Journal of Assisted Reproduction and Genetics
https://doi.org/10.1007/s10815-025-03471-z

ASSISTED REPRODUCTION TECHNOLOGIES

Pharmacogenetic analysis using artificial intelligence (AI) to identify polymorphisms associated with sub-optimal ovarian response and hyper-response

José A. Ortiz[1] · B. Lledó[1] · L. Luque[4] · R. Morales[1] · S. Myles[1] · María Pérez[1] · J. Guerrero[2] · A. Bernabeu[3,5]

Received: 5 February 2025 / Accepted: 27 March 2025

Abstract

Purpose To identify genetic variants associated with an increased likelihood of sub-optimal ovarian response or hyper-response by machine learning.

Methods This retrospective observational study, conducted between March 2018 and April 2022, analyses 495 ovarian stimulations in oocyte donors. Only each donor's first ovarian stimulation was considered. The egg donors were healthy women aged 18 to 35 years. Donor characteristics and ovarian stimulation data were recorded, as well as genotypes of 31 polymorphisms previously identified as modulators of ovarian response.

Models to predict the type of ovarian response (sub-optimal, normal, or hyper-response) were performed using 5 different classification machine-learning algorithms. The most important variables were determined by SHAP (Shapley-Additive-exPlanations) values.

Results Despite being young with good ovarian reserves and using similar stimulation protocols, 15.15% of oocyte donors had a sub-optimal response (4–9 oocytes), while 27.27% showed a hyper-response (over 20 oocytes).

The best predictive model was random forest, with an AUC of 0.822. Six significant genetic polymorphisms were identified: three in hormone receptors—oestrogen receptor (ESR2; c.*39G > A, c.984G > A), follicle-stimulating hormone receptor (FSHR; p.Asn680Ser, c.-29G > A), and AMH receptor (AMHR2; c.622-6C > T) and one in growth differentiation factor 9 (GDF9; c.398-39G > C). Four polymorphisms (ESR2, FSHR) were linked to sub-optimal response, while two (AMHR2, GDF9) were associated with hyper-response.

Conclusions By using a predictive model to asses ovarian response, we identified six genetic polymorphisms associated with ovarian response. Women who carry these genetic variants may be suitable candidates for personalised ovarian stimulation treatments to help prevent inadequate responses.

Keywords Sub-optimal ovarian response · Hyper-response · Polymorphisms · Machine learning · SHAP value

Introduction

Controlled ovarian stimulation (COS) is the first step in assisted reproduction treatments (ART). The goal of COS is to promote the simultaneous maturation of multiple follicles through the administration of exogenous gonadotropins. The number of oocytes retrieved after COS is a critical determinant of the success of assisted reproductive treatments [1]. Generally, a higher number of oocytes, and consequently embryos, enhance the likelihood of achieving a successful pregnancy.

A substantial proportion of women who would be expected to have a normal ovarian response ("normo-responders") have a "sub-optimal" oocyte recovery range,

✉ José A. Ortiz
jaortiz@institutobernabeu.com

1 Molecular Biology, Instituto Bernabeu, Avda. Albufereta 31, 03016 Alicante, Spain

2 Reproductive Biology, Instituto Bernabeu, Alicante, Spain

3 Reproductive Medicine, Instituto Bernabeu, Alicante, Spain

4 Reproductive Medicine, Instituto Bernabeu, Albacete, Spain

5 Cátedra de Medicina Comunitaria y Salud Reproductiva, Miguel Hernández University, Alicante, Spain

Published online: 02 April 2025

between 4 and 9 oocytes, to the detriment of treatment success [2, 3], or, conversely, an exaggerated response ("hyper-responders") causing discomfort, or more severe clinical complications associated with ovarian hyperstimulation syndrome (OHSS) [4] and without improving live birth rates [5, 6].

Pharmacogenetics is the study of the response to drugs that everyone has according to their genotype. Its aim is to optimize and personalise treatment in the safest and most effective way. Pharmacogenetics is being applied to ovarian stimulation with the aim of using the most appropriate protocol for the patient's genotype [7, 8]. This is why pharmacogenetics has become important in the field of reproductive medicine [9]. In this way, the genetic study of patients confers added value when it comes to predicting ovarian response and thus achieving safer and more effective treatment [10, 11]. In recent years, many genetic variants associated with ovarian response have been described and characterised [12].

Artificial Intelligence (AI) has facilitated the development of various applications in medicine that are highly valuable for diagnosis and treatment optimization. Its impact is particularly notable in reproductive medicine, where the vast amount of data collected enables the creation of models to optimize every stage of assisted reproduction treatment, ultimately aiming to achieve a successful pregnancy. AI has also been applied to ovarian stimulation to optimize the trigger day [13–15], the starting dose of stimulation [16–20], and the most appropriate stimulation protocol [21] and predict the number of oocytes retrieved [22–26]. Interestingly, only one article uses the genotype of patients undergoing ovarian stimulation as a predictor [27].

In this context, we conducted a study aimed at identifying genetic variants associated with an increased likelihood of both types of abnormal ovarian response (sub-optimal and hyper-response) using AI tools.

Material and methods

Study design

A prospective observational cohort study was conducted, including a total of 495 egg donors from March 2018 to April 2022. Only the first oocyte donor ovarian stimulation cycles were included in the study. This patient population serves as an excellent model for assessing the impact of different genetic variants, as it comprises individuals of similar age (18–35 years) and normal ovarian function (Table 1). Oocyte donor candidates were selected based on the criteria established by our clinic's donation program and the ASRM and ESHRE guidelines for oocyte donation. This selection process included a comprehensive physical, psychological, gynaecological, and fertility evaluation, as well as complete blood analyses and chromosomal and genetic studies, including an expanded carrier screening.

Ovarian stimulation in oocyte donors

The primary protocol used was progesterone-primed ovarian stimulation (PPOS) and a fixed daily regimen of GnRH antagonists starting on day 5 of stimulation. A smaller group of patients followed a short antagonist protocol. The gonadotrophin used was of recombinant origin in all treatments. The initial gonadotropin dose was selected to optimize

Table 1 Descriptive of patients and COS

Characteristic	Overall[1] N=495	Sub-optimal[1] (4–9 oocytes) N=75 (15.15%)	Normal[1] (10–20 oocytes) N=285 (57.58%)	Hiper-response[1] (>20 oocytes) N=135 (27.27%)	p-value[2]
Age (years)	25.3 (22.3, 28.9)	26.9 (22.7, 29.5)	25.0 (22.2, 29.1)	25.3 (21.8, 28.0)	0.2
AFC	19 (14, 24)	14 (12, 17)	18 (14, 23)	23 (18, 30)	<0.001
BMI (kg/m^2)	22.70 (20.46, 24.86)	23.23 (20.85, 25.72)	22.76 (20.41, 24.89)	22.50 (20.35, 24.18)	0.10
Stimulation protocols					0.040
Short antagonist	20 (4.0%)	4 (5.3%)	15 (5.3%)	1 (0.7%)	
Progesterone-primed (PPOS)	475 (96%)	71 (95%)	270 (95%)	134 (99%)	
Stimulation days	10.00 (9.00, 11.00)	10.00 (8.00, 11.00)	9.00 (8.00, 10.00)	10.00 (9.00, 11.00)	0.3
Gonadotrophin dose (IU)	2025 (1613, 2475)	2400 (2100, 3000)	2025 (1800, 2500)	1800 (1425, 2063)	<0.001
Oocytes retrieved	16 (11, 21)	8 (6, 8)	15 (12, 17)	26 (23, 31)	<0.001
MII retrieved	12 (9, 17)	6 (4, 8)	11 (9, 14)	20 (18, 25)	<0.001

1 Median (IQR)/n (%)

2 Kruskal–Wallis rank sum test; Pearson's chi-squared test

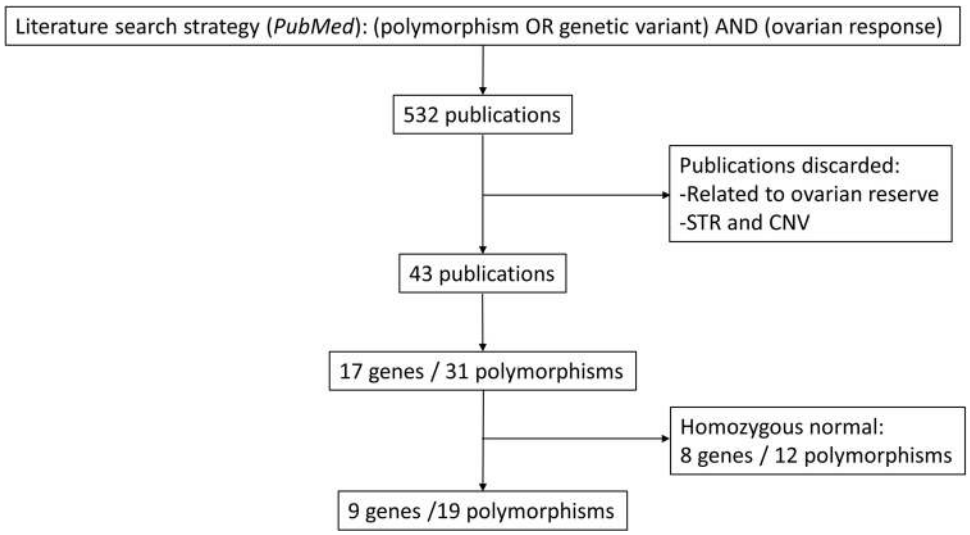

Fig. 1 Literature search strategy. Search strategy and results for literature review of published works

follicular recruitment while minimising the risk of a high response. In summary, the recommended optimal dose was 150 IU for donors with an antral follicle count (AFC) greater than 14, while a dose of 225 IU was considered appropriate for donors with 10 to 14 antral follicles. For donors with fewer than 10 follicles, a dose of 300 IU was prescribed. It is important to note that, at the clinician's discretion, these doses could be adjusted based on the donor's BMI. Donors were monitored via transvaginal ultrasound every 2 to 3 days starting from day 5 or 6 of stimulation.

Final follicular maturation was induced using a GnRH agonist (0.2 mg) when at least three follicles greater than 17 mm in diameter were observed. Follicular aspiration was performed 36 h later via ultrasound-guided transvaginal needle aspiration. Cycles cancelled due to low ovarian response were excluded from the study, as were ovarian stimulations from which fewer than four oocytes were collected. A database was created to collect different variables from the stimulation protocols, which were then used as predictors in machine learning models.

Literature search strategy

A literature review was conducted in search of genetic polymorphisms associated with ovarian response. The PubMed database (https://pubmed.ncbi.nlm.nih.gov/) was consulted by applying the search strategies (polymorphism OR genetic variant) AND (ovarian response), and 532 publications were identified. Only English-language transcripts were reviewed. This search was last updated in May 2022. Forty-three articles were finally selected. We discarded those that were only related to ovarian reserve and not to ovarian response and those variants that could not be analysed by NGS such as STRs and CNVs. The complete review protocol is outlined in (Fig. 1). The polymorphisms and genes selected are described in Table 2.

Whole-exome sequencing and genomic variant identification

Genomic DNA was extracted from peripheral blood using the MagMAX DNA Multi-Sample Ultra 2.0 kit (Thermo Fisher Scientific, Colchester, UK) on a KingFisher™ Duo Prime system (Thermo Fisher Scientific, Colchester, UK), following the manufacturer's instructions. Whole exome sequencing (WES) of genomic DNA was performed using the Trusight One expanded sequencing panel (Illumina®). Sequenced data were aligned to human genome 19 (hg19) for identification of single-nucleotide polymorphisms (SNP) with the help of a bioinformatics application (*Variant Interpreter*-Illumina-).

Machine learning predictive model variables

The predictors used in the different models were the donor's characteristics and their ovarian stimulation, as well as the genotypes of the polymorphisms previously identified in the literature as modulators of ovarian response.

The response variable used in the predictive models was the number of oocytes retrieved, but divided into three categories (multinomial). Ovarian responses were considered normal in the range of 10 to 20 oocytes retrieved. Ovarian hyper-response was considered hyper-response if the number of oocytes retrieved was greater than 20 and sub-optimal between 4 and 9.

Table 2 List and description of the polymorphisms selected as predictors

Gene	rs ID (SNP identifier)	DNA change	Protein change	Reference
AMH	rs10407022	146G > T	Ser49Ile	(Cerra et al., 2016; Karagiorga et al., 2015)
AMHR2	rs2002555	−482A > G	-	(Cerra et al., 2016; Karagiorga et al., 2015; Lazaros et al., 2016)
	rs2071558	622-6C > T	-	
BMP15	rs58995369	−673C > T	-	(Hanevik et al., 2011; Morón et al., 2006)
	rs3810682	−9C > G	-	
	rs3897937	328 + 905A > G	-	
CYP19A1	rs10046	*19C > T	-	(Binder et al., 2008; Song et al., 2019)
ESR1	rs2234693	453-397 T > C	-	(Altmäe et al., 2007; Ayvaz et al., 2009; de Mattos et al., 2014; Lledó et al., 2019)
	rs9340799	453-351A > G	-	
ESR2	rs1256049	984G > A	Val328 =	(Altmäe et al., 2007; de Mattos et al., 2014)
	rs4986938	*39G > A	-	
FSHB	rs10835638	−281G > T	-	(Trevisan et al., 2019)
FSHR	rs6166	2039A > G	Asn680Ser	
	rs6165	919A > G	Thr307Ala	
	rs1394205	−29G > A	-	(Desai et al., 2013; Lledó et al., 2016; Motawi et al., 2017)
GDF9	rs10491279	546G > A	Glu182 =	(Bilibio et al., 2020; Lledó et al., 2019)
	rs254286	447C > T	Thr149 =	
	rs254285	398-39G > C	-	
HRG	rs9898	610C > T	Pro204Ser	(Bakay et al., 2022; Nordqvist et al., 2015)
LHB	rs1800447	82 T > C	Trp28Arg	(Alviggi et al., 2011)
	rs34349826	104 T > C	Ile135Thr	
LHCGR	rs4073366	161 + 28G > C	-	(G A et al., 2018; Lledó et al., 2019; O'Brien et al., 2013; Yin et al., 2018)
	rs2293275	935A > G	Asn312Ser	
	rs13405728	161 + 4491 T > C	-	
PGR	rs10895068	−413G > A	-	(Ghaderian et al., 2019)
SOD2	rs4880	47 T > C	Val16Ala	(Ruiz-Sanz et al., 2011)
TERT	rs2075786	2654 + 269 T > C	-	(Dai et al., 2019)
	rs2853677	1574-4455C > T	-	
	rs2853691	*893A > G	-	
TP53	rs1042522	215C > G	Arg72Pro	(Boudjenah et al., 2012)
TP73	rs4648551	430-5855G > A	-	(Bakay et al., 2022)

Univariate analysis of variables

This analysis will depend on the type of characteristic. Qualitative variables are described by frequency and percentage. In the case of quantitative variables, the descriptive analysis has been carried out using the median and the interquartile range.

For the univariate statistical analysis of qualitative variables, the chi-square test or Fisher's exact test will be used. For evaluation of normal distributions, the Shapiro–Wilk's test was performed. Depending on whether the variable has a normal distribution, the comparison between means was carried out using ANOVA test or Kruskal–Wallis rank sum test. Values of $p < 0.05$ will be considered statistically significant.

Data preprocessing

Before starting the analysis, the database was anonymised. Only 0.06% of the data was missing and imputed. The missing data were imputed by the MICE algorithm—***M**ultiple **I**mputation by Chained **E**quations* [28]. In our case, we chose the imputation using the classification trees (cart) among the different options of the algorithm.

No outliers were detected in the database. In the case of highly correlated variables, only one of them was kept in the predictive model. On the other hand, variables with almost

zero variance were not included due to their low predictive value.

Before training the models, class balancing was performed for the variables to be predicted. Before training the different algorithms, the database was randomly divided into a training set (80% of the database) and a test set (20% of the database).

Hyperparameter optimisation of classification models

Five machine learning (classification) algorithms, including support vector machines, *k*-nearest neighbours, random forest, multilayer neural networks and eXtreme Gradient Boosting (XGBoost), were applied.

To guarantee the independence of the data and to be able to properly evaluate the models, the fivefold cross-validation technique with adjustment of the different hyperparameters was applied. During the optimisation process, different performance metrics were calculated: AUC (area under the ROC curve), mean-sensitivity, mean-specificity, and accuracy.

The ROC curve was computed for each class of the response variable, and the overall mean was calculated using a so-called macro method, which involves averaging the results of all groups (one versus the rest) through linear interpolation between the points of the ROC curves.

Final predictive model

The best model was selected based primarily on the AUC, a parameter that measures the model's ability to discriminate the dependent variable.

Important variables: SHAP values

The key predictor variables in the final model were identified using SHAP values, which help explain the outcomes of machine learning models. The theoretical foundation of SHAP values is rooted in game theory [29, 30]. Machine learning algorithms assign a SHAP value to each predictor for each instance, indicating the variable's contribution to the final prediction. This value serves as a metric to assess whether a variable's impact on the prediction is positive or negative. One of the notable features of SHAP values is their additivity, which allows for the decomposition of a model's prediction into the sum of the individual SHAP values for each variable.

Different types of graphs using SHAP values provide valuable insights for interpreting machine learning models:

1. **Top influencers**: This graph displays the ten most significant predictors in the model, calculated based on the mean of the absolute SHAP values for each variable. It highlights the features that have the greatest impact on the model's predictions.
2. **Impact of directionality**: This graph illustrates how various factors influence the model's predictions. It includes a summary plot that shows the relationship between the identified polymorphisms (colour-coded: yellow for homozygous/alternative heterozygous, purple for wild-type homozygous) and their impact on the prediction. Positive values for a given variant are associated with an increased risk of ovarian hyper-response and negative values with a predisposition to sub-optimal response.

Multivariate logistic regression

To quantitatively evaluate the contribution of the selected predictor variables, a logistic regression analysis was conducted using the ten best variables identified by the machine learning models with the highest AUC values. Values of $p < 0.05$ will be considered statistically significant.

Statistical and machine learning analysis was carried out using SPSS (v23.0) and R (v. 4.2.0) statistical software.

Results

Baseline characteristics

A retrospective, observational study was conducted involving the first ovarian stimulation of 495 oocyte donors (March 2018–April 2022). The variables age, body mass index and antral follicle count, as well as the type of protocol, the number of days of treatment and the type and dose of gonadotrophins administered were recorded and used as predictor variables in the different machine learning models. The oocyte donors had a median age of 25.3 (IQR: 22.3–28.9) and a high ovarian reserve (median AFC: 19.0; IQR: 14–24). The majority of the ovarian stimulation protocols were progesterone-primed (96.0%) followed by short-antagonist (4.0%). Ovarian stimulation lasted for a median of 10 days (IQR: 9–11) with a median gonadotropin consumption of 2025 IU (IQR: 1750–247). The median number of oocytes retrieved after ovarian stimulation was 15.0 (IQR: 11–21), of which 12.0 (IQR: 9–17) were mature (Table 1).

The response variable used in the predictive models was the number of oocytes retrieved, but categorised. 57.58% of the responses were classified as normal (10–20 oocytes), 27.27% hyper-response (greater than 20 oocytes) and only 15.15% sub-optimal (4–9 oocytes).

The comparison of predictor variable values between all categories is presented in Table 1. No significant differences were observed in patient ages ($p = 0.2$), BMI ($p = 0.1$), or days of stimulation ($p = 0.3$). However, there were logical changes in the antral follicle count (AFC): sub-optimal (14.0), normal (18.0), and hyper-response (23.0; $p < 0.001$). Additionally, the stimulation protocol showed that 99.0% of stimulations in the hyper-response group were PPOS ($p = 0.040$), and the gonadotropin dose decreased with increasing ovarian response: sub-optimal (2400 IU), normal (2025 IU), and hyper-response (1800 IU; $p < 0.001$).

Literature search

A literature review was conducted in search of genetic polymorphisms associated with ovarian response following the search criteria detailed in Materials and methods. This search was last updated in May 2022. The complete review protocol is outlined in (Fig. 1). The polymorphisms and genes selected are described in Table 2. These genes code for proteins are hormones or hormone receptors, proteins involved in folliculogenesis, cell cycle, DNA damage and detoxification.

Genotyping

The donor genotype was determined for all selected polymorphisms. Supplementary Table 1 shows the genotyping results of the 31 selected variants. For 12 of the polymorphisms (8 genes), the genotype of all donors analysed was wild type homozygous. These variants were not included as predictors in the AI models because they lacked variability.

The remaining 19 polymorphisms were used as predictor variables in the machine learning models along with variables related to donor characteristics and ovarian stimulation protocols.

Prediction model

Five different supervised classification machine learning algorithms (multi-layer perceptron, support vector-machines, *k*-nearest neighbours, random forest, and eXtreme Gradient Boosting (XGBoost)) were used to establish a prediction model ovarian response (three categories: Sub-optimal, normal and hyper-response).

The criterion for selecting the best predictive model was the area under the ROC curve (AUC). The model with the lowest AUC value was the multilayer neural network (AUC = 0.698). The rest of the models had an AUC value above 0.75 (*k*-nearest neighbours (AUC = 0.778), and the support vector machines (AUC = 0.784) and eXtreme Gradient Boosting (XGBoost) (AUC = 0.797). Finally, the best model with the highest AUC value was random forest (AUC = 0.822) (Table 3). This predictive model maximises not only the value of AUC but also that of the other model performance parameters analysed, such as mean sensitivity (0.603), mean specificity (0.802) and accuracy (0.603). All these model performance parameter values were obtained from the test dataset (20% of the initial data). Therefore, these data were not used to train the models, so they are new and unknown data for the different models.

Model interpretation: SHAP values

The most important prediction variables of the random forest model were determined from the SHAP values. Figure 2A shows the 10 most relevant predictors ("Top influencers") of our best model and their corresponding mean SHAP values. The 4 most important variables are "Antral follicle count," "Gonadotrophin dose," "Body mass index (BMI)" and "Female age." On the other hand, 6 genetic variants associated with the risk of a sub-optimal or ovarian hyper-response have been identified, corresponding to the following genes: oestrogen receptor 2 (*ESR2*), anti-Mullerian hormone receptor type 2 (*AMHR2*), follicle-stimulating hormone receptor (*FSHR*) and growth differentiation factor 9 (*GDF9*).

In Fig. 2B ("Directionality impact"), we can see how the identified factors influence the ovarian response using the SHAP value. Four polymorphisms are associated with sub-optimal ovarian response (the alternative allele has negative values of SHAP): the variants c.*39G > A (rs4986938) and c.984G > A (rs1256049) of the *ESR2* gene, along with the p.Asn680Ser (rs6166) and c.−29G > A (rs1394205) polymorphisms of the *FSHR* gene. Two of the identified genetic variants predispose individuals to ovarian hyper-response

Table 3 Comparison between the different metrics of the final models

Model	AUC	Mean Sensitivity	Mean Specificity	Accuracy
Multi-layer perceptron	0.698	0.490	0.745	0.490
Support vector machines with radial basis function Kernel	0.784	0.585	0.793	0.586
k-nearest neighbours	0.778	0.550	0.775	0.551
Random forest	0.822	0.603	0.802	0.603
eXtreme Gradient Boosting (XGBoost)	0.797	0.589	0.794	0.588

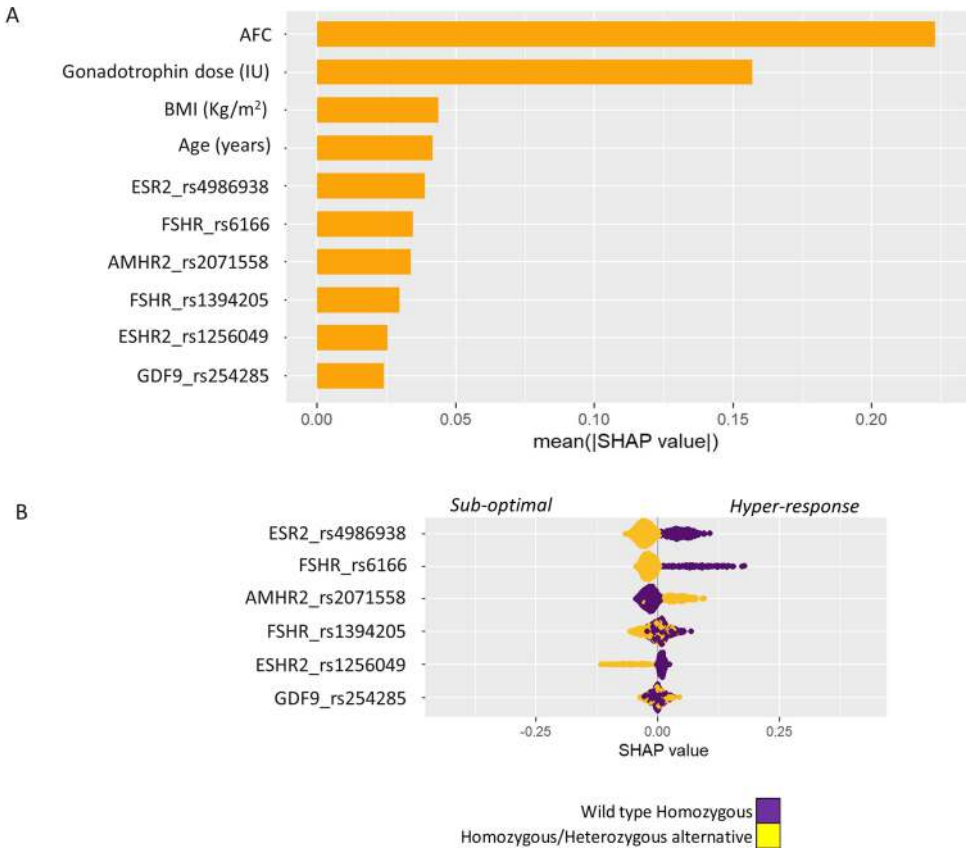

Fig. 2 SHAP (SHapley Additive exPlanations) plots. **A** ***Top Influencers graph*** shows the ten most important predictors for prediction in the random forest model, calculated from the mean of the absolute SHAP values for each feature. **B** ***Directionality impact graph***. The *x*-axis represents the SHAP value, and the y-axis contains the polymorphisms ordered according to their influence on the model prediction (random forest). Each point on the graph is a SHAP value for one prediction and one feature. Yellow indicates the highest value of the feature (homozygous/heterozygous alternative). Purple indicates the lowest value of the feature (wild type homozygous). The distribution of the yellow and purple points distribution gives a general idea of the features directionality impact. Positive values for a given variant are associated with an increased risk of ovarian hyper-response, while negative values indicate a predisposition to sub-optimal response

(the alternative allele has positives values of SHAP): the c.622-6C > T (rs2071558) variant of the *AMHR2* gene and the c.398-39G > C (rs254285) variant of the *GDF9* gene. All identified genetic variants exhibited dominant behaviour as a single copy of the alternative allele was sufficient to observe their effect.

The SHAP values provided a qualitative assessment of the most relevant variables for predicting ovarian response in our patients. However, to quantitatively evaluate the contribution of polymorphisms to the risk of experiencing either a sub-optimal response or hyper-response, we conducted a multivariate logistic regression using the ten best predictors from the random forest model, which included six genetic polymorphisms Fig. 3. The variables associated with a sub-optimal response correspond to the *ESR2* and *FSHR* genes. All four variants exhibited odds ratios (OR) less than one, with only two reaching statistical significance: *ESR2* (c.*39G > A; rs4986938): OR 0.51 (95% CI: 0.31–0.81; $p = 0.005$) and *FSHR* (p.Asn680Ser; rs6166): OR 0.29 (95% CI: 0.15–0.55; $p < 0.001$). For ovarian hyper-response, only the *AMHR2* polymorphism (c.622-6C > T; rs2071558) demonstrated statistical significance, with an odds ratio (OR) of 2.03 (95% CI: 1.25–3.35; $p = 0.005$). In contrast, the *GDF9* gene variant (c.398-39G > C; rs254285) did not show statistical significance in the multivariate logistic regression model, with an OR of 0.89 (95% CI: 0.57–1.40; $p = 0.620$) (Fig. 3).

Discussion

Precision medicine tailors the diagnosis, treatment, and prevention of diseases to individual characteristics, with a patient's genetic profile playing a crucial role. Genetic variants can influence drug responses, and this approach is also applied in reproductive medicine, where ovarian stimulation protocols are sometimes designed based on the patient's genetics [31, 32].

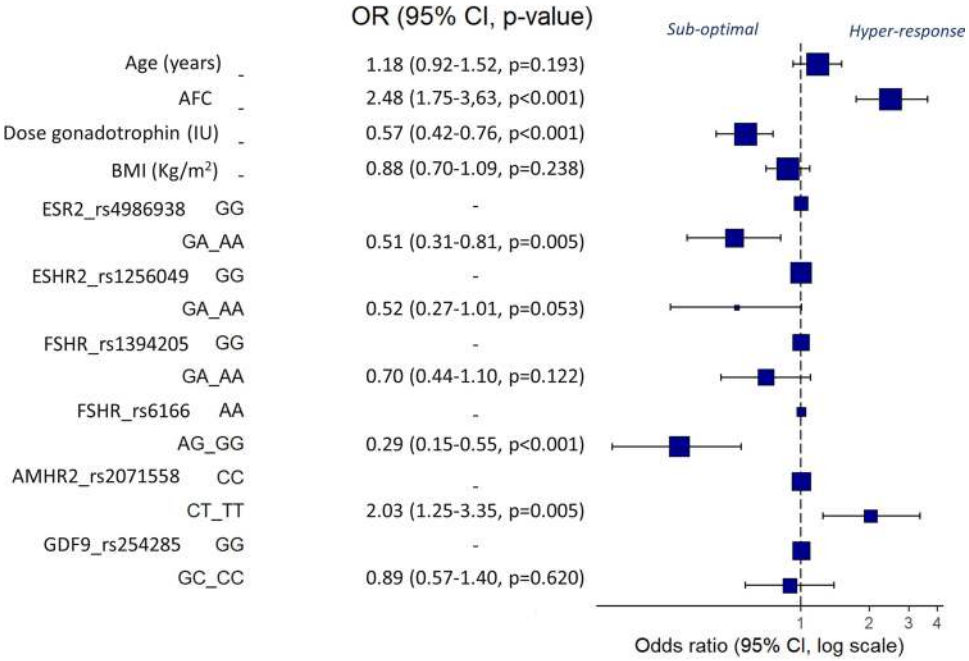

Fig. 3 Multivariate logistic regression. Logistic regression analysis was conducted using the ten most important predictors in the random forest model. The Odds ratio (OR) and its 95% confidence interval (CI) are plotted for each variable. OR greater than one is associated with an ovarian hyper-response and less than one with a sub-optimal ovarian response

This study aims to develop a predictive model for ovarian response using data from oocyte donors, their ovarian stimulations, and genetic information. Our study focuses on young patients with good ovarian reserve undergoing similar stimulation protocols, though some experience sub-optimal or high responses, posing associated risks. We have decided that the variable to predict will be the categorised number of oocytes retrieved (sub-optimal, normal, and hyper-response), rather than the number of mature oocytes (MII). The goal is to identify polymorphisms associated specifically with ovarian response. If we were to use mature oocytes as the response variable, the results could be influenced by genetic variants involved in oocyte maturation, potentially distorting our objective, which is focused on ovarian response, not maturation. To investigate the genetic factors related to oocyte maturation could be a valuable objective for future studies.

Several studies have been described in the literature to optimize and/or predict multiple aspects of ovarian stimulation using AI but only one of them takes into account patients' genetic profiles as predictors [27]. These authors developed a machine-learning-based predictive model for ovarian response, using both clinical characteristics and genetic information (specifically, 22 polymorphisms in 7 genes) derived from a dataset of 516 ovarian stimulations. The contribution of genotype to the prediction is quantified and represents one third of the best predictor, which is AMH levels. The difference with our predictive model is mainly that we have developed a classification model since the response variable is a categorical variable with 3 classes (sub-optimal ovarian response, normal or hyper-response), whereas Zieliński et al. have developed a regression model in which the response variable is the number of mature oocytes.

Our study is a pioneering analysis of factors associated with ovarian response, including genotype, using AI. Unlike traditional approaches that aim to create predictive models through machine learning, our primary objective was to identify the most significant genotypes influencing ovarian response in young patients with good ovarian reserve and to understand their effects. We aimed to uncover key genetic factors associated with ovarian stimulation in patients expected to be normo-responders, utilising machine learning techniques to reveal relationships that traditional methods might overlook.

We employed five diverse machine learning algorithms, encompassing neural networks, bagging, boosting, and *k*-nearest neighbours, for a comprehensive analysis. Among these, the random forest algorithm demonstrated superior performance, achieving an area under the curve (AUC) of 0.822 along with other high-quality performance metrics. This algorithm, introduced by [33], generates multiple decision trees trained on random subsets of samples drawn with replacement.

Machine learning models exhibit remarkable accuracy and strong predictive capabilities. However, one significant drawback is their resemblance to black boxes, wherein the roles of input variables in generating predictions remain obscure. The interpretability of models holds paramount importance, particularly in life sciences. SHAP (SHapley Additive exPlanations) values have emerged as a valuable

tool for unravelling model intricacies at the feature level [29, 30].

In the context of the random forest prediction model, the four most important variables calculated based on the mean of the absolute SHAP values of each feature are antral follicle count (AFC), gonadotrophin dose administered during ovarian stimulation, female age and BMI. These findings align with those from other COS predictive models, such as those reported by Zieliński et al., where markers of ovarian reserve like AMH and AFC on the day of stimulation, along with maternal age, are identified as main predictors. Zieliński et al. also incorporate data on patients' prior ovarian stimulations, including the number of cumulus denuded and mature oocytes, as additional predictive variables.

Among the ten variables with the highest predictive power, six were genetic polymorphisms. These included five in hormone receptors: the oestrogen receptor (*ESR2*; c.*39G > A and c.984G > A), *FSHR* (p.Asn680Ser and c.−29G > A) and AMHR2 (c.622-6C > T) and one in a protein involve in folliculogenesis c.398-39G > C variant of the *GDF9* gene.

The c.*39G > A (rs4986938) variant in the *ESR2* gene requires only one alternative allele to produce the effect (dominant model). The findings presented are consistent with previous research documented in the literature. Specifically, the reference allele G, in contrast to the alternative allele A of the c.*39G > A variant in the *ESR2* gene, is linked to an enhanced yield of oocytes COS and an elevated risk of ovarian hyperstimulation syndrome, as detailed by de Mattos et al. (2014).

Previous studies on the *ESR2* c.984G > A (rs1256049) polymorphism have mainly indicated that the G allele is associated with a higher gonadotropin dose requirement during stimulation [34, 35]. In contrast, our findings demonstrate that individuals with the GA or AA genotypes (dominant effect) are predisposed to a sub-optimal response. There may be several reasons for these differences: the number of patients analysed in our study is higher (495 vs. 136) with a lower median age (25.3 vs. 33). In addition, the machine learning analysis method can identify behaviours that are not identified by conventional statistical analysis.

The p.Asn680Ser (rs6166) variant in the *FSHR* gene is the most studied and well-characterised polymorphism related to ovarian response. FSH is a critical factor in human reproduction, playing a pivotal role, along with its receptor (FSHR), in follicular development and the regulation of steroidogenesis within the ovary (Dupakuntla and Mahale, 2010). Clinical studies have revealed that the p.Asn680Ser (rs6166) polymorphism influences ovarian response to FSH stimulation in patients undergoing IVF treatment (Altmäe et al., 2011; Laan et al., 2012; Yao et al., 2011). Individuals with the Asn/Asn genotype at position 680 require a lower amount of FSH for COS. Conversely, patients with the Ser680 allele require higher doses of FSH during the stimulation phase, indicating reduced sensitivity to exogenous FSH [36, 37]. In our machine learning model this *FSHR* polymorphism appears among the top 10 predictors as do other predictive models of ovarian response [27]. Our findings regarding the p.Asn380Ser polymorphism of the *FSHR* align with previous studies. In our analysis, this variant is associated with a predisposition to sub-optimal ovarian response in a dominant fashion, as demonstrated by both the SHAP value analysis and the multivariate logistic regression model (OR 0.29; 95% CI: 0.15–0.55; $p < 0.001$).

The c.−29G > A (rs1394205) variant of *FSHR* is associated with a sub-optimal response, as shown in both the present study and previous research, where lower stimulation efficiency is observed despite increased gonadotropin consumption [38]. This variant is located in the regulatory region of the gene and it appears that the A variant causes a reduction in *FSHR* expression [39, 40] which may explain the lower ovarian response of carriers.

Two of the identified polymorphisms predispose individuals to ovarian hyper-response (positive SHAP values). The variant c.622-6C > T (rs2071558) of *AMHR2* show a dominant effect on ovarian hyper-response. The *AMHR2* variant, the result is consistent with previous work by Lazaros et al. 2016, where the c.622-6C > T variant is associated with increased ovarian response. In the same sense, previous studies have demonstrated that the c.398-39G > C (rs254285) variant of the *GDF9* gene is associated with a high ovarian response [41, 42]. Our findings support this observation. However, this effect is no longer evident in the logistic regression model. The variant is ranked tenth in the predictor ranking, suggesting that its contribution to ovarian response may be relatively minor.

This study has identified genetic variants that are relevant to ovarian stimulation in young women with a good ovarian reserve. Expanding the analysis to include additional polymorphisms could be valuable, especially by identifying new genetic variants associated with ovarian response through various biostatistical methods. These predictive models are promising; once validated in prospective studies, they could significantly enhance stimulation protocols and personalise treatment for specific groups of female patients. This would be particularly beneficial for patients with low ovarian response, where maximising the effectiveness of ovarian stimulation is crucial for the success of assisted reproduction treatments.

A machine learning model has been established for the prediction of sub-optimal response and hyper-response in young patients with good ovarian reserve. In this model, six genetic polymorphisms stand out as predictors of ovarian response. Patients carrying these genetic variants are

candidates for personalised ovarian stimulation treatment to avoid inadequate responses.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10815-025-03471-z.

Acknowledgements We would like to thank Arancha Estañ for her valuable help in the preparation and revision of the English manuscript.

Author contribution SM and MP: collection of data and critical review of article. BL, JG, and RM: study conception, design, interpretation of the data and critical review of article. LL: study conception, recruitment of patients, interpretation of the data and critical review of article. AB: recruitment of patients, interpretation of the data and critical review of article. J.A.O.: study conception, design, interpretation, and data analysis and writing the article.

Data availability Data will be made available on request to the authors.

Declarations

Ethics approval The data included in this study were within the framework of routine clinical activity. All work was conducted with formal approval of the Institutional Review Board, and it follows the principles of the Declaration of Helsinki.

Conflict of interest The authors declare no competing interests.

References

1. Vaughan DA, Leung A, Resetkova N, Ruthazer R, Penzias AS, Sakkas D, et al. How many oocytes are optimal to achieve multiple live births with one stimulation cycle? The one-and-done approach. Fertil Steril. 2017;107:397-404.e3.
2. Abu-Musa A, Haahr T, Humaidan P. Novel physiology and definition of poor ovarian response; clinical recommendations. Int J Mol Sci. 2020;21:2110.
3. Polyzos NP, Sunkara SK. Sub-optimal responders following controlled ovarian stimulation: an overlooked group? Hum Reprod. 2015;30:2005–8.
4. Manau D, Balasch J, Arroyo V, Jiménez W, Fábregues F, Casamitjana R, et al. Circulatory dysfunction in asymptomatic in vitro fertilization patients. Relationship with hyperestrogenemia and activity of endogenous vasodilators. J Clin Endocrinol Metab. 1998;83:1489–93.
5. Steward RG, Lan L, Shah AA, Yeh JS, Price TM, Goldfarb JM, et al. Oocyte number as a predictor for ovarian hyperstimulation syndrome and live birth: an analysis of 256,381 in vitro fertilization cycles. Fertil Steril. 2014;101:967–73.
6. Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. Hum Reprod. 2011;26:1768–74.
7. Lledo B, Llácer J, Ortiz JA, Martinez B, Morales R, Bernabeu R. A pharmacogenetic approach to improve low ovarian response: the role of CAG repeats length in the androgen receptor gene. Eur J Obstet Gynecol Reprod Biol. 2018;227:41–5.
8. Lledó B, Llácer J, Turienzo A, Ortiz JA, Guerrero J, Morales R, et al. Androgen receptor CAG repeat length is associated with ovarian reserve but not with ovarian response. Reprod Biomed Online. 2014;29:509–15.
9. Loutradis D, Vlismas A, Drakakis P, Antsaklis A. Pharmacogenetics in ovarian stimulation–current concepts. Ann N Y Acad Sci. 2008;1127:10–9.
10. Alviggi C, Humaidan P, Ezcurra D. Hormonal, functional and genetic biomarkers in controlled ovarian stimulation: tools for matching patients and protocols. Reprod Biol Endocrinol. 2012;10:9.
11. Conforti A, Esteves SC, Iorio GG, Carbone L, Alviggi C. Pharmacogenomic approach in controlled ovarian stimulation: time to take the plunge? Gynecol Obstet Invest. 2022;87:173–5.
12. Lledo B, Ortiz JA, Llacer J, Bernabeu R. Pharmacogenetics of ovarian response. Pharmacogenomics. 2014;15:885–93.
13. Fanton M, Nutting V, Solano F, Maeder-York P, Hariton E, Barash O, et al. An interpretable machine learning model for predicting the optimal day of trigger during ovarian stimulation. Fertil Steril. 2022;118:101–8.
14. Hariton E, Chi EA, Chi G, Morris JR, Braatz J, Rajpurkar P, et al. A machine learning algorithm can optimize the day of trigger to improve in vitro fertilization outcomes. Fertil Steril. 2021;116:1227–35.
15. Reuvenny S, Youngster M, Luz A, Hourvitz R, Maman E, Baum M, et al. An artificial intelligence based approach for selecting the optimal day for triggering in antagonist protocol cycles. Reprod Biomed Online. 2024;48(1):103423.
16. Correa N, Cerquides J, Arcos JL, Vassena R, Popovic M. Personalizing the first dose of FSH for IVF/ICSI patients through machine learning: a non-inferiority study protocol for a multi-center randomized controlled trial. Trials. 2024;25:38.
17. Correa N, Cerquides J, Arcos JL, Vassena R. Supporting first FSH dosage for ovarian stimulation with machine learning. Reprod Biomed Online. 2022;45:1039–45.
18. Fanton M, Nutting V, Rothman A, Maeder-York P, Hariton E, Barash O, et al. An interpretable machine learning model for individualized gonadotrophin starting dose selection during ovarian stimulation. Reprod Biomed Online. 2022;45:1152–9.
19. Guo X, Zhan H, Zhang X, Pang Y, Xu H, Zhang B, et al. Predictive models for starting dose of gonadotropin in controlled ovarian hyperstimulation: review and progress update. Hum Fertil (Camb). 2023;26(6):1609–1616.
20. Hua L, Zhe Y, Jing Y, Fujin S, Jiao C, Liu L. Prediction model of gonadotropin starting dose and its clinical application in controlled ovarian stimulation. BMC Pregnancy Childbirth. 2022;22:810.
21. Letterie G, Mac DA. Artificial intelligence in in vitro fertilization: a computer decision support system for day-to-day management of ovarian stimulation during in vitro fertilization. Fertil Steril. 2020;114:1026–31.
22. Ferrand T, Boulant J, He C, Chambost J, Jacques C, Pena C-A, et al. Predicting the number of oocytes retrieved from controlled ovarian hyperstimulation with machine learning. Hum Reprod. 2023;38(10):1918–1926.
23. Garg A, Bellver J, Bosch E, Remohí JA, Pellicer A, Meseguer M. Machine learning tool to predict number of metaphase II oocytes and trigger day at the start of ovarian stimulation cycle: step towards personalized ovarian stimulation treatment. Reprod BioMed Online. 2025;50(2):104441.
24. Hariton E, Pavlovic Z, Fanton M, Jiang VS. Applications of artificial intelligence in ovarian stimulation: a tool for improving efficiency and outcomes. Fertil Steril. 2023;120:8–16.
25. Liu L, Shen F, Liang H, Yang Z, Yang J, Chen J. Machine learning-based modeling of ovarian response and the quantitative evaluation of comprehensive impact features. Diagnostics (Basel). 2022;12:492.
26. Yan S, Jin W, Ding J, Yin T, Zhang Y, Yang J. Machine-intelligence for developing a potent signature to predict

ovarian response to tailor assisted reproduction technology. Aging. 2021;13:17137–54.
27. Zieliński K, Pukszta S, Mickiewicz M, Kotlarz M, Wygocki P, Zieleń M, et al. Personalized prediction of the secondary oocytes number after ovarian stimulation: a machine learning model based on clinical and genetic data. PLoS Comput Biol. 2023;19:e1011020.
28. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30:377–99.
29. Arrow KJ, Barankin EW, Blackwell D, Bott R, Dalkey N, Dresher M, et al. Contributions to the theory of games (AM-28), Volume II . Princeton University Press; 1953.
30. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst. 2014;41:647–65.
31. Hajirasouliha I, Elemento O. Precision medicine and artificial intelligence: overview and relevance to reproductive medicine. Fertil Steril. 2020;114:908–13.
32. Sigman M. Introduction: personalized medicine: what is it and what are the challenges? Fertil Steril. 2018;109:944–5.
33. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
34. Altmäe S, Haller K, Peters M, Hovatta O, Stavreus-Evers A, Karro H, et al. Allelic estrogen receptor 1 (ESR1) gene variants predict the outcome of ovarian stimulation in in vitro fertilization. Mol Hum Reprod. 2007;13:521–6.
35. de Mattos CS, Trevisan CM, Peluso C, Adami F, Cordts EB, Christofolini DM, et al. ESR1 and ESR2 gene polymorphisms are associated with human reproduction outcomes in Brazilian women. J Ovarian Res. 2014;7:114.
36. Behre HM, Greb RR, Mempel A, Sonntag B, Kiesel L, Kaltwasser P, et al. Significance of a common single nucleotide polymorphism in exon 10 of the follicle-stimulating hormone (FSH) receptor gene for the ovarian response to FSH: a pharmacogenetic approach to controlled ovarian hyperstimulation. Pharmacogenet Genomics. 2005;15:451–6.
37. Perez Mayorga M, Gromoll J, Behre HM, Gassner C, Nieschlag E, Simoni M. Ovarian response to follicle-stimulating hormone (FSH) stimulation depends on the FSH receptor genotype. J Clin Endocrinol Metab. 2000;85:3365–9.
38. Desai SS, Achrekar SK, Paranjape SR, Desai SK, Mangoli VS, Mahale SD. Association of allelic combinations of FSHR gene polymorphisms with ovarian response. Reprod Biomed Online. 2013;27:400–6.
39. Cai J, Lou H, Dong M, Lu X, Zhu Y, Gao H, et al. Poor ovarian response to gonadotropin stimulation is associated with low expression of follicle-stimulating hormone receptor in granulosa cells. Fertil Steril. 2007;87:1350–6.
40. Nakayama T, Kuroi N, Sano M, Tabara Y, Katsuya T, Ogihara T, et al. Mutation of the follicle-stimulating hormone receptor gene 5'-untranslated region associated with female hypertension. Hypertension. 2006;48:512–8.
41. Bilibio JP, Meireles AJC, Conto ED, Lorenzzoni PL, Nascimento FCdo, Cunha-Filho JSda. GDF9 polymorphisms: influence on ovarian response in women undergoing controlled ovarian hyperstimulation. JBRA Assist Reprod. 2020;24:447–53.
42. Lledó B, Llácer J, Blanco L, Ortiz JA, Morales R, Fabregat A, et al. The relevance of the individual screening for genetic variants in predicting ovarian response. Pharmacogenet Genomics. 2019;29:216–23.