Spring 2018 – Systems Biology of Reproduction
Discussion Outline – Reproductive Endocrinology Systems
Michael K. Skinner – Biol 475/575
CUE 418, 10:35-11:50 am, Tuesday & Thursday
April 5, 2018
Week 13

Reproductive Endocrinology Systems

Primary Papers:

1. Stotzel, et al. (2012) Theriogenology 78:1415-1428
2. Mereness, et al (2016) Endocrinology 517:913-927
3. Barban, et al (2016) Nat Genetics 48:1462

Discussion

Student 7: Reference 1 above

- What endocrine parameters were synchronized and what regulatory agent tested?
- What experimental model was used?
- What model was established and validated?

Student 8: Reference 2 above

- What was the experimental design and technology used?
- What did the comparison of granulosa and theca cell observations suggest?
- What was identified regarding the hormone regulated phasic sensitivity to LH in the ovary?

Student 10: Reference 3 above

- What was the experimental design and technology used?
- What reproductive factors were used and what traits were associated?
- What conclusions can be drawn on genomic control of reproduction?

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Theriogenology 78 (2012) 1415–1428

Theriogenology

www.theriojournal.com

Advances in modeling of the bovine estrous cycle: Synchronization with PGF2α

C. Stötzel[a,*], J. Plöntzke[b], W. Heuwieser[b], S. Röblitz[a]

[a] *Computational Systems Biology Group, Numerical Analysis and Modeling, Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany*
[b] *Clinic for Animal Reproduction, Faculty of Veterinary Medicine, Freie Universität Berlin, Königsweg 65, 14163 Berlin, Germany*

Received 6 October 2011; received in revised form 25 April 2012; accepted 26 April 2012

Abstract

Our model of the bovine estrous cycle is a set of ordinary differential equations which generates hormone profiles of successive estrous cycles with several follicular waves per cycle. It describes the growth and decay of the follicles and the corpus luteum, as well as the change of the key reproductive hormones, enzymes and processes over time. In this work we describe recent developments of this model towards the administration of prostaglandin F2α. We validate our model by showing that the simulations agree with observations from synchronization studies and with measured progesterone data after single dose administrations of synthetic prostaglandin F2α.

Keywords: Cow; Reproduction; Hormone patterns; Differential equations; Systems biology; Synchronization

1. Introduction

Throughout the last decades, increasing amounts of biological data have become available, and mathematical modeling aims at relating and interpreting these data with the help of powerful computational tools and efficient algorithms. As part of the rapidly expanding research field systems biology, this approach is highly interdisciplinary, and requires a close collaboration between mathematicians and natural scientists. Although a few models are already available [1–4], systems biology modeling in animal sciences is still at an early stage of development [5].

The purpose of this paper is to demonstrate how dynamic mathematical models can enhance the understanding of synchronization in the bovine estrous cycle. The model used in this article is based on a model that was presented in previous studies [4,6]. It describes the key feedback mechanisms within the bovine estrous cycle, and generates hormonal profiles of successive estrous cycles over time. The long-term goal of developing such a model of the endocrine mechanisms in the bovine is to assist with research and application in drug testing, management decision making, or therapeutic strategies. On the one hand, we want to contribute to a better understanding of the complex biological mechanisms that drive the estrous cycle, and on the other hand we aim to exploit the predictive abilities of such a mathematical model, e.g., by simulating external influences. The short-term goal and current aim of this article is to validate and improve the existing model, thus to make a step towards a reliable model, which makes predictions more adequate.

The starting point of our work was the validation of the model used in [6]. For this purpose we investigated

* Corresponding author. Tel: +49 30 84185 335; Fax: +49 30 84185 107
E-mail address: stoetzel@zib.de (C. Stötzel).

0093-691X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.theriogenology.2012.04.017

how adequately the model simulations match available information. Experimental data required for model validation would for example consist of measured hormonal concentrations of healthy, untreated, individual cows at different stages of estrous cycle. Unfortunately, measurements published in literature are rare and do often not meet the requirements for validation; observed time scales are often too small or too coarse, or too few substances are measured. Therefore, we were looking for alternative information that we could monitor with the model. We wanted to check the correctness of the model for a specific scenario where the system answer is known. More precisely, synchronization protocols [7] drew our attention to prostaglandin F2α ($PGF_{2\alpha}$). In veterinary medicine, $PGF_{2\alpha}$ and its analogues are administered to cows mainly to make use of their luteolytic action, e.g., in estrus synchronization protocols. It is known that the sudden rise of this hormone at certain stages of the estrous cycle results in an immediate decay of the responsive corpus luteum (CL), and an immediate fall of progesterone levels in plasma. In a first validation step of our model, we simulated the administration of $PGF_{2\alpha}$ and compared the outcome of our simulation with known responses reported in literature. Furthermore, we used measurements taken after single injections of $PGF_{2\alpha}$ to validate our simulation.

The above described validation is an iterative process. Whenever simulations with the original model were not satisfactory, adjustments to the model were made. In particular, during simulations for the administration of hormones, we figured out that this model still had some shortcomings regarding the growth and decay of the CL. A replacement of mechanisms involved in ovulation and the refinement of luteolysis became necessary.

The objective of this paper is to describe the adjustments to the original model, and to show that the advanced model captures the known effects after $PGF_{2\alpha}$ administration.

2. Materials and methods

Our modeling approach aims to reproduce how the various components involved in the bovine estrous cycle function together. Instead of only investigating individual parts, we describe the biological system on the whole-organism level, in order to capture the most important dynamic feedback mechanisms. We mostly consider regulatory mechanisms (inhibitory and stimulatory effects of hormones), and model them as continuous functions over time. The model in [4] was constructed by first defining a number of key components of the system and their interactions, which were then represented in a flow chart. Subsequently, a set of differential equations was derived to describe the relations within the flow chart mathematically. For every model component, e.g., a hormonal concentration or the follicular capacity to produce hormones, one ordinary differential equation (ODE) was developed. The dependencies between the components are described in the right hand sides of the ODEs. Hill functions are used to model inhibitory and stimulatory relations. A slightly modified version of the model in [4] with improved luteolysis was used in [6], which we will refer to as the original model in the following. The model derived in this paper is the result of several adjustment of the original model, which will be described in detail. An overview of the final model is given in Figure 4. As we aim to validate our model with the help of external manipulations, we will describe the background of this administration first, before going into the details of the model advancements.

2.1. Background of $PGF_{2\alpha}$ administration - estrus synchronization protocols

Protocols of estrus synchronization consist of consecutive administration of different hormones or their analogues following a certain order. They have the goal to synchronize the estrus of individual females in order to facilitate timing of following artificial insemination, indepently of estrous cycle stage at the start of the protocol. They are commonly used in cattle and in other domestic and non-domestic species [8]. Our approach to validate the model of the bovine estrous cycle was to include synchronization protocols as described in [7,9]. Here, we restrict ourselves to single and double administrations of $PGF_{2\alpha}$, which is used in these protocols as single or double injection. We model the administration of $PGF_{2\alpha}$ at different stages of the estrous cycle.

The $PGF_{2\alpha}$ is responsible for the onset of luteolysis in the cow. With luteolysis the luteal phase of the cycle ends and a new estrus can take place. The $PGF_{2\alpha}$ induces functional luteolysis by reducing progesterone production followed by structural luteolysis with tissue degeneration and cell death [10,11]. The $PGF_{2\alpha}$ is synthetized in the endometrium and released in pulses, regulated by estradiol (E_2), progesterone (P_4) and oxytocin (OT) during the estrous cycle [12–14]. In animal production, administration of synthetical analogues of $PGF_{2\alpha}$ (e.g., Cloprostenol, Luprostiol, Tiaprost) or original $PGF_{2\alpha}$ (e.g., Dinoprost) is used for various pur-

poses in the cow, such as induction of estrus or synchronization protocols. The effect of the treatment depends on the stage of estrous cycle which determines the responsiveness of the CL on the luteolytic effect of $PGF_{2\alpha}$ [11]. At midluteal stage of the estrous cycle administration of $PGF_{2\alpha}$ leads to luteolysis within a few hours. This results in a decrease of P_4 concentration, increase of E_2, a peak of the Luteinizing Hormone (LH) and ovulation [15].

Virtual administration of $PGF_{2\alpha}$ to the cow model was conducted on various days of the estrous cycle. With the original model, the simulation outcome of the model after $PGF_{2\alpha}$ application was not as expected, which gave us a starting point to improve the model.

2.2. Advancements in the model

To improve the model with respect to the expected effects of a single $PGF_{2\alpha}$ injection, we introduced some new features which are described in this section. A list of the Hill functions—sigmoidal functions to model inhibitory or stimulatory effects as described in [4]—can be found in Appendix B. Here H^+ and H^- denote scaled positive, respectively negative Hill functions. Parameter values are specified in Appendix C.

In the former model, the equation for the CL described the change of the capacity of the CL to produce P_4. For reasons described later in this section, we now interpret this equation as the development of the size (e.g., diameter) of the CL over the cycle. This is also advantageous as soon as we deal with ultrasound measurements for the corpus luteum. Likewise, the equation for the follicles (Foll) now describes the development of the total size of all follicles.

2.2.1. Drug administration

We add an additional model component for analogues of $PGF_{2\alpha}$, denoted PGF_{syn} in the following. It is reported that $PGF_{2\alpha}$ and its analogues have a very short half-life [16,17], thus we chose to model PGF_{syn} with a rapid decay. Furthermore, it is known that $PGF_{2\alpha}$ analogues have an up to three times higher biological activity than original $PGF_{2\alpha}$ [17]. Even low doses of PGF_{syn} cause a peak in $PGF_{2\alpha}$ that exceeds the natural level [18]. Due to this high potency of PGF_{syn}, we chose to model PGF_{syn} with a three times higher relative level compared to normal $PGF_{2\alpha}$ levels. We model the effect of the synthetical analogue by summing the level of PGF_{syn} to the normal $PGF_{2\alpha}$ level.

To model the rise of PGF_{syn} in the system, we take a function which is zero before dosing time (t_D), and has a sharp left-skewed peak with maximum shortly after t_D. This leads to a slight delay in the effect of the injection. As suggested in [19] and based on techniques described in

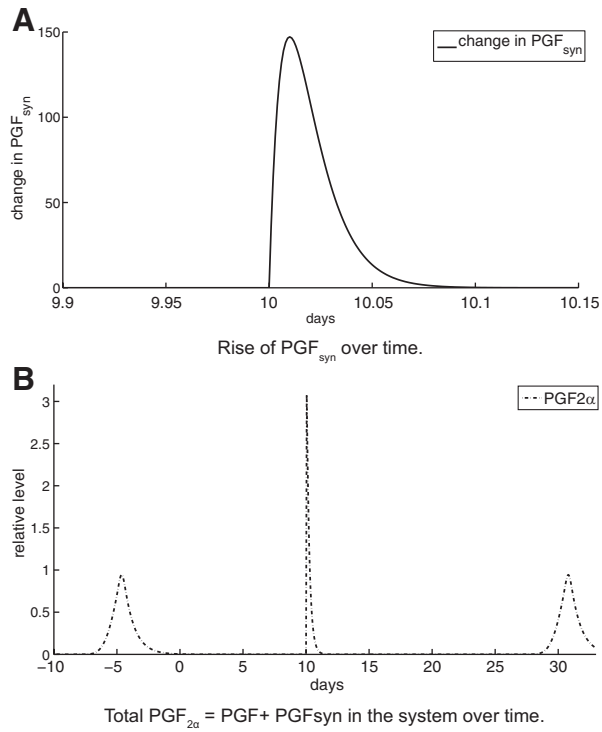

Fig. 1. Change in PGF_{syn} and Administration of PGF_{syn} at time $t_D=10$. Parameters are $D = 4$ and $\beta=100$. Maximum rise of PGF_{syn} is at $t = t_D + 1/\beta=10.01$. The level of $PGF_{2\alpha}$ is the result of summing PGF_{syn} levels to naturally arising $PGF_{2\alpha}$.

[20], we take the probability density function of the Gamma-distribution with fixed shape parameter $\alpha = 2$, and inverse scale parameter β leading to a left-skewed curve which has its maximum at $t = 1/\beta$. The change of concentration of synthetic $PGF_{2\alpha}$ is calculated as

$$\frac{d}{dt}PGF_{syn}(t) = D \cdot \beta^2 \cdot t_{mod}(t) \cdot \exp(-\beta \cdot t_{mod}(t)) - c_{PGF_{syn}} \cdot PGF_{syn}(t).$$

The parameter D represents the amount of administered drug scaled to obtain the designated height of the relative level of PGF_{syn}, see Figure 1. The parameter c_{PGFsyn} denotes the clearance rate constant of PGF_{syn}. The modified time function t_{mod} denotes time after dosing, and is given by

$$t_{mod}(t) = \max(0, t - t_D)$$

The rise of PGF_{syn} is large right after dosing time and approaches zero quickly thereafter, leading to a rapid decay of the function $PGF_{syn}(t)$.

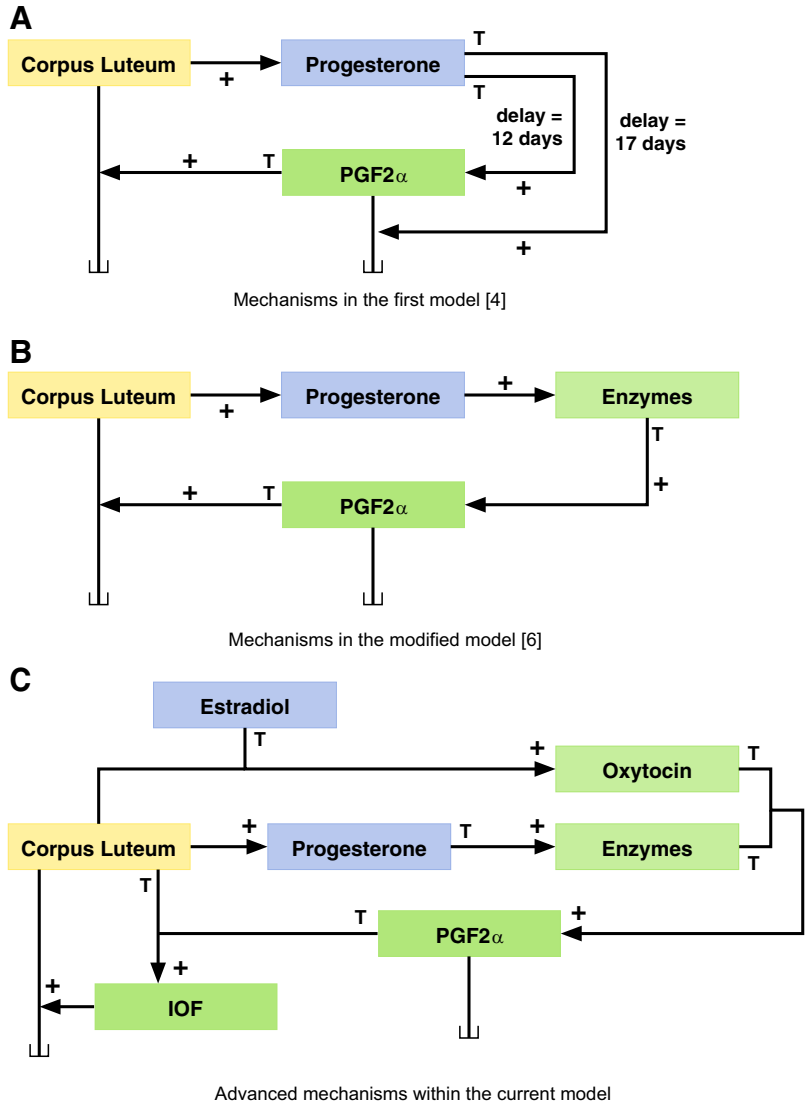

Fig. 2. Changes in the mechanisms involved in luteolysis. "+" marks a stimulatory effect, "T" denotes a threshold within a Hill function. No description means a transition, and "⊔" marks a degraded substance. In (A), P_4 was driving luteolysis with large time delays. In (B), enzymes were responsible for the rise of $PGF_{2\alpha}$. In the advanced model (C), we have added oxytocin and interovarian factors to the drivers of luteolysis.

2.2.2. *Improvement of luteolysis*

In [4], the rise of $PGF_{2\alpha}$ triggering the decay of the CL was modeled as a black box, depending with large delays on P_4 only. In [6], this was improved as enzymes were introduced that stimulate $PGF_{2\alpha}$, and the model became more robust. However, in simulating the administration of $PGF_{2\alpha}$ we detected that the modeling of luteolysis still had some shortcomings. It is known that after the administration of $PGF_{2\alpha}$ the responsive CL decays immediately [21]. In the original model, complete luteolysis took too much time after administration of PGF_{syn}, and also after rise of the regular $PGF_{2\alpha}$. Simply accelerating the CL decay lead to an accelerated fall of P_4 levels, and thus to very low P_4 levels already at the beginning of the second follicular wave, which had undesired consequences on the rest of the model output. Thus, since P_4 levels should stay on a high level for the duration of the first two follicular waves, we required the CL still being active at this time. Therefore, the starting point of its decay must occur later. This means we need the initiator of luteolysis,

$PGF_{2\alpha}$, to appear a couple of days later compared to the original model. To account for this effect, we now model the mechanisms that lead to a rise in $PGF_{2\alpha}$ differently. The development of the model regarding luteolysis is illustrated in Figure 2.

Instead of leaving only the enzymes (Enz) being responsible for $PGF_{2\alpha}$ levels as in [6], we now also include OT as another initiator of $PGF_{2\alpha}$ [22]. The E_2 stimulates OT synthesis in the granulosa cells [23] and the effect of OT on $PGF_{2\alpha}$ [13]. We assume that OT production depends on the surface of the CL and thus quadratically on CL size, and that it is cleared with constant rate c_{OT}. The equation for the rise and fall of OT is now

$$\frac{d}{dt}OT(t) = H_{17}^{+}(E2) \cdot CL(t)^2 - c_{OT} \cdot OT(t).$$

The OT together with Enz are now responsible for the rise of $PGF_{2\alpha}$. With the function $H_{16}^{+}(Enz\ \&\ OT)$, which represents a stimulatory effect if the levels of Enz and OT are both high, and the constant clearance rate c_{PGF}, the equation for $PGF_{2\alpha}$ becomes

$$\frac{d}{dt}PGF(t) = H_{19}^{+}(Enz\&OT) - c_{PGF} \cdot PGF(t).$$

In the former model, $PGF_{2\alpha}$ triggered luteolysis directly, independent of estrous stage. However, it is known that the CL is resistant to the action of $PGF_{2\alpha}$ at early luteal stage. We therefore remodeled the action of $PGF_{2\alpha}$ on the CL. According to [11], the direct action of $PGF_{2\alpha}$ on the CL is mediated by local factors: endothelin-1-system, cytokines, and nitric oxide. The expression of these interovarian substances is upregulated by $PGF_{2\alpha}$, and strictly depends on the stage of the CL. Summarizing these local factors, we introduce a new component to our model and call it interovarian factors (IOF). IOF is stimulated by $PGF_{2\alpha}$ only if the CL has reached a certain size, and is cleared with constant rate c_{IOF},

$$\frac{d}{dt}IOF(t) = H_{18}^{+}(PGF\&CL) - c_{IOF} \cdot IOF(t).$$

The rise of the interovarian factors now induces luteolysis.

2.2.3. Improvement of ovulation

In the original model, LH was the initiator of ovulation, responsible for decay of the dominant follicle, and at the same time the initiator of the rise of the CL 4.5 days after the LH peak. A delay differential equation was needed to model this effect. The atretic follicles disappeared from the system, and the CL emerged independently of the size of the just ovulated dominant follicle.

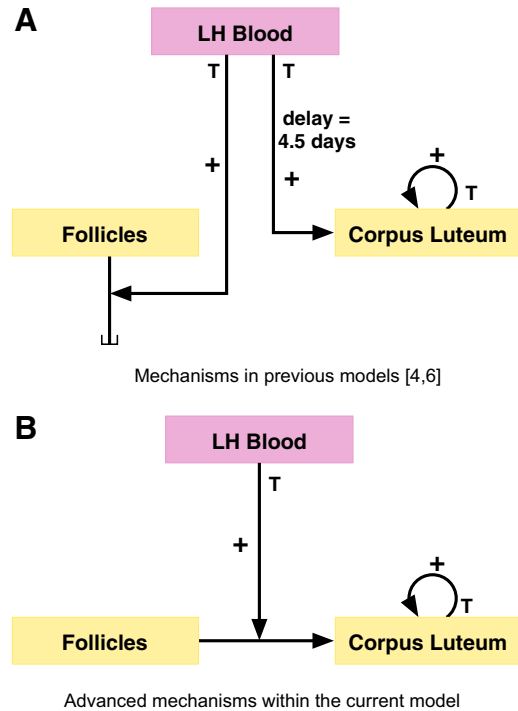

Fig. 3. Changes in the mechanisms involved in ovulation. "+" marks a stimulatory effect, "T" denotes a threshold within a Hill function. Arrows without description mean a transition, and "⊔" marks a degraded substance. In both models, ovulation includes the degradation of the follicles, and the formation of the CL, triggered by LH. In previous models (A), the CL arose independently from the size of the follicles, only regulated by the timepoint of the LH peak. In (B), a new dependency is introduced as the degraded follicles transform to the newly arising CL, making thus the size of the CL directly dependent on the follicles.

However, it is known that thecal and granulosa cells of the ruptured follicle transform to small and large luteal cells which form the rising CL [24]. Therefore, to make the model more realistic and to be able to account for different sizes of the dominant follicle [25], we changed the involved mechanisms. The ovulatory follicle now directly influences the initiation of CL growth, and no further delay differential equation is needed. The old and new mechanisms are illustrated in Figure 3. The equations for the follicular size (denoted Foll) and the CL are modified as follows:

$$\frac{d}{dt}Foll(t) = \widetilde{H}_{11}^{+}(FSH_{Bld}) - (H_{12}^{+}(P4) + H_{13}^{+}(LH_{Bld})) \cdot Foll(t),$$

$$\frac{d}{dt}CL(t) = SF \cdot H_{13}^{+}(LH_{Bld}) \cdot Foll(t) + H_{14}^{+}(CL) - H_{15}^{+}(IOF) \cdot CL(t).$$

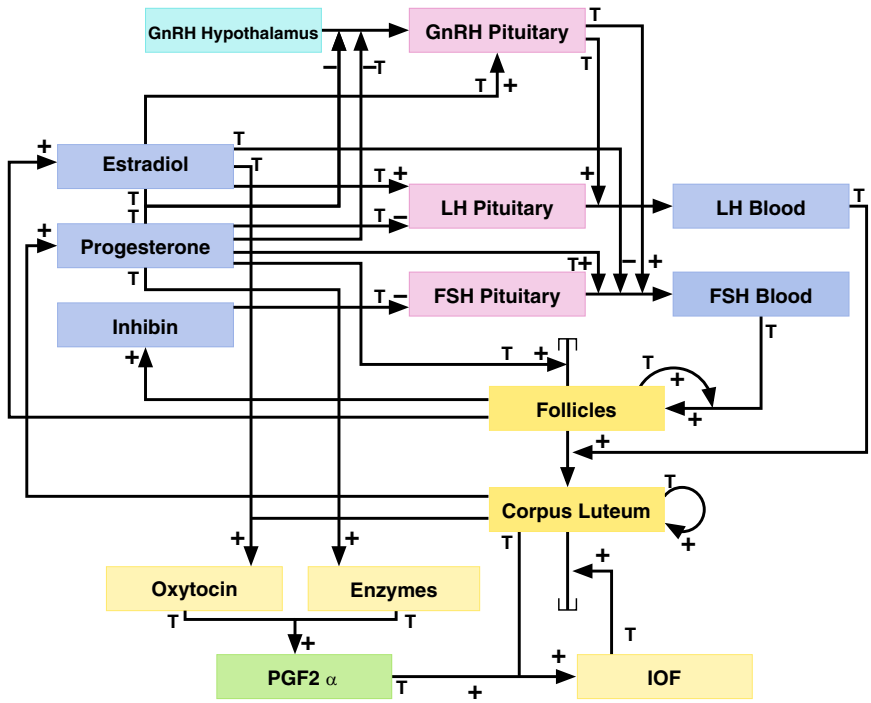

Fig. 4. Schematic representation of the components and relations in the herein presented advanced model of the bovine estrous cycle. Each box represents a substance for which a differential equation is derived, the color resp. grayscale of a box indicates the different physiological compartments where the substance mainly occurs: hypothalamus (GnRH), pituitary (GnRH, LH, FSH), blood (LH, FSH, estradiol, progesterone, inhibin), ovaries (follicles, corpus luteum, oxytocin, enzymes, IOF), uterus ($PGF_{2\alpha}$). The arrows represent the regulations between the substances in the model, where "+" marks a stimulatory effect, "T" denotes a threshold within a Hill function. No description means a transition, and '⊔' marks a degraded substance.

In the model, the part of the follicles decaying due to LH, i.e., the ovulated follicle, is now preserved in the system, forming the rising CL. The scaling factor *SF* is included to keep the relative levels of the substances between 0 and 1. Further growth of the CL is still modeled by a self-growth, i.e., a positive influence of the CL on its own size from a certain size on. Since the CL therefore starts to grow earlier now, the threshold and rate of self-growth have been adjusted.

2.2.4. Further modifications

Because the development of the CL depends on three mechanisms (an initiating impulse from LH, a self-growth and the decay caused by the interovarian factors), the level of the CL changes as follows: Right after ovulation the CL starts to grow, reaches the size needed for self-growth, and then grows with constant rate until the rise of $PGF_{2\alpha}$. In the original model, with the former mechanisms of delayed rise of the CL after ovulation, the course of CL was interpreted as "capacity to produce P_4", and thus the P_4 profile looked similar. Available experimental data for P_4 lead us to modify its production in the model. In particular, data from a study with single dose $PGF_{2\alpha}$ administration showed that the P_4 profile stayed low for about 10 days after the $PGF_{2\alpha}$ administration. Therefore, and since it is known that P_4 production of the CL is not absolutely proportional to the CL size [26], the mechanisms leading to the rise of P_4 were adjusted to obtain a P_4 production which is lower at start of CL growth compared to later luteal stages. We now interpret the equation for CL as development of the "size of the CL", and assume that production of P_4 depends on the surface size and thus quadratically on CL. The equation for P_4 becomes

$$\frac{d}{dt}P4(t) = c_{CL}^{P4} \cdot CL(t)^2 - c_{P4} \cdot P4(t).$$

In the former model, the capacity of the follicles to produce E_2 and inhibin (Inh) was described in one equation, and E_2 and Inh levels were proportional to the relative level of this component. For consistency reasons we now also assume a quadratic relationship between the follicles and E_2, respectively Inh, and the corresponding equations become

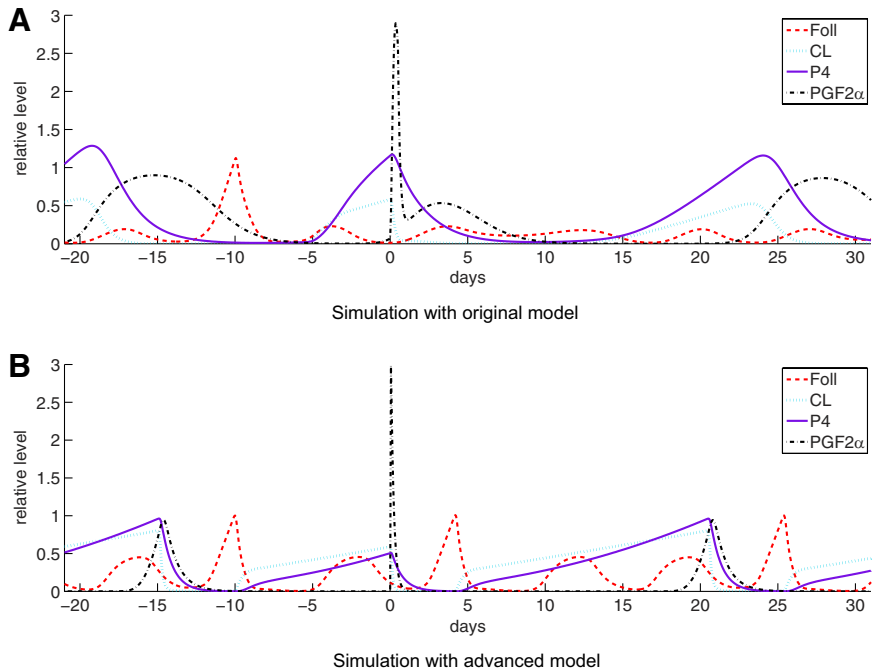

Fig. 5. Comparison of the simulation output of the previous model (shown in A) and the advanced model (shown in B) after a single administration of PGF_{syn} on day zero. Under investigation are Foll, CL, P_4, and total $PGF_{2\alpha}$. A high peak of Foll indicates ovulation. In both plots, the last ovulation has occurred ten days before administration. The original model does not capture the known effects after PGF_{syn} application, but the improved model in particular shows a rapid CL and P_4 decline and ovulation in the next follicular wave.

$$\frac{d}{dt}E2(t) = c_{Foll}^{E2} \cdot Foll(t)^2 - c_{E2} \cdot E2(t),$$

$$\frac{d}{dt}Inh(t) = c_{Foll}^{Inh} \cdot Foll(t)^2 - c_{Inh} \cdot Inh(t).$$

Diminishing the former delay for inhibin on FSH has been possible by augmenting the threshold for inhibin until its negative influence on FSH synthesis arises, at the same time steepening the regulatory effect on FSH. Moreover, the production rate of inhibin as well as its clearance rate have been lowered in order to defer the simulated inhibin curve. The FSH threshold for its influence on the follicles has also been increased. The fact that we were able to dispose of this delay without changing the differential equation at all was only possible because the delay was quite short (1.41 days in [6]).

A flow chart of the complete mechanisms of the model is shown in Figure 4.

3. Results and discussion

The here presented advanced model of the bovine estrous cycle consists of 15 ordinary differential equations and 60 parameters, generating successive estrous cycles of 21 days with three follicular waves per cycle. It does not contain time delays anymore. Therefore, there is no longer a need for a delay differential equation solver. We now use a linear implicit Euler method with extrapolation, implemented in the code LIMEX [27]. Parameters are identified with the software NLSCON developed at the Zuse Institute. This software uses subtle mathematical techniques, such as affine covariant Gauss-Newton methods that take into account sensitivities and linear dependencies of the parameters [28,29].

Our model of the bovine estrous cycle is dimensionless in the sense of [30], i.e., the numerical values of the components are independent of the standard of measurement. Simulated hormone levels and ovarian components have been scaled to be between 0 and 1 by dividing the equation by its maximum output level. Once we have measurement data available we will scale the functions to the corresponding quantities by scaling the involved parameters. This can be done because until now none of the parameters has a fixed value verified by experiments. We refer to the simulated dimensionless output functions as relative levels.

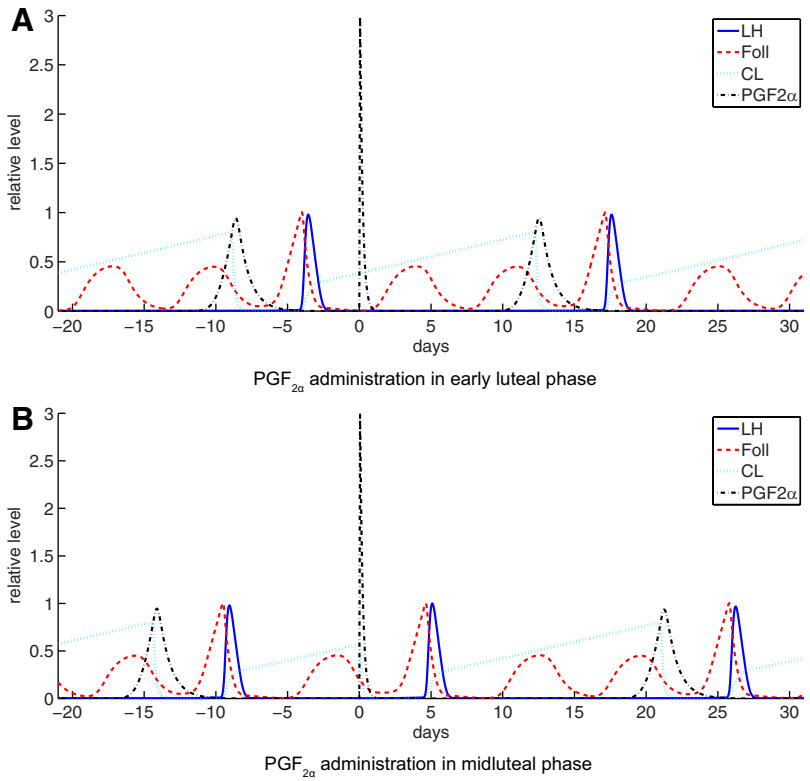

Fig. 6. Simulation results with the advanced model for Foll, CL, and LH, and total $PGF_{2\alpha}$ after administration of PGF_{syn} on different stages of the cycle. A high peak of Foll indicates ovulation. In (A) it is shown that there is no response if ovulation has occurred only four days before administration, in (B) we can observe that giving $PGF_{2\alpha}$ ten days after ovulation leads to a decay in CL, an LH peak and ovulation within five days after administration.

3.1. Simulation of single administration of $PGF_{2\alpha}$ at different stages of the cycle

The changes in the model described in the previous section have led to the following changes in the simulation results (Fig. 5). In contrast to the previous model, after administration of $PGF_{2\alpha}$ on Day 10 after ovulation the CL now decays immediately to zero. P_4 levels follow shortly after. Right after administration, $PGF_{2\alpha}$ does not have high levels right anymore, but stays low for 21 days. The most important difference between the outcome of the former and the advanced model can be observed in the follicles. Before, the administration of $PGF_{2\alpha}$ did not affect regular function, anovulatory waves stayed anovulatory. Now, the next arising follicular wave does not decay but continues to rise, leading to ovulation.

In Figure 6 we can observe that virtual administration of $PGF_{2\alpha}$ in the early luteal stage does not lead to a decay of the CL, while at later time points of the cycle it results in an immediate decay of the responsive CL, an LH peak, and ovulation during the following follicular wave.

3.2. Simulation of repeated administration of $PGF_{2\alpha}$ at different stages of the cycle

As described in Section 2.1, protocols of estrus synchronization often contain two administrations of $PGF_{2\alpha}$. It is known that the success of these protocols depends on the time interval between the two doses [31]. We thus investigated several time intervals with our model. In Figure 7, the effect of this application on the follicles is shown. On day zero, $PGF_{2\alpha}$ is administered to six cows in different phases of their cycles. A high peak in the course of Foll of a cow corresponds to the ovulatory wave of this cow. As can be observed in Figure 7A, a single dose of $PGF_{2\alpha}$ results in an ovulation in the next follicular wave in some cows (cow3, cow4, cow5, cow6), while in other cows (cow1 and cow2) it does not have any effect. In Figure 7B, a second dose of $PGF_{2\alpha}$ is given to the same cows seven days after the first application. Here too, some cows are affected by this administration (cow1 and cow2 are responsive), while others are not affected. This is due to the mechanisms driving luteolysis, where the

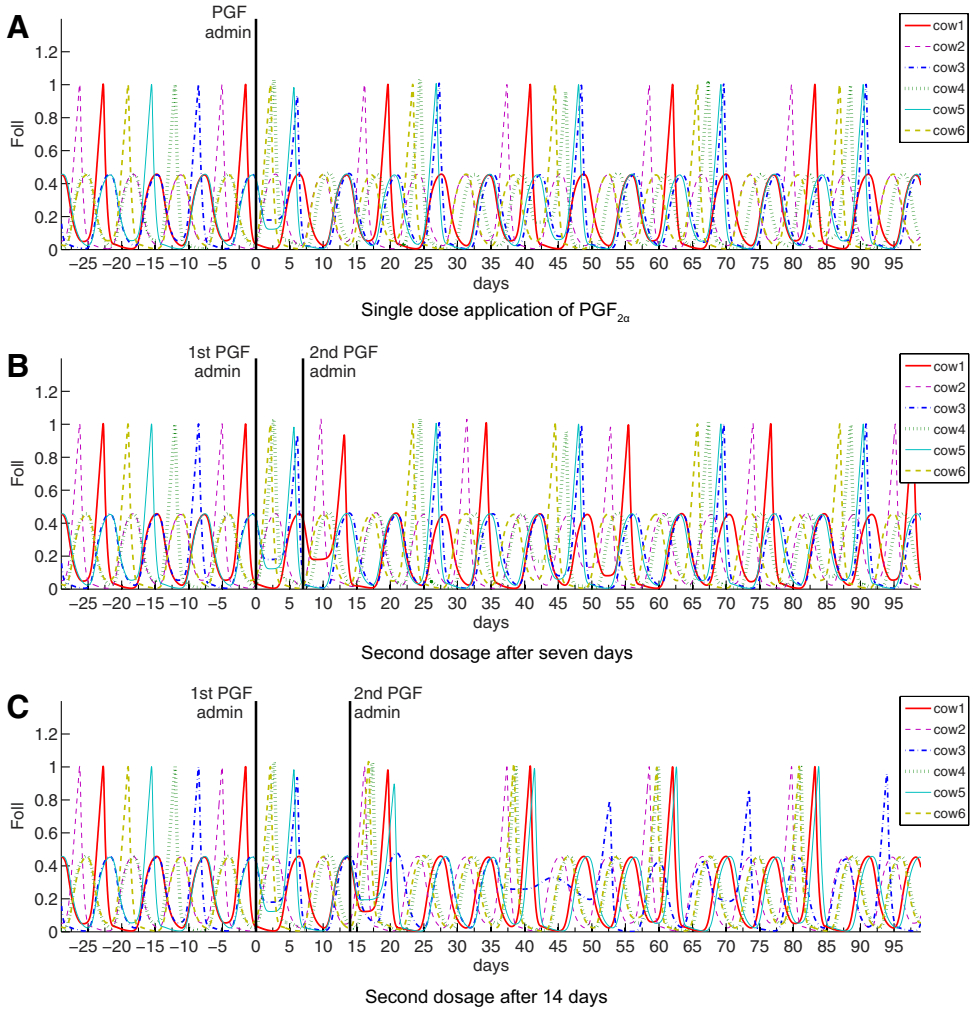

Fig. 7. Simulation results for the follicles of six different cows during virtual administration of $PGF_{2\alpha}$. Day zero denotes the day of first administration, when the cows are each in a different stage of their estrous cycle, the high peaks of the curves denoting the timepoints of ovulation. In (A), a single dose of $PGF_{2\alpha}$ is given, which impacts the cycle of at least two cows (cow3 and cow4). In (B), a second dose is given seven days after the first dosage, now influencing the cycle of two other cows (cow1 and cow2). In (C), the second dose is given 14 days after the first dose, now influencing four cows (cow3, cow4, cow5 and cow6).

CL is not responsive to $PGF_{2\alpha}$ early after its rise. The cows affected by the first dosage are now in the early luteal stage and thus not responsive to the second dosage, while the cows not responsive at the first administration are now in a later luteal stage where they respond to $PGF_{2\alpha}$. This suggests giving the second application when the cows affected by the first dosage are again in the phase of their estrous cycle where they have a responsive CL. In Figure 7C, the second dose is given 14 days after the first dose, leading to a synchronization effect within almost all of the cows under investigation.

Before the first administration to the cows in Figure 7, the ovulation timepoints of the cows were evenly distributed, while after double administration within 14 days they were brought much closer together, such that the cows ovulate within a timespan of 4.5 days. This is consistent with the double injection 11 to 14 days apart in common synchronization protocols as reviewed in [31]. Only for some cows in a certain stage of their cycle (e.g., cow 5 in Fig. 7C), the synchronization effect in our model is not as desired, which is in line with observations after presynch protocols [32]. Nevertheless, this result validates our model in the sense that it captures the known effects after this certain external manipulation of the bovine estrous cycle.

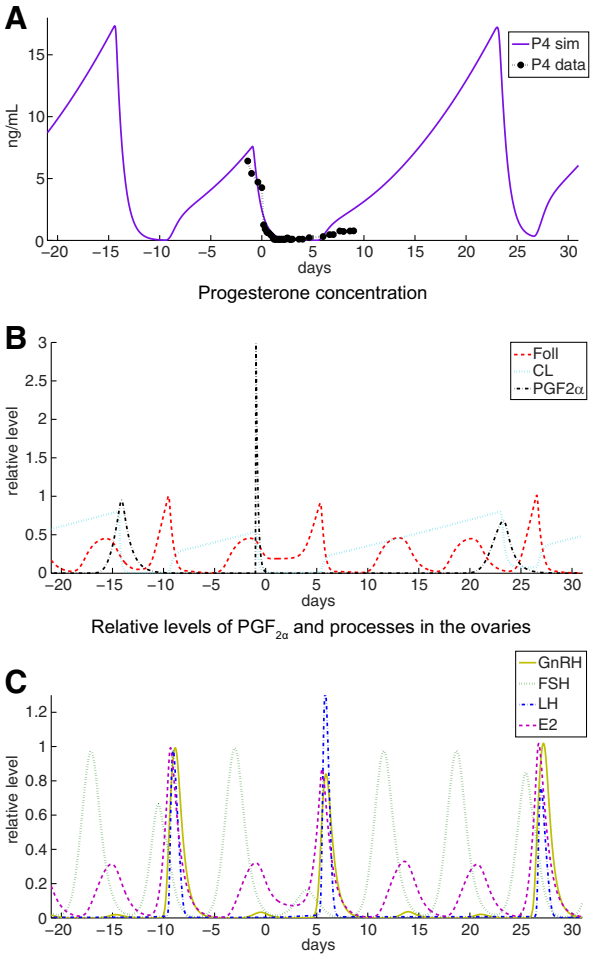

Fig. 8. Simulation of a single dose of PGFα and its impact on other components of the model. Parameters have been fitted such that the simulated P_4 levels match the given experimental P_4 data. With the set of identified parameters, we can investigate the course of the other model components. In particular, we can observe that ovulation occurs six days after injection. (C) Relative levels of selected other hormones.

3.3. Comparison with progesterone measurements after single administration of $PGF_{2\alpha}$

In a recent study performed at the institute of animal reproduction, Department of Veterinary Medicine at Freie Universität Berlin, a single dose of 5 mg $PGF_{2\alpha}$ was injected to seven cows, and plasma progesterone concentrations were measured before and after the administration. In particular, blood was collected every morning (8:00h) and evening (17:00h) before the injection, every four h after the injection, and twice a day after ovulation, detected by ultrasound.

Model parameters have been identified so that the simulated P_4 level matches the given data. Note that we now observe and simulate concentrations instead of relative levels for progesterone. Certain parameter units therefore have to be adapted adequately. In Figure 8A an example of measured P_4 concentrations for one of the examined cows is shown, together with the simulated P_4 concentration. Ovulation has been detected by ultrasound a couple of days after the PG injection. This is well captured in the simulation. Not only does this approve our model, we can also observe substances that are not measured within the experiment, and our simulation gives us insight into the development of these substrates after a single $PGF_{2\alpha}$ injection. For example, in Figure 8C we observe a GnRH peak after administration of $PGF_{2\alpha}$, which can be understood as an increase in pulse frequency and is in the scope of expected observations.

4. Conclusions

In this work we have enhanced the model of the bovine estrous cycle which was introduced in [4,6] towards the simulation of synchronization protocols. We have replaced the mechanisms regarding ovulation and refined the modeling of luteolysis. The new components representing oxytocin and interovarian factors have been introduced, integrated, and connected to the other components of the model. To eliminate time delays, certain growth and decay rates, as well as several thresholds and steepness factors have been adjusted. To account for effects observed in experimental data, the relationship between CL growth and the rise of P_4 levels has been modified, the action of Foll has been adjusted accordingly. We have validated our model by capturing the synchronization effect of double $PGF_{2\alpha}$ administration. In our model, the responsiveness of the corpus luteum is the decisive factor for synchronization effects after $PGF_{2\alpha}$ injection.

We have shown simulation results for cows with three follicular waves per cycle. Different parameterizations can also lead to cycles with different wave numbers or irregular wave patterns, but those simulations would go beyond the scope of this work. The original motivation of developing a model of the bovine estrous cycle [4] was to better understand the underlying biological mechanisms and dynamics, but this motivation is not exclusive.

In the future, the model of the bovine estrous cycle could be used within study planning or evaluation, for educational purposes or to assist management decisions. The model could be integrated into existing models, e.g., metabolic networks [1]. New models could also be developed for other aspects interacting with the cycle, e.g., stress, negative energy balance, or milk

production, and then coupled to the model presented in this work. Further applications could be the modeling of pathologic situations, e.g., cystic ovarian disease, anestrous, or inflammation. The model could be used to deeper investigate their interaction with fertility hormones of the cow. Further, an optimal control problem could be formulated to design synchronization protocols regarding optimal dosing and frequency. A future refinement could require the inclusion of reactions that take place on single-cell level, e.g., receptor binding mechanisms as in [33]. The level of detail will be adjusted according to the applications. The direction of future research will highly depend on future collaborations and input from animal reproduction experts. The prospectives will thus depend crucially on the available experimental data and interest of animal scientists in systems biology methods.

Acknowledgments

The authors would like to thank Marike Boer for discussions about luteolysis, Vishal Suthar for providing the progesterone data from the $PGF_{2\alpha}$ study, and Peter Deuflhard for the provided support. Susanna Röblitz and Claudia Stötzel have been supported by the DFG Research Center MATHEON 'Mathematics for Key Technologies' in Berlin, Germany.

References

[1] Baldwin RL. Modeling Ruminant Digestion and Metabolism. Chapman & Hall 1995.
[2] Hill SR, Knowlton KF, Kebreab E, France J, Hanigan MD. A model of phosphorus digestion and metabolism in the lactating dairy cow. J Dairy Sci 2008;91:2021–32.
[3] Dijkstra J, Neal HD, Beever DE, France J. Simulation of nutrient digestion, absorption and outflow in the rumen: model description. J Nutr 1992;122:2239–56.
[4] Boer HMT, Stötzel C, Röblitz S, Deuflhard P, Veerkamp RF, Woelders H. A simple mathematical model of the bovine estrous cycle: follicle development and endocrine interactions. J Theor Biol 2011;278:20–31.
[5] Woelders H, Te Pas MFW, Bannink A, Veerkamp RF, Smits MA. Systems biology in animal sciences. Animal 2011;5:1036–47.
[6] Boer HMT, Röblitz S, Stötzel C, Veerkamp RF, Kemp B, Woelders H. Mechanisms regulating follicle wave patterns in the bovine estrous cycle investigated with a mathematical model. J Dairy Sci 2011;94:5987–6000.
[7] Stevenson JL, Dalton JC, Santos JE, Sartori R, Ahmadzadeh A, Chebel RC. Effect of synchronization protocols on follicular development and estradiol and progesterone concentrations of dairy heifers. J Dairy Sci 2008;91:3045–56.
[8] Schiewe MC, Bush M, Phillips LG, Citino S, Wildt DE. Comparative aspects of estrus synchronization, ovulation induction, and embryo cryopreservation in the scimitar-horned oryx, bongo, eland, and greater kudu. J Exp Zool 1991;258:75–88.
[9] Lauderdale JW. ASAS centennial paper: contributions in the journal of animal science to the development of protocols for breeding management of cattle through synchronization of estrus and ovulation. J Anim Sci 2009;87:801–12.
[10] McCracken JA, Custer EE, Lamsa JC. Luteolysis: A neuroendocrine-mediated event. Physiol Rev 1999;79:263–323.
[11] Skarzynski DJ, Ferreira-Dias G, Okuda K. Regulation of luteal function and corpus luteum regression in cows: hormonal control, immune mechanisms and intercellular communication. Reprod Domest Anim 2008;43:57–65.
[12] Silvia WJ, Lewis GS, McCracken JA, Thatcher WW, Wilson L Jr. Hormonal regulation of uterine secretion of prostaglandin F2α during luteolysis in ruminants. Biol Reprod 1991;45:655–63.
[13] Asselin E, Goff AK, Bergeron H, Fortier MA. Influence of sex steroids on the production of prostaglandins F2α and E2 and response to oxytocin in cultured ephithelial and stromal cells of the bovine endometrium. Biol Reprod 1996;54:371–99.
[14] Xiao CW, Liu JM, Sirois J, Goff AK. Regulation of cyclooxygenase-2 and prostaglandin F synthase gene expression by steroid hormones and interferon-τ in bovine endometrial cells. Endocrinology 1998;139:2293–9.
[15] Schallenberger E, Schams D., Bullermann B, Walters DL. Pulsatile secretion of gonadotrophins, ovarian steroids and ovarian oxytocin during prostaglandin-induced regression of the corpus luteum in the cow. J Reprod Fertil 1984;71:493–501.
[16] Stellflug JN, Louis TM, Hafs HD, Seguin BE. Luteolysis, estrus and ovulation, and blood prostaglandin F after intramuscular administration of 15, 30 or 60 mg prostaglandin F2α. Prostaglandins 1975;9:609–15.
[17] Kroker R. Beeinflussung der Uterusfunktion KR. In: Löscher W, Ungemach F, Kroker R, editors. Pharmakotherapie bei Haus- und Nutztieren; 6th ed. Buchverlag: Parey; 2003, p. 168–71.
[18] Ginther OJ, Araujo RR, Palhão MP, Rodrigues BL, Beg MA. Necessity of sequential pulses of prostaglandin F2α for complete physiologic luteolysis in cattle. Biol Reprod 2009;80:641–8.
[19] Reinecke I. Mathematical modeling and simulation of the female menstrual cycle. Ph.D. Thesis; Freie Universität Berlin; 2009.
[20] Keenan DM, Veldhuis JD. A biomathematical model of time-delayed feedback in the human male hypothalamic-pituitary-leydig cell axis. Am J Physiol Endocrinol - Endocrinol Metab 1998;275:E157–76.
[21] Hixon JE, Hansel W. Evidence for preferential transfer of prostaglandin F2α to the ovarian artery following intrauterine administration in cattle. Biol Reprod 1974;11:543–52.
[22] Kotwica J, Skarzynski D, Miszkiel G, Melin P, Okuda K. Oxytocin modulates the pulsatile secretion of prostaglandin F2α in initiated luteolysis in cattle. Res Vet Sci 1999;66:1–5.
[23] Voss AK, Fortune JE. Estradiol-17β has a biphasic effect on oxytocin secretion by bovine granulosa cells. Biol Reprod 1993;48:1404–9.
[24] Rathbone MJ, Kinder JE, Fike K, Kojima F, Clopton D, Ogle CR, Bunt C. Recent advances in bovine reproductive endocrinology and physiology and their impact on drug delivery system design for the control of the estrous cycle in cattle. Adv Drug Deliv Rev 2001;50(**3**);50:277–320.

[25] Rodgers RJ, Irving-Rodgers HF. Morphological classification of bovine ovarian follicles. Reproduction 2010;139:309–18.
[26] Kastelic JP, Bergfelt DR, Ginther OJ. Relationship between ultrasonic assessment of the corpus luteum and plasma progesterone concentration in heifers. Theriogenology 1990;33: 1269–78.
[27] Deuflhard P, Nowak U. Extrapolation integrators for quasilinear implicit ODEs. Progr Sci Comp 1987;7:37–50.
[28] Nowak U, Deuflhard P. Numerical identification of selected rate constants in large chemical reaction systems. Appl Numer Math 1985;1:59–75.
[29] Deuflhard P, Nowak U. Efficient numerical simulation and identification of large chemical reaction systems. Berichte der Bunsenges Bunsengesellschaft für Phys Chem 1986;90:940–6.
[30] Lin CC, Segel LA. Mathematics Applied to Deterministic Problems in the Natural Sciences; vol. 11. Society for Industrial Mathematics; 1988.
[31] Murugavel K, Yániz JL, Santolaria P, López-Béjar M, López-Gatius F. Prostaglandin based estrus synchronization in postpartum dairy cows: an update. J Appl Res Vet Med 2003;1:51–65.
[32] Stevenson JS, Phatak AP. Rates of luteolysis and pregnancy in dairy cows after treatment with cloprostenol or dinoprost. Theriogenology 2010;73:1127–38.
[33] Röblitz S, Stötzel C, Deuflhard P, Jones H, Azulay DO, van der Graaf P, Martin S. A Mathematical Model of the Human Menstrual Cycle for the Administration of GnRH Analogues. ZIB-Report 11-16. Berlin: Zuse Institute Berlin; 2011, p. 11–6. Available at: http://vs24.kobv.de/opus4-zib/frontdoor/index/index/docId/1273.

Appendix A. Equations

The model describing the bovine estrous cycle without external manipulation consists of 15 ordinary differential equations with 60 parameters. For virtual administration of $PGF_{2\alpha}$ we use one additional ordinary differential equation (ODE) containing three extra parameters.

GnRH:

$$\frac{d}{dt}GnRH_{Hypo}(t) = Syn_{GnRH}(t) - Rel_{GnRH}(t) \quad (1)$$

$$Syn_{GnRH}(t) = c_{GnRH,1} \cdot \left(1 - \frac{GnRH_{Hypo}(t)}{GnRH_{Hypo}^{\max}}\right)$$

$$Rel_{GnRH}(t) = (H_1^-(P4\,\&\,E2) + H_2^-(P4)) \cdot GnRH_{Hypo}(t)$$

$$\frac{d}{dt}GnRH_{Pit}(t) = Rel_{GnRH}(t) \cdot H_3^+(E2) - c_{GnRH,2} \cdot GnRH_{Pit}(t) \quad (2)$$

FSH:

$$\frac{d}{dt}FSH_{Pit}(t) = Syn_{FSH}(t) - Rel_{FSH}(t) \quad (3)$$

$$Syn_{FSH}(t) = H_4^-(Inh)$$

$$Rel_{FSH}(t) = (b_{FSH} + H_5^+(P4) + H_6^-(E2) + H_7^+(GnRH_{Pit})) \cdot FSH_{Pit}(t)$$

$$\frac{d}{dt}FSH_{Bld}(t) = Rel_{FSH}(t) - c_{FSH} \cdot FSH_{Bld}(t) \quad (4)$$

LH:

$$\frac{d}{dt}LH_{Pit}(t) = Syn_{LH}(t) - Rel_{LH}(t) \quad (5)$$

$$Syn_{LH}(t) = H_8^+(E2) + H_9^-(P4)$$

$$Rel_{LH}(t) = (b_{LH} + H_{10}^+(GnRH_{Pit})) \cdot LH_{Pit}(t)$$

$$\frac{d}{dt}LH_{Bld}(t) = Rel_{LH}(t) - c_{LH} \cdot LH_{Bld}(t) \quad (6)$$

Follicles and corpus luteum:

$$\frac{d}{dt}Foll(t) = \widetilde{H}_{11}^+(FSH_{Bld}) - (H_{12}^+(P4) + H_{13}^+(LH_{Bld})) \cdot Foll(t) \quad (7)$$

$$\frac{d}{dt}CL(t) = SF \cdot H_{13}^+(LH_{Bld}) \cdot Foll(t) + H_{14}^+(CL) - H_{15}^+(IOF) \cdot CL(t) \quad (8)$$

Hormones produced in the ovaries:

$$\frac{d}{dt}P4(t) = c_{CL}^{P4} \cdot CL(t)^2 - c_{P4} \cdot P4(t) \quad (9)$$

$$\frac{d}{dt}E2(t) = c_{Foll}^{E2} \cdot Foll(t)^2 - c_{E2} \cdot E2(t) \quad (10)$$

$$\frac{d}{dt}Inh(t) = c_{Foll}^{Inh} \cdot Foll(t)^2 - c_{Inh} \cdot Inh(t) \quad (11)$$

Enzymes, oxytocin and inter-ovarian factors:

$$\frac{d}{dt}Enz(t) = H_{16}^+(P4) - c_{Enz} \cdot Enz(t) \quad (12)$$

$$\frac{d}{dt}OT(t) = H_{17}^+(E2) \cdot CL(t)^2 - c_{OT} \cdot OT(t) \quad (13)$$

$$\frac{d}{dt}IOF(t) = H_{18}^+(PGF\,\&\,CL) - c_{IOF} \cdot IOF(t) \quad (14)$$

$PGF_{2\alpha}$ and synthetic prostaglandin

$$\frac{d}{dt}PGF(t) = H_{19}^{+}(Enz\,\&\,OT) \; - \; c_{PGF} \; \cdot \; PGF(t) \qquad (15)$$

$$\frac{d}{dt}PGF_{syn}(t) = \; D \; \cdot \; \beta^2 \; \cdot \; t_{mod}(t) \; \cdot \; \exp(-\beta \; \cdot \; t_{mod}(t) \;) \; - \; c_{PGF_{syn}} \; \cdot \; PGF_{syn}(t)$$

$$t_{mod}(t) : \; = \; \max(0, t - t_D)$$

Appendix B. List of hill functions

Positive resp. negative Hill functions are defined as

$$h^{+}(S(t);T,n) : = \frac{S(t)^n}{T^n + S(t)^n}, \qquad h^{-}(S(t);T,n) : = \frac{T^n}{T^n + S(t)^n}.$$

The Hill functions listed below are the full notations of the Hill functions mentioned in Section 2.2, in Appendix A. They represent a majority of the mechanisms shown in Fig. 4.

$$H_1^{-}(P4\,\&\,E2) = m_{P4\&E2}^{GnRH} \; \cdot \; (h^{-}(P4(t);\, T_{P4}^{GnRH,1}, 2) \; + \; h^{-}(E2(t),\; T_{E2}^{GnRH,1},\; 2) \; - h^{-}(P4(t);\, T_{P4}^{GnRH,1}, 2) \; \cdot \; h^{-}(E2(t),\; T_{E2}^{GnRH,1}, 2))$$

$$H_2^{-}(P4) = m_{P4}^{GnRH,2} \; \cdot \; h^{-}(P4(t),\; T_{P4}^{GnRH,2}, 2)$$

$$H_3^{+}(E2) = m_{E2}^{GnRH,2} \; \cdot \; h^{+}(E2(t),\; T_{E2}^{GnRH,2}, 5)$$

$$H_4^{-}(Inh) = m_{Inh}^{FSH} \; \cdot \; h^{-}(Inh(t),\; T_{Inh}^{FSH}, 5)$$

$$H_5^{+}(P4) = m_{P4}^{FSH} \; \cdot \; h^{+}(P4(t);\, T_{P4}^{FSH}, 2)$$

$$H_6^{-}(E2) = m_{E2}^{FSH} \; \cdot \; h^{-}(E2(t);\, T_{E2}^{FSH}, 2)$$

$$H_7^{+}(GnRH_{Pit}) = m_{GnRH}^{FSH} \; \cdot \; h^{+}\,(GnRH_{Pit}(t);\, T_{GnRH}^{FSH}, 1)$$

$$H_8^{+}(E2) = m_{E2}^{LH} \; \cdot \; h^{+}(E2(t);\, T_{E2}^{LH}, 2)$$

$$H_9^{-}(P4) = m_{P4}^{LH} \; \cdot \; h^{-}(P4(t);\, T_{P4}^{LH}, 2)$$

$$H_{10}^{+}(GnRH_{Pit}) = \; m_{GnRH}^{LH} \; \cdot \; h^{+}(GnRH_{Pit}(t);\, T_{GnRH}^{LH}, 5)$$

$$\widetilde{H}_{11}^{+}(FSH_{Bld}) = \; m_{FSH}^{Foll} \; \cdot \; h^{+}(FSH_{Bld}(t);\, \widetilde{T}_{FSH}^{Foll}(t), 2), \;\; \widetilde{T}_{FSH}^{Foll}(t) = T_{FSH}^{Foll} \; \cdot \; h^{-}(Foll(t);\, T_{Foll}^{FSH},\; 2)$$

$$H_{12}^{+}(P4) = m_{P4}^{Foll} \; \cdot \; h^{+}(P4(t);\, T_{P4}^{Foll}, 5)$$

$$H_{13}^{+}(LH_{Bld}) = m_{LH}^{Ovul} \; \cdot \; h^{+}(LH_{Bld}(t);\, T_{LH}^{Ovul},\; 2)$$

$$H_{14}^{+}(CL) = \; m_{CL}^{CL} \; \cdot \; h^{+}(CL(t),\; T_{CL}^{CL},\; 2)$$

$$H_{15}^{+}(IOF) = \; m_{IOF}^{CL} \; \cdot \; h^{+}(IOF(t);\, T_{IOF}^{CL},\; 5)$$

$$H_{16}^{+}(P4) = m_{P4}^{Enz} \; \cdot \; h^{+}(P4(t);\, T_{P4}^{Enz},\; 5)$$

$$H_{17}^{+}(E2) = m_{E2}^{OT} \; \cdot \; h^{+}(E2(t);\, T_{E2}^{OT},\; 2)$$

$$H_{18}^{+}(PGF\,\&\,CL) = m_{PGF\&CL}^{IOF} \cdot \; h^{+}(PGF(t) \; + \; PGF_{syn}(t);\, T_{PGF}^{IOF},\; 5) \; \cdot \; h^{+}(CL(t);\, T_{CL}^{IOF}, n_{CL}^{IOF},\; 10)$$

$$H_{19}^{+}(Enz\,\&\,OT) = m_{Enz\&OT}^{PGF} \cdot \; h^{+}(Enz(t);\, T_{Enz}^{PGF},\; 5) \; \cdot \; h^{+}(OT(t);\, T_{OT}^{PGF},\; 2)$$

Appendix C. Parameter values

[·] stands for the unit of the substance, usually a concentration, and can be specified from measurements. Typical units are [FSH] = [LH] = IU/l, [P4] = ng/mL, and [E2] = pg/ml. If units of FSH and luteinizing hormone (LH) differ in pituitary and blood, release-terms have to be scaled adequately. t denotes "'time'"; in our model [·] stands for "'days'"

Par. No.	Symbol	Value	Unit
1	$GnRH_{Hypo}^{\max}$	16	$[GnRH_{Hypo}]$
2	$c_{GnRH,1}$	2.75	$\frac{[GnRH_{Hypo}]}{[t]}$
3	$m_{P4\&E2}^{GnRH}$	2.05	1/[t]
4	$T_{E_2}^{GnRH,1}$	0.0972	[E2]
5	$T_{P_4}^{GnRH,1}$	0.35	[P4]
6	$m_{P_4}^{GnRH,2}$	1.91	1/[t]
7	$T_{P_4}^{GnRH,2}$	0.252	[P4]
8	$m_{E_2}^{GnRH,2}$	0.99	$\frac{[GnRH_{Pit}]}{[GnRH_{Hypo}]}$
9	$T_{E_2}^{GnRH,2}$	0.648	[E2]
10	$c_{GnRH,2}$	1.63	1/[t]
11	m_{Inh}^{FSH}	4.21	[FSH]/[t]
12	T_{Inh}^{FSH}	0.118	[Inh]
13	b_{FSH}	0.948	1/[t]
14	$m_{P_4}^{FSH}$	0.293	1/[t]
15	$T_{P_4}^{FSH}$	0.152	[P4]
16	$m_{E_2}^{FSH}$	0.396	1/[t]
17	$T_{E_2}^{FSH}$	0.312	[E2]
18	m_{GnRH}^{FSH}	1.23	1/[t]
19	T_{GnRH}^{FSH}	0.0708	$[GnRH_{Pit}]$
20	c_{FSH}	2.73	1/[t]
21	$m_{E_2}^{LH}$	0.376	[LH]/[t]
22	$T_{E_2}^{LH}$	0.243	[E2]

Par. No.	Symbol	Value	Unit
23	$m_{P_4}^{LH}$	2.71	[LH]/[t]
24	$T_{P_4}^{LH}$	0.0269	[P4]
25	b_{LH}	0.0141	1/[t]
26	m_{GnRH}^{LH}	2.22	1/[t]
27	T_{GnRH}^{LH}	0.69	[$GnRH_{Pit}$]
28	c_{LH}	12.0	1/[t]
29	m_{FSH}^{Foll}	0.562	[Foll]/[t]
30	T_{FSH}^{Foll}	0.57	[FSH]
31	T_{Foll}^{FSH}	0.22	[Foll]
32	$m_{P_4}^{Foll}$	1.1	1/[t]
33	$T_{P_4}^{Foll}$	0.126	[P4]
34	m_{LH}^{Ovul}	3.49	1/[t]
35	T_{LH}^{Ovul}	0.171	[LH]
36	SF	0.2	[CL]/[t]
37	m_{CL}^{CL}	0.0353	[CL]/[t]
38	T_{CL}^{CL}	0.1	[CL]
39	m_{IOF}^{CL}	41.39	1/[t]
40	T_{IOF}^{CL}	1.32	[IOF]
41	$c_{CL}^{P_4}$	2.25	$\frac{[P_4]/[CL]^2}{[t]}$
42	cP_4	1.41	1/[t]
43	$c_{Foll}^{E_2}$	2.19	$\frac{[E2]/[Foll]^2}{[t]}$
44	c_{E2}	1.23	1/[t]
45	c_{Foll}^{Inh}	1.41	$\frac{[Inh]/[Foll]^2}{[t]}$
46	c_{Inh}	0.475	1/[t]
47	m_{P4}^{Enz}	3.58	[Enz]/[t]
48	T_{P4}^{Enz}	0.77	[P4]
49	c_{Enz}	2.98	1/[t]
50	m_{E2}^{OT}	1.59	$\frac{[OT]/[CL]^2}{[t]}$
51	T_{E2}^{OT}	0.143	[E2]
52	c_{OT}	0.644	1/[t]
53	$m_{PGF\&CL}^{IOF}$	39.68	[IOF]/[t]
54	T_{PGF}^{IOF}	1.22	[PGF]
55	T_{CL}^{IOF}	0.6	[CL]
56	c_{IOF}	0.298	1/[t]
57	$m_{Enz\&OT}^{PGF}$	53.91	[PGF]/[t]
58	T_{Enz}^{PGF}	1.43	[Enz]
59	T_{OT}^{PGF}	1.087	[OT]
60	c_{PGF}	1.23	1/[t]
	D	3.7	[PGF]
	β	100	1/[t]
	$c_{PGF_{syn}}$	5.5	1/[t]

Conditional Deletion of *Bmal1* in Ovarian Theca Cells Disrupts Ovulation in Female Mice

Amanda L. Mereness, Zachary C. Murphy, Andrew C. Forrestel, Susan Butler, CheMyong Ko, JoAnne S. Richards, and Michael T. Sellix

Department of Medicine (A.L.M., Z.C.M., A.C.F., S.B., M.T.S.), Division of Endocrinology, Diabetes and Metabolism, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642; Department of Comparative Biosciences (C.K.), College of Veterinary Medicine, University of Illinois Urbana-Champaign, Urbana, Illinois 61802; and Department of Molecular and Cellular Biology (J.S.R.), Baylor College of Medicine, Houston, Texas 77030

Rhythmic events in female reproductive physiology, including ovulation, are tightly controlled by the circadian timing system. The molecular clock, a feedback loop oscillator of clock gene transcription factors, dictates rhythms of gene expression in the hypothalamo-pituitary-ovarian axis. Circadian disruption due to environmental factors (eg, shift work) or genetic manipulation of the clock has negative impacts on fertility. Although the central pacemaker in the suprachiasmatic nucleus classically regulates the timing of ovulation, we have shown that this rhythm also depends on phasic sensitivity to LH. We hypothesized that this rhythm relies on clock function in a specific cellular compartment of the ovarian follicle. To test this hypothesis we generated mice with deletion of the *Bmal1* locus in ovarian granulosa cells (GCs) (Granulosa Cell *Bmal1* KO; GCKO) or theca cells (TCs) (Theca Cell *Bmal1* KO; TCKO). Reproductive cycles, preovulatory LH secretion, ovarian morphology and behavior were not grossly altered in GCKO or TCKO mice. We detected phasic sensitivity to LH in wild-type littermate control (LC) and GCKO mice but not TCKO mice. This decline in sensitivity to LH is coincident with impaired fertility and altered patterns of LH receptor (*Lhcgr*) mRNA abundance in the ovary of TCKO mice. These data suggest that the TC is a pacemaker that contributes to the timing and amplitude of ovulation by modulating phasic sensitivity to LH. The TC clock may play a critical role in circadian disruption-mediated reproductive pathology and could be a target for chronobiotic management of infertility due to environmental circadian disruption and/or hormone-dependent reprogramming in women. **(*Endocrinology* 157: 913–927, 2016)**

Precise timing of physiological events is critical for normal reproductive physiology. In female mammals rhythmic events in the reproductive tract are normally kept tightly entrained to (synchronized with) the light: dark (L:D) cycle (1, 2). A salient feature of the mammalian reproductive cycle is the preovulatory increase or "surge" in serum LH (3–6). In rats and mice, the LH surge occurs during the late afternoon and early evening of proestrus (3, 7). The daily pattern of ovulation is thought to be reliant on the timing of this gonadotropin surge (6, 8, 9). Rhythmic LH secretion depends on the activity of pacemaker neurons in the central clock or suprachiasmatic nucleus (SCN) (10, 11), GnRH neurons in the preoptic area and positive feedback from ovarian estradiol (12). Lesions that destroy the SCN or mutations that disrupt the activity of SCN neurons are known to affect both the timing and amplitude of the LH surge, block ovulation, and disrupt reproductive cycles (13, 14).

The substrate for circadian rhythms is a feedback loop of interacting transcriptional regulators or clock genes referred to collectively as the molecular clock. The core pacemaker includes the transcriptional activator BMAL1, its

ISSN Print 0013-7227 ISSN Online 1945-7170
Printed in USA

Received July 24, 2015. Accepted December 8, 2015.
First Published Online December 15, 2015

Abbreviations: CET, Cetrotide; COC, cumulus oocyte complex; CT, circadian time; DD, constant darkness; eCG, equine gonadotropin; eLH, equine LH; Fshr, FSH receptor; GC, granulosa cell; H+E, hematoxylin/eosin; HPO, hypothalamo-pituitary-ovarian; KO, knockout; LC, littermate control; L:D, light:dark; LHR, LH receptor; P4, progesterone; PND, postnatal day; SC, stromal cell; SCN, suprachiasmatic nucleus; SF-1, steroidogenic factor 1; TC, theca cell; ZT, zeitgeber time.

binding partner CLOCK and the repressors PERIOD (*Per1–Per3*) and CRYPTOCHROME (*Cry1* and *Cry2*) (15). Rhythms of *Bmal1* expression are maintained by the balanced activity of the transcriptional activator RORα and the repressor REV-ERBα, both regulated by the BMAL1:CLOCK complex (16, 17). Mutations altering or abolishing clock gene expression, especially *Bmal1*, have considerable negative impacts on fertility (1, 2, 18, 19). Although each tissue of the hypothalamo-pituitary-ovarian (HPO) axis is composed of cell-autonomous circadian oscillators, physiological function has been best described in the pituitary gland (20–22) and ovary (23). Clock function in the ovarian follicle has been linked to folliculogenesis, with rhythms of clock gene expression only appearing during the late stages of follicular maturation (preantral-antral; Ref. 24, 25). Rhythms of clock gene expression in the follicle are dependent on FSH and proliferation of gap junction proteins (26, 27) during follicular maturation. In granulosa cells (GCs), the clock drives rhythms of clock-controlled gene expression, including the LH receptor (LHR) (*Lhcgr*), prostaglandin synthase (*Ptgs2* or *Cox2*), and steroidogenic factors like steroidogenic acute regulatory protein (*StAR*) and 3β-hydroxysteroid dehydrogenase (*3β-HSD*) (27–32). Less evidence has accrued for clock function in theca cells (TCs), although rhythms of clock gene expression have been reported. The ovarian clock is also critical for maintaining adequate progesterone (P4) secretion from luteal cells (18, 30) with both global (18) and targeted (30) deletion of the *Bmal1* locus having been shown to impair uterine implantation in mice.

Growing evidence suggests that the ovarian clock plays a role in the timing of ovulation. We have reported phasic sensitivity to gonadotropins in the ovary (33) and that the timing of ovulation depends in part on the timing of this period of increased sensitivity (34). Equine LH (eLH), given to rats after suppression of endogenous hormone with a systemic GnRH receptor antagonist, produces a light-entrained circadian rhythm of oocyte release marked by a nighttime peak of sensitivity (34). Thus, the ovarian clock may dictate, either in concert with the SCN or independently, the timing of ovulation by regulating sensitivity of the preovulatory follicle to gonadotropins. However, these experiments failed to adequately define the influence of temporal cues that originate in the SCN. It is possible that rhythmic patterns of systemic hormone secretion (eg, corticosterone, leptin, etc.,) or inputs from descending autonomic nerves (parasympathetic nerves) convey timing cues to the ovary and singularly drive the rhythm of sensitivity to gonadotropins (35). To address this notion we developed transgenic mice with targeted deletion of the *Bmal1* gene in specific cellular compartments of the ovarian follicle. Female mice expressing *Cyp17-iCre* (TCs) or *Cyp19-Cre* (GCs) transgenes were bred with mice homozygous for floxed alleles of the *Bmal1* locus to generate ovary cell-specific conditional knockout (KO) mice (*Cyp17-Cre;Bmal1$^{flx/flx}$* [TCKO] and *Cyp19-Cre;Bmal1$^{flx/flx}$* [GCKO]). We hypothesized that animals with conditional KO of *Bmal1* would be subfertile and display a marked change in the daily rhythm of sensitivity to gonadotropins. We predicted that this effect would be the result of circadian disruption within a specific cell-type in the ovary, most likely resulting in altered expression of clock-controlled genes important for phasic sensitivity to gonadotropins. Our results reveal that the clock in ovarian TCs regulates the timing of clock-controlled genes, including the LHR, and plays a considerable role in the timing of ovulation by maintaining a stable rhythm of ovarian sensitivity to gonadotropins.

Materials and Methods

Animal welfare assurance

All experiments and procedures were conducted according to the National Institutes of Health Guidelines for the Care and Use of Animals and were approved by the University Committee for Animal Resources at the University of Rochester School of Medicine and Dentistry.

Experimental animals

For our initial ovulation timing experiments, we used adult or juvenile (21–31 d old) female C57BL6/J mice. Ovarian cell type-specific *Bmal1* knockout mice were developed in our laboratory by combining existing strains of cell type-specific CRE-expressing mice with mice carrying floxed alleles of the *Bmal1* locus (generated by Dr Christopher Bradfield, University of Wisconsin-Madison) provided to us by Dr Joseph Takahashi (University of Texas Southwestern Medical Center). Briefly, Cyp17-iCRE (17α-hydroxylase-CRE; TC specific, provided by Dr Cheymong Ko, University of Illinois) (36, 37) and Cyp19-CRE (aromatase-CRE; GC specific, provided by Dr JoAnne S. Richards, Baylor College of Medicine) (38) males were bred with *Bmal1$^{flx/flx}$* females to generate heterozygous *Cyp17* or *Cyp19-Cre$^{+/-}$;Bmal1$^{flx/-}$* offspring. These mice were then bred to homozygosity for the floxed *Bmal1* alleles. For the generation of experimental mice male carriers of the *Cyp17/19-Cre* transgenes that were homozygous for floxed *Bmal1* were then bred to *Cre null;Bmal1$^{flx/flx}$* females to generate confirmed *Cyp17/19-Cre$^{+/-}$;Bmal1$^{flx/flx}$* mice. For all experiments we used female *Cyp17-Cre$^{+/-}$;Bmal1$^{flx/flx}$* (referred to as TCKO mice), *Cyp19-Cre$^{+/-}$;Bmal1$^{flx/flx}$* (referred to as GCKO mice) or littermate controls (LCs) (*Cre-null;Bmal1$^{flx/flx;}$* referred to as LC throughout). All mice were generated on a congenic C57BL6/J background.

Generation of transgenic fluorescent reporter mice and ovarian morphological analysis

To confirm the cell type specificity of the GCKO and TCKO mice both strains (GCKO [Cyp19cre/−;Bmal1flx/flx] and TCKO [Cyp17cre/−;Bmal1flx/flx]) were bred with B6.129(Cg)-

Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J transgenic reporter mice (referred to as TOM-GFP mice) purchased from The Jackson Laboratory (stock number 007676). TOM-GFP mice possess floxed alleles of the membrane targeted tdTomato (mT) cassette and express red fluorescence in all cells and tissues. In cells expressing CRE recombinase the mT cassette is deleted allowing for free expression of the mG or eGFP locus just downstream. A combined ROSA and β-actin promoter drives the fluorescent reporters leading to enhanced membrane-associated fluorescent signal. For histological analyses tissues including the ovary, liver and pituitary gland from adult (3–4 mo old) transgenic mice (TCKO;TOM-GFP or GCKO;TOM-GFP) were recovered in 10% neutral-buffered formalin and fixed in the dark for 48 hours. Tissues were encased in low melting point agarose at 37°C (8% in HEPES-buffered sterile saline and sectioned with a vibratome at 500-μm thickness). Whole mounts of fixed tissue were placed on a slide with HEPES-buffered sterile saline, cover slipped, and imaged at $\times$5 with a Nikon C-DSD115 fluorescence compound microscope (Nikon USA).

To confirm normal ovarian morphology in transgenic mice ovaries were recovered from 3- to 4-month-old adult GCKO and TCKO mice, fixed in 10% Neutral Buffered Formalin, paraffin embedded, sectioned at 5-μm thickness, and stained with hematoxylin/eosin (H+E) (University of Rochester Pathology Core). Brightfield images of H+E-stained ovary tissue from GCKO and TCKO mice were taken at 5$\times$ magnification with a Zeiss Primo Star compound microscope (Carl Zeiss, Inc). To determine whether follicular development is altered in TCKO mice, we measured ovarian morphology in 4-week-old equine gonadotropin (eCG)-primed LC and TCKO mice as described in (39). Brightfield images of H+E-stained ovarian tissue (5-μm sections every 30 μm) from LC and TCKO mice were used to count structures including primordial follicles, primary follicles, preantral follicles, antral follicles, atretic follicles, and corpora lutea across the entire rostro-caudal extent of the tissue. A labeling system was developed in MS Paint (Microsoft) to track and record the structures and data were expressed as percentage of the total structures counted. By recording images and carefully examining the image sequence we were able to avoid duplicate counts. Total ovarian area was also estimated by setting the scale for ovary cross-section images in ImageJ after micrometer scope calibration. A polygon tool was used to trace each section and the area measurement was calculated by ImageJ and integrated into mm^2. The distance between cross-sections and the number of cross-sections were multiplied by this value to approximate a prism model of the ovary.

Measuring ovarian sensitivity to gonadotropins in vivo with an acute ovulation induction assay

We measured ovarian sensitivity to LH with our acute ovulation induction assay as previously described with slight modification (34). This approach allowed us to measure the acute and time-dependent response of the ovarian follicles in vivo to exogenous LH with (adults) or without (juvenile mice) necessary suppression of the endogenous LH and FSH surges. Regardless of age and previous treatment this approach relies on examination of the oviducts 14–16 hours after acute treatment with a rupture-inducing injection of eLH. In this method, the animal is treated only once with eLH and euthanized for oocyte counting, thus precluding repeated measurements. Adult mice (>60 d old) were housed in running wheel cages and reproductive cycles were monitored by vaginal lavage for a minimum of 14 days before treatment (see Supplemental Materials and Methods). Mice received a single injection of the GnRH receptor antagonist Cetrotide (CET; AEterna Zentaris) (500 μg; im under light isoflurane anesthesia) at zeitgeber time (ZT)5 (ZT is in reference to the 12:12 L:D cycle, where ZT0 = time of lights on) on the morning of proestrus. After CET treatment, groups of mice were treated with a single injection of eLH (200 IU/mouse in 100-μL sterile saline, ip) or 100 μL of sterile saline vehicle every 3 hours from ZT12 (time of lights off) on proestrus to ZT9 on estrus (the next day). For experiments in constant darkness (DD) mice of the same age were released into DD and injected with CET at circadian time (CT)5 (CT is a measure dependent on the activity of the animal, CT12 = onset of the activity period) on proestrus followed by eLH at CT12, CT18, CT24/0, and CT6 (the midpoint of the inactivity period) the next day (estrus). In both experiments, mice were euthanized 14–16 hours after eLH treatment, and the oviducts were removed followed by recovery and counting of cumulus oocyte complexes (COCs).

Juvenile mice were weaned on postnatal day (PND)23 and transferred to a wheel running cage. For experiments in a 12:12 L:D cycle, animals were injected with eCG (5 IU ip in sterile saline; Sigma) between ZT11.5 and ZT12 on PND29. Mice were then injected 37 hours later at ZT0.5–ZT1 with CET (500 μg im under light isoflurane anesthesia) followed by eLH (200 IU ip) or saline vehicle every 3 hours from ZT12 on to ZT9 the next day. For experiments in DD, mice were released into DD on PND25, primed with eCG on PND29, and treated with CET (500 μg im) at CT0.5 on PND31. Mice were then treated with eLH on PND31 at CT12, CT18, CT24/0, and CT6.

For experiments with transgenic mice carrying cell type-specific deletion of *Bmal1* in the ovary, we repeated the experiment in adult LC, GCKO, and TCKO mice maintained under a 12:12 L:D cycle. Mice were treated with CET at ZT5 on proestrus followed by eLH (200 IU) at ZT18 or ZT6 on the next day (estrus). As before, mice were euthanized and oviducts were recovered for COC counts 14–16 hours later. Because we determined that the rhythm of ovarian sensitivity occurred in eCG-primed juvenile mice with a similar phase as cycling adults on proestrus we conducted subsequent experiments on juvenile-primed LC, GCKO, and TCKO mice. Because prepubertal mice fail to produce a significant preovulatory LH or FSH surge, and we are providing these mice with exogenous chorionic gonadotropins to stimulate follicular growth, we opted not to provide CET-mediated suppression to this group. This approach allowed us to eliminate the confounding effect of attenuated gonadotrophic support. Juvenile (24–30 d old) LC, GCKO, and TCKO mice were primed with eCG at ZT11.5–ZT12 as above and treated with eLH (200 IU ip) beginning 36 hours later every 3 hours from ZT0 to ZT18. Mice were sacrificed 14–16 hours after treatment and COCs were recovered and counted.

Quantitative real-time PCR analysis of ovarian clock gene and clock-controlled gene mRNA abundance in the ovary

Reproductive cycles were monitored in adult TCKO, GCKO, and LC mice for a minimum of 17 days and only mice with 2 confirmed cycles were included in our analysis. Animals were euthanized on proestrus at ZT0, ZT6, ZT12, and ZT18 and on

estrus at ZT24/0 (ZT0 = lights on). Ovarian tissue was recovered and frozen in liquid nitrogen. After extraction with TRIzol (Sigma) 250 ng of total RNA was used for cDNA synthesis using the Maxima First Strand cDNA Synthesis kit with DNase treatment (Thermo Scientific). Quantitative real-time PCR was carried out using the Luminaris Color HiGreen qPCR Master Mix (Thermo Scientific) according to the manufacturer's instructions. Samples were analyzed with gene-specific primers for core clock genes (*Bmal1*, *Clock*, *Cry1*, *Per1*, and *Rev-erbα*) and clock-controlled genes, including the LH (*Lhcgr*) and FSH receptor (Fshr) (see Supplemental Table 1 for primer sequences). β-Actin was included as a housekeeping gene. Primer sequences were drawn from the Harvard/Massachusetts General Hospital Center for Computational and Integrative Biology PrimerBank database (pga.mgh.harvard.edu) and were independently verified by NCBI BLAST sequence analysis. Data are presented as change in mRNA abundance over time and were analyzed using the ΔΔCT threshold cycle method. Data were normalized to the housekeeping gene β-actin and analyzed relative to the level of mRNA abundance in tissues from LC mice.

Data analysis

For in vivo ovarian sensitivity assays the number of COCs as a function of time were analyzed with one-factor ANOVA followed by Neuman-Keuls post hoc tests. Data were also analyzed as a function of time [times genotype and their interaction with a two-factor ANOVA followed by Bonferroni post hoc comparisons. Absolute levels of LH at ZT12 on proestrus after CET-mediated suppression were analyzed with an unpaired *t* test. The percentage of ovarian structures was analyzed as a function of genotype [times structure with two-factor ANOVA followed by Bonferroni post hoc tests. Ovarian tissue volume was analyzed as a function of genotype using an unpaired student *t* test. The dose-response effects of eLH on proestrus were analyzed as a function of dose × time with a two-factor ANOVA followed by Bonferroni post hoc tests. Rhythms of serum LH and P4 were analyzed as a function of time [times genotype and their interaction with a two-factor ANOVA followed by Bonferroni post hoc tests. Rhythms of relative mRNA abundance were analyzed with CircWave (R. A. Hut, University of Groningen, The Netherlands) nonlinear regression software to test for significant rhythmicity and determine the "center of gravity" or peak of mRNA abundance. CircWave is an F-tested forward harmonic regression process that automatically detects the number of suitable harmonics using F-test criterion and is a well-established method for statistical confirmation of rhythmicity (40). CircWave analyses were used as a statistical criterion for referring to a give data series as being rhythmic. Quantitative PCR data were also analyzed with two-factor ANOVA followed by Bonferroni post hoc tests for pairwise comparison with confirm between genotype effects. With the exception of CircWave all other plotting and statistics were done using GraphPad Prism software (GraphPad). All data are presented as mean ± SEM and the threshold for significance was set at $P < .05$.

Results

Circadian rhythms of ovarian sensitivity to LH in adult C57BL6/J mice

We have previously reported a light-entrained and free-running rhythm of ovarian sensitivity to eLH in female rats after suppression of endogenous LH (34). We observed an increased percentage of ovulating rats and COCs recovered during the night (ZT12–ZT24) and subjective night (CT12–CT24). We reasoned that this rhythm depends on the activity of the clock in cells of the ovarian follicle. To examine the role of the ovarian clock in the timing of ovulation in mice it was first necessary to confirm a rhythm of sensitivity to LH in cycling female mice. To that end we applied our acute ovulation induction assay with slight modification. As we have reported in rats, circulating LH levels were suppressed in mice after a single injection of Cetrorelix depot on the late morning of proestrus (Figure 1A). We next verified the dose-response to exogenous gonadotropins by injecting cycling, CET-treated mice with eLH (50, 200, or 800 IU) on proestrus at ZT18 or the next day at ZT6 (estrus). As expected, mice showed a greater response at ZT18 on proestrus, regardless of dose (Figure 1B).

As in female rats, adult female C57BL6/J mice displayed significant diurnal rhythms of ovulation in response to phasic eLH (Figure 1D). One-factor ANOVA confirmed a significant effect of time ($F = 7.07$, $df = 7$, $P < .001$) with peak oocyte release in response to eLH given between ZT12 and ZT21 ($P < .05$ vs ZT3–ZT9). This peak was followed by a precipitous decline in oocyte release in response to eLH by ZT3 on estrus (Figure 1C). Although not reaching significance, there appears to be a slight increase in the response to eLH at ZT9 (Figure 1C). We also observed a significant rhythm of induced ovulation in mice maintained in DD for at least one full cycle ($F = 6.43$, $df = 3$, $P < .05$). Similar to the data from entrained mice, those free-running in DD showed a peak ovulatory response to eLH between CT12–CT24 followed by a considerable decline by CT6 the next day (estrus) (Figure 1D).

Circadian rhythms of sensitivity to eLH in juvenile gonadotropin-primed C57BL6/J mice

Evidence suggests that clock function depends on gonadotropin-stimulated follicular maturation (24, 26). Thus, we hypothesized that the timing of ovarian sensitivity to gonadotropins set by the ovarian clock is an innate property of the mature ovarian follicle. To address this hypothesis we examined the rhythm of induced ovulation in juvenile eCG-primed female mice. As in adult cycling mice, we observed a robust rhythm of ovarian sensitivity to eLH ($F = 8.53$, $df = 7$, $P < .0001$). This rhythm peaked during the dark phase at ZT18 ($P < .05$ when compared with ZT0–ZT9 on estrus) (Figure 1E). We also observed a rhythm of ovarian sensitivity to eLH in juvenile eCG-primed mice maintained in DD for several days before eLH

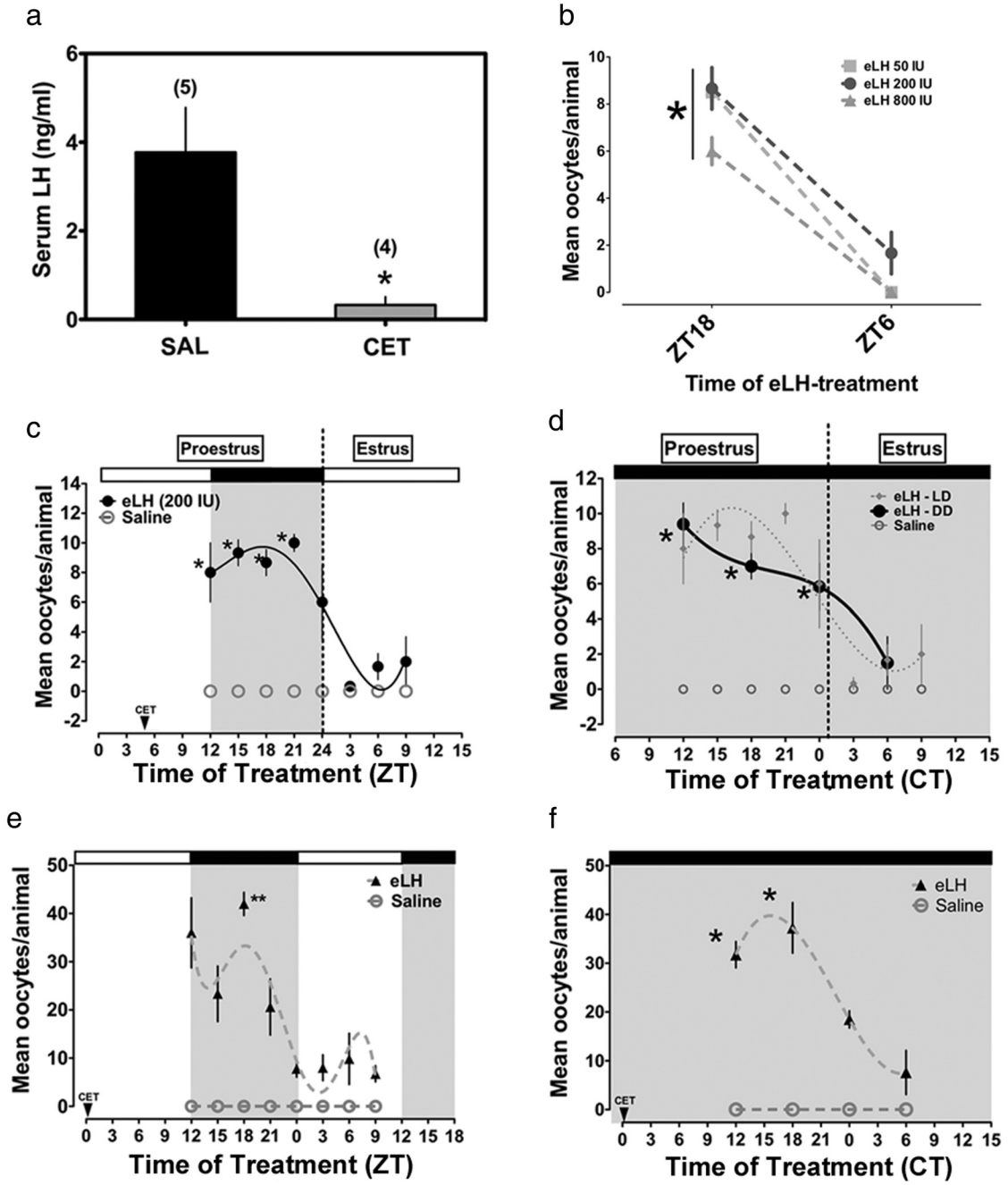

Figure 1. Circadian rhythms of ovarian sensitivity to phasic eLH treatment in adult and juvenile eCG-primed C57BL6/J mice. A, Serum levels of LH at ZT12 on proestrus were suppressed by treatment with CET (n = 4; *, $P < .05$ vs saline [n = 5]). B, Dose-response curve for eLH treatment. All 3 doses (50, 200, or 800 IU; n = 3–4 per treatment) stimulated more oocyte release at ZT18 (*, $P < .001$). C, Rhythm of ovarian sensitivity to eLH in adult C57BL6/J mice (n = 3–6 per treatment time; *, $P < .05$ vs ZT3, ZT6, and ZT9). D, Free-running circadian rhythm of sensitivity to LH in C57BL6/J mice (*, $P < .05$ vs ZT6; n = 3–6 per treatment time). E, A diurnal rhythm of sensitivity to eLH in juvenile eCG-primed mice (**, $P < .001$ vs ZT0–ZT9 on estrus, n = 3–6 per treatment time). F, A circadian rhythm of responsiveness to eLH in juvenile-eCG-primed mice housed in DD (*, $P < .05$ vs the nadir at CT6 on estrus, n = 3–4 per treatment time). In F, the time of CET injection (CT5) is not shown due to scaling of the x-axis. In C–F, saline failed to stimulate ovulation. Data are presented as mean ± SEM and fit to a nonlinear regression (see Materials and Methods). The shaded areas in C–F indicate the dark phase.

treatment ($F = 12.10$, $df = 3$, $P < .001$) (Figure 1F). As in adult cycling mice, oocyte release in response to eLH peaked during the subjective night between CT12 and CT18 in juvenile eCG-primed mice ($P < .05$ when compared with CT0 and CT6 the next day) (Figure 1F).

Conditional KO of Bmal1 in ovarian TCs abolishes the rhythm of ovarian sensitivity to eLH

Our data suggest that the ovarian clock may play a role in the timing of ovulation by setting a window of sensitivity to gonadotropins. We hypothesized that this critical

period of sensitivity depends on rhythms of clock-controlled gene expression and enzyme activity in one (or both) of the primary endocrine cell compartments of the follicle (23). Admittedly, it is also possible that this window of sensitivity is determined by inputs from the SCN, of either a neural or humoral nature (23). In an effort to address this hypothesis we generated ovarian cell type-specific conditional *Bmal1* KO mice to selectively disturb clock function in the 2 primary endocrine cell compartments of the ovarian follicle (GCs and TCs). The cell-type specificity of *Bmal1* deletion was confirmed by breeding GCKO and TCKO mice with a fluorescent reporter strain wherein every cell normally expresses membrane associated tdTomato but only express enhanced Green Fluorescent Protein in the presence of CRE recombinase (TOM-GFP mice). As shown in Figure 2, expression of CRE-driven Green Fluorescent Protein was limited to GCs (both mural and cumulus) lining the inside of the follicle in GCKO mice (Figure 2A) and TC/stromal cells (SCs) on the outside border of the follicle and the interstitial space in TCKO mice (Figure 2B). Signal in TCKO mice appears sparse due to the less compact and defined distribution of TCs/SCs in the ovary (37). Imaging of tissue sections from the pituitary gland (Figure 2C) and the liver (Figure 2D) confirm that CRE expression was largely confined to the ovary, although some cells did express CRE recombinase in these tissues (<5% of those observed; GCKO tissues shown). This pattern of expression is nearly identical to that originally reported in both *Cyp17-iCre* and *Cyp19-Cre* mice (36–38). These data suggest that CRE-driven deletion of the floxed Bmal1 alleles occurred only in the TC/SC or GC compartments of the ovarian follicle.

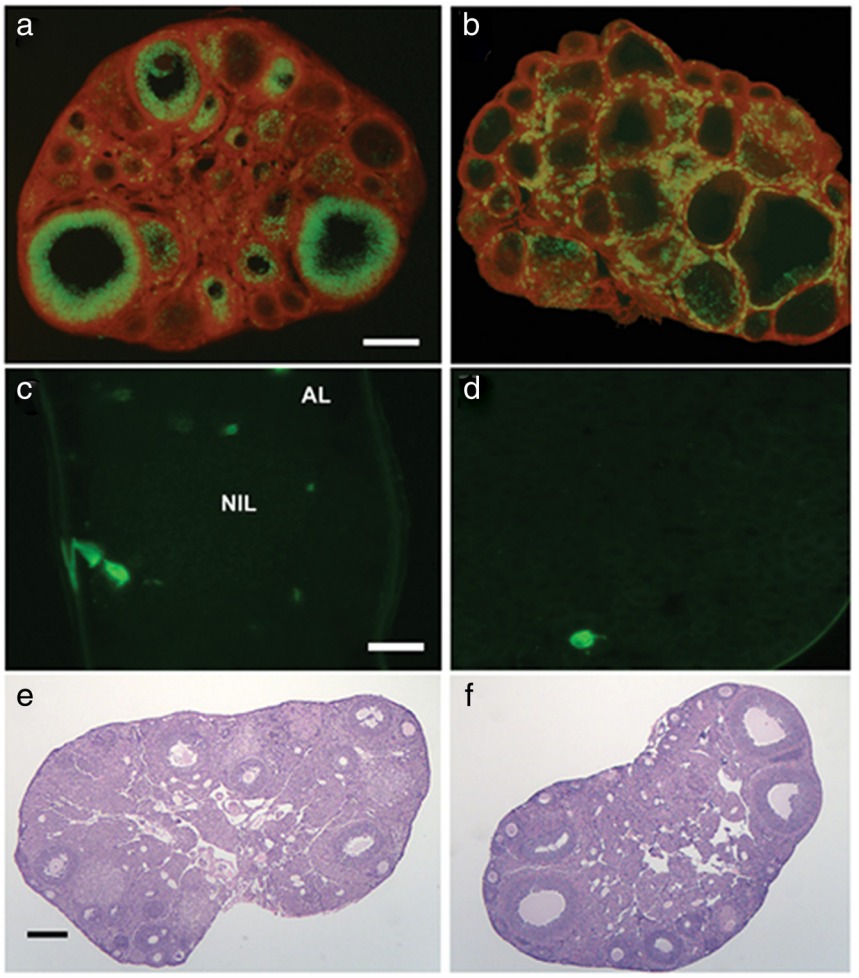

Figure 2. Targeted deletion of *Bmal1* in specific cellular compartments in the ovary of GCKO and TCKO mice. Targeted deletion of *Bmal1* in granulosa (GCKO) and theca/stromal (TCKO) cells of the ovary using the CRE:LOX system. Transgenic GCKO (*Cyp19-Cre;Bmal1*$^{flx/flx}$) and TCKO (*Cyp17-Cre*;*Bmal1*$^{flx/flx}$) mice were crossed with TOM-GFP mice to produce TCKO-TOM-GFP and GCKO-TOM-GFP transgenic mice (see Supplemental Methods). As shown in A, CRE-driven eGFP expression in GCKO-TOM-GFP is nearly completely limited to the mural and cumulus GC layer with little to any GFP expression detected in the TC/SC compartment or the ovarian surface epithelium (OSE). B, In TCKO-TOM-GFP mice, CRE-driven GFP is more dispersed and spotty in the interstitial cells and limited the TC layer of larger antral follicles with no signal detected in the GC compartment or surface epithelium. Representative images of whole mount (C) pituitary gland and (D) liver from a GCKO mouse. A small number of cells in these tissues from both transgenic strains were positive for GFP (<5%). Scale bars: 0.25 μm shown in A (A and B), 0.05 mm shown in C (C and D), and 0.25 mm shown in E (E and F). Images are representative of 3–4 mice per genotype (LC, GCKO, and TCKO).

Ovarian tissue from adult GCKO and TCKO mice failed to reveal any gross morphological abnormalities (Figure 2, E and F). We confirmed this conclusion with a more detailed analysis of ovarian morphology in juvenile eCG-primed LC and TCKO mice (Supplemental Figure 1). We detected a significant effect of structure analyzed ($F = 172.9$, $P < .001$) and an interaction between structure [times genotype ($F = 5.55$, $P < .05$), but we did not detect a significant effect of genotype alone ($F = 3.04 \times 10^{-7}$, $P > .05$). When compared with LCs, the percentage of primordial follicles ($P < .05$) was reduced and the percentage of preantral follicles ($P < .01$) was increased in TCKO mice. The remaining structures, including antral and atretic follicles, were not affected in TCKO mice. Most importantly, we did not detect a difference in the number of

CL and the overall tissue volume between TCKO and LC mice (Supplemental Figure 1B).

It is notable that we did not detect significant eGFP signal in the SCN of TCKO or GCKO mice, indicating that CRE-mediated deletion of Bmal1 did not disturb central clock function (data not shown). This assertion is supported by the observation that both GCKO and TCKO mice maintained normal, light entrained rhythms of locomotor activity when compared with LCs (Supplemental Figure 2A). Moreover, BMAL1 staining within SCN neurons was evident in both GCKO (Supplemental Figure 2B, top) and TCKO (Supplemental Figure 2B, bottom) mice. Also in agreement with the lack of an influence of the ovarian KO on central clock function, we found that reproductive cycles (Figure 3, A and B), serum LH levels (Figure 3C), and serum P4 levels (Figure 3D) on proestrus were normal in both GCKO and TCKO mice.

To determine the effects of targeted molecular clock disruption in ovarian follicular cells, we applied our acute ovulation induction assay to adult female GCKO, TCKO, and LCs. As shown for adult female C57BL6/J mice (see Figure 1C), we found a rhythm of ovarian sensitivity to eLH in LC mice characterized by greater oocyte release in response to eLH given at ZT18 when compared with a basal response at ZT6 ($P < .05$) (Figure 4A, left panel). We detected a similar pattern of ovarian sensitivity in GCKO mice, although the amplitude was diminished ($P < .05$ ZT18 vs ZT6) (Figure 4A, middle panel). Notably, we did not detect a day-night difference of oocyte release in response to eLH in TCKOs ($P > .05$ ZT18 vs ZT6) (Figure 4A, right panel). Analysis of ovarian sensitivity in juvenile eCG-primed mice provided additional confirmation of our finding in adult cycling mice. In both GCKO ($F = 3.57$, $df = 6$, $P < .01$) and LC ($F = 3.77$, $df = 6$, $P < .01$) mice we detected a significant diurnal rhythm of ovarian sensitivity to eLH with peaks at ZT12 and ZT9, respectively (Figure 4B). However, we did not observe a significant rhythm of ovarian sensitivity to eLH in TCKO mice ($F = 2.02$, $df = 6$, $P = .09$). Comparisons with two-factor ANOVA (time [times genotype; time, $F = 4.99$, $df = 6$, $P < .001$ and genotype, $F = 6.62$, $df = 2$, $P < .01$) reveal that peak oocyte release in response to exogenous LH was blunted in TCKOs relative to LCs ($P < .05$ at ZT9 and ZT12) and GCKO ($P < .05$ at ZT6 and ZT12) mice (Figure 4B). In agreement with the marked decline in rhythmic sensitivity to gonadotropins in TCKO mice, we found these mice to be subfertile relative to both GCKO mice and LCs (see Table 1). TCKO mice produced fewer viable litters (only 56% of crosses with a nonsibling *Cre* null:*Bmal1*$^{flx/flx}$ males) and approximately half as many pups per litter as LC and GCKO mice (TCKO 3.3 ± 1.8 vs GCKO 5.7 ± 2.2 and LCs 6.4 ± 1.8, $P < .05$ by ANOVA) (Table 1).

Altered rhythms of clock gene and clock-controlled mRNA abundance in the ovaries of conditional Bmal1 knockout mice

We hypothesized that the timing of ovarian sensitivity to gonadotropins depends on rhythms of CCG expression in follicular cells. In particular we had predicted, based on the literature and on our own work that GCs were the primary "gonadotropin-sensitive" oscillator in the follicle (23). Thus, we predicted that clock-controlled gene expression in GCs, including the rhythmic expression of LHRs, was critical for the maintenance of phasic sensitivity to gonadotropins. To our surprise, our data reveal that the timing of ovulation may depend more on clock function in the ovarian TC (Figure 4 and Table 1). To better define the role of the molecular clock in each ovarian cell type, we measured the abundance of clock and clock-controlled gene mRNA in ovaries recovered from GCKO, TCKO, and LC mice every 6 hours from ZT0 to ZT24 on proestrus. As shown in Figure 5, we detected diurnal rhythms of clock gene mRNA abundance, including *Bmal1*, *Per1*, and *Rev-erbα* (all $P < .05$ by CircWave as denoted in Figure 5) in ovaries from LC mice. In LC mice *Bmal1* mRNA levels peaked in the late portion of the dark phase (ZT21.55 ± 1.6). This rhythm of *Bmal1* expression was attenuated in both GCKO and TCKO mice (both $P > .05$ by CircWave) (Figure 5). Two-factor ANOVA (genotype × time; genotype $F = 26.35$, $df = 2$, $P < .0001$ and time $F = 6.87$, $df = 4$, $P < .001$) confirmed a significant decline in *Bmal1* expression at ZT18 in both GCKO ($P < .001$) and TCKO ($P < .001$) mice compared with the same time in LCs. Further, the center of gravity or "peak" of *Bmal1* mRNA abundance was delayed in the ovaries of both GCKO (ZT0.74 ± 1.6) and TCKO (ZT3.25 ± 1.5) mice. *Per1* mRNA abundance was also rhythmic in the ovary from LCs, with a peak in the middle of the light phase (ZT8.49 ± 1.67). This rhythm persisted ($P < .05$ by CircWave) in targeted KO mice although the rhythm was slightly phase advanced and attenuated relative to LCs (GCKO mice, ZT4.13 ± 1.4; TCKO mice, ZT4.44 ± 0.96) (Figure 5). Again, two-factor ANOVA (genotype $F = 25.92$, $df = 2$, $P < .001$ × time $F = 6.71$, $df = 4$, $P < .001$) confirmed a significant decline in ovarian *Per1* mRNA levels in both GCKO and TCKO mice near the peak (ZT6–ZT12) of *Per1* mRNA abundance in LC mice ($P < .05$ for both). CircWave analysis did not detect a significant rhythm of *Cry1* mRNA abundance in our LC or KO mice. Like *Per1*, *Rev-erbα* mRNA abundance was also rhythmic in LC mice with a peak at midday (ZT5.64 ± 1.7) followed by a nadir in the latter portion of the dark phase as determined by CircWave. As expected, the peak of *Rev-erbα* mRNA level preceded (and was nearly antiphase to) the peak of *Bmal1* abundance. We did

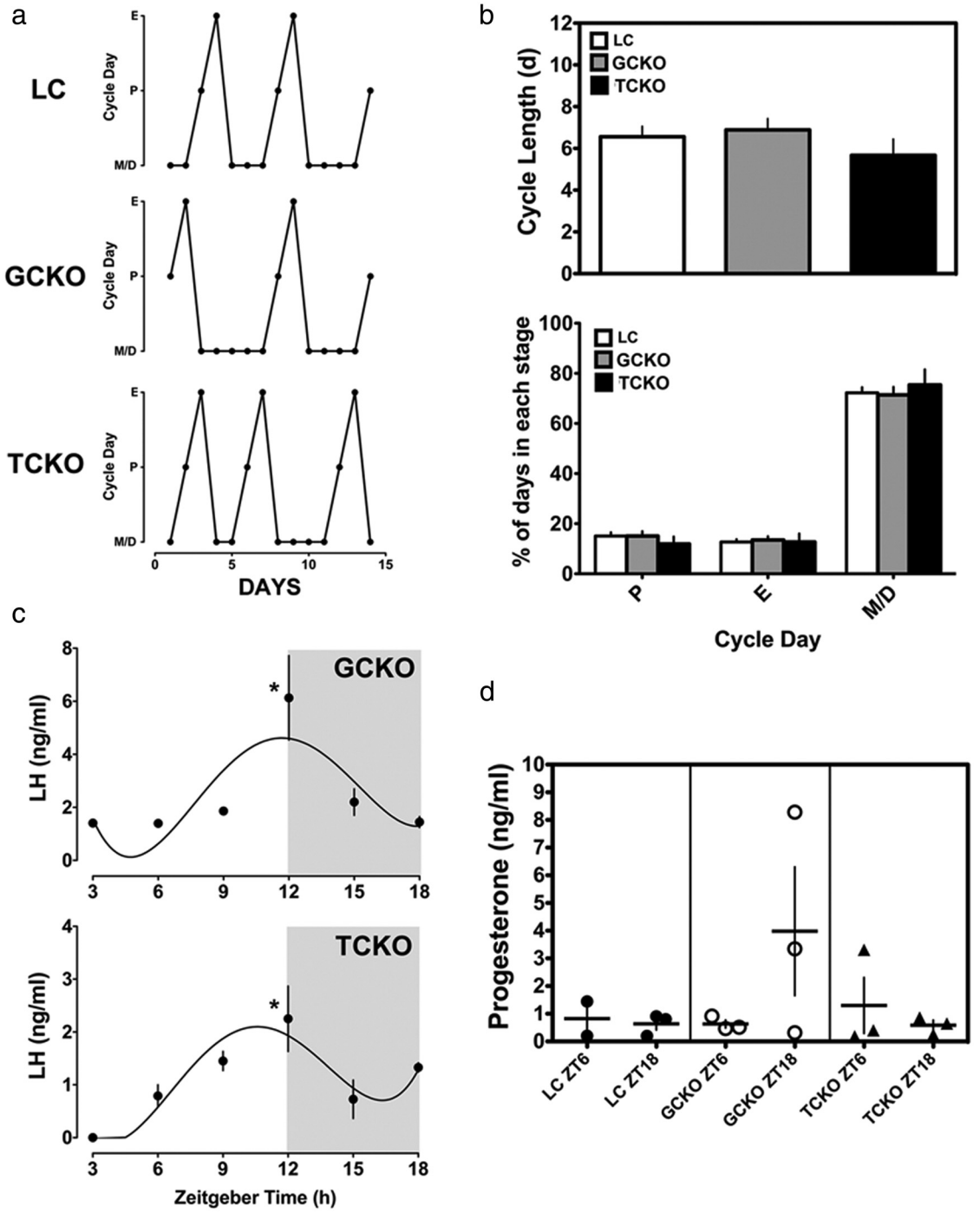

Figure 3. Reproductive cycles, serum LH, and serum P4 levels on proestrus are not altered by conditional *Bmal1* deletion in the ovary of GCKO and TCKO mice. A, Representative reproductive cycle graphs from LC, GCKO, and TCKO transgenic mice. B, Analysis of (top) cycle length and (bottom) percentage of time spent in each day of the cycle reveals no significant difference between LC, GCKO, and TCKO transgenic mice (n = 9 per group). C, Serum LH measured on proestrus from ZT3–ZT18 reveals that both GCKO and TCKO transgenic mice had appreciable surges of serum LH with both rhythms peaking at ZT12 (*, $P < .05$, n = 3–4 samples per time point). D, Levels of serum P4 at ZT6 and ZT18 on proestrus were not significantly affected by targeted deletion of *Bmal1* in either GCKO or TCKO transgenic mice (n = 3 samples per time point except for ZT6 in LC mice [n = 2]). Data are labeled according to genotype and time of serum collection. All data are presented as mean ± SEM.

not detect a statistically significant effect of genotype ($F = 0.23$, $df = 2$, $P = .80$) or time ($F = 1.96$, $df = 4$, $P = .13$) on *Rev-erbα* mRNA abundance, due largely to an apparent increase in variance across time points in TCKO and GCKO mice (Figure 5). Analysis with CircWave confirmed results of ANOVA, indicating an absence of rhythmicity in both conditional KO strains. We also did not see an effect of targeted *Bmal1* KO on the phase of *Rev-erbα* mRNA abundance. Two-factor ANOVA (genotype $F = 7.59$, $df = 2$, $P < .01$ and time $F = 6.99$, $df = 4$, $P < .001$)

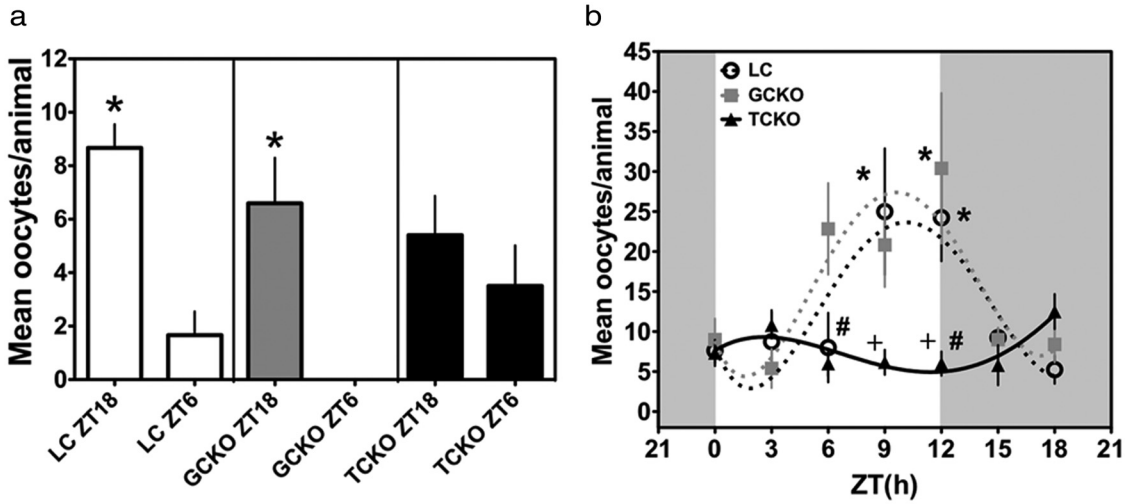

Figure 4. Phasic sensitivity to LH is disrupted in adult and juvenile eCG-primed conditional Bmal1 TCKO mice. A, Rhythm of sensitivity to eLH in cycling adult LC (n = 3; both ZT18 and ZT6), GCKO (ZT18 n = 5, ZT6 n = 4), and TCKO (ZT18 n = 5, ZT6 n = 6) mice. Both LC and GCKO transgenic mice displayed strong diurnal rhythms of eLH sensitivity with peak responses to eLH treatment at ZT18 (*, $P < .05$ vs ZT6) for both. In contrast, adult cycling TCKO mice did not maintain a rhythm of sensitivity to eLH ($P > .05$ ZT18 vs ZT6). B, Diurnal rhythms of sensitivity to LH in juvenile-eCG-primed LC (n = 4–8 per time of treatment), GCKO (n = 5–6 per time of treatment), and TCKO (n = 4–6 per time of treatment) mice. Both LC (open circles) and GCKO (gray squares) mice displayed diurnal rhythms of responsiveness with significant peaks of responsiveness to eLH treatment at ZT9–ZT12 (*, $P < .05$ vs troughs at ZT0, ZT3, and ZT18). TCKO mice (black triangles) did not show a significant rhythm in response to phasic eLH. Data are presented as mean ± SEM and in B data are fit to a nonlinear regression (see Materials and Methods). In B, # indicates difference between TCKO and GCKO, and + indicates a significant difference between TCKO and LC mice.

and CircWave ($P < .05$) confirmed a significant rhythm of *Clock* mRNA abundance in TCKO mice peaking during the middle of the subjective day (ZT4.83 ± 1.51). We did not detect similar rhythms in the ovary of LC or GCKO mice.

We next determined whether the altered rhythms of clock gene mRNA abundance we observed in TCKO mice resulted in irregular patterns of clock-controlled gene expression in the ovary. We measured the level of LH (*Lhcgr*) and FSH (*Fshr*) receptor mRNAs in whole ovaries collected every 6 hours from ZT0–ZT24 on proestrus. As shown in Figure 6, we detected a significant diurnal rhythm of *Lhcgr* mRNA abundance in LC mice that peaked during the early portion of the light phase (ZT2.4 ± 1.4) (Figure 6). Unlike the rhythm of *Bmal1* mRNA abundance, the rhythm of *Lhcgr* expression persisted in GCKO mice with a peak close to lights on (ZT1.98 ± 1.59) (Figure 6). This rhythm was abolished in TCKO mice, as confirmed by both CircWave ($P > .05$) and two-factor ANOVA (genotype, $F = 5.63$, $df = 2$, $P < .01 \times$ time, $F = 6.53$, $df = 4$, $P < .001$). Dampening of this rhythm was associated with a decline of the daytime peak and a spike of *Lhcgr* mRNA levels at ZT24 in TCKO mice, which was not detected in GCKO mice or LCs ($P < .05$ vs both) (Figure 6). We did not detect a significant phase shift of *Lhcgr* expression in either GCKO or TCKO mice (when compared with LCs). Although *Fshr* mRNA appeared to fluctuate across the day in all 3 strains, we could only confirm a rhythm in GCKO mice ($P < .05$ by CircWave). As with *Lhcgr*, we did not detect a significant effect of genotype on *Fshr* mRNA abundance in the ovary in any strain at any of the time points examined (Figure 6). This result was also confirmed by two-factor ANOVA (genotype, $F = 1.28$, $df = 2$, $P = .29 \times$ time, $F = 3.80$, $df = 4$, $P < .05$).

Discussion

The influence of the timing system on female reproductive physiology is well known, with the SCN playing a pivotal role in the timing of events in the hypothalamo-pituitary-

Table 1. Fertility of Targeted *Bmal1* KO Mice

Genotype	Dams	Average Matings per Dam[a]	% Viable Litters (Total)	Mean (SEM) Pups/Litter
LC	12	2	88% (24)	6.4 (1.8)
GCKO	12	1.6	94% (18)	5.7 (2.2)
TCKO	12	1.2	56% (16)	3.3 (1.8)[b]

[a] Mated with *Cre null;Bmal1*$^{flx/flx}$ non-sibling male.

[b] $P < .05$ by ANOVA vs LC and GCKO.

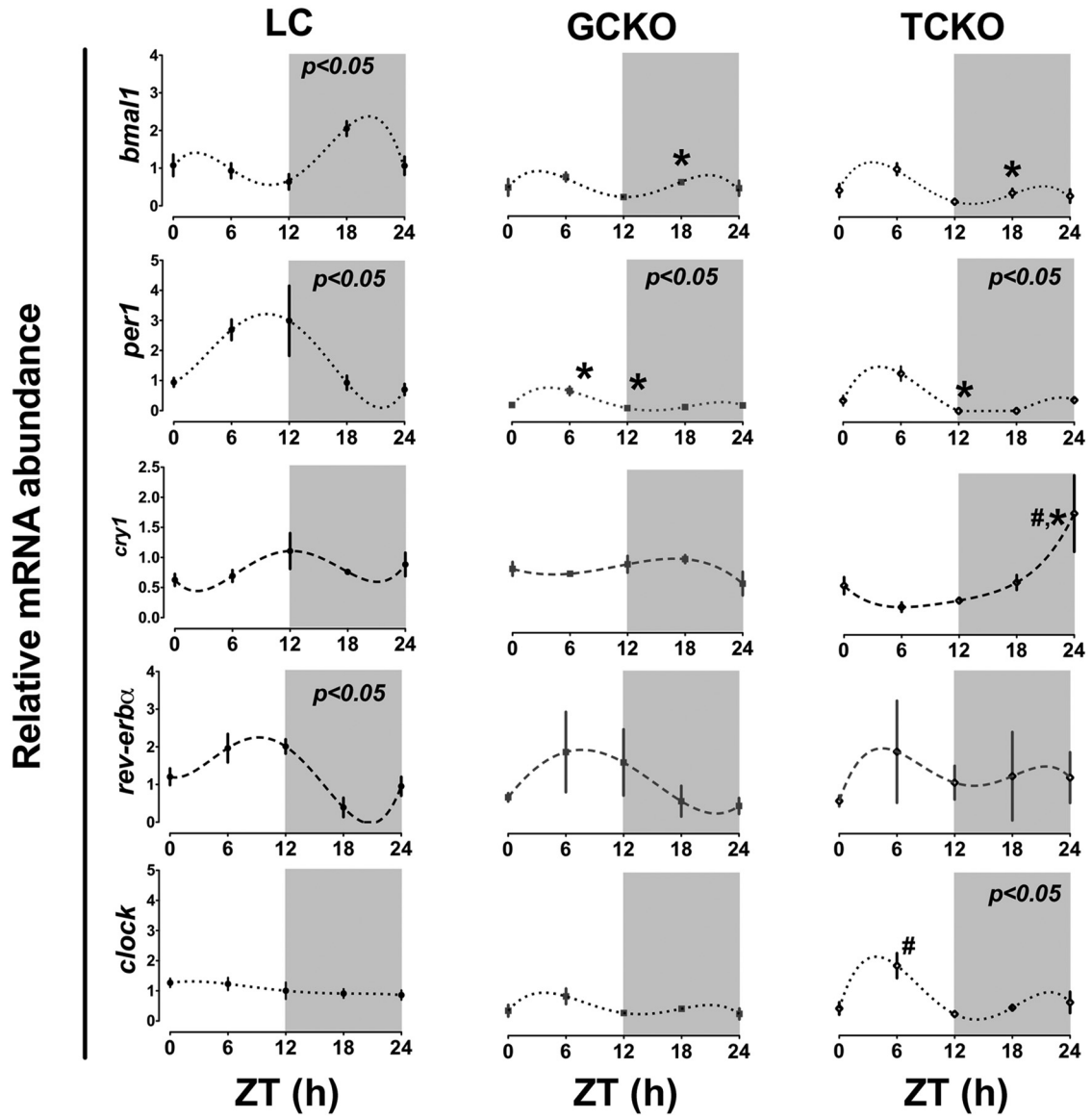

Figure 5. Rhythms of clock gene mRNA abundance in whole ovary are disrupted in both GCKO and TCKO transgenic mice. Abundance of clock gene mRNA in ovaries recovered from LC (n = 3), GCKO (n = 3), and TCKO (n = 3) mice on proestrus. LC mice displayed diurnal rhythms of *Bmal1*, *Per1*, and *Rev-erbα* mRNA abundance ($P < .05$ indicated in top right of each panel). Rhythms of *Bmal1* and *Rev-erbα* mRNA abundance were disrupted in both GCKO and TCKO mice. Although still rhythmic, the absolute level of *Per1* mRNA was suppressed in both GCKO and TCKO mice. Although not significant by CircWave analysis, *Cry1* mRNA abundance oscillated at low amplitude with a peak near midday in LC mice and was altered in both GCKO and TCKO mice. *Clock* mRNA displayed a low amplitude rhythm in TCKO mice that was not detected in LC or GCKO mice. Data are presented as mean ± SEM; n = 3 samples per time point except for ZT24 in TCKO mice (n = 2). (#, $P < .05$ GCKO vs TCKO; *, $P <$ .05 LC vs GCKO or TCKO). Values for target mRNA abundance were calculated using the ΔΔCT method as described in Materials and Methods with β-actin as the reference gene. Data were fit to a nonlinear regression (see Materials and Methods).

ovarian axis (10, 41). Evidence has also accumulated in support of a contribution from peripheral oscillators including the ovary (42). We have previously reported a circadian rhythm of ovarian sensitivity to gonadotropins in rats which is independent of the timing of the endogenous LH surge (34). This rhythm, not surprisingly, peaks during the night. These data suggest that the ovarian clock may contribute to the timing of ovulation by setting a window of LH sensitivity. Although the follicle receives multiple timing cues from the SCN (23), logic dictates that a critical physiological system (eg, the reproductive axis) with strict temporal requirements might rely on redundant activity of central and peripheral oscillators. However, the weight of these contributions remains largely unknown (43). We applied our ovulation induction assay to mice and confirmed rhythmic sensitivity to LH, with a peak response also limited to the early/middle portion of the night when mating is likely to occur (44). These data extend our previous results and support our general hypothesis. It has been suggested that LH sensitivity is pro-

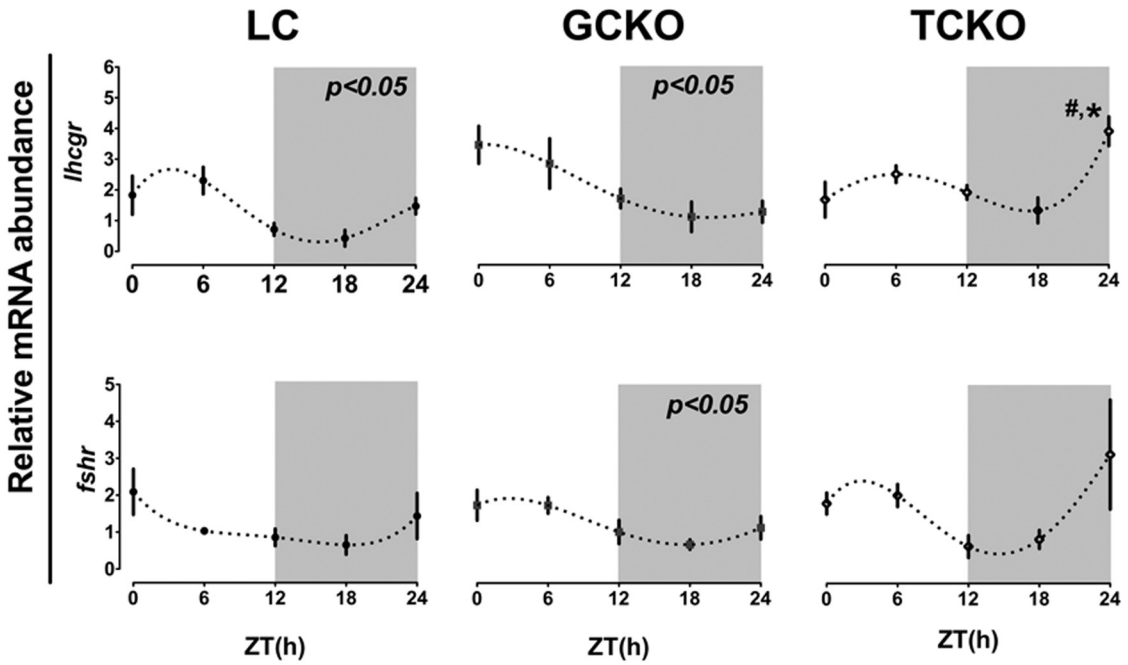

Figure 6. Conditional KO of *Bmal1* in TCs alters rhythms of LHR mRNA abundance in the ovary. Rhythms of *Lhcgr* and *Fshr* expression in whole ovary tissue recovered from LC (n = 3), GCKO (n = 3), and TCKO (n = 3) transgenic mice. In both LC and GCKO mice we detected a significant diurnal rhythm of *Lhcgr* mRNA abundance ($P < .05$ for both) that was abolished in TCKO transgenic mice. There was a marked increase in *Lhcgr* expression at ZT24 in TCKO mice (#, $P < .05$) when compared with the same time point in both LC and GCKO mice. A low amplitude oscillation of *Fshr* mRNA expression was detected in GCKO transgenic mice with a peak in the morning (ZT1.99). Daily variation of *Fshr* mRNA abundance was not significant in either LC or TCKO transgenic mice. Data are presented as mean ± SEM. n = 3 ovaries from 3 different mice per time point. #, $P < .05$ GCKO vs TCKO; *, $P < .05$ LC vs GCKO or TCKO. Only those rhythms labeled directly with a *P* value were considered significant by CircWave analyses. Values for target mRNA abundance were calculated using the ΔΔCT method as described in Materials and Methods, with β-actin as the reference gene. Data were fit to a nonlinear regression (see Materials and Methods).

grammed during puberty and that the rhythm we observed depends almost entirely on the persistence (or absence) of gonadotrophic support. It is true that treatment with CET leads to a precipitous decline in gonadotrophic support for the developing follicle (45, 46). To address this issue we measured the timing of ovarian sensitivity in juvenile (<31 d old) mice primed with eCG. Chorionic gonadotropin has a long half-life in rodent serum ranging from 40–72 hours (47) and is used to stimulate folliculogenesis 48–72 hours before a "trigger" injection of LH (48, 49). Because gonadotropin secretion is low in the juvenile mouse this CET treatment was unnecessary, but initially included to control for the effects of blocking endogenous LH (3, 50). Juvenile eCG-primed mice displayed a rhythm of sensitivity to eLH that declined in the face of persistent gonadotrophic support for the ovarian follicles (ZT0–ZT6 on the second day after priming). Moreover, this rhythm showed peak and trough phases akin to those detected in adults, indicating that the phase of sensitivity is established early in development and may be an innate property of the mature follicle.

To identify the ovarian pacemaker responsible for this rhythm, we used the CRE-LOX system to generate ovarian cell type-specific conditional *Bmal1* knockout mice. BMAL1 is a core component of the oscillator and the only single locus whose deletion leads to near complete clock disruption (17). We used CRE-driven fluorescent reporter mice to verify that CRE expression was largely restricted to the intended ovarian cell type and that BMAL1 expression in the SCN and rhythms of behavior were not affected in GCKO or TCKO mice. We found very limited expression of CRE-driven GFP outside of the ovary, including the anterior pituitary, liver (see Figure 2) and oviduct (not shown). Our findings are in agreement with the initial reports describing the development of both the *Cyp17-iCre* and *Cyp19-Cre* transgenic lines (36, 38). We found that TCKO mice were subfertile when compared with both GCKO and LC mice marked by mating difficulties and fewer offspring per litter, although most reproductive functions (eg, estrous cycles, P4, and LH levels) were unaffected. A limitation of our approach is that P4 levels were assessed on proestrous rather than metestrus or diestrus. Thus, it is possible that there is an undetected drop in P4 levels during the brief luteal phase in our TCKO and GCKO mice. Because we did not measure implantation rate (30), we cannot make strong conclusions regarding luteal P4 levels. However, the decline in offspring we observed matches the reduction in response to exogenous LH

we detected in cycling TCKO mice, again supporting the conclusion that conditional *Bmal1* KO in TCs affects ovulation but not luteal P4 levels or implantation.

Although we measured serum LH in both TCKO and GCKO mice, we did not directly measure this hormone in our *Bmal1*$^{flx/flx}$; *Cre* null LC mice. However, the timing and amplitude of LH secretion during the mouse estrous cycle is well known and has recently been described in the literature (3, 50–52). Because we saw no gross deficits in fertility among our LC mice (Table 1) we can safely assume they produce a normal LH surge on proestrus. We also did not detect any gross deficits in folliculogenesis among TCKO mice. We did observe a slight decline in the number of primordial follicles and a modest increase in the percentage of preantral follicles. It is possible that an undetected increase in early atresia in TCKO mice or more rapid progression to the preantral stage accounts for this apparent decline in follicular reserve. Nonetheless, this does not appear to account for the decline in fertility we observed. Although able to reproduce, TCKO KO mice were less successful and produced on average half as many offspring. To directly assess the role of the clock in TCs we measured ovarian sensitivity to eLH across the 24-hour day. Both LC and GCKO mice displayed rhythmic sensitivity to eLH peaking at ZT18, but this rhythm was blunted in TCKO mice. To eliminate the confounding influences of CET treatment and a mature HPO axis we examined this rhythm in juvenile eCG-primed mice. In this experiment CET treatment was avoided to eliminate any confounding effects of the drug on *Cyp19* expression in GCs (53). Juvenile eCG-primed LC and GCKO mice also displayed rhythmic sensitivity to eLH. We did not detect a rhythm of LH sensitivity in TCKO mice, suggesting that the oscillator in TCs is necessary for phasic responsiveness to LH. Because clock gene expression rhythms in the ovarian follicle appear to be linked to follicular development (ie, only more mature preantral and antral follicles express a rhythm of clock gene expression), we cannot feasibly dissociate the timing of ovarian sensitivity from the development of the follicle either across the cycle or in response to eCG-priming. Our results simply support the conclusion that the clock in the ovarian follicle, at that time when said follicles reach their peak of maturity and are ready to respond, will do so in a rhythmic fashion that is independent of any temporal cues from the SCN. This conclusion is drawn largely from our data in juvenile eCG-primed LC mice that failed to ovulate after approximately 40 hours of priming (ZT0–ZT3) and the dampened rhythm in TCKO mice primed with eCG. Further, the acute drop in ovulation in response to eLH at ZT15–ZT18 in these mice, notably only 51–54 hours after eCG priming when mature follicles are certainly still able to respond, supports our assertions.

LH, acting through its G protein-coupled receptor, initiates a cascade of cellular signaling events leading to cellular differentiation and altered gene expression (54). Previous studies have examined the daily variation in LHR gene (*Lhcgr*) expression in the ovary and isolated GCs (27, 30, 55). Combined, these data indicate that the clock drives rhythmic *Lhcgr* expression in luteinized GCs. We hypothesized that the blunted rhythm of sensitivity to LH we observed in TCKO mice was due to an attenuated rhythm of clock-controlled gene expression in TCs. We observed significant diurnal rhythms of *Bmal1*, *Per1*, and *Rev-erbα* mRNA abundance in ovaries from LC mice. These rhythms were suppressed in GCKO and TCKO mice, indicating that deletion of *Bmal1* leads to altered or even abolished rhythms of clock gene expression in both GC and TC. We next examined the influence of *Bmal1* KO on the timing of *Lhcgr* and *Fshr* mRNA abundance in the ovary. In both LC and GCKO mice we observed significant diurnal rhythms of *Lhcgr* mRNA abundance with peaks at ZT2.4 and ZT1.98, respectively. This rhythm was altered in TCKO transgenic mice that displayed a large increase in *Lhcgr* mRNA at ZT24. The altered rhythm of *Lhcgr* mRNA abundance we have observed in TCKO mice is likely due to the suppression of *Bmal1* mRNA in these cells. Because we did not analyze gene expression directly in isolated TCs or theca interna tissue it is possible that the rhythm of *Lhcgr* and *Fshr* mRNA abundance we measured in TCKO mice is generated by GCs that are indirectly influenced by clock disruption in TCs. Further targeted analysis of the effects of *Bmal1* KO in isolated GCs and TCs is warranted. In mammals the TC produces androgens that are critical for normal folliculogenesis (39, 56). The LHR is expressed at high levels in the TC wherein LH stimulates steroidogenesis through up-regulation of steroid acute regulatory protein (*StAR*) and *Cyp17* gene expression (57, 58). LH enhances follicular development through its impacts on TC androgen secretion (59). Data from bovine and rodent models confirm considerable changes in gene expression across multiple loci in the TC after exposure to LH (60, 61). The expression of LHR is under the direct control of the molecular clock, as treatment with PER2 or CLOCK siRNA leads to down-regulation of *Lhcgr* expression (62). Conversely, LH and FSH can alter the timing of clock gene expression in the ovary (for review see Ref. 23). Our data suggest that suppression of ovarian sensitivity to LH in TCKO is the result of abnormal rhythms of *Lhcgr* expression in TCs. Although this is certainly a reasonable interpretation, there are several possible alternatives. Because we did not measure ovarian sensitivity across the entire 24-hour day

or in DD it is possible that TCKO dramatically shifts the peak of sensitivity to a fine window in the late night. It is also possible, as previously alluded to, that deletion of the clock in TCs leads to indirect deficits in GC function which in turn leads to the decline in sensitivity in TCKO mice.

Although both novel and significant, our results conflict with a recent study indicating that deletion of *Bmal1* in steroid-producing cells of the ovary using a steroidogenic factor 1 (SF-1)-*Cre*;*Bmal1*$^{flx/flx}$ transgenic model leads to marked reduction in P4 secretion and implantation failure (30). Their findings leave little doubt that deficits in implantation are due to a decline of ovarian P4 secretion. How then can we explain this discrepancy? It may be resolved by understanding differences in temporal/spatial expression of the ovary-targeted CRE drivers used. Both *Cyp19* and *Cyp17* mRNA expression levels vary during folliculogenesis (63, 64). *Cyp19* and *Cyp17* expression levels increase nearly 40-fold during follicular development, starting low in primary follicles and reaching a peak in large preovulatory follicles (63, 65). Further, evidence indicates that both *Cyp17* and *Cyp19* expression is dramatically down-regulated after the LH surge (66, 67). In contrast, SF-1 expression is more ubiquitously expressed with significant levels reported in primary, secondary and tertiary follicles (68–70). Hinshelwood et al reported weak SF-1 expression limited to the TC/SC compartment and exceedingly low levels of SF-1 in luteal cells, suggesting that SF-1 is not the primary regulator of P4 secretion during pregnancy (70). This finding is supported by genome wide studies showing a near 3-fold decline in SF-1 expression after luteinization (60). In contrast, others have recorded significant SF-1 signal in both ovarian cell types and confirmed that SF-1 regulates estradiol synthesis in GCs (69, 71). These data lead us to conclude that deletion of *Bmal1* in ovarian follicles is enhanced in our transgenic mice during late folliculogenesis, resulting in robust circadian disruption in preovulatory follicles that is potentially silenced after luteinization.

Our data reveal, for the first time, that circadian clock function in the TC appears to be necessary and sufficient for a normal rhythm of ovarian sensitivity to gonadotropins and thus precise timing of ovulation in mice. Moreover, we have determined that this rhythm of sensitivity may be directly linked to rhythmic expression of LHR mRNA in the TC. A change in the diurnal rhythm of *Lhcgr* mRNA abundance is associated with an attenuated rhythm of sensitivity and an overall decline in fertility in TCKO mice. The TC is the primary androgen-producing cell in the ovary and androgen secretion from TCs represents a tipping point in the balance between normal folliculogenesis (low levels of androgen) and abnormal follicular growth (excess androgen). Our own data reveal that excess androgen exposure during sexual development, a treatment that produces a Polycystic Ovary Syndrome phenotype in mice, leads to considerable circadian disruption, irregular reproductive cycles and infertility (72, 73). Here we reveal that TCs, acting as the ovarian pacemaker for gonadotropin sensitivity, could play a critical part in the etiology of infertility due to environmental and/or genetic influences.

Acknowledgments

We thank Londyn Cullifer, Drew Phillips, and Lindsay Marchetti for technical assistance.

Address all correspondence and requests for reprints to: Michael T. Sellix, PhD, Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, University of Rochester School of Medicine and Dentistry, Room 3–5752C, 601 Elmwood Avenue, Box 693, Rochester, NY 14642. E-mail: michael_sellix@urmc.rochester.edu.

Disclosure Summary: The authors have nothing to disclose.

References

1. **Kennaway DJ.** The role of circadian rhythmicity in reproduction. *Hum Reprod Update*. 2005;11:91–101.
2. **Boden MJ, Kennaway DJ.** Circadian rhythms and reproduction. *Reproduction*. 2006;132:379–392.
3. **Bronson FH, Vom Saal FS.** Control of the preovulatory release of luteinizing hormone by steroids in the mouse. *Endocrinology*. 1979; 104:1247–1255.
4. **Goldman BD.** The circadian timing system and reproduction in mammals. *Steroids*. 1999;64:679–685.
5. **Legan SJ, Karsch FJ.** A daily signal for the LH surge in the rat. *Endocrinology*. 1975;96:57–62.
6. **Moenter SM, DeFazio AR, Pitts GR, Nunemaker CS.** Mechanisms underlying episodic gonadotropin-releasing hormone secretion. *Front Neuroendocrinol*. 2003;24:79–93.
7. **Everett JW, Sawyer CH.** A 24-hour periodicity in the "LH-release apparatus" of female rats, disclosed by barbiturate sedation. *Endocrinology*. 1950;47:198–218.
8. **Kriegsfeld LJ, Silver R.** The regulation of neuroendocrine function: timing is everything. *Horm Behav*. 2006;49:557–574.
9. **Karsch FJ, Bowen JM, Caraty A, Evans NP, Moenter SM.** Gonadotropin-releasing hormone requirements for ovulation. *Biol Reprod*. 1997;56:303–309.
10. **de la Iglesia HO, Schwartz WJ.** Minireview: timely ovulation: circadian regulation of the female hypothalamo-pituitary-gonadal axis. *Endocrinology*. 2006;147:1148–1153.
11. **Mohawk JA, Green CB, Takahashi JS.** Central and peripheral circadian clocks in mammals. *Annu Rev Neurosci*. 2012;35:445–462.
12. **Legan SJ, Coon GA, Karsch FJ.** Role of estrogen as initiator of daily LH surges in the ovariectomized rat. *Endocrinology*. 1975;96:50–56.
13. **Miller BH, Olson SL, Levine JE, Turek FW, Horton TH, Takahashi JS.** Vasopressin regulation of the proestrous luteinizing hormone surge in wildtype and clock mutant mice. *Biol Reprod*. 2006;75: 778–784.
14. **Wiegand SJ, Terasawa E, Bridson WE, Goy RW.** Effects of discrete

lesions of preoptic and suprachiasmatic structures in the female rat. Alterations in the feedback regulation of gonadotropin secretion. *Neuroendocrinology*. 1980;31:147–157.
15. **Albrecht U.** Timing to perfection: the biology of central and peripheral circadian clocks. *Neuron*. 2012;74:246–260.
16. **Sato TK, Panda S, Miraglia LJ, et al.** A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron*. 2004;43:527–537.
17. **Bunger MK, Wilsbacher LD, Moran SM, et al.** Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell*. 2000;103:1009–1017.
18. **Ratajczak CK, Boehle KL, Muglia LJ.** Impaired steroidogenesis and implantation failure in Bmal1−/− mice. *Endocrinology*. 2009;150:1879–1885.
19. **Ratajczak CK, Asada M, Allen GC, et al.** Generation of myometrium-specific Bmal1 knockout mice for parturition analysis. *Reprod Fertil Dev*. 2012;24:759–767.
20. **Olcese J, Sikes HE, Resuehr D.** Induction of PER1 mRNA expression in immortalized gonadotropes by gonadotropin-releasing hormone (GnRH): involvement of protein kinase C and MAP kinase signaling. *Chronobiol Int*. 2006;23:143–150.
21. **Resuehr D, Wildemann U, Sikes H, Olcese J.** E-box regulation of gonadotropin-releasing hormone (GnRH) receptor expression in immortalized gonadotrope cells. *Mol Cell Endocrinol*. 2007;278:36–43.
22. **Resuehr HE, Resuehr D, Olcese J.** Induction of mPer1 expression by GnRH in pituitary gonadotrope cells involves EGR-1. *Mol Cell Endocrinol*. 2009;311:120–125.
23. **Sellix MT.** Circadian clock function in the mammalian ovary. *J Biol Rhythms*. 2015;30:7–19.
24. **Gräs S, Georg B, Jorgensen HL, Fahrenkrug J.** Expression of the clock genes Per1 and Bmal1 during follicle development in the rat ovary. Effects of gonadotropin stimulation and hypophysectomy. *Cell Tissue Res*. 2012;350:539–548.
25. **Fahrenkrug J, Georg B, Hannibal J, Hindersson P, Gräs S.** Diurnal rhythmicity of the clock genes Per1 and Per2 in the rat ovary. *Endocrinology*. 2006;147:3769–3776.
26. **Chen H, Zhao L, Chu G, et al.** FSH induces the development of circadian clockwork in rat granulosa cells via a gap junction protein Cx43-dependent pathway. *Am J Physiol Endocrinol Metab*. 2013;304:E566–E575.
27. **Chen H, Zhao L, Kumazawa M, et al.** Downregulation of core clock gene Bmal1 attenuates expression of progesterone and prostaglandin biosynthesis-related genes in rat luteinizing granulosa cells. *Am J Physiol Cell Physiol*. 2013;304:C1131–C1140.
28. **Chen H, Chu G, Zhao L, Yamauchi N, Shigeyoshi Y, Hashimoto S, Hattori MA.** Rev-erbα regulates circadian rhythms and StAR expression in rat granulosa cells as identified by the agonist GSK4112. *Biochem Biophys Res Commun*. 2012;420:374–379.
29. **Chu G, Misawa I, Chen H, et al.** Contribution of FSH and triiodothyronine to the development of circadian clocks during granulosa cell maturation. *Am J Physiol Endocrinol Metab*. 2012;302:E645–E653.
30. **Liu Y, Johnson BP, Shen AL, et al.** Loss of BMAL1 in ovarian steroidogenic cells results in implantation failure in female mice. *Proc Natl Acad Sci USA*. 2014;111:14295–14300.
31. **Alvarez JD, Sehgal A.** The thymus is similar to the testis in its pattern of circadian clock gene expression. *J Biol Rhythms*. 2005;20:111–121.
32. **Alvarez JD, Hansen A, Ord T, et al.** The circadian clock protein BMAL1 is necessary for fertility and proper testosterone production in mice. *J Biol Rhythms*. 2008;23:26–36.
33. **Yoshikawa T, Sellix M, Pezuk P, Menaker M.** Timing of the ovarian circadian clock is regulated by gonadotrophins. *Endocrinology*. 2009;150:4338–4347.
34. **Sellix MT, Yoshikawa T, Menaker M.** A circadian egg timer gates ovulation. *Curr Biol*. 2010;20:R266–R267.
35. **Menaker M, Murphy ZC, Sellix MT.** Central control of peripheral circadian oscillators. *Curr Opin Neurobiol*. 2013;23:741–746.
36. **Bridges PJ, Koo Y, Kang DW, et al.** Generation of Cyp17iCre transgenic mice and their application to conditionally delete estrogen receptor α (Esr1) from the ovary and testis. *Genesis*. 2008;46:499–505.
37. **Lee S, Kang DW, Hudgins-Spivey S, et al.** Theca-specific estrogen receptor-α knockout mice lose fertility prematurely. *Endocrinology*. 2009;150:3855–3862.
38. **Fan HY, Liu Z, Cahill N, Richards JS.** Targeted disruption of Pten in ovarian granulosa cells enhances ovulation and extends the life span of luteal cells. *Mol Endocrinol*. 2008;22:2128–2140.
39. **Sen A, Hammes SR.** Granulosa cell-specific androgen receptors are critical regulators of ovarian development and function. *Mol Endocrinol*. 2010;24:1393–1403.
40. **Hwang JW, Sundar IK, Yao H, Sellix MT, Rahman I.** Circadian clock function is disrupted by environmental tobacco/cigarette smoke, leading to lung inflammation and injury via a SIRT1-BMAL1 pathway. *FASEB J*. 2014;28:176–194.
41. **Boden MJ, Varcoe TJ, Kennaway DJ.** Circadian regulation of reproduction: from gamete to offspring. *Prog Biophys Mol Biol*. 2013;113:387–397.
42. **Sellix MT.** Clocks underneath: the role of peripheral clocks in the timing of female reproductive physiology. *Front Endocrinol (Lausanne)*. 2013;4:91.
43. **Wiegand SJ, Terasawa E.** Discrete lesions reveal functional heterogeneity of suprachiasmatic structures in regulation of gonadotropin secretion in the female rat. *Neuroendocrinology*. 1982;34:395–404.
44. **Davidson AJ, Menaker M.** Birds of a feather clock together–sometimes: social synchronization of circadian rhythms. *Curr Opin Neurobiol*. 2003;13:765–769.
45. **Reissmann T, Schally AV, Bouchard P, Riethmiiller H, Engel J.** The LHRH antagonist cetrorelix: a review. *Hum Reprod Update*. 2000;6:322–331.
46. **Dierich A, Sairam MR, Monaco L, et al.** Impairing follicle-stimulating hormone (FSH) signaling in vivo: targeted disruption of the FSH receptor leads to aberrant gametogenesis and hormonal imbalance. *Proc Natl Acad Sci USA*. 1998;95:13612–13617.
47. **Smith PL, Bousfield GR, Kumar S, Fiete D, Baenziger JU.** Equine lutropin and chorionic gonadotropin bear oligosaccharides terminating with SO4–4-GalNAc and Sia α 2,3Gal, respectively. *J Biol Chem*. 1993;268:795–802.
48. **Montgomery V, Loutradis D, Tulchinsky D, Kiessling A.** FSH-induced ovulation in intact and hypophysectomized mice. *J Reprod Fertil*. 1988;84:1–6.
49. **Pakarainen T, Zhang FP, Nurmi L, Poutanen M, Huhtaniemi I.** Knockout of luteinizing hormone receptor abolishes the effects of follicle-stimulating hormone on preovulatory maturation and ovulation of mouse graafian follicles. *Mol Endocrinol*. 2005;19:2591–2602.
50. **Bronson FH.** The regulation of luteinizing hormone secretion by estrogen: relationships among negative feedback, surge potential, and male stimulation in juvenile, peripubertal, and adult female mice. *Endocrinology*. 1981;108:506–516.
51. **Turgeon JL, Waring DW.** Luteinizing hormone secretion from wild-type and progesterone receptor knockout mouse anterior pituitary cells. *Endocrinology*. 2001;142:3108–3115.
52. **Jayes FL, Burns KA, Rodriguez KF, Kissling GE, Korach KS.** The naturally occurring luteinizing hormone surge is diminished in mice lacking estrogen receptor β in the ovary. *Biol Reprod*. 2014;90:24.
53. **Winkler N, Bukulmez O, Hardy DB, Carr BR.** Gonadotropin releasing hormone antagonists suppress aromatase and anti-Mullerian hormone expression in human granulosa cells. *Fertil Steril*. 2010;94:1832–1839.
54. **Espey LL, Richards JS.** Ovulation. In: Neill J, ed. *Physiology of*

Reproduction. vol 1. 3rd ed. St. Louis, MO: Elsevier Academic Press; 2006:425–474.
55. **Chu G, Yoshida K, Narahara S, et al.** Alterations of circadian clockworks during differentiation and apoptosis of rat ovarian cells. *Chronobiol Int*. 2011;28:477–487.
56. **Leung PCK, Adashi EY.** *The Ovary*. 2nd ed. Amsterdam, Boston: Elsevier; 2004.
57. **Orisaka M, Hattori K, Fukuda S, et al.** Dysregulation of ovarian follicular development in female rat: LH decreases FSH sensitivity during preantral-early antral transition. *Endocrinology*. 2013;154: 2870–2880.
58. **Murayama C, Miyazaki H, Miyamoto A, Shimizu T.** Luteinizing hormone (LH) regulates production of androstenedione and progesterone via control of histone acetylation of StAR and CYP17 promoters in ovarian theca cells. *Mol Cell Endocrinol*. 2012;350: 1–9.
59. **Sen A, Prizant H, Light A, et al.** Androgens regulate ovarian follicular development by increasing follicle stimulating hormone receptor and microRNA-125b expression. *Proc Natl Acad Sci USA*. 2014; 111:3008–3013.
60. **Christenson LK, Gunewardena S, Hong X, Spitschak M, Baufeld A, Vanselow J.** Research resource: preovulatory LH surge effects on follicular theca and granulosa transcriptomes. *Mol Endocrinol*. 2013;27:1153–1171.
61. **Richards JS, Russell DL, Robker RL, Dajee M, Alliston TN.** Molecular mechanisms of ovulation and luteinization. *Mol Cell Endocrinol*. 1998;145:47–54.
62. **Shimizu T, Hirai Y, Murayama C, Miyamoto A, Miyazaki H, Miyazaki K.** Circadian clock genes Per2 and clock regulate steroid production, cell proliferation, and luteinizing hormone receptor transcription in ovarian granulosa cells. *Biochem Biophys Res Commun*. 2011;412:132–135.
63. **West-Farrell ER, Xu M, Gomberg MA, Chow YH, Woodruff TK, Shea LD.** The mouse follicle microenvironment regulates antrum formation and steroid production: alterations in gene expression profiles. *Biol Reprod*. 2009;80:432–439.
64. **Hunzicker-Dunn M, Maizels ET.** FSH signaling pathways in immature granulosa cells that regulate target gene expression: branching out from protein kinase A. *Cell Signal*. 2006;18:1351–1359.
65. **Parakh TN, Hernandez JA, Grammer JC, et al.** Follicle-stimulating hormone/cAMP regulation of aromatase gene expression requires β-catenin. *Proc Natl Acad Sci USA*. 2006;103:12435–12440.
66. **Nimz M, Spitschak M, Fürbass R, Vanselow J.** The pre-ovulatory luteinizing hormone surge is followed by down-regulation of CYP19A1, HSD3B1, and CYP17A1 and chromatin condensation of the corresponding promoters in bovine follicles. *Mol Reprod Dev*. 2010;77:1040–1048.
67. **Vanselow J, Spitschak M, Nimz M, Fürbass R.** DNA methylation is not involved in preovulatory down-regulation of CYP11A1, HSD3B1, and CYP19A1 in bovine follicles but may have a role in permanent silencing of CYP19A1 in large granulosa lutein cells. *Biol Reprod*. 2010;82:289–298.
68. **Caron KM, Clark BJ, Ikeda Y, Parker KL.** Steroidogenic factor 1 acts at all levels of the reproductive axis. *Steroids*. 1997;62:53–56.
69. **Logan KA, Juengel JL, McNatty KP.** Onset of steroidogenic enzyme gene expression during ovarian follicular development in sheep. *Biol Reprod*. 2002;66:906–916.
70. **Hinshelwood MM, Repa JJ, Shelton JM, Richardson JA, Mangelsdorf DJ, Mendelson CR.** Expression of LRH-1 and SF-1 in the mouse ovary: localization in different cell types correlates with differing function. *Mol Cell Endocrinol*. 2003;207:39–45.
71. **Falender AE, Lanz R, Malenfant D, Belanger L, Richards JS.** Differential expression of steroidogenic factor-1 and FTF/LRH-1 in the rodent ovary. *Endocrinology*. 2003;144:3598–3610.
72. **Sellix MT, Murphy ZC, Menaker M.** Excess androgen during puberty disrupts circadian organization in female rats. *Endocrinology*. 2013;154:1636–1647.
73. **Mereness AL, Murphy ZC, Sellix MT.** Developmental programming by androgen affects the circadian timing system in female mice. *Biol Reprod*. 2015;92:88.

nature genetics

Genome-wide analysis identifies 12 loci influencing human reproductive behavior

The genetic architecture of human reproductive behavior—age at first birth (AFB) and number of children ever born (NEB)—has a strong relationship with fitness, human development, infertility and risk of neuropsychiatric disorders. However, very few genetic loci have been identified, and the underlying mechanisms of AFB and NEB are poorly understood. We report a large genome-wide association study of both sexes including 251,151 individuals for AFB and 343,072 individuals for NEB. We identified 12 independent loci that are significantly associated with AFB and/or NEB in a SNP-based genome-wide association study and 4 additional loci associated in a gene-based effort. These loci harbor genes that are likely to have a role, either directly or by affecting non-local gene expression, in human reproduction and infertility, thereby increasing understanding of these complex traits.

Human reproductive behavior—AFB and NEB—has been associated with human development[1,2], infertility[3,4] and neuropsychiatric disorders[5]. Reproductive tempo (AFB) and quantum (NEB) are cross-cutting topics in the medical, biological, evolutionary and social sciences and are central in national and international policies[6]. Advanced societies have experienced a rapid postponement of AFB, with the mean AFB of women now being 28–29 years in many countries[7]. This increase in AFB has been linked to lower fertility rates, unprecedented rates of childlessness (~20%) and infertility, which affects 10 to 15% of couples[8]. An estimated 48.5 million couples worldwide are infertile, with a large part of subfertility, particularly in men, remaining unexplained[9]. Although infertility has been related to advanced AFB, ovulation defects, failure of spermatogenesis, and single-gene or polygenic defects, the causal effects for these factors remain unsubstantiated[10].

Recently, genetic and clinical research has focused on proximal infertility phenotypes[3,4,10,11]. AFB and NEB represent accurate measures of complex reproductive outcomes, are frequently recorded and consistently measured, and are key parameters for demographic population forecasting[12]. There is evidence of a genetic component underlying reproduction, with heritability estimates of up to 50% for AFB and NEB (**Supplementary Fig. 1**)[6]. A recent study attributed 15% of the variance in AFB and 10% of the variance in NEB to common genetic variants[13]. There are also sex-specific differences in human reproduction, related to the timing of fertility, fecundability and sex–genotype interactions (**Supplementary Note**). Researchers have given less attention to traits such as NEB because of an erroneous and frequently repeated misinterpretation of Fisher's fundamental theorem of natural selection[14] that the additive genetic variance in fitness should be close to zero. This misreading of the theorem had a naively intuitive appeal: genes that reduce fitness should be passed on less frequently. Fisher, however, actually argues that fitness is moderately heritable in human populations (for a discussion, see the **Supplementary Note**). As no established genes are currently available for clinical testing of infertility[10], isolating genetic loci and their causative effects has the potential to provide new insights into the etiology of reproductive diseases and novel diagnostic and clinical technologies for infertility treatment.

RESULTS

We report a large meta-analysis of genome-wide association studies (GWAS) of 251,151 individuals for AFB and 343,072 individuals for NEB from a total of 62 cohorts of European ancestry. We identify 12 independent loci (10 of which are new and 2 of which were previously identified in a study on age at first sexual intercourse[11]) that are significantly associated with AFB and/or NEB in men, women or both sexes combined (**Table 1**). Follow-up analyses identified a number of genetic variants and genes that likely drive the GWAS associations. We also quantified the genetic overlap with biologically adjacent reproductive, developmental, neuropsychiatric and behavioral phenotypes. A detailed description of all materials and methods is available in the **Supplementary Note**.

Meta-analysis of GWAS

Associations of AFB (mean ± s.d., 26.8 ± 4.78 years) and/or NEB (mean ± s.d., 2.3 ± 1.43 children) with SNPs imputed from NCBI Build 37 HapMap phase 2 data were examined in 62 cohorts using multiple linear regression under an additive model, in men and women separately (**Supplementary Note**). Associations were adjusted for principal components, to reduce confounding by population stratification[15], as well as for the birth year of the respondent and its square and cube to control for nonlinear birth cohort effects (**Supplementary Tables 1** and **2**, and **Supplementary Note**). NEB was assessed only for those who had completed their reproductive period (age ≥45 years for women and ≥55 years for men), while AFB was only assessed for those who were parous. Quality control was conducted in two independent centers using QCGWAS[16] and EasyQC[17] (**Supplementary Note**). Results were subsequently submitted to meta-analysis for the 2.4 million SNPs that passed quality control filters (**Supplementary Note**) and are reported for men and women combined and separately.

A full list of authors and affiliations appears at the end of the paper.

Received 12 June; accepted 22 September; published online 31 October 2016; doi:10.1038/ng.3698

Table 1 GWAS meta-analysis results for independent loci that are genome-wide significantly ($P < 5.0 \times 10^{-8}$) associated with AFB or NEB in either the combined or sex-specific meta-analysis

SNP	Chr.	Position (bp)	Nearest gene	Annotation	Effect allele/ other allele	EAF	β	*P* value	*n* (pooled)	β (men)	*P* value (men)	β (women)	*P* value (women)
Age at first birth													
rs10908557	1	153,927,052	*CRTC2*	N, R, ctQ, ctM	C/G	0.695	0.091	**5.59×10^{-10}**	249,025	0.185	2.98×10^{-7}	0.076	5.38×10^{-6}
rs1160544	2	100,832,218	*LINC01104*	R, cQ, cM	A/C	0.395	−0.082	**2.90×10^{-9}**	250,330	−0.042	2.12×10^{-1}	−0.092	**5.00×10^{-9}**
rs2777888	3	49,898,000	*CAMKV*	N, R, ctQ, ctM	A/G	0.507	0.106	**4.58×10^{-15}**	250,941	0.155	2.40×10^{-6}	0.095	**6.07×10^{-10}**
rs6885307	5	45,094,503	*MRPS30, HCN1*	R, ctQ, cM	A/C	0.799	−0.107	**2.32×10^{-10}**	248,999	−0.131	2.07×10^{-3}	−0.104	**3.90×10^{-8}**
rs10056247	5	133,898,136	*JADE2*	cQ, cM	T/C	0.289	0.082	**4.37×10^{-8}**	249,429	0.050	1.68×10^{-1}	0.089	1.28×10^{-7}
rs2347867	6	152,229,850	*ESR1*	cM	A/G	0.649	0.091	**1.38×10^{-10}**	248,039	0.098	4.69×10^{-3}	0.097	**1.80×10^{-9}**
rs10953766	7	114,313,218	*FOXP2*	cM	A/G	0.429	0.087	**1.82×10^{-10}**	248,039	0.106	1.31×10^{-3}	0.089	**8.41×10^{-9}**
rs2721195	8	145,677,011	*CYHR1*	R, cQ, ctM	T/C	0.469	−0.073	6.25×10^{-7}	250,493	−0.014	6.85×10^{-1}	−0.099	**6.13×10^{-9}**
rs293566	20	31,097,877	*NOL4L*	cQ, cM	T/C	0.650	0.081	**1.41×10^{-8}**	245,995	0.110	1.47×10^{-3}	0.079	1.31×10^{-6}
rs242997	22	34,503,059	*LARGE1, ISX*		A/G	0.613	−0.084	**3.38×10^{-9}**	238,002	−0.139	8.51×10^{-5}	−0.076	1.82×10^{-6}
Number of children ever born													
rs10908474	1	153,753,725	*SLC27A3, GATAD2B*		A/C	0.384	0.020	**2.08×10^{-8}**	342,340	0.021	8.10×10^{-4}	0.020	7.89×10^{-6}
rs13161115	5	107,050,002	*EFNA5, FBXL17*	cM	C/G	0.234	−0.041	1.34×10^{-2}	341,737	−0.041	**1.37×10^{-8}**	0.005	3.29×10^{-1}
rs2415984	14	46,873,776	*LINC00871*	cM	A/G	0.470	−0.020	**2.34×10^{-8}**	315,167	−0.029	2.41×10^{-6}	−0.016	3.71×10^{-4}

The rows in bold correspond to the independent signals reaching $P < 5 \times 10^{-8}$ in the meta-analysis. Annotation shows for each of the 12 independent lead SNPs (excluding rs10908474 on chromosome 1) whether it is (i) in strong LD ($r^2 > 0.8$) with a nonsynonymous variant (N) or one or more variants prioritized by RegulomeDB (R) with evidence of having functional consequences (defined by a score <4); (ii) associated with an eQTL in *cis* and/or *trans* (ctQ); and (iii) associated with an meQTL in *cis* and/or *trans* (ctM). EAF, effect allele frequency of the pooled meta-analysis; β, effect size in the AFB and NEB analyses. All *P* values are from the sample-size-weighted fixed-effects meta-analysis.

We applied a single genomic control at the cohort level and performed meta-analysis of results using a sample-size-weighted fixed-effect method in METAL (**Supplementary Note**). The PLINK clumping function isolated 'lead SNPs'—those with the lowest *P* value for association that are independently associated—using an r^2 threshold of 0.1 and a distance threshold of 500 kb. For AFB, we identified ten loci associated at genome-wide significance ($P < 5 \times 10^{-8}$ for combined results and $P < 1.67 \times 10^{-8}$ for sex-specific results adjusted for multiple testing), of which 9 were significantly associated in both sexes combined and 1 was associated in women only (n = 154,839) (**Fig. 1a** and **Table 1**). Three loci were significantly associated with NEB: two in both sexes combined and one in men only (n = 103,736) (**Fig. 1b**, **Table 1** and **Supplementary Note**). One locus on chromosome 1 reached significance for association with both AFB and NEB with $r^2 = 0.57$ between the two lead SNPs, suggesting a shared genetic basis for the two traits (**Table 2**). A statistical test of sex-specific effects confirmed that differences are mainly due to variation in sample size and not variation in effect size (**Supplementary Note**).

As for other complex traits[18], the quantile–quantile plots of the meta-analyses exhibited strong inflation of low *P* values (**Fig. 2**), suggesting that, although cohorts controlled for the top principal components and cohort-level genomic control was applied (**Supplementary Note**), residual population stratification may remain. However, the LD Score intercept method[19] as well as a series of individual and within-family regression analyses using polygenic scores as predictors[20,21] (**Supplementary Note**) indicated that the observed inflation was almost entirely attributable to a true polygenic signal, rather than population stratification.

Gene-based GWAS

To increase the power to find statistically significant associations and causal genes, we additionally performed a gene-based GWAS using VEGAS[22,23]. The results confirmed top hits from the single-SNP analyses. For AFB, seven loci from the SNP-based GWAS were also represented in the gene-based analysis (**Supplementary Table 3**), and three additional loci emerged, represented by *SLF2* (chromosome 10), *ENO4* (chromosome 10) and *TRAF3-AMN* (chromosome 14). For NEB, one locus from the SNP-based GWAS was represented in the gene-based analysis—*GATAD2B* (chromosome 1)—and one new locus on chromosome 17 was identified (**Supplementary Table 4**).

Causal variants

To identify functional and potentially causal variants, both coding and regulatory, within loci identified in the SNP-based GWAS (**Table 1**), we first performed an *in silico* sequencing annotation analysis using the post-GWAS pipeline reported by Vaez *et al.*[24]. This showed that rs10908557 on chromosome 1 is in high linkage disequilibrium (LD) with nonsynonymous SNPs in *CRTC2* (rs11264680; $r^2 = 0.98$) and *CREB3L4* (rs11264743; $r^2 = 0.94$) (**Supplementary Table 5**). Interestingly, rs11264743 is considered 'deleterious' and 'probably damaging' by SIFT and PolyPhen, respectively (Ensembl release 83). In addition, rs2777888 on chromosome 3 is in high LD with two nonsynonymous SNPs in *MST1R* (rs2230590, $r^2 = 0.95$ and rs1062633, $r^2 = 0.95$) (**Table 1** and **Supplementary Table 5**).

We subsequently performed a comprehensive analysis using results from the Encyclopedia of DNA Elements (ENCODE)[25] and Roadmap Epigenomics[26] projects, as integrated in RegulomeDB[27], to identify variants that likely influence downstream gene expression via regulatory pathways. Among all SNPs that reached $P < 5 \times 10^{-8}$ in the meta-analyses (n = 322), 50 SNPs in five loci showed the most evidence of having functional consequences (**Table 1** and **Supplementary Table 6**).

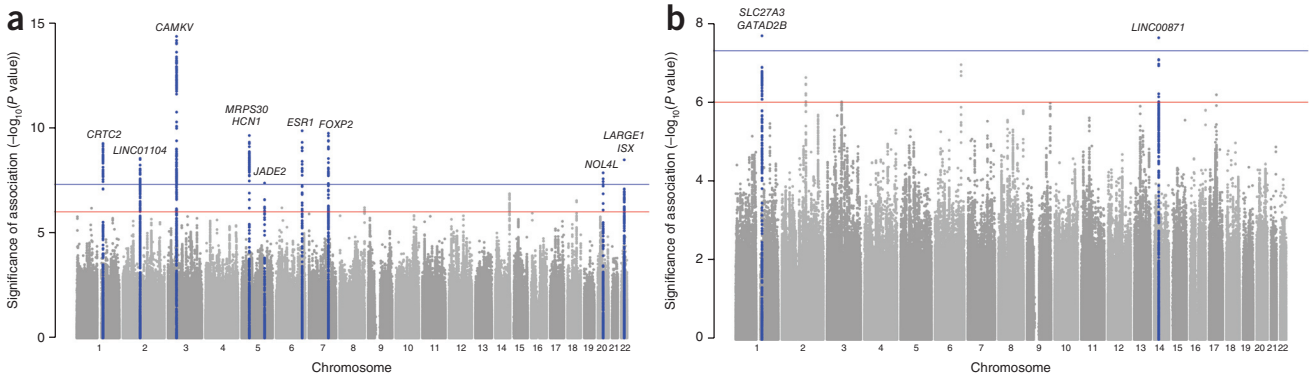

Figure 1 Manhattan plots of SNPs for age at first birth and number of children ever born in single-genomic-control meta-analysis. (**a**,**b**) SNPs are plotted on the *x* axis according to their position on each chromosome against association with AFB (**a**) and NEB (**b**). The solid blue line indicates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$), and the red line represents the threshold for suggestive hits ($P < 5 \times 10^{-6}$). Blue points represent SNPs in a 100- kb region centered on genome-wide significant hits. Loci are annotated with the names of the genes closest to the significant SNPs.

Two sets of SNPs on chromosome 1 (18 SNPs) and chromosome 3 (25 SNPs) stand out in particular. The most promising SNP in the chromosome 1 locus (rs6680140) is located in a site of acetylation of histone H3 at lysine 27 (H3K27ac), often found near active regulatory elements, and lies in a DNase I hypersensitivity cluster where eight proteins are anticipated to bind. One of these proteins is cAMP responsive element binding (CREB)-binding protein, encoded by *CREBBP*. In the chromosome 3 locus, rs2526397 is located in a transcription factor binding site and is an expression quantitative trait locus (eQTL) for *HYAL3* in monocytes, while rs2247510 and rs1800688 are located in H3K27ac sites and DNase I hypersensitivity clusters where a large number of transcription factors are expected to bind (**Supplementary Table 6**). An analysis using Haploplotter showed that rs2247510 and rs7628058 in the chromosome 3 locus are among the 5% of signals showing the most evidence of positive selection in the population. The same applies to the lead SNP of the chromosome 14 locus for NEB (rs2415984).

Causal genes

Information on the function and anticipated relevance of genes in the 12 loci identified in the SNP-based GWAS that are most likely to be causal on the basis of all evidence discussed below is provided in **Table 2**.

Cis- and *trans*-eQTL and meQTL analyses

Identifying alterations in gene methylation status and/or expression levels in relation to GWAS-identified variants may help prioritize causal genes. We examined associations with gene expression and methylation status for the 12 independent lead SNPs in whole-blood BIOS eQTL ($n = 2{,}116$) and methylation quantitative trait locus (meQTL; $n = 3{,}841$) databases in *cis* and *trans*[28,29]. Seven SNPs were associated in *cis* with the expression of 54 unique genes (**Table 1** and **Supplementary Table 7**). Five of these seven SNPs were in high LD ($r^2 > 0.8$) with the strongest eQTL for at least one of the genes within the corresponding locus, indicating that the SNP associated with AFB or NEB and the SNP most significantly associated with expression tag the same functional site: rs10908557 (associated with the expression of *CRTC2* and *SLC39A1*), rs1160544 (*AFF3*), rs2777888 (*RBM6*, *RNF123* and *RBM5*), rs2721195 (*CYHR1*, *GPT*, *RECQL4* and *PPP1R16A*) and rs293566 (*NOL4L*). Three SNPs were associated with the expression of a total of eight genes in *trans* (**Table 1** and **Supplementary Table 8**). Of these SNPs, only rs2777888 was in high LD ($r^2 > 0.8$) with the strongest eQTL for three of its five associated genes: *LRFN1*, *LAMP2* and *FGD3*.

The meQTL analysis showed that 11 of the 12 independent lead SNPs were associated with DNA methylation of a total of 131 unique genes in *cis* (**Table 1** and **Supplementary Table 9**). Seven of the 11 SNPs were in high LD ($r^2 > 0.8$) with the strongest meQTL for one of the corresponding methylation sites: rs10908557 (associated with methylation of *CRTC2*, *SLC39A1*, *CREB3L4*, *DENND4B* and *RAB13*), rs1160544 (*AFF3*), rs2777888 (*CAMKV*), rs6885307 (*C5orf34*), rs10056247 (*JADE2*), rs2721195 (*CYHR1*) and rs13161115 (*EFNA5*). Three of the SNPs were associated with the same genes for both methylation and gene expression in *cis*: rs10908557 (*CRTC2*), rs1160544 (*AFF3*) and rs2721195 (*CYHR1*) (**Supplementary Tables 7** and **9**). Three SNPs were associated with methylation of a total of ten genes in *trans* (**Table 1** and **Supplementary Table 10**). Of these SNPs, only rs2777888 was in high LD ($r^2 > 0.8$) with the strongest meQTL for a corresponding methylation site (*ASAP3*). Of note, rs2777888 was also a *trans*-eQTL.

Gene prioritization

We used four publicly available bioinformatics tools to systematically identify genes that are more likely than neighboring genes to cause the associations identified by our GWAS. Of all genes that reached $P < 0.05$ in analyses using Endeavor[30], MetaRanker[31] and ToppGene[32], eight genes were prioritized for both AFB and NEB: *TPM3*, *GRM7*, *TKT*, *MAGI2*, *PTPRD*, *PTPRM*, *RORA* and *WT1*. DEPICT—a fourth comprehensive and unbiased recently described gene prioritization tool[33]—identified three genes in GWAS significant loci as likely being causal for AFB (*MON1A*, *RBM6* and *U73166.2*) (**Supplementary Tables 11** and **12**).

Gene-based results from RegulomeDB

An analysis using RegulomeDB identified 50 SNPs in five loci that most likely have regulatory consequences (**Supplementary Table 6**). Eighteen and 25 of these SNPs are within the chromosome 1 and chromosome 3 loci, respectively. Among the genes that, at a protein level, bind at the site of one or more of the 18 variants in the locus on chromosome 1 are *CREBBP*, *HNF4A*, *CDX2* and *ERG*. These genes may act upstream in the causal pathway and influence the expression of causal genes at this locus. Of the 25 SNPs on chromosome 3, 10 were eQTLs for *RBM6* in monocytes and 7 were eQTLs for *HYAL3* in monocytes. Among the genes that, at a protein level, bind at rs2247510 and rs1800688 in the chromosome 3 locus are *ARID3A*, *REST* and *TFAP2C*, as well as *HNF4A* and *CDX2*, which also bind at the chromosome 1 locus.

Table 2 Function and potential relevance for genes in GWAS-identified loci that are most likely causal on the basis of all available evidence

Lead SNP	Gene	Chr.	Evidence	Gene function and potential role in reproduction and (in)fertility	Ref.
rs10908557	*CRTC2*	1	G, V, ctQ, ctM, Q lymph. (R)	Functions as a Ca^{2+}- and cAMP-sensitive coincidence sensor; promotes CREB target gene expression; signal mediator in FSH and TGF-β1 steroidogenesis in ovarian granulosa cells	42
rs10908557	*CREB3L4*	1	N, V, cQ, cM	Has a role in protein maturation; involved in spermatid differentiation and male germ cell development; expressed in prostate, oocytes, fallopian tube and mammary gland	44,45
rs10908557	*GATAD2B*	1	V, Q monoc. (R)	Transcriptional repressor and a component of nucleosome remodeling complex Mi2/NuRD; increased expression in endometriosis; linked to a common gynecological disorder that causes pelvic pain and infertility	58,59
rs10908557	*SLC39A1*	1	V, cQ, cM	Zinc uptake transporter; major zinc regulator in prostate cells; involved in the regulation of zinc homeostasis by cumulus cells in the oocyte	60,61
rs10908557	*DENND4B*	1	cM	Paralog of *DENN1A*, which has been implicated in polycystic ovary syndrome; expressed at the protein level in the cervix	46,62
rs1160544	*AFF3*	2	cQ, cM	Lymphoid nuclear transcriptional activator implicated in tumorigenesis; same family as *AFF3* and *AFF4* (*FMR2* family member 4); transcriptional regulator in testicular somatic cells; essential for male germ cell differentiation and survival in mice	63,64
rs1160544	*LINC01104*	2	G, V	Unknown	
rs2777888	*HYAL3*	3	cM, Q monoc. (R)	Hyaluronidases, including HYAL3, are involved in degradation of hyaluronan, a major glycosaminoglycan of the extracellular matrix; acquired during sperm maturation in the epididymis and involved in sperm function and the acrosome reaction; required for *in vitro* cumulus penetration in mice, although its absence is not associated with infertility (perhaps compensated for by other hyaluronidases)	65
rs2777888	*RBM6*	3	V, cQ, cM, DEPICT, Q monoc. (R)	Involved in RNA splicing	66
rs2777888	*RNF123*	3	V, cQ, cM, Q liver (R)	Has a role in cellular transitioning from quiescence to a proliferative state through its E3 ubiquitin ligase activity toward cyclin-dependent kinase inhibitor 1B, which controls cell cycle progression in G1 phase	66–68
rs2777888	*RBM5*	3	V, cQ	Involved in cell cycle regulation; regulator of pre-mRNA splicing; involved in spermatogenesis and fertility in mice	47
rs2777888	*MON1A*	3	V, cM, DEPICT	Involved in the movement and trafficking of proteins (for example, ferroportin) through the secretory apparatus	69
rs2777888	*U73166.2*	3	DEPICT	Unknown	
rs2777888	*MST1R*	3	N, V, cM, MetaRanker, ToppGene and Endeavor	Cell surface receptor for MSP with tyrosine kinase activity, expressed on ciliated epithelia of the mucociliary transport apparatus of the lung; involved in host defense, expressed in sperm; may act in a regulatory system of ciliary motility, together with MSP, which sweeps eggs along the oviduct; expressed in mucous membrane and mammary gland	70
rs10056247	*JADE2*	5	G, V, cM	Involved in histone acetylation	
rs13161115	*EFNA5*	5	cM	Involved in development and differentiation of the nervous system and folliculogenesis regulation; required for normal fertility in female mice; expressed in embryonic stem cells and embryoid bodies	50
rs6885307	*HCN1*	5	G, cM	Hyperpolarization-activated cation channel that contributes to the native pacemaker current in, for example, neurons; HCN1 channels are present in kisspeptin (Kiss1) neurons in the rostral periventricular area of the third ventricle (RP3V), which provide an excitatory drive to gonadotropin-releasing hormone (GnRH)-expressing neurons that control fertility	71
rs2347867	*ESR1*	6	G, cM, binds at rs4851269 on chr. 2 (R)	Transcription factor involved in estrogen-responsive gene expression; essential for sexual development and reproductive function in women; genetic variants in *ESR1* may influence susceptibility to endometriosis or female fertility in patients with endometriosis; involved in male fertility by transferring estrogen effect; expressed in myometrium, endometrium, oocytes, uterus and fallopian tube	51,52, 72–74
rs10953766	*FOXP2*	7	G, cM, binds at rs6997 on chr. 3 (R)	Transcription factor expressed in fetal and adult brain that is involved in speech and language development; involved in regulation of neuronal development in the embryonic forebrain; expressed in mucous membrane and myometrium	75
rs2721195	*CYHR1*	8	cQ, cM	Histidine-cysteine-rich protein involved in spermatogenesis	53
rs2721195	*GPT*	8	V, cQ, cM, Q monoc. (R)	Involved in intermediary metabolism of glucose and amino acids; may have a role in spermatogenesis via testicular glucose metabolism, which is pivotal for the normal occurrence of spermatogenesis; levels in the normal range are positively associated with metabolic and endocrine abnormalities in women of reproductive age and negatively associated with FSH levels, independent of obesity	76,77
rs2721195	*RECQL4*	8	V, cQ, cM	Processing of aberrant DNA structures that arise during DNA replication and repair; predominantly expressed in testis	78
rs2721195	*PPP1R16A*	8	V, cQ, cM, Q monoc. (R)	Regulator of protein phosphatase PP1β; present in the sperm tail where it interacts with proteins that are important in sperm structure and motility; expressed in mammary gland and fallopian tube	79
rs293566	*NOL4L*	20	cQ, cM	Component of cytoplasm and nucleoplasm; expressed in umbilical vein	

Evidence categories include the nearest gene (G), nonsynonymous variants (N), gene-based GWAS performed in VEGAS (V), eQTLs in *cis* and/or *trans* (ctQ), meQTLs in *cis* and/or *trans* (ctM), eQTLs (Q) in lymphoblasts (lymph), monocytes (monoc) or liver based on RegulomeDB (R), gene prioritization using either DEPICT or MetaRanker, ToppGene and Endeavor, and protein binding at SNPs based on RegulomeDB (R). Chr., human chromosome on which the gene is located; FSH, follicle-stimulating hormone; CREB, cAMP response element–binding protein; TGF-β1, transforming growth factor β1; MSP, macrophage-stimulating protein. SNIPPER was used for the literature search, with the search terms "fertility," "sperm," "ovum" and "reproduction."
Gene Network was used to find the tissue or organ with high expression for a given gene (AUC > 0.8).

Five genes encode proteins that bind at the site of both SNPs on chromosome 2 that reach $P < 5 \times 10^{-8}$ in the meta-analysis of GWAS. One of these is *REST*; another one, *ESR1*, is the most likely causal gene in the chromosome 6 locus.

Functional network and enrichment analyses

Functional network analysis using five prioritized candidate gene sets as input (**Supplementary Note**) showed no significantly enriched biological function. No biological function was significantly enriched after correction for multiple testing using the Benjamini–Hochberg procedure. Similarly, no reconstituted gene sets and cell or tissue types were significantly enriched in the GWAS meta-analysis results based on results from the DEPICT analysis (**Supplementary Tables 13–20**). The lack of significantly enriched genes, tissue sets and biological functions reflects the need for a larger sample size but also the distal nature of the phenotypes, which are influenced by a mixture of biological, psychological and socioenvironmental factors.

Polygenic prediction

To assess the predictive power of our results, we produced polygenic scores for AFB and NEB with sets of SNPs whose nominal *P* values ranged from $P < 5 \times 10^{-8}$ (using only genome-wide significant SNPs) to 1 (using all SNPs that passed quality control) using PRSice[34] (**Supplementary Note**). We then performed a series of four different out-of-sample predictions in four independent cohorts: HRS, LifeLines, STR and TwinsUK. Across the four cohorts, the mean predictive power of a polygenic score constructed from all measured SNPs is 0.9% for AFB and 0.2% for NEB (**Supplementary Fig. 2**). Despite the low predictive power of the polygenic scores, the results showed that an increase of 1 s.d. in the NEB polygenic score is associated with a 9% (95% confidence interval (CI) = 5–13%) decrease in the probability of women remaining childless, with no significant association in men (**Supplementary Table 21**). When we controlled for right-censored data using a survival model for AFB, we found that an increase of 1 s.d. in the AFB polygenic score was associated with an 8% (95% CI = 7–10%) reduction in the hazard ratio of reproduction in women and a 3% (95% CI = 1–5%) reduction in men (**Supplementary Table 22**). As an additional test, we examined whether the aforementioned polygenic scores for AFB and NEB could predict related fertility traits such as age at menopause and age at menarche (**Supplementary Table 23**). Our estimates indicated that an increase of 1 s.d. in the AFB polygenic score was associated with a 3% decrease in the probability of natural menopause at any age (95% CI = 1–5%) and a 20-d increase in age at menarche (95% CI = 0.4–40 d).

Genetic association with related traits and diseases

Several loci for which the associations with AFB and NEB reached genome-wide significance are associated with behavioral and reproductive phenotypes. The lead SNPs in the chromosome 2 and chromosome 3 loci have been associated with educational attainment[35] and the locus on chromosome 5 has been associated with age at menarche[2], while the locus on chromosome 6 has recently been associated with age at first sexual intercourse[11] (**Supplementary Table 24**). Some of the 38 loci for age at first sexual intercourse that were recently identified[11] in 125,667 UK Biobank participants were also associated with AFB (in or near *RBM6–SEMA3F* and *ESR1*) and NEB (in or near *CADM2* and *ESR1*). The lead SNPs for *RBM6–SEMA3F* (rs2188151) and *ESR1* (rs67229052) are in LD with our lead SNPs for AFB on chromosome 3 ($r^2 = 0.44$) and chromosome 6 ($r^2 = 0.94$), respectively. An *in silico* pleiotropy analysis showed that our lead SNP in the chromosome 3 locus (rs2777888) is in LD ($r^2 = 0.59$) with rs6762477, which has been

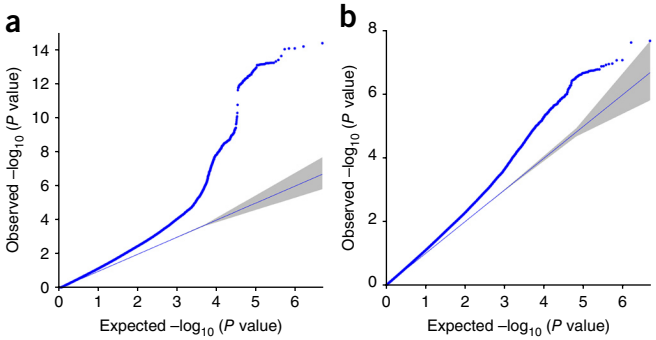

Figure 2 Quantile–quantile plots. (**a**,**b**) SNPs for AFB (**a**) and NEB (**b**) in single-genomic-control meta-analysis. The gray-shaded areas in the quantile–quantile plots represent the 95% confidence bands around the *P* values under the null hypothesis.

associated with age at menarche[2], while other SNPs in the same locus have been associated with HDL cholesterol[36] (rs2013208, $r^2 = 0.81$) and body mass index (BMI)[37] (rs7613875, $r^2 = 0.81$) (**Supplementary Table 5**). We next performed an exploratory analysis using the proxy phenotype method[38] to examine whether the SNPs strongly associated with AFB in women are empirically plausible candidate SNPs for related traits such as age at menarche and age at menopause (**Supplementary Note**). After controlling for multiple testing, we identified three AFB-associated SNPs near rs2777888 on chromosome 3 (rs9589, rs6803222 and rs9858889) that are also associated with age at menarche ($P < 4.10 \times 10^{-4}$). None of the AFB- or NEB-associated SNPs are associated with age at menopause.

We performed a bivariate LD score regression analysis[39] to estimate the pairwise genetic correlation with 27 publicly available GWAS results for traits associated with human reproductive behavior (**Supplementary Note**). AFB showed significant and positive genetic correlation with the human (reproductive) developmental traits of age at menarche, voice breaking, age at menopause, birth weight and age at first sexual intercourse, as well as with years of education. Conversely, having more AFB-increasing alleles was associated with a lower genetic risk of smoking (ever, number of cigarettes and later onset) and with lower insulin-resistance-related phenotypes, that is, BMI, waist–hip ratio adjusted for BMI, fasting insulin, triglyceride levels and risk of diabetes (**Fig. 3** and **Supplementary Table 25**). All genetic correlations remained significant after Bonferroni correction for multiple testing ($P < 2.6 \times 10^{-3}$). Years of education ($P = 6.6 \times 10^{-14}$) and age at first sexual intercourse ($P = 1.14 \times 10^{-15}$) are the only traits that showed significant and negative genetic correlation with NEB. We also observed significant genetic correlations of $r_g = 0.86$ (standard error (SE) = 0.052) for AFB and $r_g = 0.97$ (SE = 0.095) for NEB between men and women, implying that most genetic effects on reproductive behavior resulting from common SNPs are shared by both sexes.

DISCUSSION

This GWAS is a large-scale genetic epidemiological discovery effort for human reproduction, with implications for population fitness and physiological mechanisms linking hypothesized genes and observed phenotypes. Related studies previously focused on reproductive life span[40,41], age at first sexual intercourse[11] and more proximal infertility phenotypes[2–4], largely overlooking AFB and NEB. The rapid postponement of AFB and increased infertility and involuntary childlessness in many societies[7] make it important to uncover the genetic and biological architecture of reproduction. We identify ten

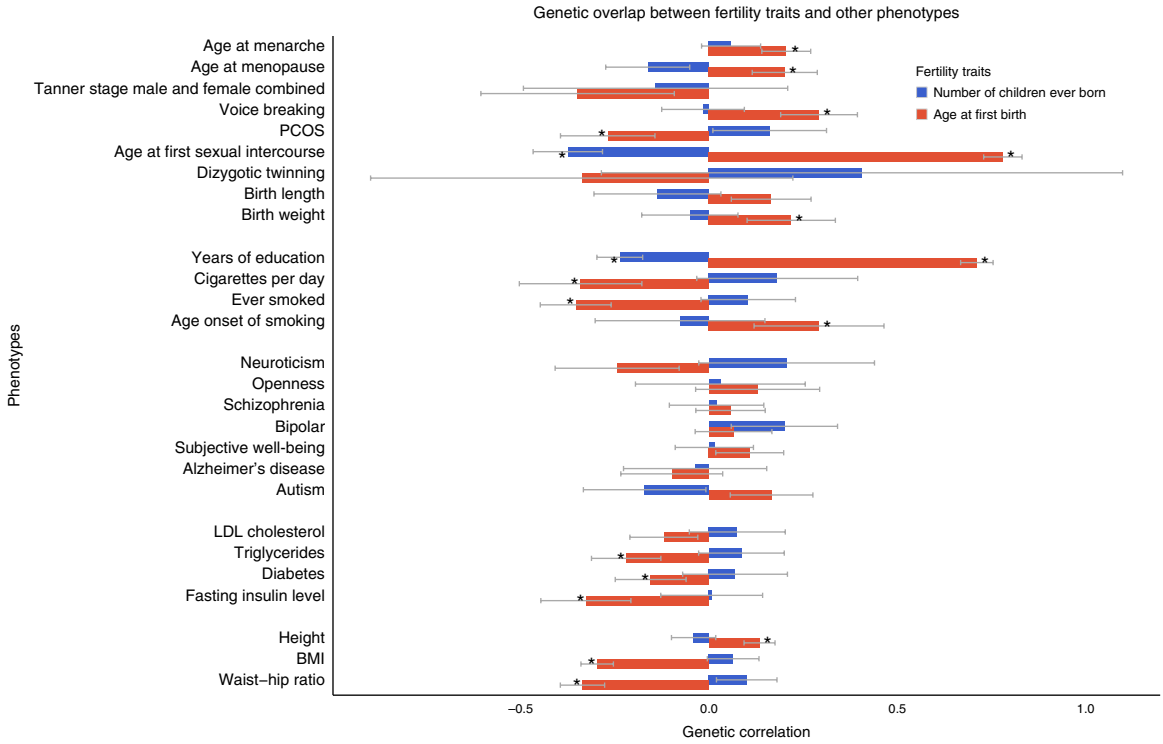

Figure 3 Genetic overlap between AFB or NEB and other related traits. Results from LD Score regressions show estimates of genetic correlation with developmental, reproductive, behavioral, neuropsychiatric and anthropometric phenotypes for which GWAS summary statistics were available in the public domain. The lengths of the bars correspond to estimates of genetic correlation. Gray error bars represent 95% confidence intervals. An asterisk indicates that the estimate of genetic correlation is statistically significant after controlling for multiple testing ($P < 0.05/27 = 1.85 \times 10^{-3}$).

new and confirm two recently identified genetic loci that are robustly associated with AFB and NEB, as well as variants and genes within these loci that potentially drive these associations. Four additional loci were identified in a gene-based GWAS.

Two loci that show interesting results in follow-up analyses are located on chromosomes 1 and 3. The lead SNPs of the chromosome 1 locus for AFB and NEB are in LD with likely functional nonsynonymous SNPs in genes encoding (i) CREB-regulated transcription co-activator 2 (*CRTC2*), which at the protein level acts as a critical signal mediator in follicle-stimulating hormone (FSH) and transforming growth factor (TGF)-β1–stimulated steroidogenesis in ovarian granulosa cells[42]; and (ii) CREB protein 3 like 4 (*CREB3L4*)[43], which in humans is highly expressed in the prostate, ovaries, uterus, placenta and testis and has a role in spermatid differentiation[44] and male germ cell development[45]. The lead SNP and/or functional variants in LD with it are also associated with the methylation status of these two genes and expression of *CRTC2* in whole blood and lymphocytes. Three promising functional variants in the chromosome 1 locus reside in binding sites of the transcriptional co-activator CREBBP. In addition to a direct effect of the above-mentioned nonsynonymous SNPs on protein function, the associations of AFB and NEB with variants in the locus on chromosome 1 may thus be mediated by alterations in cAMP responsive element binding in men and women. The locus on chromosome 1 also harbors *DENND4B*, a paralog of *DENND1A*, implicated in polycistic ovary syndrome (PCOS)[46]. Whereas *DENND1A* is expressed at the protein level in the ovary and testis, DENND4B is expressed in the cervix and its function and role are less well understood.

The lead SNP of the locus on chromosome 3 (rs2777888) is associated with methylation and expression of several genes, in *cis* and *trans*, that are known to have a role in cell cycle progression and/or sperm function. First, rs2777888 is associated with the expression of *RNF123* (or *KPC1*) in *cis*, which has a role in cellular transition from quiescence to a proliferative state. Second, rs2777888 or functional variants in LD with it may influence the cell cycle by altering the expression of *RBM5* and *RBM6* (RNA-binding motif proteins 5 and 6). The former has a role in cell cycle arrest and apoptosis induction and regulates haploid male germ cell pre-mRNA splicing and fertility in mice. *Rmb5*-mutant mice exhibit spermatid differentiation arrest, germ cell sloughing and apoptosis, leading to lack of sperm in ejaculation[47]. Third, rs2777888 affects expression of *LAMP2* in *trans*, which is located on the X chromosome and encodes a lysosomal membrane protein involved in the acrosome reaction, that is, the enzymatic drill allowing sperm to penetrate and fertilize ova[48]. *LAMP2* is expressed at the protein level in male and female reproductive organs with an effect size of rs2777888 for *LAMP2* mRNA expression that is almost twice as large in women than it is in men (**Supplementary Fig. 3**). This suggests an important role for the lysosome in both sperm and ova. Finally, functional variants in the chromosome 3 locus are associated with the mRNA expression of *HYAL3* (hyaluronoglucosaminidase 3) in monocytes. The encoded protein degrades hyaluronan, which also has an important role in sperm function and the acrosome reaction[47,49].

Functional follow-up experiments could disentangle the potential interplay between many candidate genes in the loci on chromosomes 1 and 3 in reproductive behavior and, given our *in silico* results, infertility. This can be extended to candidate genes in the remaining loci identified in the present study, some of which are relevant for fertility: mice lacking *Efna5* (chromosome 5 NEB locus) are subfertile[50], *ESR1* on chromosome 6 encodes an estrogen receptor[51,52] and *CYHR1* on chromosome 8 is involved in spermatogenesis[53]. Such experiments would help in understanding whether binding of estrogen receptor 1,

encoded by *ESR1* in the locus on chromosome 6, at the site of functional variants in the locus on chromosome 2 drives or mediates the association with AFB in the chromosome 2 locus, as well as to identify and characterize causal genes. Recent developments in high-throughput, multiplex mutagenesis using CRISPR/Cas9 allow such experiments to be performed using *in vivo* model systems[54].

AFB and NEB are not only driven by biological processes, but are also subject to individual choice and personal characteristics such as personality traits, as well as by the historical, cultural, economic and social environment (for example, effective contraception and childcare availability). Demographic research has shown a strong positive association between AFB and educational attainment[12]. We show that the associations between fecundity, reproductive behavior and educational attainment are partly driven by genetic factors and identified loci that are associated with AFB as well as with, for example, age at first sexual intercourse[11] and educational attainment[35].

Our findings could lead to insights into how postponing reproduction may be more detrimental for some, on the basis of their genetic make-up, than others, fuel experiments to determine 'how late can you wait' (ref. 55) and stimulate reproductive awareness. Causal genes in the loci we identified could potentially serve as novel drug targets, to prevent or delay age-related declines in fertility and sperm quality or to increase assisted reproductive technology efficiency, but further characterization is needed. Our study examines the genetics of reproductive behavior in both men and women, and, to our knowledge, it is the first that is adequately powered to identify loci in both women and men. We also provide support for Fisher's theorem that fitness is moderately heritable in human populations. Although the effect sizes of the identified common variants are small, there are examples of GWAS-identified loci of small effect that end up leading to important biological insights[56,57]. Many of our findings suggest a role for sperm quality, which is one lead for researchers to pursue. Because current sperm tests remain rudimentary, our findings could serve as a basis for 'good quality' sperm markers. We identified both coding and regulatory variants that are potentially causal, as well as a set of genes that could underlie the associations we identified. Follow-up experiments in animal models are required to confirm and characterize the causal genes in these loci.

URLs. Analysis plan predeposited at the Open Science Framework website, https://osf.io/53tea/; Gene Network, http://129.125.135.180:8080/GeneNetwork/; ReproGen, http://www.reprogen.org/data_download.html; Sociogenome, http://www.sociogenome.com/; Social Science Genetic Association Consortium, http://thessgac.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

For full acknowledgments, see the **Supplementary Note**. Funding to lead this consortium was provided by grants awarded to M.C.M.: ERC Consolidator Grant SOCIOGENOME (615603), a Dutch Science Foundation (NWO) grant (VIDI grant 452-10-012), a UK ESRC/NCRM SOCGEN grant (ES/N011856/1), the European Union's FP7 Families And Societies project (320116), and the Wellcome Trust ISSF and John Fell Fund. M.d.H. was supported by grants from the Swedish Research Council (2015-03657) and the Swedish Heart-Lung Foundation (20140543). Research was carried out in collaboration with the Social Science Genetic Association Consortium (SSGAC), with funding from the US National Science Foundation (EAGER: 'Workshop for the Formation of a Social Science Genetic Association Consortium'), a Supplementary grant from the National Institutes of Health Office of Behavioral and Social Science Research, the Ragnar Söderberg Foundation (E9/11), the Swedish Research Council (421-2013-1061), the Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810 and T32-AG000186-23 to NBER and R01-AG042568-02 to the University of Southern California. X.S. was supported by a grant from the Swedish Research Council (537-2014-371). We thank X. Ding for research assistance, N. Pirastu, K. Coward and L. Layman for valuable comments, and the University of Oxford Advanced Research Computing (ARC) facility (http://dx.doi.org/10.5281/zenodo.22558). This research has been conducted using the UK Biobank Resource.

AUTHOR CONTRIBUTIONS

Senior investigators who led writing, analysis and study design: M.C.M., H. Snieder and M.d.H. Senior investigators who participated in study design: P.D.K., D.J.B. and D.C. Junior investigator who contributed to the study design and management: N. Barban. Population stratification: N. Barban and F.C.T. Genetic correlations and polygenic score prediction: N. Barban. Meta-analysis and quality control: N. Barban, R.d.V., J.J.M. and I.M.N. Biological annotation: R.J., M.d.H. and A.V. Sex-specific genetic effects: N. Barban and F.C.T. Bivariate and conditional analysis of the two fertility traits: X.S., J.F.W. and D.I.C. Gene-based analysis V.T. and S.W.v.d.L. Authors not listed contributed to recruitment, genotyping or data processing for the meta-analysis (**Supplementary Table 43**).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Elks, C.E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
2. Perry, J.R.B. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
3. Rahmioglu, N. *et al.* Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. *Hum. Reprod. Update* **20**, 702–716 (2014).
4. Day, F.R. *et al.* Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat. Commun.* **6**, 8464 (2015).
5. Mehta, D. *et al.* Evidence for genetic overlap between schizophrenia and age at first birth in women. *JAMA Psychiatry* **73**, 497–505 (2016).
6. Mills, M.C. & Tropf, F.C. The biodemography of fertility: a review and future research frontiers. *Kolner Z. Soz. Sozpsychol.* **67** (Suppl. 1), 397–424 (2015).
7. Mills, M., Rindfuss, R.R., McDonald, P. & te Velde, E. Why do people postpone parenthood? Reasons and social policy incentives. *Hum. Reprod. Update* **17**, 848–860 (2011).
8. Boivin, J., Bunting, L., Collins, J.A. & Nygren, K.G. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum. Reprod.* **22**, 1506–1512 (2007).
9. Mascarenhas, M.N., Flaxman, S.R., Boerma, T., Vanderpoel, S. & Stevens, G.A. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med.* **9**, e1001356 (2012).
10. Venkatesh, T., Suresh, P.S. & Tsutsumi, R. New insights into the genetic basis of infertility. *Appl. Clin. Genet.* **7**, 235–243 (2014).
11. Day, F.R. *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* **48**, 617–623 (2016).
12. Balbo, N., Billari, F.C. & Mills, M.C. Fertility in advanced societies: a review of research. *Eur. J. Popul.* **29**, 1–38 (2012).
13. Tropf, F.C. *et al.* Human fertility, molecular genetics, and natural selection in modern societies. *PLoS One* **10**, e0126821 (2015).
14. Fisher, R.A. *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
15. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
16. van der Most, P.J. *et al.* QCGWAS: a flexible R package for automated quality control of genome-wide association results. *Bioinformatics* **30**, 1185–1186 (2014).
17. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
18. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
19. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
20. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
21. Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
22. Liu, J.Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).

23. Mishra, A. & Macgregor, S. VEGAS2: software for more flexible gene-based testing. *Twin Res. Hum. Genet.* **18**, 86–91 (2015).
24. Vaez, A. *et al. In silico* post genome-wide association studies analysis of C-reactive protein loci suggests an important role for interferons. *Circ Cardiovasc Genet* **8**, 487–497 (2015).
25. ENCODE Project Consortium. ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
26. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
27. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
28. Zhernakova, D. *et al.* Hypothesis-free identification of modulators of genetic risk factors. Preprint at *bioRxiv* http://dx.doi.org/10.1101/033217 (2015).
29. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. Preprint at *bioRxiv* http://dx.doi.org/10.1101/033084 (2015).
30. Tranchevent, L.C. *et al.* ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res.* **36**, W377–W384 (2008).
31. Pers, T.H., Dworzyński, P., Thomas, C.E., Lage, K. & Brunak, S. MetaRanker 2.0: a web server for prioritization of genetic variation data. *Nucleic Acids Res.* **41**, W104–W108 (2013).
32. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
33. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
34. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
35. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
36. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
37. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
38. Rietveld, C.A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. USA* **111**, 13790–13794 (2014).
39. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
40. Day, F.R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and *BRCA1*-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
41. Perry, J.R. *et al.* A genome-wide association study of early menopause and the combined impact of identified variants. *Hum. Mol. Genet.* **22**, 1465–1472 (2013).
42. Fang, W.-L. *et al.* CREB coactivator CRTC2/TORC2 and its regulator calcineurin crucially mediate follicle-stimulating hormone and transforming growth factor β1 upregulation of steroidogenesis. *J. Cell. Physiol.* **227**, 2430–2440 (2012).
43. Cao, G. *et al.* Molecular cloning and characterization of a novel human cAMP response element–binding (CREB) gene (*CREB4*). *J. Hum. Genet.* **47**, 373–376 (2002).
44. El-Alfy, M. *et al.* Stage-specific expression of the Atce1/Tisp40α isoform of CREB3L4 in mouse spermatids. *J. Androl.* **27**, 686–694 (2006).
45. Adham, I.M. *et al.* Reduction of spermatogenesis but not fertility in *Creb3l4*-deficient mice. *Mol. Cell. Biol.* **25**, 7657–7664 (2005).
46. McAllister, J.M. *et al.* Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. *Proc. Natl. Acad. Sci. USA* **111**, E1519–E1527 (2014).
47. O'Bryan, M.K. *et al.* RBM5 is a male germ cell splicing factor and is required for spermatid differentiation and male fertility. *PLoS Genet.* **9**, e1003628 (2013).
48. Tsukamoto, S. *et al.* Functional analysis of lysosomes during mouse preimplantation embryo development. *J. Reprod. Dev.* **59**, 33–39 (2013).
49. Szucs, M., Osvath, P., Laczko, I. & Jakab, A. Adequacy of hyaluronan binding assay and a new fertility index derived from it for measuring of male fertility potential and the efficacy of supplement therapy. *Andrologia* **47**, 519–524 (2015).
50. Buensuceso, A.V. *et al.* Ephrin-A5 is required for optimal fertility and a complete ovulatory response to gonadotropins in the female mouse. *Endocrinology* **157**, 942–955 (2016).
51. Jisa, E. & Jungbauer, A. Kinetic analysis of estrogen receptor homo- and heterodimerization *in vitro*. *J. Steroid Biochem. Mol. Biol.* **84**, 141–148 (2003).
52. O'Donnell, L., Robertson, K.M., Jones, M.E. & Simpson, E.R. Estrogen and spermatogenesis. *Endocr. Rev.* **22**, 289–318 (2001).
53. Ly-Huynh, J.D. *et al.* Importin α2–interacting proteins with nuclear roles during mammalian spermatogenesis. *Biol. Reprod.* **85**, 1191–1202 (2011).
54. Varshney, G.K. *et al.* CRISPRz: a database of zebrafish validated sgRNAs. *Nucleic Acids Res.* **44**, D1, D822–D826 (2016).
55. Menken, J. Age and fertility: how late can you wait? *Demography* **22**, 469–483 (1985).
56. Manolio, T.A., Brooks, L.D. & Collins, F.S. A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118**, 1590–1605 (2008).
57. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
58. Okkelman, I.A., Sukaeva, A.Z., Kirukhina, E.V., Korneenko, T.V. & Pestov, N.B. Nuclear translocation of lysyl oxidase is promoted by interaction with transcription repressor p66β. *Cell Tissue Res.* **358**, 481–489 (2014).
59. Joshi, N.R. *et al.* Altered expression of microRNA-451 in eutopic endometrium of baboons (*Papio anubis*) with endometriosis. *Hum. Reprod.* **30**, 2881–2891 (2015).
60. Franklin, R.B. *et al.* Human ZIP1 is a major zinc uptake transporter for the accumulation of zinc in prostate cells. *J. Inorg. Biochem.* **96**, 435–442 (2003).
61. Lisle, R.S., Anthony, K., Randall, M.A. & Diaz, F.J. Oocyte–cumulus cell interactions regulate free intracellular zinc in mouse oocytes. *Reproduction* **145**, 381–390 (2013).
62. Shan, B. *et al.* Association of *DENND1A* gene polymorphisms with polycystic ovary syndrome: a meta-analysis. *J. Clin. Res. Pediatr. Endocrinol.* **8**, 135–143 (2016).
63. Impera, L. *et al.* A novel fusion *5′AFF3/3′BCL2* originated from a t(2;18)(q11.2;q21.33) translocation in follicular lymphoma. *Oncogene* **27**, 6187–6190 (2008).
64. Urano, A. *et al.* Infertility with defective spermiogenesis in mice lacking AF5q31, the target of chromosomal translocation in human infant leukemia. *Mol. Cell. Biol.* **25**, 6834–6845 (2005).
65. Reese, K.L. *et al.* Acidic hyaluronidase activity is present in mouse sperm and is reduced in the absence of SPAM1: evidence for a role for hyaluronidase 3 in mouse and human sperm. *Mol. Reprod. Dev.* **77**, 759–772 (2010).
66. Heath, E., Sablitzky, F. & Morgan, G.T. Subnuclear targeting of the RNA-binding motif protein RBM6 to splicing speckles and nascent transcripts. *Chromosome Res.* **18**, 851–872 (2010).
67. Kamura, T. *et al.* Cytoplasmic ubiquitin ligase KPC regulates proteolysis of $p27^{Kip1}$ at G1 phase. *Nat. Cell Biol.* **6**, 1229–1235 (2004).
68. Kato, J.Y., Matsuoka, M., Polyak, K., Massagué, J. & Sherr, C.J. Cyclic AMP–induced G1 phase arrest mediated by an inhibitor ($p27^{Kip1}$) of cyclin-dependent kinase 4 activation. *Cell* **79**, 487–496 (1994).
69. Bagley, D.C., Paradkar, P.N., Kaplan, J. & Ward, D.M. Mon1a protein acts in trafficking through the secretory apparatus. *J. Biol. Chem.* **287**, 25577–25588 (2012).
70. Sakamoto, O. *et al.* Role of macrophage-stimulating protein and its receptor, RON tyrosine kinase, in ciliary motility. *J. Clin. Invest.* **99**, 701–709 (1997).
71. Zhang, C. *et al.* Molecular mechanisms that drive estradiol-dependent burst firing of Kiss1 neurons in the rostral periventricular preoptic area. *Am. J. Physiol. Endocrinol. Metab.* **305**, E1384–E1397 (2013).
72. Ponglikitmongkol, M., Green, S. & Chambon, P. Genomic organization of the human oestrogen receptor gene. *EMBO J.* **7**, 3385–3388 (1988).
73. de Mattos, C.S. *et al. ESR1* and *ESR2* gene polymorphisms are associated with human reproduction outcomes in Brazilian women. *J. Ovarian Res.* **7**, 114 (2014).
74. Lamp, M. *et al.* Polymorphisms in *ESR1*, *ESR2* and *HSD17B1* genes are associated with fertility status in endometriosis. *Gynecol. Endocrinol.* **27**, 425–433 (2011).
75. Chiu, Y.-C. *et al.* Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Dev. Neurobiol.* **74**, 723–738 (2014).
76. Alves, M.G. *et al.* Metabolic fingerprints in testicular biopsies from type 1 diabetic patients. *Cell Tissue Res.* **362**, 431–440 (2015).
77. Mojiminiyi, O.A., Safar, F.H., Al Rumaih, H. & Diejomaoh, M. Variations in alanine aminotransferase levels within the normal range predict metabolic and androgenic phenotypes in women of reproductive age. *Scand. J. Clin. Lab. Invest.* **70**, 554–560 (2010).
78. Van Maldergem, L. *et al.* Revisiting the craniosynostosis-radial ray hypoplasia association: Baller–Gerold syndrome caused by mutations in the *RECQL4* gene. *J. Med. Genet.* **43**, 148–152 (2016).
79. Ruan, Y., Cheng, M., Ou, Y., Oko, R. & van der Hoorn, F.A. Ornithine decarboxylase antizyme Oaz3 modulates protein phosphatase activity. *J. Biol. Chem.* **286**, 29417–29427 (2011).

Nicola Barban[1,173], Rick Jansen[2], Ronald de Vlaming[3–5], Ahmad Vaez[6,7], Jornt J Mandemakers[8], Felix C Tropf[1], Xia Shen[9–11], James F Wilson[10,11], Daniel I Chasman[12], Ilja M Nolte[6], Vinicius Tragante[13], Sander W van der Laan[14], John R B Perry[15], Augustine Kong[16,17], BIOS Consortium[18], Tarunveer S Ahluwalia[19–21], Eva Albrecht[22], Laura Yerges-Armstrong[23], Gil Atzmon[24–26], Kirsi Auro[27,28], Kristin Ayers[29], Andrew Bakshi[30], Danny Ben-Avraham[25], Klaus Berger[31], Aviv Bergman[32], Lars Bertram[33,34], Lawrence F Bielak[35], Gyda Bjornsdottir[17], Marc Jan Bonder[36], Linda Broer[37], Minh Bui[38], Caterina Barbieri[39], Alana Cavadino[40,41], Jorge E Chavarro[42–44], Constance Turman[44], Maria Pina Concas[45], Heather J Cordell[29], Gail Davies[46,47], Peter Eibich[48], Nicholas Eriksson[49], Tõnu Esko[50,51],

Joel Eriksson[52], Fahimeh Falahi[6], Janine F Felix[4,53,54], Mark Alan Fontana[55], Lude Franke[36], Ilaria Gandin[56], Audrey J Gaskins[42], Christian Gieger[57,58], Erica P Gunderson[59], Xiuqing Guo[60], Caroline Hayward[10], Chunyan He[61], Edith Hofer[62,63], Hongyan Huang[44], Peter K Joshi[11], Stavroula Kanoni[64], Robert Karlsson[9], Stefan Kiechl[65], Annette Kifley[66], Alexander Kluttig[67], Peter Kraft[44,68], Vasiliki Lagou[69–71], Cecile Lecoeur[72], Jari Lahti[73–75], Ruifang Li-Gao[76], Penelope A Lind[77], Tian Liu[78], Enes Makalic[38], Crysovalanto Mamasoula[29], Lindsay Matteson[79], Hamdi Mbarek[80,81], Patrick F McArdle[23], George McMahon[82], S Fleur W Meddens[5,83], Evelin Mihailov[50], Mike Miller[84], Stacey A Missmer[44,85], Claire Monnereau[4,53,54], Peter J van der Most[6], Ronny Myhre[86], Mike A Nalls[87], Teresa Nutile[88], Ioanna Panagiota Kalafati[89], Eleonora Porcu[90,91], Inga Prokopenko[92–94], Kumar B Rajan[95], Janet Rich-Edwards[43,44,96], Cornelius A Rietveld[3–5], Antonietta Robino[97], Lynda M Rose[12], Rico Rueedi[98,99], Kathleen A Ryan[23], Yasaman Saba[100], Daniel Schmidt[38], Jennifer A Smith[35], Lisette Stolk[37], Elizabeth Streeten[23], Anke Tönjes[101], Gudmar Thorleifsson[17], Sheila Ulivi[97], Juho Wedenoja[102], Juergen Wellmann[31], Peter Willeit[65,103,104], Jie Yao[60], Loic Yengo[72,105], Jing Hua Zhao[15], Wei Zhao[35], Daria V Zhernakova[36], Najaf Amin[4], Howard Andrews[106], Beverley Balkau[72], Nir Barzilai[24], Sven Bergmann[98,99], Ginevra Biino[107], Hans Bisgaard[21], Klaus Bønnelykke[21], Dorret I Boomsma[80,81], Julie E Buring[12], Harry Campbell[11], Stefania Cappellani[97], Marina Ciullo[88,108], Simon R Cox[46,47], Francesco Cucca[90,91], Daniela Toniolo[39], George Davey-Smith[109], Ian J Deary[46,47], George Dedoussis[89], Panos Deloukas[64,110], Cornelia M van Duijn[4], Eco J C de Geus[80,81], Johan G Eriksson[75,111–114], Denis A Evans[95], Jessica D Faul[115], Cinzia Felicita Sala[39], Philippe Froguel[72], Paolo Gasparini[56,116], Giorgia Girotto[56,116], Hans-Jörgen Grabe[117], Karin Halina Greiser[118], Patrick J F Groenen[5,119], Hugoline G de Haan[76], Johannes Haerting[67], Tamara B Harris[120], Andrew C Heath[121], Kauko Heikkilä[28], Albert Hofman[4,5,44], Georg Homuth[122], Elizabeth G Holliday[123,124], John Hopper[38], Elina Hyppönen[40,125,126], Bo Jacobsson[86,127], Vincent W V Jaddoe[4,53,54], Magnus Johannesson[128], Astanand Jugessur[86], Mika Kähönen[129], Eero Kajantie[130–132], Sharon L R Kardia[35], Bernard Keavney[29,133], Ivana Kolcic[134], Päivikki Koponen[135], Peter Kovacs[136], Florian Kronenberg[137], Zoltan Kutalik[99,138], Martina La Bianca[97], Genevieve Lachance[139], William G Iacono[84], Sandra Lai[90], Terho Lehtimäki[140], David C Liewald[46], LifeLines Cohort Study[18], Cecilia M Lindgren[93,94,141,142], Yongmei Liu[143], Robert Luben[144], Michael Lucht[111], Riitta Luoto[145], Per Magnus[86], Patrik K E Magnusson[9], Nicholas G Martin[77], Matt McGue[84,146], Ruth McQuillan[11], Sarah E Medland[77], Christa Meisinger[58,147], Dan Mellström[52], Andres Metspalu[50,148], Michela Traglia[39], Lili Milani[50], Paul Mitchell[66], Grant W Montgomery[121,149], Dennis Mook-Kanamori[76,150,151], Renée de Mutsert[76], Ellen A Nohr[152], Claes Ohlsson[52], Jørn Olsen[153], Ken K Ong[15], Lavinia Paternoster[109], Alison Pattie[47], Brenda W J H Penninx[2,81], Markus Perola[27,28,50], Patricia A Peyser[35], Mario Pirastu[45], Ozren Polasek[134], Chris Power[40], Jaakko Kaprio[27,28,102], Leslie J Raffel[154], Katri Räikkönen[73], Olli Raitakari[155], Paul M Ridker[12], Susan M Ring[109], Kathryn Roll[60], Igor Rudan[11], Daniela Ruggiero[88], Dan Rujescu[156], Veikko Salomaa[27], David Schlessinger[157], Helena Schmidt[100], Reinhold Schmidt[62], Nicole Schupf[158], Johannes Smit[2,81], Rossella Sorice[88,108], Tim D Spector[139], John M Starr[46,159], Doris Stöckl[158], Konstantin Strauch[22,160], Michael Stumvoll[101,136], Morris A Swertz[36], Unnur Thorsteinsdottir[17,161], A Roy Thurik[3,5,162], Nicholas J Timpson[109], Joyce Y Tung[49], André G Uitterlinden[3,5,37], Simona Vaccargiu[45], Jorma Viikari[163], Veronique Vitart[10], Henry Völzke[164], Peter Vollenweider[165], Dragana Vuckovic[56,116], Johannes Waage[21], Gert G Wagner[166], Jie Jin Wang[66], Nicholas J Wareham[15], David R Weir[115], Gonneke Willemsen[80,81], Johann Willeit[65], Alan F Wright[10], Krina T Zondervan[167,168], Kari Stefansson[17,161], Robert F Krueger[84], James J Lee[84], Daniel J Benjamin[55,169], David Cesarini[170,171], Philipp D Koellinger[3,5,83], Marcel den Hoed[172,173], Harold Snieder[6,173] & Melinda C Mills[1,173]

[1]Department of Sociology and Nuffield College, University of Oxford, Oxford, UK. [2]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. [3]Department of Applied Economics, Erasmus School of Economics, Rotterdam, the Netherlands. [4]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [5]Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, the Netherlands. [6]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [7]Research Institute for Primordial Prevention of Non-Communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran. [8]Sociology of Consumption and Households, Wageningen University Research, Wageningen, the Netherlands. [9]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [10]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [11]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [12]Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [13]Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. [14]Laboratory of Experimental Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands. [15]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [16]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [17]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [18]A list of members and affiliations appears in the **Supplementary Note**. [19]Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [20]Steno Diabetes Center, Gentofte, Denmark. [21]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

[22]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [23]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [24]Department of Medicine, Institute for Aging Research and the Diabetes Research Center, Albert Einstein College of Medicine, Bronx, New York, USA. [25]Department of Genetics, Institute for Aging Research and the Diabetes Research Center, Albert Einstein College of Medicine, Bronx, New York, USA. [26]Department of Natural Science, University of Haifa, Haifa, Israel. [27]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [28]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland. [29]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK. [30]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [31]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [32]Departments of Systems and Computational Biology, of Pathology and of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, USA. [33]Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. [34]School of Public Health, Faculty of Medicine, Imperial College, London, UK. [35]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA. [36]Department of Genetics, Genomics Coordination Center, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [37]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [38]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. [39]Division of Genetics and Cell Biology, San Raffaele Research Institute, Milan, Italy. [40]Population, Policy and Practice, UCL Institute of Child Health, London, UK. [41]Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventative Medicine, Queen Mary University of London, London, UK. [42]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [43]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [44]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [45]Institute of Genetic and Biomedical Research, National Research Council, UOS of Sassari, Sassari, Italy. [46]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [47]Department of Psychology, University of Edinburgh, Edinburgh, UK. [48]Health Economics Research Centre, University of Oxford, Oxford, UK. [49]23andMe, Inc., Mountain View, California, USA. [50]Estonian Genome Center, University of Tartu, Tartu, Estonia. [51]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [52]Centre for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [53]Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [54]Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [55]Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA. [56]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy. [57]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [58]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [59]Cardiovascular and Metabolic Conditions Section, Division of Research, Kaiser Permanente Northern California, Oakland, California, USA. [60]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA. [61]Department of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA. [62]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. [63]Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. [64]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [65]Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria. [66]Centre for Vision Research, Department of Ophthalmology and Westmead Institute for Medical Research, University of Sydney, Westmead, New South Wales, Australia. [67]Institute of Medical Epidemiology, Biostatistics and Informatics, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [68]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [69]Department of Neurosciences, KU Leuven, Leuven, Belgium. [70]Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium. [71]Translational Immunology Laboratory, VIB, Leuven, Belgium. [72]University of Lille, CNRS, Institut Pasteur of Lille, Lille, France. [73]Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland. [74]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland. [75]Folkhälsan Research Centre, Helsinki, Finland. [76]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [77]Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Herston Brisbane, Queensland, Australia. [78]Center for Lifespan Psychology, Max Planck Institute for Human Development and Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [79]Minnesota Center for Twin and Family Research, Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA. [80]Department of Biological Psychology, VU University, Amsterdam, the Netherlands. [81]EMGO+ Institute for Health and Care Research, Amsterdam, the Netherlands. [82]School of Social and Community Medicine, University of Bristol, Bristol, UK. [83]Complex Trait Genetics, VU University, Amsterdam, the Netherlands. [84]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA. [85]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [86]Department of Genetics and Bioinformatics, Area of Health Data and Digitalization, Institute of Public Health, Oslo, Norway. [87]Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [88]Institute of Genetics and Biophysics 'A. Buzzati-Traverso', CNR, Naples, Italy. [89]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [90]Istituto di Ricerca Genetica e Biomedica, CMR, Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy. [91]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy. [92]Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK. [93]Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [94]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [95]Rush University Medical Center, Chicago, Illinois, USA. [96]Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [97]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [98]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [99]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [100]Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria. [101]Department of Medicine, University of Leipzig, Leipzig, Germany. [102]Department of Public Health, University of Helsinki, Helsinki, Finland. [103]King's British Heart Foundation Centre, King's College London, London, UK. [104]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [105]Centre for Neurogenetics and Statistical Genomics, University of Queensland, Brisbane, Queensland, Australia. [106]Data Coordinating Center, New York State Psychiatric Institute, New York, New York, USA. [107]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. [108]IRCCS Neuromed, Pozzilli, Isernia, Italy. [109]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [110]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [111]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [112]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [113]Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. [114]Vasa Central Hospital, Vasa, Finland. [115]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA. [116]Division of Experimental Genetics, Sidra, Doha, Qatar. [117]Department of Psychiatry, University Medicine Greifswald, Greifswald, Germany. [118]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. [119]Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [120]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Bethesda, Maryland, USA. [121]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [122]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [123]School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia. [124]Hunter Medical Research Institute, Newcastle, New South Wales, Australia. [125]Centre for Population Health Research, Sansom Institute of Health Research and School of Health Sciences, University of South Australia, Adelaide, South Australia, Australia. [126]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [127]Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden. [128]Department of Economics, Stockholm School of Economics, Stockholm, Sweden. [129]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. [130]Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. [131]Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland. [132]Department of Obstetrics and Gynecology, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. [133]Institute of Cardiovascular Sciences, University of Manchester, Manchester, UK. [134]Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia. [135]Health Monitoring Unit, National Institute for Health and Welfare, Helsinki, Finland. [136]IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany. [137]Division of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. [138]Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [139]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [140]Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland. [141]NIHR Oxford Biomedical Research Centre, Oxford, UK. [142]Li Ka Shing Centre for Health Information and Discovery, Big Data Institute, University of Oxford, Oxford, UK. [143]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [144]Strangeways Research Laboratory, University of Cambridge, Cambridge, UK. [145]UKK Institute for

Health Promotion,Tampere, Finland. [146]Danish Aging Research Center and Danish Twin Registry, Institute of Public Health, University of Southern Denmark, Odense, Denmark. [147]MONICA/KORA Myocardial Infarction Registry, Central Hospital of Augsburg, Augsburg, Germany. [148]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [149]Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [150]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [151]Department of BESC, Epidemiology Section, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. [152]Research Unit for Gynecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense, Denmark. [153]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark. [154]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [155]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [156]Department of Psychiatry, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [157]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. [158]Departments of Epidemiology and Psychiatry, Columbia University Medical Center, New York, New York, USA. [159]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [160]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians Universität, Munich, Germany. [161]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [162]Montpellier Business School, Montpellier, France. [163]Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland. [164]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [165]Department of Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [166]German Socio-Economic Panel Study (SOEP), Max Planck Institute for Human Development and Berlin University of Technology (TUB), Berlin, Germany. [167]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [168]Endometriosis CaRe Centre, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK. [169]National Bureau of Economic Research, Cambridge, Massachusetts, USA. [170]Department of Economics, New York University, New York, New York, USA. [171]Research Institute for Industrial Economics, Stockholm, Sweden. [172]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [173]These authors contributed equally to this work. Correspondence should be addressed to M.C.M. (melinda.mills@nuffield.ox.ac.uk), N.B. (nicola.barban@sociology.ox.ac.uk), H.S. (h.snieder@umcg.nl) or M.d.H. (marcel.den_hoed@medsci.uu.se).

ONLINE METHODS

GWAS of reproductive behavior study design in brief. Genome-wide association analyses of AFB and NEB were performed at the cohort level according to a prespecified analysis plan (**Supplementary Note**). Cohort-uploaded results were imputed using the HapMap 2 CEU (r22.b36) or 1000 Genomes Project reference sample. Cohorts were asked to only include participants of European ancestry, with no missing values for all relevant covariates (sex, birth year and cohort-specific covariates), who were successfully genotyped over the whole genome and passed cohort-specific quality control filters. We followed the quality control protocol of the GIANT Consortium's recent study of human height[20] and employed QCGWAS[16] and EasyQC[17] software, which allowed us to harmonize the files and identify possible sources of error in association results.

Cohort association results (after applying the quality control filters) were combined using sample-size-weighted meta-analysis with genomic control correction within each study, implemented in METAL[80]. SNPs were considered genome-wide significant at P values smaller than 5×10^{-8} ($\alpha = 5\%$, Bonferroni corrected for 1 million tests). The meta-analyses were carried out by two independent analysts. Detailed results for each genome-wide significant locus are shown in in **Supplementary Figures 4–29**.

The total sample size of the meta-analyses is $n = 251{,}151$ for AFB pooled and $n = 343{,}072$ for NEB pooled. The PLINK clumping function[81] was used to identify the most significant SNPs in associated regions (termed lead SNPs). Detailed cohort descriptions, information about cohort-level genotyping and imputation procedures, cohort-level measures and quality control filters are shown in **Supplementary Tables 26** and **27** and discussed in the **Supplementary Note**.

Dominant genetic variation in fertility. We applied a method recently developed by Zhu *et al.*[82] to estimate dominant genetic effects on the basis of the genetic relatedness of unrelated individuals. Our results, based on combined TwinsUK and LifeLines samples, showed no evidence of dominant genetic effects for either NEB (1.0×10^{-7}, SE = 0.07; $P = 0.45$) or AFB (0.02, SE = 0.08; $P = 0.43$). Results are shown in **Supplementary Table 28** and discussed in the **Supplementary Note**.

Bivariate and conditional analyses. As joint analysis of correlated traits may boost power for mapping functional loci, we applied a recently developed multiple-trait analysis method[83] to test the association between each variant and the two correlated traits AFB and NEB simultaneously using multivariate analysis of variance (MANOVA) (**Supplementary Table 29** and **Supplementary Note**). The analysis was performed on the basis of the genome-wide meta-analysis summary statistics for each single trait. Although this analysis did not identify additional genome-wide significant loci ($\lambda = 0.995$), it did account for the correlation between the two phenotypes, thus improving the strength of two signals on chromosomes 1 and 5, indicating possible pleiotropic architecture for AFB and NEB (**Supplementary Fig. 30**). The analysis also provided a conditional association test of the genetic effect of each variant on AFB including NEB as a covariate and the genetic effect on NEB including AFB as a covariate (**Supplementary Fig. 31**).

Population stratification. We used two methods to assess whether our GWAS results exhibited signs of population stratification (**Supplementary Note**). First, we used the LD Score intercept method described in Bulik-Sullivan *et al.*[19] to test whether inflation in χ^2 statistics was due to confounding biases such as cryptic relatedness and population stratification. In all six cases, the intercept estimates were not significantly different from 1, suggesting no appreciable inflation of the test statistics attributable to population stratification. Second, we conducted a series of individual and within-family regressions using polygenic scores as predictors[20,21,38] on a data set of dizygotic twins (STR and TwinsUK). The regression analyses showed that within-family regression coefficients for both AFB and NEB were statistically different from 0 when the P-value threshold was sufficiently high (**Supplementary Figs. 32** and **33**, and **Supplementary Tables 30** and **31**).

Sex-specific effects. In addition to the pooled GWAS for which results are presented in the main text, we also ran sex-specific GWAS meta-analyses for AFB and NEB (**Supplementary Note**). The sample sizes for sex-specific analysis were as follows: AFB in women, $n = 189{,}656$; AFB in men, $n = 48{,}408$; NEB in women, $n = 225{,}230$; NEB in men, $n = 103{,}909$. Our results identified six genome-wide significant ($P < 5 \times 10^{-8}$) independent SNPs for AFB in women and one genome-wide significant independent SNP for NEB in men (**Supplementary Figs. 34** and **35**, and **Supplementary Table 32**). We also used LD Score bivariate regression and GREML bivariate analysis to estimate the genetic correlation between men and women on the basis of the sex-specific summary statistics from the AFB and NEB meta-analyses. Our estimates based on LD bivariate regression indicated genetic correlations between the sexes of $r_g = 0.86$ (SE = 0.052) for AFB and $r_g = 0.97$ (SE = 0.095) for NEB. Results are shown in **Supplementary Tables 33** and **34** and discussed in the **Supplementary Note**.

Polygenic score prediction. We performed out-of-sample prediction and calculated polygenic scores for AFB and NEB, on the basis of genome-wide association meta-analysis results, and used regression models to predict the same phenotypes in four independent cohorts: HRS, LifeLines, STR and TwinsUK (**Supplementary Fig. 2** and **Supplementary Note**). We ran ordinary least-squares (OLS) regression models and report R^2 as a measure of goodness of fit for the model. In addition, we tested how well our polygenic scores for NEB could predict childlessness at the end of the reproductive period (using age 45 for women and 55 for men; **Supplementary Table 21**). Because AFB is observed only in parous women, we adopted an additional statistical model to account for censoring (Cox proportional hazard model; **Supplementary Table 22**) and selection (Heckman selection model; **Supplementary Table 35**). We additionally tested the predictive value of our polygenic scores for AFB on age at menarche (TwinsUK) and age at menopause (LifeLines) (**Supplementary Table 23**). Finally, we examined whether variants associated with menopause are associated with AFB. We calculated a polygenic score for age at menopause based on recent GWAS results from Day *et al.*[40] and applied the predictor to the LifeLines and TwinsUK cohorts (**Supplementary Table 36**).

Genetic correlations. We used information from 27 publicly available GWAS data sets to estimate the number of genetic correlations between AFB or NEB and related traits (**Fig. 3** and **Supplementary Table 25**) via LD Score bivariate regression. Details on these phenotypes are provided in the **Supplementary Note**. A conservative Bonferroni-corrected P-value threshold of $P < 1.85 \times 10^{-3}$ (= 0.05/27) was used to define significant associations. We also tested the correlation between NEB and AFB using bivariate GREML analysis on the Women's General Health Study (WGHS; $n = 40{,}621$).

Lookups and proxy phenotypes. Following up on the results of genetic overlap with other phenotypes, we tested in a quasi-phenotype replication setting whether the SNPs strongly associated with AFB in women were empirically plausible candidate SNPs for age at menarche and age at menopause (**Supplementary Note**). We used a two-stage approach applied in other contexts[38,84]. In the first stage, we conducted a meta-analysis of AFB excluding cohorts that were part of the meta-analysis for the phenotype we intended to replicate. We merged the SNPs from this meta-analysis with the publically available association results for the most recent GWAS on age at menarche[2] and age at menopause[40] from the ReproGen consortium website[1]. SNPs that were not present in both studies considered were dropped from the analysis. We aligned alleles and directions of effect using EasyStrata software[85]. We then selected independent SNPs with $P < 1 \times 10^{-5}$, using the clump procedure in PLINK[81] (window size of 1,000 kb and LD threshold of $r^2 > 0.1$) to identify the most significant SNPs in the associated regions included in both files. We defined 'prioritized SNP associations' as those that passed the Bonferroni correction for the number of SNPs tested ($0.05/122 = 4.10 \times 10^{-4}$, for both age at menarche and age at menopause). Our results identified three SNPs after Bonferroni correction that could be used as good candidates for age at menarche. We did not find any clear 'candidate SNP' for age at menopause (**Supplementary Fig. 36**).

Gene-based GWAS analysis. We performed gene-based testing with the full GWAS set (~2.5 million HapMap-imputed SNPs) for both phenotypes using VEGAS (**Supplementary Tables 3** and **4**, and **Supplementary Note**)[22,23].

This software has the advantage of accounting for LD structure and allowing a gene to be defined as a range with boundaries beyond the edges of the gene to include intergenic regions in the analysis. We defined genes including an additional 50-kb window around each gene. We considered every SNP for the gene-based analysis, ran the analyses for each chromosome with up to 10^6 permutations and considered $P < 2.5 \times 10^{-6}$ as the threshold for significance (0.05/~20.000 genes).

eQTL and meQTL analyses. For each of the 12 SNPs identified in the GWAS, local (*cis*; exons or methylation sites <1 Mb from the SNP) and genome-wide (*trans*; exons or methylation sites >5 Mb from the SNP) effects were identified by computing Spearman rank correlations between SNPs and local or global exons and methylation sites (**Supplementary Note**). Bonferroni correction for multiple testing was performed for the 12 SNPs tested ($P < 2.5 \times 10^{-6}$ for *cis*-meQTL analysis, $P < 1 \times 10^{-8}$ for *trans*-meQTL analysis, $P < 1.2 \times 10^{-6}$ for *cis*-eQTL analysis, $P < 1.3 \times 10^{-8}$ for *trans*-eQTL analysis). For each of the significant associations, the corresponding exons or methylation sites were selected, the strongest eQTLs were identified for these elements and the LD between the strongest eQTLs and the corresponding SNP identified in the GWAS was computed. LD was computed using BIOS genotypes (genotypes used for eQTL and meQTL mapping).

Functional variant analysis using RegulomeDB. We used RegulomeDB[27] to identify variants among the 322 SNPs that reached $P < 5 \times 10^{-8}$ for association with AFB and/or NEB in the meta-analysis of GWAS results that likely influence regulation of gene expression (**Supplementary Note**). RegulomeDB integrates results from the Roadmap Epigenomics[26] and ENCODE[86] projects. SNPs showing the most evidence of being functional—defined by having a RegulomeDB score <4—were subsequently examined in more detail in terms of effects on gene expression (eQTLs) and their protein-binding capacity (**Supplementary Table 6**).

Gene prioritization. Potentially causal genes for the associations identified by GWAS were identified using four previously described bioinformatics tools: ToppGene[4], Endeavor[5], MetaRanker[6] and DEPICT[7]. To this end, we first retrieved positional coordinates for all lead SNPs according to GRCh37/hg19 using Ensembl BioMart. These coordinates were used to extract all genes located within 40 kb of lead SNPs from the UCSC table browser. The identified genes then served as input for ToppGene and Endeavor. Genes with established roles in fertility served as training genes in this procedure, that is, *BRCA1*, *EGFR*, *ERBB2*, *ERBB3*, *ERBB4*, *HSD17B1*, *RBM5*, *ESR1*, *ESR2* and *FSHB*. For MetaRanker, we provided SNPs that reached $P < 5 \times 10^{-4}$ and their chromosomal positions as input, together with the above set of training genes. Because ToppGene, Endeavor and MetaRanker are biased toward larger and well-described genes, we also performed a gene prioritization procedure using DEPICT[7]. All SNPs that reached $P < 5 \times 10^{-4}$ in the meta-analysis served as input, and information on prioritized genes, gene set enrichment, and tissue and cell type enrichment was extracted. Genes were subsequently prioritized if they (i) reached $P < 0.05$ in DEPICT or (ii) reached $P < 0.05$ in ToppGene, Endeavor and MetaRanker (**Supplementary Table 37**).

Functional network and enrichment analyses. DEPICT was used to identify gene set, cell type and tissue enrichment, using the GWAS-identified SNPs with $P < 5 \times 10^{-4}$ as input (**Supplementary Note**). Because of the relatively small number of identified loci, DEPICT was only able to perform these analyses for AFB and NEB pooled and for AFB in women. To construct a functional association network, we combined five prioritized candidate gene sets into a single query gene set that was then used as input for functional network analysis[24]. We applied the GeneMANIA algorithm together with its large set of accompanying functional association data[87]. We used the Cytoscape software platform[88], extended by the GeneMANIA plugin (data version 8/12/2014, accessed 24 April 2016)[89]. All the genes in the composite network, from either the query or resulting gene sets, were then used for functional enrichment analysis against Gene Ontology (GO) terms[90] to identify the most relevant terms, using the same plugin[89].

Gene–environment interactions. Previous research based on twin studies shows differential heritability of fertility behavior across birth cohorts[91,92]. We used the Swedish Twin Register (STR) to examine whether the effect of a polygenic score for AFB or NEB varied across birth cohort. We followed the analysis presented in the recent GWAS of education[35] and divided the sample into six groups on the basis of year of birth. Each group spanned five birth years, with the oldest ranging from 1929–1933 and the youngest born from 1954–1958. **Supplementary Table 38** reports the estimated coefficients from these regressions. The results indicate a U-shaped trend in AFB and a linear decline in NEB, but they do not provide any clear evidence of interaction effects between the polygenic scores and birth cohort. We additionally tested the interaction effects for educational level and the polygenic scores for AFB and NEB in three different samples (LifeLines, STR and HRS). **Supplementary Table 39** reports the estimated coefficients from these regressions. The results indicate that years of education are positively associated with AFB in both the LifeLines and STR cohorts and negatively associated with NEB in the HRS cohort. With the exception of NEB in the HRS cohort, we found no evidence of gene–environment effects with education.

Robustness checks. To estimate the robustness of our results for AFB, we conducted two additional analyses. First, we estimated how the coefficients changed if we controlled for educational attainment. Using data from deCODE, we ran an additional association analysis using the ten loci that were genome-wide significant in the meta-analysis ($P < 5 \times 10^{-8}$). The analysis was restricted to individuals born between 1910 and 1975 who also had data available on completed education. The total sample size was 42,187 (17,996 males and 24,191 females). The analysis was adjusted for sex, year of birth (linear, squared and cubed), interaction between sex and year of birth, and the first ten principal components. Education is measured by years of education, ranging between 10 and 20 years. **Supplementary Table 40** reports the association results before and after adjusting for educational attainment. Our analysis shows that effect sizes shrink after including educational attainment as a covariate, with an average reduction of around 15%. We also estimated the effect of a polygenic risk score for AFB calculated from meta-analysis data excluding the deCODE cohort. The polygenic risk score remained highly significant. The effect of 1 s.d. for the AFB score decreased from 0.19 years (69 d) without controlling for education to 0.16 years (59 d) when we controlled for years of education. Second, we estimated how the coefficients changed after controlling for educational attainment and age at first sexual intercourse using the UK Biobank cohort ($n = 50{,}954$). We ran two association models: the first followed the GWAS analysis plan with no additional covariates, and the second added years of education and age at first sexual intercourse as covariates. The results are presented in **Supplementary Figure 37** and **Supplementary Table 41**. Our analysis shows that the effect sizes of our top hits are highly concordant ($R^2 = 0.94$). The inclusion of educational attainment and age at first sexual intercourse as covariates weakened the effect sizes on average by 40% and increased the *P* values of the estimated coefficients. Overall, we interpret this additional analysis as a robustness test that confirms that the top hits from our meta-analysis are robust to the inclusion of the confounding factors of educational attainment and age at first sexual intercourse.

Positive selection. We performed Haploplotter analysis[93] to examine whether lead SNPs and/or functional variants identified using RegulomeDB showed evidence of positive selection. Three variants showed standardized integrated haplotype scores <–2 or >2, indicating that these variants represent the top 5% of signals in the population. These SNPs are (i) rs7628058 on chromosome 3 for AFB, an eQTL for *RBM6* in monocytes; (ii) rs2247510 on chromosome 3 for AFB, an eQTL for *RBM6* and *HYAL3* in monocytes and a binding site for a range of transcription factors; and (iii) rs2415984, the lead SNP in the chromosome 14 locus for NEB. Results are presented in **Supplementary Table 42**.

Data availability. Results can be downloaded from the SOCIOGENOME and SSGAC website. Data come from multiple studies, most of which are subject to a MTA, and are listed in the **Supplementary Note**. Correspondence and requests for materials should be addressed to the corresponding authors or info@sociogenome.com.

doi:10.1038/ng.3698

80. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
81. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
82. Zhu, Z. *et al*. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).
83. Shen, X. *et al*. Simple multi-trait analysis identifies novel loci associated with growth and obesity measures. Preprint at *bioRxiv* http://dx.doi.org/10.1101/022269 (2015).
84. Okbay, A. *et al*. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
85. Winkler, T.W. *et al*. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
86. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
87. Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* **9** (Suppl. 1), S4 (2008).
88. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nat. Methods* **9**, 1069–1076 (2012).
89. Montojo, J. *et al*. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927–2928 (2010).
90. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
91. Kohler, H.-P., Rodgers, J.L. & Christensen, K. Is fertility behavior in our genes? Findings from a Danish twin study. *Popul. Dev. Rev.* **25**, 253–288 (1999).
92. Tropf, F.C., Barban, N., Mills, M.C., Snieder, H. & Mandemakers, J.J. Genetic influence on age at first birth of female twins born in the UK, 1919–68. *Popul. Stud. (Camb.)* **69**, 129–145 (2015).
93. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).