (2019).

53. Nelson, V. R. et al. Transgenerational epigenetic effects of the Apobec1 cytidine deaminase deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc. Natl Acad. Sci. U. S. A.* **109**, E2766–E2773 (2012).
54. Liu, W. M. et al. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc. Natl Acad. Sci. U. S. A.* **109**, 490–494 (2012).
55. Yuan, S. et al. mir-34b/c and mir-449a/b/c are required for spermatogenesis, but not for the first cleavage division in mice. *Biol. Open* **4**, 212–223 (2015).
56. Conine, C. C. et al. Small RNAs gained during epididymal transit of sperm are essential for embryonic development in mice. *Dev. Cell* **46**, 470–480 (2018).
57. Suganuma, R., Yanagimachi, R. & Meistrich, M. L. Decline in fertility of mouse sperm with abnormal chromatin during epididymal passage as revealed by ICSI. *Hum. Reprod.* **20**, 3101–3108 (2005).
58. Cheng, H. S. et al. Increased susceptibility of post-weaning rats on high-fat diet to metabolic syndrome. *J. Adv. Res.* **8**, 743–752 (2017).
59. Sarker, G. et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. *Proc. Natl Acad. Sci. U. S. A.* **116**, 10547–10556 (2019).
60. Rechavi, O. & Lev, I. Principles of transgenerational small RNA inheritance in *Caenorhabditis elegans*. *Curr. Biol.* **27**, R720–R730 (2017).
61. Yu, R., Wang, X. & Moazed, D. Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* **558**, 615–619 (2018).
62. Motamedi, M. R. et al. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
63. Ost, A. et al. Paternal diet defines offspring chromatin state and intergenerational obesity. *Cell* **159**, 1352–1364 (2014).
64. Daxinger, L. et al. Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. *Sci. Rep.* **6**, 25004 (2016).
65. Morgan, H. D. et al. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314–318 (1999).
66. Schorn, A. J. et al. LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* **170**, 61–71 (2017).
67. Genenncher, B. et al. Mutations in cytosine-5 tRNA methyltransferases impact mobile element expression and genome stability at specific DNA repeats. *Cell Rep.* **22**, 1861–1874 (2018).
68. Martinez, G. tRNA-derived small RNAs: new players in genome protection against retrotransposons. *RNA Biol.* **15**, 170–175 (2018).
69. Zhang, Y., Shi, J. & Chen, Q. tsRNAs: new players in mammalian retrotransposon control. *Cell Res.* **27**, 1307–1308 (2017).
70. Schorn, A. J. & Martienssen, R. Tie-break: host and retrotransposons play tRNA. *Trends Cell Biol.* **28**, 793–806 (2018).
71. Horie, M. et al. An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. *Sci. Rep.* **6**, 25873 (2016).
72. Yang, Q. et al. Highly sensitive sequencing reveals dynamic modifications and activities of small RNAs in mouse oocytes and early embryos. *Sci. Adv.* **2**, e1501482 (2016).
73. Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
74. Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
75. van de Werken, C. et al. Paternal heterochromatin formation in human embryos is H3K9/HP1 directed and primed by sperm-derived histone modifications. *Nat. Commun.* **5**, 5868 (2014).
76. Genuth, N. R. & Barna, M. The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell* **71**, 364–374 (2018).
77. Shi, Z. et al. Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* **67**, 71–83 (2017).
78. Li, X. et al. Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nat. Rev. Mol. Cell Biol.* **19**, 563–578 (2018).
79. Chen, Q. et al. Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. *Nat. Commun.* **9**, 1819 (2018).
80. Shi, J. et al. Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. *Development* **142**, 3468–3477 (2015).
81. Skinner, M. K. et al. Alterations in sperm DNA methylation, non-coding RNA and histone retention associate with DDT-induced epigenetic transgenerational inheritance of disease. *Epigenetics Chromatin* **11**, 8 (2018).
82. Schuster, A., Skinner, M. K. & Yan, W. Ancestral vinclozolin exposure alters the epigenetic transgenerational inheritance of sperm small noncoding RNAs. *Environ. Epigenet.* **2**, dvw001 (2016).
83. Donkin, I. et al. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab.* **23**, 369–378 (2016).
84. Stanford, K. I. et al. Paternal exercise improves glucose metabolism in adult offspring. *Diabetes* **67**, 2530–2540 (2018).
85. Ingerslev, L. R. et al. Endurance training remodels sperm-borne small RNA expression and methylation at neurological gene hotspots. *Clin. Epigenet.* **10**, 12 (2018).
86. Kwok, C. K. et al. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844 (2016).
87. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
88. Zhang, Y. & Chen, Q. The expanding repertoire of hereditary information carriers. *Development* **146**, dev170902 (2019).
89. Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
90. Guo, F. et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* **27**, 967–988 (2017).
91. Bian, S. et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018).
92. Li, X. et al. GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
93. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
94. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
95. Rodriguez, K. F. et al. Effects of in utero exposure to arsenic during the second half of gestation on reproductive end points and metabolic parameters in female CD-1 mice. *Environ. Health Perspect.* **124**, 336–343 (2016).
96. Dias, B. G. & Ressler, K. J. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* **17**, 89–96 (2014).
97. Sun, W. et al. Cold-induced epigenetic programming of the sperm enhances brown adipose tissue activity in the offspring. *Nat. Med.* **24**, 1372–1383 (2018).
98. Le, Q. et al. Drug-seeking motivation level in male rats determines offspring susceptibility or resistance to cocaine-seeking behaviour. *Nat. Commun.* **8**, 15527 (2017).
99. Yan, W. et al. Birth of mice after intracytoplasmic injection of single purified sperm nuclei and detection of messenger RNAs and MicroRNAs in the sperm nuclei. *Biol. Reprod.* **78**, 896–902 (2008).
100. Lalancette, C. et al. Paternal contributions: new functional insights for spermatozoal RNA. *J. Cell. Biochem.* **104**, 1570–1579 (2008).
101. Hamatani, T. Human spermatozoal RNAs. *Fertil. Steril.* **97**, 275–281 (2012).
102. Al-Dossary, A. A. et al. Oviductosome-sperm membrane interaction in cargo delivery: detection of fusion and underlying molecular players using three-dimensional super-resolution structured illumination microscopy (SR-SIM). *J. Biol. Chem.* **290**, 17710–17723 (2015).
103. Liu, Y. & Chen, Q. 150 years of Darwin's theory of intercellular flow of hereditary information. *Nat. Rev. Mol. Cell Biol.* **19**, 749–750 (2018).

Acknowledgements
The authors thank members of the Chen laboratory for extensive discussions on the manuscript. Y.Z. is supported by the Army Medical University (18JS008 and 2018XLC2018). F.T. is supported by the Institute of Genetics and Biophysics A. Buzzati-Traverso, C.N.R., Italy. M.R. is supported by La Fondation Nestlé France. Research in the Q.C. laboratory is supported by the NIH (R01HD092431).

Author contributions
Q.C., Y.Z. and J.S. developed the concept and wrote the manuscript. M.R. and F.T. provided a substantial contribution to discussion of the content, and reviewed and edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Endocrinology thanks R. Barrès, M.-E. Patti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.