Reproductive Toxicology 45 (2014) 77–86

Reproductive Toxicology

journal homepage: www.elsevier.com/locate/reprotox

More than just hormones: H295R cells as predictors of reproductive toxicity

Jodi M. Maglich[a], Max Kuhn[b], Robert E. Chapin[c], Mathew T. Pletcher[d,*]

[a] Compound Safety Prediction, Pfizer Global Research and Development, Pfizer Inc., Cambridge, MA 02420, United States
[b] Non-Clinical Statistics, Pfizer Global Research Development, Pfizer Inc., Groton, CT 06340, United States
[c] Developmental and Reproductive Toxicology Center of Expertise, Pfizer Global Research and Development, Pfizer Inc., Groton, CT 06340, United States
[d] Rare Diseases Research Unit, Pfizer Global Research Development, Pfizer Inc., Cambridge, MA 02420, United States

ARTICLE INFO

Article history:
Received 23 August 2013
Received in revised form
20 November 2013
Accepted 18 December 2013
Available online 13 January 2014

Keywords:
Steroidogenesis
Reproductive toxicology
H295R
Predictive screening

ABSTRACT

Many of the commonly observed reproductive toxicities associated with therapeutic compounds can be traced to a disruption of the steroidogenic pathway. We sought to develop an *in vitro* assay that would predict reproductive toxicity and be high throughput in nature. H295R cells, previously validated as having an intact and functional steroidogenic pathway, were treated with 83 known-positive and 79 known-negative proprietary and public-domain compounds. The assay measured the expression of the key enzymes *STAR*, 3*βHSD2*, *CYP17A1*, *CYP11B2*, *CYP19A1*, *CYP21A2*, and *CYP11A1* and the hormones DHEA, progesterone, testosterone, and cortisol. We found that a Random Forest model yielded a receiver operating characteristic area under the curve (ROC AUC) of 0.845, with sensitivity of 0.724 and specificity of 0.758 for predicting *in vivo* reproductive toxicity with this *in vitro* assay system.

1. Introduction

Toxicology is in the middle of a profound change, from a descriptive science to a predictive science, mediated by the visionary Toxicity Testing in the 21st Century publication [1]. This approach has certainly found fertile ground in pharmaceuticals [2]. Some form of safety evaluation often occurs very early in the development a of compound series. Selection of key compounds for further development uses data from multiple assays, each for an individual pathway of toxicity or key biological process [3,4]. The body depends on steroid hormones to regulate or influence the immune system, response to stress, gluconeogenesis, components of behavior, and the many complex parts of male and female reproduction. Based on the importance of this pathway in many physiological processes, it would be benificial to be able to choose between two candidate drug molecules, one of which showed an unwanted impact on steroidogenesis, and the other of which did not. To accommodate the needs of a screening program early in the candidate selection process, such an assay should require minimal amount of test compound (*de novo* synthesized drug candidate molecules), be at least relatively high-throughput, and not require extended exposures.

In steroidogenesis, cholesterol is first shuttled to the inner mitochondrial membrane in a rate-limiting step by steroidogenic acute regulatory protein (*StAR*) [5] (Fig. 1). Cholesterol is then converted to pregnenolone by side chain cleavage enzyme (*CYP11A1*). Progesterone is produced by 3 beta-hydroxysteroid dehydrogenase (*3βHSD2*) action on pregnenolone. *CYP17A1* catalyzes the 17-hydroxylation of pregnenolone and progesterone to 17α-hydroxy intermediates and the 17,20 lyase reactions leading to DHEA and along with 17beta-hydroxysteroid dehydrogenase (*17βHSD*) activity, to testosterone. Cortisol is synthesized from the 17α-hydroxy intermediates by the enzymes 3βHSD, 21-hydrolase (*CYP21A2*), and 11-beta hydroxylase (*CYP11B1*). Estradiol is converted from testosterone by the enzyme aromatase (*CYP19A1*). Estradiol can alternatively be converted by *17βHSD* from estrone, a hormone produced by aromatase activity on androstenedione.

H295R cells are a transformed human adrenal cell line which secretes all the steroid intermediates of the steroidogenesis pathway, and has been found useful for studying steroidogenesis [6–9]. These cells are zonally undifferentiated, *i.e.*, they produce the steroids of each of the three zones normally segregated in the adult adrenal cortex [10–12]. Because H295R cells uniquely express all of the enzymes in the steroidogenesis pathway, they allow the

* Corresponding author. Tel.: +1 617 665 5463.
E-mail address: Mathew.Pletcher@pfizer.com (M.T. Pletcher).

0890-6238
http://dx.doi.org/10.1016/j.reprotox.2013.12.009

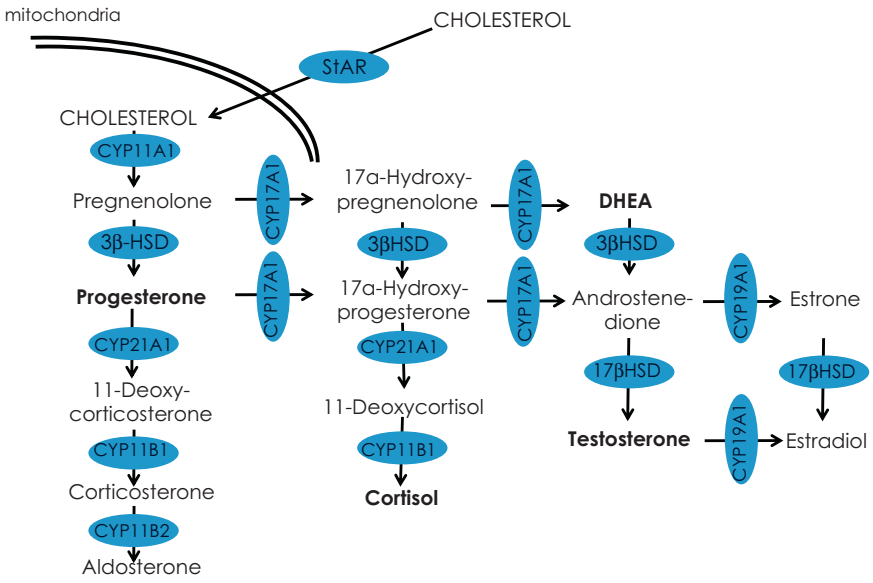

Fig. 1. The steroidogenesis pathway. Enzymes and transport proteins are represented in ovals. Measures hormone endpoints are presented in bold.

simultaneous testing of all components of steroidogenesis to the test exposure *in vitro*. These cells have been widely used in other testing efforts: the US EPA endocrine disruptor screening program (EDSP), ToxCast, European REACH, as well as the global Organization for Economic Cooperation and Development (OECD) [13–15] have all employed hormone measurement in the H295R cell model for identification of endocrine disrupting chemicals.

The objective of our work reported here was to develop a multi-parameter assay to detect chemical disruption of the steroidogenesis pathway. The assay would measure quantitative differences in the expression of enzymes in the steroidogenic pathway and levels of hormones secreted into the media following treatment with known *in vivo* toxicants and non-toxic compounds. This approach then allowed us to explore whether there was any relationship between changes in these H295R endpoints and male or female reproductive system pathology in rats in preclinical safety studies. Thus, we posed the question "how well does one or a combination of steroidogenic endpoints *in vitro* predict any male or female reproductive toxicity *in vivo*?"

2. Materials and methods

2.1. Compound selection

Compounds that had previously been evaluated *in vivo* for their adverse effects on reproductive tissues were selected from Pfizer's internal library and from published *in vivo* studies (Table 1). We took an unbiased approach and chose compounds that displayed reproductive toxicity in either male or females, and incorporated all reproductive findings, not just those considered to be hormonally driven. Not requiring a known mechanism of action for the *in vivo* reproductive toxicity was necessary to ensure sufficient power for the planned model building. It also allowed for us to determine if an assay focusing on steroidogenic regulation could play a role in predicting more general reproductive toxicity especially since most toxicity testing ends at that level of pathological resolution. The vast majority of these findings were of structural abnormalities (*i.e.*, pathology or lesions) noted in reproductive organs after dosing for periods ranging from 2 weeks to 6 months. We compiled 83 compounds with adverse reproductive findings and 79 compounds with no *in vivo* adverse finding. Forskolin (Sigma) and prochloraz (Sigma) served as positive controls for the assay. All compounds were dissolved in DMSO and DMSO (1%) was used as vehicle control.

2.2. Cell culture and toxicity assay

H295R (ATCC) cells were cultured at 37 °C in 5% CO_2 in DMEM:F12 (Invitrogen) medium supplemented with 2.5% Nu-Serum (BD Biosciences), 1% ITS+ Premix (BD Biosciences), L-glutamine (Invitrogen), and penicillin-streptomycin (Invitrogen). Exposures were conducted in 2 phases: dose-range finding, and the response phase. For dose-range finding, cells were plated at 10×10^4 per well in 96 well plates for 48 h before treating with compounds dissolved in DMSO for 30 h in 8 point curves (2-fold dilutions, starting at 300 μM). Compound effect on cell viability was determined by using the Cell Titer Glo (Promega) assay kit on compound-treated cells and results were graphed using the IDBS *XLFit* version *4.2.1* plug-in for Microsoft Excel to determine a TC_{10} (toxic concentration at which 10% cell death is observed) for the concentration used in the response experiment. If no cytotoxicity was observed, the cells were treated at 300 μM. The response phase exposure was run with cells plated at 10×10^4 cells per well in 96-well plates for 48 h before compound exposure at the TC_{10} for an additional 30 h.

We wanted to determine if the readout from our assay was correlated with cytotoxicity. We employed an assay which is commonly used in our lab as a general readout of cytotoxicity: the transformed human liver epithelium cell line [16,17]. THLE-2 cells (ATCC: CRL2706) were cultured at 37 °C in 5% CO_2 in BEBM (Lonza) supplemented with 10% HI-FBS, 5 ng/ml hEGF, 70 ng/ml phosphoethanolamine, and the supplied BEBM bullet kit (Lonza). Cells were plated at 2.5×10^3 cells/well in 384-well plates for 24 h before compound treatment. Compounds were added dissolved in DMSO in 10-point curves with 300 μM as the highest concentration. After a 72-h incubation, Cell Titer Glo reagent (Promega) was added. Luminescence values were read on a luminometer and graphed using the IDBS *XLFit* version *4.2.1* plug-in for Microsoft Excel to determine an IC_{50} values for cytotoxicity.

2.3. RNA isolation and quantitative RT-PCR analysis

At time of harvest, the cell supernatants were removed and frozen, and 100 μl lysis buffer was added to each well in the 96-well plate. The RNA was then extracted using the SV96 RNA kit (Promega) according to manufacturer's instructions. cDNA was generated using HiCapacity RT kit (ABI #4368813) according to

Table 1
Model compounds with adverse *in vivo* reproductive outcomes. *In vivo* findings occurred in rat studies unless otherwise noted.

Compound	On market name	*In vivo* observed toxicity	Compound	On market name	*In vivo* observed toxicity
Compound002		Mucification of vagina, subnuclear vacuoles in uterus	Compound085		inflammation in epididymides
Compound004	Clomiphene	Seminal vesicle and prostate gland weights decreased, reduced testosterone [38]	Compound092		Inflammation in epididymides
Compound007	Cefoperazone	Testicular degeneration [39]	Compound093		Degeneration in testis, abnormal content in epididymides
Compound009	Vincristine	Malformation of late spermatids and arrest of cell division of spermatocytes and spermatogonia	Compound094		Dilatation tubules in testis
Compound010		Hypertrophy of interstitial cells in the testis	Compound096		Estrous cycle disruption, mammary hyperplasia, decreased seminal vesicles
Compound012		Testicular degeneration	Compound098		Inflammation in Epididymides
Compound013		Testicular degeneration, decreased ovary weight	Compound099		Low testosterone
Compound017	Colchicine	Degraded microtubules in the testis, caused abnormalities of the head and acrosome of testicular spermatids [40]	Compound100		Inflammation in Epididymides
Compound018		Inhibits sexual differentiation in gonads	Compound101		Testicular degeneration (dog), cervial polyps
Compound019		Testicular atrpophy, disturbance in rat estrus cylce	Compound102		Degenerative spermatid cell debris, Atrophy, delayed estrus cycle
Compound022		Degeneration of seminiferous tubules, germinal cell degeneration/depletion, hypospermia, interstitial cell hyperplasia in the testis, Decrease ovary weights	Compound104		Testicular lesions: germ cell degeneration and necrosis
Compound024		Testicular degeneration/atrophy	Compound105		Testicular lesions: germ cell degeneration and necrosis
Compound026		Vaginal mucosal atrophy, Vaginal parakeratosis due to senescence	Compound110		Decreased epididymides and testes weights
Compound027		Abnormal luminal content, increased giant cells in seminiferous, necrosis of corpora lutea in ovaries	Compound111		Testis/epididymis, seminiferous tubules, spermatocytic degeneration, necrosis, mammary increases
Compound028	Acyclovir	Testicular atrophy, spermatogenesis (mouse) [41]	Compound112		Decreased prostate/seminal vesicle wt, delayed sperm release, delayed estrus cycle
Compound029	Nifedipine	Reduced weight of the testis and epididymis, reduced sperm count [42]	Compound116		Grossly small testes and epididymides spermatogenic arrest in seminiferous tubules, spermatogenic epithelium was attenuated
Compound031	Spironolactone	Low testosterone levels [43]	Compound118		Testis/epididymis, seminiferous tubules, spermatocytic degeneration, necrosis
Compound032		Low testosterone, leydig and sertoli cells were decreased	Compound119		Testicular seminiferous tubular degeneration, epididymal sperm granulomas, oligospermia and increased intralumenal cell debris in epididymides
Compound037	Chlorambucil	Damage spermatogenesis and cause testicular damage [44]	Compound122		Seminiferous tubule and epididymal lesions
Compound039	Flutamide	Increases in plasma testosterone level and Leydig cell hyperplasia, seminiferous tubular atrophy and degeneration [45]	Compound126		Arteriopathy in Epididymides (dog)
Compound042	Chlorpromazine	Decrease in testicular weight [46]	Compound128		Testicular giant cells, epdidymal spermatic granuloma
Compound048	Divalproex	Reduced spermatogenesis and testicular atrophy [47]	Compound129		Germ cell degeneration and necrosis degeneration of seminiferous tubulesluminal cellular debris in epididymes
Compound053	Busulfan	Testes, tubules and germinal epithelia were decreased significantly, disrupt spermatogenesis through affecting both germ and somatic cells [48]	Compound130		Abnormal testis and Epididymides
Compound057	Theobromine	Testicular atrophy with extensive spermatogenic cell degeneration and necrosis [49]	Compound135		Seminiferous tubular degenerat,decreased numbers of spermatids, sloughing of degenerate and necrotic cells in the tubular lumen, presence of giant cells and vacuolation of Sertoli cells
Compound059		Vacuolation in testis	Compound137		Seminiferous tubule degeneration
Compound060		Spermatic Retention in testis, Abnormal content in Epididymides	Compound138		Testicular degeneration
Compound061		Testicular degeneration	Compound139		Testicular degeneration and muscle necrosis
Compound063		Multifocal dilation of the seminiferous tubules, stops rat cycles	Compound140		Testicular degeneration, Seminiferous tubule degeneration
Compound064		Decreased uterine weights, subnuclear vacuolation, Increased estrous cycle length, increased number of animals in metestrous decreased conception, corpora lutea	Compound142		Delayed sperm release

Table 1 (*Continued*)

Compound	On market name	*In vivo* observed toxicity	Compound	On market name	*In vivo* observed toxicity
Compound066		Increase in accumulation of cellular debris in the lumen of the epididymal ducts	Compound143		Lesions in epididymis, bilateral sperm granuloma
Compound067	Bromocriptine	Decrease in testicular testosterone content and a reduction in plasma testosterone levels [50]	Compound145		Seminiferous tubule degeneration
Compound068		Testicular degeneration (mouse)	Compound147		Seminiferous tubule degeneration
Compound069	Sulfasalazine	Abnormal motion of epididymal sperm, decrease in sperm count and motility [51]	Compound148		Delayed sperm release
Compound070	Acrylamide	Destruction of seminiferous tubules, decrease in the total sperm count [52]	Compound149	Reserpine	Atrophy of the interstitial cells of the testis [53]
Compound071		Testicular atrophy (dog)	Compound150	Melphalan	Induces testicular damage and affects sperm variables [54]
Compound074		Low plasma testosterone levels	Compound153		Reduced the sperm count and motility,testicular damage and affected sperm variables
Compound075		Decreased seminal vesicles(rat), Estrous cycle disruption, mammary hyperplasia, Hypertrophy in mammary, decreased seminal vesicles	Compound155	Ganciclovir	Testicular damage and decreased sperm count [55]
Compound076		Prolonged CL's in ovaries, mammary hyperplasia in females, and vaginal mucification.	Compound159		Testicular atrophy (dog)
Compound077		Degeneration of pachytene spermatocytes in the testes	Compound160		Testis weight drop and germ cell apoptosis
Compound080		Low plasma testosterone levels, decrease epididymides	Compound161		Tubular degeneration in testis, Abnormal content in Epididymides
Compound082		Luminal fluid decreased in seminal vesicles, decreased seminal vesicles	Compound162		Increased ovarian follicular cysts, occasional absence of corpora lutea, imbalance in estrous cycle
Compound084		Hypertrophy of testis, decreased seminal vesicles			

manufacturer's instructions. Q-PCR was performed in 384-well plates using Gene Expression Master Mix (ABI#4370074) and ABI Inventoried Gene Expression assays: *GAPDH* (Hs99999905 m1) *STAR* (Hs00264912 m1), *3βHSD2* (Hs00605123 m1), *CYP17A1* (Hs00164375 m1), *CYP11B2* (Hs01597732 m1), *CYP19A1* (Hs00903413 m1), *CYP21A2* (Hs00416901 g1), *CYP11A1* (Hs00167984 m1). Values normalized to *GAPDH* are presented as fold over DMSO.

2.4. Hormone measurement

Hormones were measured using Multi-Spot 96 HB 4-Spot Custom Steroid Hormone Panel (Mesoscale Discovery) with alterations. Briefly, a custom cortisol assay was developed in conjunction with MesoScale Discovery and cortisol was added to the available 4-plex along with currently provided analytes progesterone, DHEA, and testosterone. 50 μl of spent culture medium was run per well in duplicate and values were calculated based on a standard curve. Values are represented as percent of DMSO control.

2.5. Statistical analysis

The Supplemental Data (Supplemental Table 1) contain the raw data values and computer code in the R programming language [18,19] that can be used to completely reproduce all the calculations shown in this manuscript.

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.reprotox.2013.12.009.

The assay data were pre-processed prior to modeling. First, there were 23 compounds missing at least one data point. These data were imputed using the five nearest-neighbor imputation [20]. Also, some of the analyte distributions were right-skewed and statistical transformations were applied to the data to produce a more symmetric distribution. Finally, the analytes were standardized so that they have a mean of zero and a standard deviation of one.

A variety of statistical models were explored to find the best option for predicting the probability of toxicity. A detailed description of each of these models is given in Hastie et al. [21] and Kuhn and Johnson [22]. Many of these models have "tuning parameters" which cannot be estimated directly from the data. For these, the cross-validation procedure was executed for each value of the tuning parameter(s) and the best model was selected based on the area under the ROC curve. The models, and their tuning parameters, used here were:

- Random Forests (RF) is a tree-based ensemble model [23]. Five values of the tuning parameter, the number of random selected analytes at each split, were used to optimize the model.
- Support Vector Machines (SVM) [24] is a kernel-based machine learning method (the radial basis kernel was used in these models). For this model, five values of the cost parameter were tuned. Additionally, the method of Caputo et al. [25] was used to analytically estimate the radial basis function parameter.
- Nearest Shrunken Centroids (NSC) is a linear classification model with built-in feature selection. Forty values of the tuning parameter, the shrinkage threshold, were evaluated.
- Naive Bayes (NB) uses the distribution of each analyte independently to estimate the probability of toxicity [21]. Two different methods for estimating the distributions (parametrically and non-parametrically) were tested.
- Logistic Regression is a well-known statistical technique that produces linear decision boundaries [26]. It will be demonstrated that there are many significant between-analyte correlations, which can cause instability in the model and may negatively affect performance. Because of this, a second order (L2) penalty will be used during model fitting to compensate for the collinearity of the analytes similar to classical ridge regression [27].

These models used all of the analytes data produced in the model (degree of change in gene expression for the steroid synthesis genes, as well as medium levels of steroids). Recursive feature

elimination (RFE) was also used in conjunction of several of these models (except NCS). This backwards selection routine ranks the analytes by their importance to the model [24]. After the model with all analytes was created, the least important analytes were removed one at a time, the model was re-fit and performance was tracked. Using this profile of performance, the optimal number of predictors was determined and the final model was fit using only the top analytes.

Three measures of performance were used to characterize the effectiveness of the models:

- *Sensitivity*: given that a compound induces reproductive toxicity, what is the probability that the compound is predicted as toxic (a.k.a. true positive rate).
- *Specificity*: given that a compound does not induce reproductive toxicity, what is the probability that the compound is predicted as non-toxic (true negative rate).
- *Area under the ROC curve*: using the estimated probability of reproductive toxicity, the ROC curve tracks the sensitivity and specificity over a continuum of probability cut-offs [28]. The area under the curve can be used as a measure that combines sensitivity and specificity.

To obtain useful estimates of model performance, cross-validation was used on the training set samples [29] given the size of the data set. This is a form of "resampling" where random subsets of the samples are methodically held-back from the data set. The model was fit to the majority of the data and the held-out samples were predicted. This was repeated multiple times and performance was estimated for each of the held-out samples and these values were aggregated into a single estimate of performance. In these analyses, five repeats of ten-fold cross-validation were used because of its attractive bias and variance properties. In the end, 50 different hold-out samples of roughly 10% were used to estimate performance. For recursive feature elimination, the same cross-validation scheme was utilized to estimate the uncertainty of the feature selection [30] while an internal cross-validation was used to tune the models.

3. Results

3.1. Assay performance using prototypical inducers

Development of the steroidogenesis assay in H295R cells included systematically evaluating different cell plating densities and times of sampling (Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.reprotox.2013.12.009). Because the goal of this work was to develop a cell-based screen that could be utilized for the evaluation of large compound sets, it would not be possible to use multiple timepoints in the final version of the assay. Conditions were optimized so that the two model compounds, forskolin and prochloraz, produced the greatest effect in the assay endpoints. These two compounds were then used as the assay's positive controls and included in each subsequent run. Optimal conditions were determined to be 100,000 cells plated per well plated 48 h before compound addition and exposure for an additional 30 h.

Forskolin is an adenylate cyclase activator that raises intracellular cAMP levels. It has been demonstrated to effect the enzymes of the steroidogenesis pathway as well as increase secreted progesterone and cortisol levels [31]. Treatment of cells with 10 μM forskolin for 30 h resulted in significant changes in gene expression. Robust induction of *CYP11B2* (270-fold, p-value <0.0001) was observed. Forskolin exposure also produced strong inductions of *CYP19A1* (26-fold, p-value <0.0001), *CYP21A2* (12.3-fold, p-value <0.0005), and *HSD3β2* (24.4-fold, p-value <0.0005), as well as more mild, but still significant increases in *CYP11A1* (3.3-fold, p-value <0.00001), *CYP17A1* (6.1-fold, p-value <0.0001), and *STAR* (5.1-fold, p-value <0.00001) expression (Fig. 2A). These findings are similar in rank order to those reported by others [6]. Based on such gene changes, one would expect to see increased product being secreted, and this was the case: the expected increases in the secreted levels of cortisol (290%, p-value <0.0001) and progesterone (392%, p-value <0.0005) into the culture medium were observed (Fig. 2B). Minimal effects on testosterone and DHEA were observed at this time point. Prochloraz is an antifungal compound that produces instances of male reproductive toxicity when administered to rats during sexual differentiation [32,33]. It has been reported to increase progesterone levels whilst decreasing testosterone levels *in vivo* [34]. Exposure of the cells to 4 μM prochloraz resulted in a robust induction of *CYP11A2* (97-fold, p-value <0.0001) and modest inductions of *CYP19A1* (3.4-fold, p-value <0.0001), *CYP21A1* (3.4-fold, p-value <0.0005), and *HSD3β2* (5.2-fold, p-value <0.001) (Fig. 2C). Treatment of H295R cells with prochloraz resulted in considerable increases in progesterone secretion (1838%, p-value <0.001) while substantially decreasing the secretion of cortisol (3%, p-value $<10^{-18}$), DHEA (7%, p-value $<10^{-18}$), and testosterone (9%, p-value $<10^{-17}$) after 30 h.

3.2. Compound test set administration

H295R cells were plated at 100,000 cells per well in 96 well plates for 48 h before compound administration in 8-point curves with a 300 μM starting maximal concentration. Cell Titer Glo (Promega) assay kit, which measures ATP, was used as a measure of cytotoxicity. TC_{10} (10% unspecified cytotoxicity) values were calculated from the curves and used as the treatment dose in the subsequent effect assay. A single concentration point was employed in order to enable this assay to be high throughput. The TC_{10} concentration was chosen as many known endocrine disruptors were shown to produce significant effects in the assay system at the TC_{10} concentration (Ulleras et al.) If no cytotoxicity was observed, the cells were treated with the compounds at 300 μM. Cells were treated in duplicate wells. The experiment was run under the previously described optimized conditions. Compound effects on gene expression of the enzymes in the steroidogenesis pathway as well as the level of secreted progesterone, testosterone, DHEA, and cortisol were measured and the results (see supplemental data Table 1) were used to build the statistical model.

3.3. Predictive modeling

ROC curve analysis was conducted for each analyte separately to obtain an initial indication of its utility as a predictor. For each predictor, a series of thresholds are examined to see if the toxic and non-toxic compounds can be classified using a simple cut-off point. The sensitivities and specificities resulting from the various cut-offs are then used to form the predictor's ROC curve. The area under that curve is then used to quantify the predictive ability of each assay. If the assay can differentiate between toxic and non-toxic compounds on its own, the area under the ROC curve should be close to one. The closer to 1, the greater the value that endpoint has in predicting reproductive toxicity. The areas under the ROC curves for each individual analyte were: *CYP11B1* (0.77), *CYP11A1* (0.77), *CYP21A1* (0.74), *CYP19A1* (0.7), *STAR* (0.67), *CYP11B2* (0.65), *HSD3B2* (0.62), *CYP17A1* (0.61), progesterone (0.55), testosterone (0.54), cortisol (0.54), DHEA (0.46). Here, the RT-PCR assays of gene expression showed higher individual predictivity of toxicity than the hormonal assays. However, many of the predictive models that were used here are able to model complex relationships between the outcomes and multiple predictors and, because of this, the

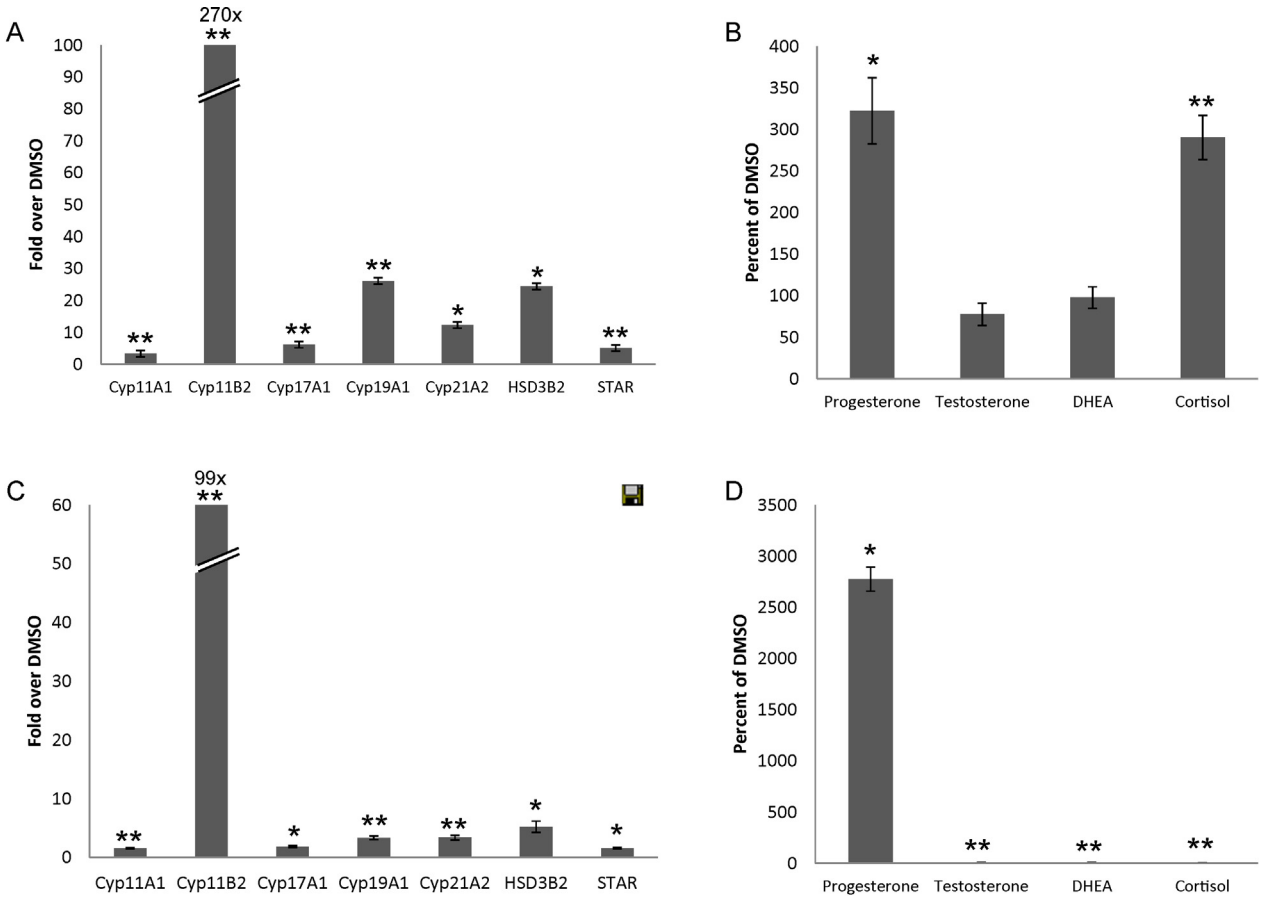

Fig. 2. Expression analysis of the enzymes in the steroidogenesis pathway. H295R cells were treated for 30 h with 10 μM forskolin (A) or 4 μM prochloraz (C). All values are normalized to GAPDH and presented as fold over DMSO. Hormone secretion into the media from cells treated with 10 μM forskolin (B) or 4 μM prochloraz (2D) was analyzed and is represented as percent of DMSO control. *p-value <0.001, **p-value <0.0001.

individual ROC curve results are poorly indicative of the relative importance of an individual assay to a final comprehensive model.

The cross-validation results are shown in Table 2. The mean and corresponding standard error were computed for the 50 estimates of performance produced during the cross-validation process. For example, the Naive Bayes model had area under the ROC curve values that were on average 0.756 but, given that the standard error of the mean was 0.0178, a 95% confidence interval for the mean of the cross-validation results was (0.720, 0.792). Many of the predictive models showed performance characteristics that were roughly equivalent, given the samples size. Unlike the other models, the Nearest Shrunken Centroid model achieved a reasonable area under the ROC curve (0.77), but did so by sacrificing model specificity (0.52) for sensitivity (0.79).

The performance profiles for the models when used in conjunction with recursive feature elimination are shown in Fig. 3. Although Random Forest shows a small decrease in performance as predictors are removed, the model with the largest area under the ROC curve (random forest) uses all the endpoints as useful predictors. The Naive Bayes model showed an increase in performance up to 3 analytes (*CYP11B1*, *CYP11A1*, and *CYP21A1*). Logistic Regression had small gains when eliminating predictors.

Of these models, we focused on the basic Random Forest model, which appeared to perform the best and Naive Bayes with feature selection classifier. Naive Bayes (RFE) was chosen as the comparator because it performs the best with fewer endpoints. This model would allow the assay to maintain predictive power while eliminating the need for hormone measurements and reducing the transcriptional endpoints, further enabling the high throughput ability of the assay. Random Forest can compute a built-in variable importance metric that quantifies the loss in performance if each predictor was coerced to be non-informative by randomly scrambling the data. For example, suppose the testosterone values were randomly scrambled and new model predictions were calculated. If performance did not drop by altering this predictor, it would not be considered important to the model. This same process would then be applied independently to each predictor. Using this process, the list of predictors from most important to least important is *CYP11B1*, *CYP11A1*, *CYP21A1*, *CYP19A1*, *STAR*, *CYP11B2*, *HSD3B2*, *CYP17A1*, progesterone, testosterone, cortisol, and DHEA. For the two other models, the number of times that each predictor was selected for each of the 1 cross-validation iterations can be used to measure the importance to the models. For Naive Bayes, the same assays were selected in each cross-validation iteration: *CYP11B1*, *CYP11A1* and *CYP21A1*.

Fig. 4 shows the distributions of the class probabilities for the Random Forest and Naive Bayes models. The histograms show the probability of reproductive toxicity for each compound that was held-out during the cross-validation process. The bottom pair of histograms shows the probability of reproductive toxicity for compounds that were truly non-toxic *in vivo*. The Random Forest model shows the largest proportion of points between 20% and 40% probability. As the predicted probability of reproductive toxicity increases, fewer non-toxic compounds are misclassified as toxic. For the Bayesian model, a large number of non-toxic compounds have low probabilities (<20% probability) of being called toxic but, compared to Random Forest; more compounds are confidently mis-predicted as toxic in the Bayesian model. The upper

Table 2
Mean performance values estimated using cross-validation.

Model	Predictors	ROC		Sensitivity		Specificity	
		Mean	Std. err.	Mean	Std. err.	Mean	Std. err.
Logistic Reg	12	0.778	0.014	0.708	0.023	0.704	0.025
Logistic Reg (RFE)	7	0.801	0.014	0.745	0.021	0.731	0.025
Naïve Bayes	12	0.756	0.018	0.657	0.023	0.754	0.023
Naïve Bayes (RFE)	3	0.812	0.015	0.655	0.019	0.780	0.022
NSC	4	0.773	0.017	0.789	0.025	0.524	0.026
Random Forest	12	0.847	0.013	0.724	0.019	0.758	0.023
Random Forest (RFE)	11	0.845	0.013	0.751	0.018	0.745	0.025
SVM	12	0.852	0.012	0.748	0.024	0.791	0.019
SVM (RFE)	10	0.837	0.011	0.753	0.020	0.771	0.020

panels show similar results for the truly reproductive toxicants. Random Forest shows a broader probability distribution for reproductive toxicants while Naive Bayes tends to confidently predict reproductive toxicity (correctly or incorrectly).

Since the Random Forest model has a wider distribution of class probability values, one potential method for improving its performance is to institute an equivocal or indeterminate zone for random forest predictions. Here, we define a range of probability values that are too uncertain to confidently predict the compounds. For example, if the predicted probability of reproductive toxicity was between 40% and 60%, our confidence in the correctness of that prediction is low and the compound would not be classified as reproductive toxicant or non-toxic. For the Random Forest model, an equivocal zone between 40% and 60% would exclude, on average, 28.9% of the compounds but would increase the area under the ROC curve from 0.838 to 0.888. This approach would not be as effective for the Naive Bayes model since fewer compounds are predicted inside of the potential equivocal zone (as indicated by the zone between the dotted lines in Fig. 3).

The value of this approach increases with the criticality of the prediction. If one or more independent techniques will be used to verify the model prediction for a given compound, the equivocal zone may increase the cost and time of a model prediction. However, if very little evidence exists related to the toxicity of the compound (as is the case with new pharma candidates), the equivocal zone can play a pivotal role in increasing our confidence in the information used to judge a compound.

3.4. Reproductive toxicity predictions are independent of cytotoxicity

General cytotoxicity assays have been utilized by the pharmaceutical industry as a tool to predict *in vivo* toxicities. One commonly used assay employs THLE transformed liver cells and measures the amount of viable cells present by quantification of ATP after treatment with compounds [17]. We wished to determine if the predictions from this assay and algorithm was being driven more by drug effects on cellular health and cytotoxicity than drug-induced effects on steroidogenesis. THLE cells were treated with the compound test set in 10-point curves starting at 300 μM. Values were normalized to DMSO and IC_{50} were determined. The H295R random forest predictions were compared to the cytotoxicity IC_{50}s in THLE cells (Fig. 5). No correlation was observed between the cytotoxicity measurement and the prediction of reproductive toxicity by random forest modeling. This comparison provides confidence that the H295R assay is predicting specific toxicity due to

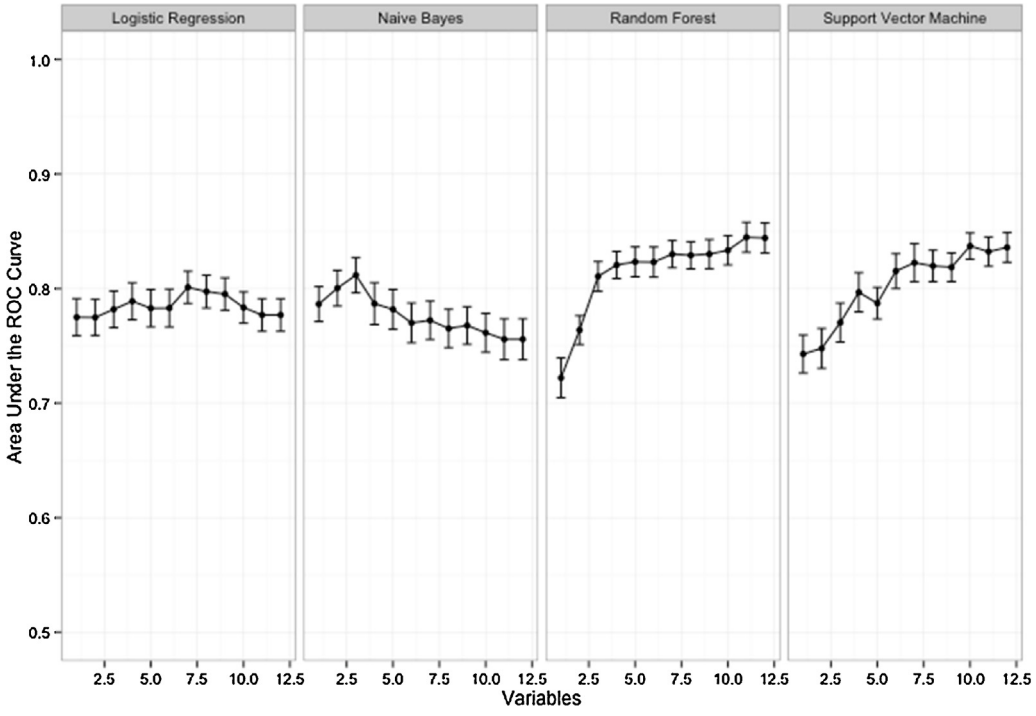

Fig. 3. Performance profiles for different models when conducting recursive feature elimination.

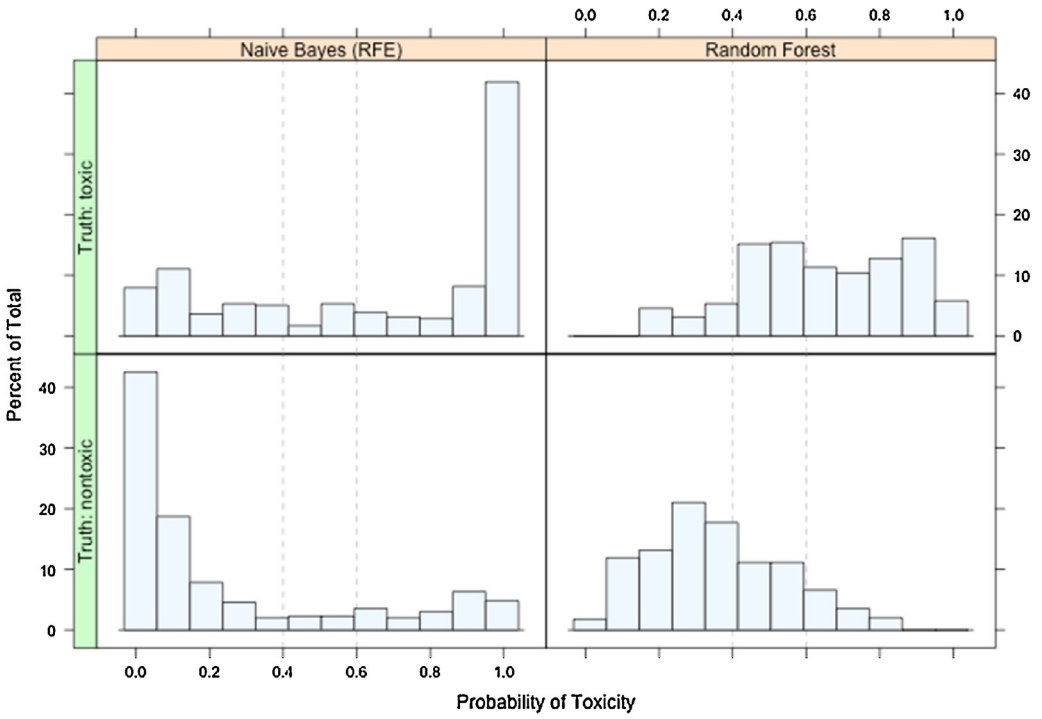

Fig. 4. Class probabilities for the two models based on the hold-out samples during cross-validation. The panels split the compounds in the training set by their true class.

alteration in the steroidogenesis pathway and not merely general cytotoxicity mechanisms.

4. Discussion

This study evaluated a mechanistic approach to identify reproductive toxicants by examining their effects on steroidogenic gene expression and selected steroid output in the H295R cell culture system. The H295R *in vitro* cell system is being employed thoughout the world for use in screening enviromental agents for possible endocrine disruption activity that may result in developmental and reproductive problems in both humans and wildilfe. The H295R cell line expresses each of the enzymes in the steroidogeneic pathway and secretes the hormones of interest at detectable levels into the culture medium making this an ideal *in vitro* model for studying the steroidogenic pathway. The steroidogenic pathway is made up of multiple enzymes and there are numerous control points in the production of steroids. The ability to simultaneously assess the effect of compounds upon multiple steps within the pathway *in vitro* and employ a statistic model to correlate these effects with a toxic/non-toxic *in vivo* prediction is of great use to those in the drug development process [35].

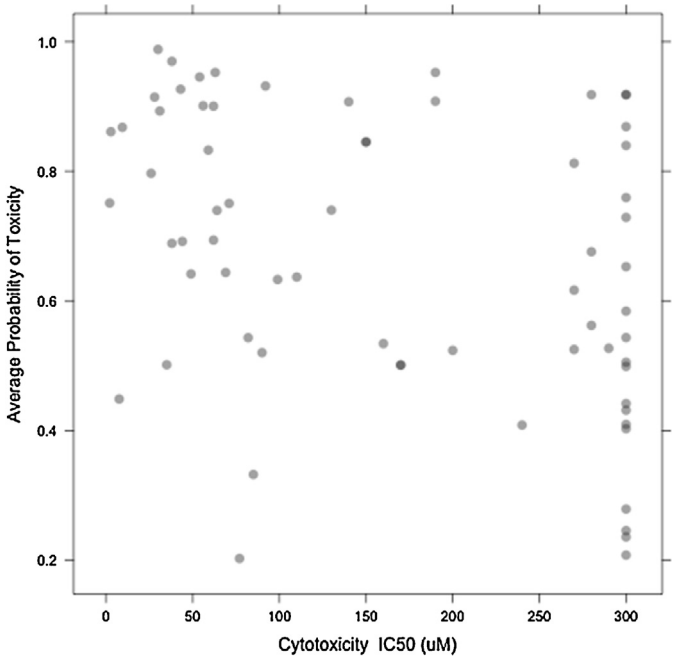

Fig. 5. Prediction of reproductive toxicity does not correlate with general cytotoxicity. Random forest prediction results from the steroidogenesis assay are compared to a compound concentration causing 50% death in a general cell cytotoxicity assay.

We used forskolin and prochloraz, two compounds well-studied for their effects on the steroidogenesis pathway, to determine the effectiveness and robustness of this cell system. Both compounds were able to differentially induce the transcriptional levels of the studied enzymes. Forskolin induced a robust increase in progesterone and cortisol levels secreted by the H295R cells. Procholaz induced progesterone, but resulted in dramatic decreases in testosterone, DHEA and cortisol. This demonstrated the cell system's ability to identify compounds that both increase and decrease hormone levels. It should be noted that modulators of enzyme expression are not the same as modulators of enzyme activity. This assay system does not detect compounds which alter the activity of an enzyme, except by inferrence from changes in steroid levels themselves.

Indeed the robust increases in progesterone observed when treating the H295R cells with prochloraz can not be explained by the transcriptional changes of the enzymes. *CYP17A1*, responsible for its conversion, is increased in expression. However, others have reported that the hydrolase activity of *CYP17A1* is severely inhibited by similar levels of prochloraz treatment in H295R cells [36]. This decreased activity is therefore likely the cause for the increases progesterone levels observed in our system.

We were able to create a model that predicts reproductive toxicants with 76% sensitivity and 72% specificity for an ROC of 0.85

using the Random Forest model with all 12 predictors. Using the Naïve Bayes model with the recursive feature elimination including only 3 predictors, the ROC drops to only 0.82, yet makes the assay easier and more amenable to a highthoughput system by elimating the hormone measurement and limiting the amount of Q-PCR required, while matintaining predictive power. Combining multiple endpoints from this assay has yielded a model which predicts better than any one endpoint alone, demonstrating the multiparamaeter approach necessary and beneficial. The majority of publications describe the use of screening H295R cells for hormone measurements only [9,14,37]. Those enlisting these cells to predict toxic outcomes without incorporating the gene expression arm are likely missing important information and could likely benefit from inclusion of a genomics assay.

Finally, we should point out that this assay does not predict only compounds with hormonal effects *in vivo*. Indeed, many of the compounds used here have no evidence of hormonal changes (Table 1). We recognize that it's possible that many of those toxicants with no detectable hormonal changes may, indeed, produce some subclinical alterations in the reproductive hormone economy, changes which would not be picked up by the relatively insensitive methods available to investigators (pathology, single-timepoint hormone measurements, organ weights). In any case, it's worth noting that we view this assay as being useful for identifying all reproductive toxicants, regardless of their degree of hormonal impact. We recognize the interpretational challenges this poses, but we feel the data are compelling.

The field of predictive toxicology has been rapidly developing. It includes epigenetics, toxicokinetics, genomic biomarkers, systems biology, *in vitro* assays, predictive modeling and computational science. These techniques are being incorporated more often in industry as we strive to reduce the costs of drug development by limiting expensive animal studies to only the most promising compounds. This has the added benefit of a reduction in animal usage. Our goal was to develop and implement an *in vitro* method to employ as early in the drug discovery process as possible to predict reproductive toxicity. We characterized the H295R system as an impressive model for reproductive toxicity that can be employed to rank order compounds and to select the safest possible compounds to be used in animal studies.

Conflict of interest

All authors were employees of Pfizer Global Research and Development, which funded this study, at the time the experimental work was conducted.

Transparency document associated to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.reprotox.2013.12.009.

References

[1] NRC. Toxicity testing in the 21st century: a vision and a strategy. Washington, D.C.: The National Academies Press; 2007.
[2] MacDonald JS, Robertson RT. Toxicity testing in the 21st century: a view from the pharmaceutical industry. Toxicol Sci 2009;110:40–6.
[3] Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. Drug Discovery Today 2005;10:1421–33.
[4] Will Y, Schroeter T. Deployment of in silico and in vitro safety assays in early-stage drug discovery. Future Med Chem 2012;4:1211–3.
[5] Kallen CB, Billheimer JT, Summers SA, Stayrook SE, Lewis M, Strauss JF. Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein. J Biol Chem 1998;273:26285–8.
[6] Hilscherova K, Jones PD, Gracia T, Newsted JL, Zhang X, Sanderson JT, et al. Assessment of the effects of chemicals on the expression of ten steroidogenic genes in the H295R cell line using real-time PCR. Toxicol Sci 2004;81:78–89.
[7] Oskarsson A, Ulleras E, Plant KE, Hinson JP, Goldfarb PS. Steroidogenic gene expression in H295R cells and the human adrenal gland: adrenotoxic effects of lindane in vitro. J Appl Toxicol 2006;26:484–92.
[8] Gracia T, Hilscherova K, Jones PD, Newsted JL, Higley EB, Zhang X, et al. Modulation of steroidogenic gene expression and hormone production of H295R cells by pharmaceuticals and other environmentally active compounds. Toxicol Appl Pharmacol 2007;225:142–53.
[9] Ulleras E, Ohlsson A, Oskarsson A. Secretion of cortisol and aldosterone as a vulnerable target for adrenal endocrine disruption – screening of 30 selected chemicals in the human H295R cell model. J Appl Toxicol 2008;28:1045–53.
[10] Gazdar AF, Oie HK, Shackleton CH, Chen TR, Triche TJ, Myers CE, et al. Establishment and characterization of a human adrenocortical carcinoma cell line that expresses multiple pathways of steroid biosynthesis. Cancer Res 1990;50:5488–96.
[11] Staels B, Hum DW, Miller WL. Regulation of steroidogenesis in NCI-H295 cells: a cellular model of the human fetal adrenal. Mol Endocrinol 1993;7:423–33.
[12] Rainey WE, Bird IM, Mason JI. The NCI-H295 cell line: a pluripotent model for human adrenocortical studies. Mol Cell Endocrinol 1994;100:45–50.
[13] Gracia T, Hilscherova K, Jones PD, Newsted JL, Zhang X, Hecker M, et al. The H295R system for evaluation of endocrine-disrupting effects. Ecotoxicol Environ Saf 2006;65:293–305.
[14] Breen MS, Breen M, Terasaki N, Yamazaki M, Conolly RB. Computational model of steroidogenesis in human H295R cells to predict biochemical response to endocrine-active chemicals: model development for metyrapone. Environ Health Perspect 2010;118:265–72.
[15] Rotroff DM, Dix DJ, Houck KA, Knudsen TB, Martin MT, McLaurin KW, et al. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ Health Perspect 2012;121:7–14.
[16] Greene N, Song M. Predicting in vivo safety characteristics using physiochemical properties and in vitro assays. Future Med Chem 2011;3:1503–11.
[17] Greene N, Aleo MD, Louise-May S, Price DA, Will Y. Using an in vitro cytotoxicity assay to aid in compound selection for in vivo safety studies. Bioorg Med Chem Lett 2010;20:5308–12.
[18] Gentleman R, Ihaka R. R: a language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.
[19] R Development Core Team. R: A Language and Environment for Statistical Computing. 2.11.1 ed. Vienna: R Foundation for Statistical Computing; 2010.
[20] Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001;17:520–5.
[21] Hastie T, Friedman TRJ. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2009. p. 745.
[22] Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer; 2013.
[23] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[24] Guyon I, Barnhill WJ, Vapnik SV. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389–422.
[25] Caputo B, Fureszjo SK, Smola FA. Appearance-based object recognition using svms: which Kernel should i use? Proceedings of NIPS workshop on statitsical methods for computational experiments in visual processing and computer vision whistler. 2002. p. 2002.
[26] Hosmer DWLS. Applied logistic regression. Hoboken, NJ: Wiley-Interscience Publication; 2000.
[27] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1–22.
[28] Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. Br Med J 1994;309:188.
[29] Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics 2005;21:3301–7.
[30] Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002;99:6562–6.
[31] Purdy SJ, Whitehouse BJ, Abayasekara DR. Stimulation of steroidogenesis by forskolin in rat adrenal zona glomerulosa cell preparations. J Endocrinol 1991;129:391–7.
[32] Noriega NC, Ostby J, Lambright C, Wilson VS, Gray Jr LE. Late gestational exposure to the fungicide prochloraz delays the onset of parturition and causes reproductive malformations in male but not female rat offspring. Biol Reprod 2005;72:1324–35.
[33] Laier P, Metzdorff SB, Borch J, Hagen ML, Hass U, Christiansen S, et al. Mechanisms of action underlying the antiandrogenic effects of the fungicide prochloraz. Toxicol Appl Pharmacol 2006;213:160–71.
[34] Blystone CR, Furr J, Lambright CS, Howdeshell KL, Ryan BC, Wilson VS, et al. Prochloraz inhibits testosterone production at dosages below those that affect androgen-dependent organ weights or the onset of puberty in the male Sprague Dawley rat. Toxicol Sci 2007;97:65–74.
[35] Greene N, Naven R. Early toxicity screening strategies. Curr Opin Drug Discov Devel 2009;12:90–7.
[36] Ohlsson A, Ulleras E, Oskarsson A. A biphasic effect of the fungicide prochloraz on aldosterone, but not cortisol, secretion in human adrenal H295R cells – underlying mechanisms. Toxicol Lett 2009;191:174–80.
[37] Hecker M, Hollert H, Cooper R, Vinggaard AM, Akahori Y, Murphy M, et al. The OECD validation program of the H295R steroidogenesis assay: phase 3, final inter-laboratory validation study. Environ Sci Pollut Res Int 2011;18:503–15.
[38] Brown JL, Chakraborty PK. Characterization of the effects of clomiphene citrate on reproductive physiology in male rats of various ages. Acta Endocrinol (Copenh) 1988;118:437–43.
[39] Comereski CR, Bregman CL, Buroker RA. Testicular toxicity of N-methyltetrazolethiol cephalosporin analogs in the juvenile rat. Fundam Appl Toxicol 1987;8:280–9.

[40] Correa LM, Nakai M, Strandgaard CS, Hess RA, Miller MG. Microtubules of the mouse testis exhibit differential sensitivity to the microtubule disruptors carbendazim and colchicine. Toxicol Sci 2002;69:175–82.
[41] Narayana K. A purine nucleoside analogue-acyclovir [9-(2-hydroxyethoxymethyl)-9h-guanine] reversibly impairs testicular functions in mouse. J Toxicol Sci 2008;33:61–70.
[42] Lee JH, Ahn HJ, Lee SJ, Gye MC, Min CK. Effects of L- and T-type Ca(2)(+) channel blockers on spermatogenesis and steroidogenesis in the prepubertal mouse testis. J Assist Reprod Genet 2011;28:23–30.
[43] Nwe KH, Morat PB, Hamid A, Fadzilah S, Khalid BA. Novel effects of deoxycorticosterone on testicular 11beta-hydroxysteroid dehydrogenase activity and plasma testosterone levels in normal and adrenalectomized rats. Exp Clin Endocrinol Diabetes 1999;107:288–94.
[44] Delic JI, Stanley JA, Harwood JR. Testicular function in adult rats treated with the alkylating agent chlorambucil. Arch Androl 1986;17:87–98.
[45] Vo TT, Jung EM, Dang VH, Jung K, Baek J, Choi KC, et al. Differential effects of flutamide and di-(2-ethylhexyl) phthalate on male reproductive organs in a rat model. J Reprod Dev 2009;55:400–11.
[46] Chatterjee A. Effect of chlorpromazine on the testicular physiology in rats. Experientia 1965;21:545–6.
[47] Cansu A, Ekinci O, Serdaroglu A, Gurgen SG, Erdogan D, Coskun ZK, et al. Effects of chronic treatment with valproate and oxcarbazepine on testicular development in rats. Seizure 2011;20:203–7.
[48] Udagawa K, Ogawa T, Watanabe T, Tamura Y, Kita K, Kubota Y. Testosterone administration promotes regeneration of chemically impaired spermatogenesis in rats. Int J Urol 2006;13:1103–8.
[49] Gans JH. Comparative toxicities of dietary caffeine and theobromine in the rat. Food Chem Toxicol 1984;22:365–9.
[50] Rao MR, Bartke A. Effects of bromocriptine on plasma testosterone and gonadotropin levels and testicular lipid fractions in adult rats. Experientia 1984;40:88–9.
[51] Fukushima T, Hamada Y, Komiyama M, Matsuno Y, Mori C, Horii I. Early changes in sperm motility, acrosome reaction, and gene expression of reproductive organs in rats treated with sulfasalazine. Reprod Toxicol 2007;23:153–7.
[52] Yang HJ, Lee SH, Jin Y, Choi JH, Han DU, Chae C, et al. Toxicological effects of acrylamide on rat testicular gene expression profile. Reprod Toxicol 2005;19:527–34.
[53] Dias PL. Histometric analysis of the effects of reserpine on the interstitial cells of the rat testis. Br J Exp Pathol 1982;63:518–21.
[54] Howell SJ, Shalet SM. Testicular function following chemotherapy. Hum Reprod Update 2001;7:363–9.
[55] Faqi AS, Klug A, Merker HJ, Chahoud I. Ganciclovir induces reproductive hazards in male rats after short-term exposure. Hum Exp Toxicol 1997;16:505–11.