General Information

Project code: PNRR-MR1-2022-12376622

Title: Role of maternal effect genes and epimutations in Beckwith-Wiedemann syndrome and adverse reproductive outcomes

Duration in months: 24

Project Classification IRG: Genes, Genomes and Genetics

Project Classification SS: Genetics of Health and Disease – GHD

Keywords: Beckwith-Wiedemann, Assisted Reproductive Technology, Infertility, Subcortical Maternal Complex

Abstract: A three-fold increase in the prevalence of Beckwith-Wiedemann syndrome (BWS) has been observed in Assisted Reproductive Technology (ART) babies. Even if rare, the association between BWS and ART offspring represents a clinical and an epidemiological concern. Molecular defects detected in BWS include global alteration of transcriptome and methylome, changes of chromosomal architectures, loss-of-imprinting at imprinted domains. Pathogenic variants in Maternal- Effect Genes (MEGs) have been identified not only in mothers of BWS babies but also in women with reproductive disturbances such as failed pregnancy attempts and recurrent pregnancy loss. Out of MEGs, the subcortical maternal complex (SCMC) is a multimeric protein complex formed in the mature mammalian oocyte. Its components are exclusively expressed from the maternal genome in oocytes and early embryos and then degraded during embryonic development without compensation by the embryonal genome. Disruption of SCMC components in mice is associated with the developmental arrest of the embryo at the two-/four-cell stage. We hypothesize that imprinting disorders and epigenetic deregulations typical of some forms of BWS might be a consequence of specific SCMC variants in oocytes playing a crucial role in "reproductive errors". Thus, this project aims to investigate the role of pathogenic variants of SCMC as drivers of reproductive problems. This aim will be pursued by (i) defining the incidence of MEGs and specifically of SCMC pathogenic variants as a cause of differences in reproductive outcomes in the infertile female population and mothers of children with BWS. Healthy women with offspring affected by BWS and peculiar reproductive history, women with an unsolved infertility after ART and with recurrent miscarriages will be recruited. To identify pathogenic variants, whole-exome and genome sequencing will be performed; (ii) identifying methylation changes in women with reproductive problems including those with offspring affected by BWS; (iii) determining the molecular mechanisms underlying female infertility and imprinting disorder associated with damaging SCMC gene variants by employing a mouse model. The model has been created based on the identification of a hypomorphic missense PADI6 variant (P632A) in compound heterozygosity with a truncating mutation in the mother of two siblings affected by BWS. This project will clarify whether the association between ART and BWS can be ascribed to the ART protocol itself or maternal genetic variants underlying infertility. The comprehension of the occurrence and entity of SCMC pathogenic variants in women's fertility will contribute to personalized reproductive and genetic counseling.

Participants:

1 - Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico

2 - IRCCS Istituto Auxologico Italiano

3 - Università degli Studi della Campania "Luigi Vanvitelli"

4 - Università degli Studi di Napoli Federico II

Description Project

Summary description: Pathogenic variants in subcortical maternal complex (SCMC) have been identified not only in mothers of Beckwith- Wiedemann syndrome (BWS) babies but also in women with reproductive disturbances such as failed pregnancy attempts and recurrent pregnancy loss. Based on the higher incidence of BWS in children born from Assisted Reproductive Technology (ART), this project aims to investigate incidence and molecular mechanism of pathogenic variants of SCMC in women with reproductive disorders. We will (i) assess the incidence of these variants as a cause of differences in reproductive outcomes in the infertile female population and mothers of children with BWS; (ii) identify methylation changes in women with reproductive problems including those with offspring affected by BWS; (iii) determine the molecular causes underlying female infertility and imprinting disorder associated with damaging SCMC gene variants by employing a mouse model.

Background / State of the art: Assisted reproductive technology (ART) is a broad field encompassing various types of infertility treatments. Worldwide, ART has been responsible for the birth of 10 million infants. Concern has prevailed whether these techniques may influence the health of newborns. Changes in DNA methylation and the frequency of imprinting disorders have been linked to ART. A three-fold increase in the prevalence of Beckwith-Wiedemann syndrome (BWS) has been observed in ART babies. The mechanisms behind this association are difficult to assess as BWS (i) is a rare condition, affecting 1:10340 newborns (Brioude et al,2018), (i) is the result not only of imprinting mechanisms but also of genetic mutations; (iii) has a wide clinical spectrum, and the definitive diagnosis may not be done prenatally or at birth. Pathogenic variants in Maternal-Effect Genes (MEGs) have been identified in mothers of children with imprinting defects and troublesome reproductive histories such as failed pregnancy attempts and recurrent pregnancy loss (Eggerman et al, 2022). Out of MEGs, the subcortical maternal complex (SCMC) is a protein complex involved in the assembly of meiotic and mitotic spindle, regulation of the metabolism of RNA expressed from maternal genome in oocyte and zygote genome activation, as well as imprinting maintenance. Genomic variants in SCMC genes were found in women displaying a spectrum of reproductive disturbances, from infertility and miscarriages to altered imprinting in the offspring.

Description and distribution of activities of each operating unit:

U.O.1 will oversee patients' recruitment since the PI is the coordinator of the ART centre, one of the five biggest Italian ART facility (>1200 cycles/year), and of the Recurrent miscarriage centre of his institution. This will guarantee the possibility to recruit the identified cohorts after approval from the local ethical committee. The CoPI is the President of the Italian Society of Human Reproduction, a scientific association dedicated to the study of infertile patients that will be involved in the recruitment of additional infertile or recurrent pregnancy loss (RPL) patients with these uncommon phenotypes.

U.O.2 will carry on genomic sequences by WES and WGS on cases selected by UO1 and UO4 and on BWS mothers. UO2 is a referring centre for the BWS diagnosis with a collection of more than 200 positive diagnoses. At least half of them displays altered methylation in various imprinting loci (Multilocus

Imprinting Disturbance, MLID). Since 2009, a collaboration with the Italian Association of BWS allowed the unit to collaborate with families, building a well structured database with clinical data and information regarding patients and mothers' reproductive history.

U.O.3 has generated a transgenic mouse line carrying the hypomorphic variant in the SCMC Padi6 gene that was found in a woman with two children affected by BWS and multi-locus imprinting disturbances. The females of this mouse line are infertile and most of the embryos generated by IVF die at the 2-cell stage. To investigate the mechanisms linking SCMC deficiency with infertility and BWS, the unit will analyze DNA methylation and RNA in oocytes and 2-cell embryos of the Padi6-mutant line. Because a few embryos develop to blastocyst in vitro, the unit will transfer them into pseudopregnant females and analyze the fetuses at mid-gestation. On the basis of the results of the NGS analysis on oocytes and embryos, the unit will select about 40 epigenetics compounds to be used in a mini-screening aiming to identify key molecules improving development of the mutant embryos. Unfertilized oocytes derived from women carriers of pathogenic variants in the prioritized genes will be also investigated by single-cell NGS analysis.

U.O.4 will carry on epigenomic analyses by EPIC array, MS-MLPA and Bis-Seq on genomic DNA to identify methylation changes in the cohorts of women with reproductive problems including those with offspring affected by BWS.U.O.4 is the referring centre for the BWS diagnosis and other imprinting disorders, the only one operating in Southern Italy. This center is currently executing diagnoses of patients from several Southern Hospitals. U.O.4 is specialized in epigenetic and epigenomic analyses for different pathologies and participates in building international databases.

Specific aim 1: The first aim of the project is to define the incidence of MEGs and specifically of SCMC pathogenic variants as a cause of differences in reproductive outcomes in the infertile female population and mothers of BWS children. We hypothesize that imprinting disorders and epigenetic deregulations might be a consequence of SCMC variants in oocytes and play a crucial role in "reproductive errors". Different pathogenic variants in SCMC genes may explain the heterogeneous phenotypes as a spectrum ranging from complete infertility to a delay in conceiving to early miscarriage, to recurrent miscarriages up to, in the mildest end, the conception of one or more children with BWS. A genotype-phenotype correlation or variable interaction of different genetic and environmental factors may be at the basis of this phenotypic spectrum. Furthermore, SCMC variants might be also associated with numerical chromosomal aberrations. The main limitation of the available studies in the literature on this topic is represented by the lack of investigations of the variants prevalence in large cohorts of infertile women considering important confounders such as age and causes of infertility. A proper selection of patients is crucial to infer on the possible impact of SCMC genetic variants on reproductive outcomes. To this aim, three cohorts of patients will be recruited:

Cohort 1: healthy women with offspring affected by BWS and peculiar reproductive history from our population of clinically and molecularly diagnosed BWS families;

Cohort 2: women under 35 undergoing ART for infertility (defined as failure to achieve a pregnancy after 12 months or more of regular unprotected sexual intercourses) and unable to obtain a live birth after three completed cycles or after the transfer of at least 6 blastocysts;

Cohort 3: women under 35, with RPL (defined as the loss of two or more pregnancies before 24 weeks of gestation).

Exclusion criteria will be the presence of conventional and molecular karyotype alterations and the occurrence of known causes that can lead to decreased fertility or recurrent abortions: disorders of the ovaries, such as polycystic ovarian syndrome and other follicular disorders, disorders of the endocrine system causing imbalances of reproductive hormones levels, autoimmune conditions, male infertility, uterine or tubal dysfunctions and malformations, thrombophilic or non- corrected thyroid dysfunctions.

For the sample size calculations, we claimed as clinically relevant the demonstration that these alterations could affect at least 20% (±10%) of women giving birth to a BWS newborn and 5% (±5%) of those with reproductive disorders. On these bases, the number of women to be recruited in cohort 1, 2 and 3 is computed to 62, 73 and 73, respectively.

To identify pathogenic variants, whole-exome sequencing (WES) will be performed as the first approach in all the recruited patients. WES analysis will be carried out following various steps. First, we will analyze different subsets of genes, belonging to:

a) Maternal effect genes as SCMC components and other related genes;

b) Genes essential in the maturation of the oocyte and zygote progression through the early phases of the embryogenesis or highly expressed at different stages of oocyte maturation;

c) Genes with known and potential roles in the establishment and control of genomic imprinting and involved in DNA methylation reactions.

Subsequently, variants with a high pathogenicity score will be analyzed, to identify any genes that may be associated with the phenomenon, but do not belong to the previously described categories of genes. Finally, we will conduct a whole- genome sequencing (WGS) analysis on a selected subgroup of BWS mothers with peculiar clinical histories and negative WES analysis, to explore all the noncoding and regulatory regions not targeted by WES.

Specific aim 2: The second aim of this project is to employ whole-genome methylation analysis to identify methylation changes in women with reproductive problems including those with offspring affected by BWS. We have recently demonstrated that a loss of function variant of the SCMC gene KHCD3L that is associated with recurrent hydatidiform mole results in deficient de novo methylation in the female germline (Demond et al., 2019). In light of this evidence, it seems conceivable that also other SCMC genes might control general DNA methylation processes. The recent availability of high-throughput screening platforms has led to the identification of specific methylation signatures associated with an increasing number of mendelian disorders, indicating that DNA methylation analysis may represent a valid tool for a better classification of diseases with overlapping clinical features and for sorting cases with ambiguous or unidentified genetic variants (Cerrato et al. 2020). Preliminary data obtained by our research units by employing the Infinium MethylationEPIC methylation array analysis of blood leukocytes DNA allowed the identification of 267 differentially methylated CpGs in the genomes of 4 women carriers of pathogenic SCMC gene variants with respect to age-matched controls (Fig. 1, upper panel). Also, the finding of accelerated epigenetic ageing in women with unsuccessful IVF (Li Piani et al. 2022) suggests that the blood whole-genome methylation profile may be used to predict the risk of generating progeny with methylation disturbances resulting in miscarriages or imprinting disorders. Specific tasks will be:

a) Determining the whole-genome methylation of blood leukocytes of the cohorts of women described in Aim 1 and comparing it with that of a similar number of sex- and age-matched controls;

b) Determining the whole-genome methylation of unfertilized oocytes derived from unsuccessful ART cycles of the same cohorts and comparing it with that of control oocytes (derived from either donation for research or public datasets).

DNA methylation will be determined in blood leukocytes by methylation array analysis and in unfertilized oocytes by single- cell BS-seq, as previously described (Demond et al., 2019). Concerning the bioinformatics analysis, particular care will be put to reduce batch-effects to a minimum, and variations due to blood cell composition and age differences will be taken into consideration, by using specific algorithms. Due to the complex distribution associated with whole-genome methylation profiles, dimensionality reduction techniques (e.g. PCA,MDS) and clustering methods (both hierarchical and centroid- based) will be used to assess the importance of specific regions (e.g. imprinted loci) to discriminate prioritized women from controls according to their methylation profiles. State-of-the art machine learning techniques (e.g. Hidden Markow models, Support vector machines, Deep learning) will be used to infer differentially methylated regions from methylation data and to classify women status based on methylation profiles. Methylation profiles will be correlated with the results of the genetic analysis described in Aim 1 to see if specific episignatures are associated with MEGs. We expect the results obtained from this part of the project will allow to identify specific methylation changes in infertile women that can be used for diagnosis and patients' stratification.

Specific aim 3: The third aim of the project is to determine the molecular mechanisms underlying female infertility and imprinting disorder associated with damaging SCMC gene variants by employing a mouse model. SCMC gene knockout in the mouse generally causes female infertility and early developmental arrest, but DNA methylation and gene expression have not been investigated or have been done only in very limited manner so far (Mahadevann et al., 2017). In humans, PADI6 mutations have been associated with female infertility and imprinting disorders, and we recently identified a hypomorphic missense PADI6 variant (P632A) in compound heterozygosity with a truncating mutation in the mother of two siblings affected by Beckwith-Wiedemann syndrome and MLID (Cubellis et al. 2021). In order to generate a mouse model with a mutation that did not completely inactivate the Padi6 gene, we have introduced in the mouse the missense variant identified in the BWS family by homologous recombination. We found that the female mice carrying the Padi6 variant in homozygosity did not produce viable offspring after ten matings. However, when their MII oocytes were fertilized in vitro by wild-type sperm, 2-cell embryos were obtained with normal efficiency. At later stages, the embryos showed in vitro developmental delay and asymmetric cell divisions, but 10% of them reached the blastocyst stage after a delay of 2-3 days with respect to the wild- type. We hypothesize that epigenetic and genetic alterations resulting from defective maternal Padi6 cause developmental delay and female infertility in the mouse similarly to what happen in the human patients, suggesting that this mouse line is a good model to study the role of PADI6 in human infertility and imprinting disorders. Specific tasks will be:

a) Determining the whole-genome methylation and RNA profiles of the Padi6 mut/mut oocytes and pre-implantation embryos obtained after IVF with wild-type sperm;

b) Transfer of the blastocysts derived from the IVF described in a) into pseudopregnant females and analysis of whole- genome DNA methylation and RNA in the derived mid-gestation embryos by BS-seq and RNA-seq;

c) Mini-screening of epigenetics compounds on the 2-cell Padi6 mut/mut embryos described in a).

DNA methylation and RNA expression of the mouse oocytes and embryos will be determined by whole-genome scBS-seq and scRNA-seq, as previously described (Demond et al 2019). Analysis of DNA methylation and expression of imprinted genes will be prioritized, as previously described (Riso et al., 2016; Acurzio et al., 2021). For the drug screening, we will use compounds with defined activity against enzymes which carry out epigenetic modifications, such as HDACs, SIRTs, Lysine demethylases, HATs, Histone methyl transferases, DNA methyltransferases as well as SIRT activators (e.g. SCREEN-WELL® Epigenetics library). At least 40 compounds will be selected on the basis of the results of NGS analysis of mutant embryos. The compounds will be added to in vitro cultured 2-cell embryos, and time and efficiency of development to blastocyst will be determined. The results of this part of the project will pave the way for new therapeutic strategies to cure female infertility.

Experimental design aim 1: The first aim of the project is to define the incidence of MEGs and specifically of SCMC pathogenic variants as a cause of differences in reproductive outcomes in the infertile female population and mothers of BWS children. We hypothesize that imprinting disorders and epigenetic deregulations might be a consequence of SCMC variants in oocytes and play a crucial role in 'reproductive errors'. Different pathogenic variants in SCMC genes may explain the heterogeneous phenotypes as a spectrum ranging from complete infertility to a delay in conceiving to early miscarriage, to recurrent miscarriages up to, in the mildest end, the conception of one or more children with BWS. A genotype-phenotype correlation or variable interaction of different genetic and environmental factors may be at the basis of this phenotypic spectrum. Furthermore, SCMC variants might be also associated with numerical chromosomal aberrations. The main limitation of the available studies in the literature on this topic is represented by the lack of investigations of the variants prevalence in large cohorts of infertile women considering important confounders such as age and causes of infertility. A proper selection of patients is crucial to infer on the possible impact of SCMC genetic variants on reproductive outcomes. To this aim, three cohorts of patients will be recruited:

Cohort 1: healthy women with offspring affected by BWS and peculiar reproductive history from our population of clinically and molecularly diagnosed BWS families;

Cohort 2: women under 35 undergoing ART for infertility (defined as failure to achieve a pregnancy after 12 months or more of regular unprotected sexual intercourses) and unable to obtain a live birth after three completed cycles or after the transfer of at least 6 blastocysts;

Cohort 3: women under 35, with RPL (defined as the loss of two or more pregnancies before 24 weeks of gestation). Exclusion criteria will be the presence of conventional and molecular karyotype alterations and the occurrence of known causes that can lead to decreased fertility or recurrent abortions: disorders of the ovaries, such as polycystic ovarian syndrome and other follicular disorders, disorders of the endocrine system causing imbalances of reproductive hormones levels, autoimmune conditions, male infertility, uterine or tubal dysfunctions and malformations, thrombophilic or non- corrected thyroid dysfunctions.

For the sample size calculations, we claimed as clinically relevant the demonstration that these alterations could affect at least 20% (±10%) of women giving birth to a BWS newborn and 5% (±5%) of those with reproductive disorders. On these bases, the number of women to be recruited in cohort 1, 2 and 3 is computed to 62, 73 and 73, respectively.

To identify pathogenic variants, whole-exome sequencing (WES) will be performed as the first approach in all the recruited patients. WES analysis will be carried out following various steps. First, we will analyze different subsets of genes, belonging to:

a) Maternal effect genes as SCMC components and other related genes;

b) Genes essential in the maturation of the oocyte and zygote progression through the early phases of the embryogenesis or highly expressed at different stages of oocyte maturation;

c) Genes with known and potential roles in the establishment and control of genomic imprinting and involved in DNA methylation reactions.

Subsequently, variants with a high pathogenicity score will be analyzed, to identify any genes that may be associated with the phenomenon, but do not belong to the previously described categories of genes. Finally, we will conduct a whole- genome sequencing (WGS) analysis on a selected subgroup of BWS mothers with peculiar clinical histories and negative WES analysis, to explore all the noncoding and regulatory regions not targeted by WES.

Experimental design aim 2: The second aim of this project is to employ whole-genome methylation analysis to identify methylation changes in women with reproductive problems including those with offspring affected by BWS. We have recently demonstrated that a loss of function variant of the SCMC gene KHCD3L that is associated with recurrent hydatidiform mole results in deficient de novo methylation in the female germline (Demond et al., 2019). In light of this evidence, it seems conceivable that also other SCMC genes might control general DNA methylation processes. The recent availability of high-throughput screening platforms has led to the identification of specific methylation signatures associated with an increasing number of mendelian disorders, indicating that DNA methylation analysis may represent a valid tool for a better classification of diseases with overlapping clinical features and for sorting cases with ambiguous or unidentified genetic variants (Cerrato et al. 2020). Preliminary data obtained by our research units by employing the Infinium MethylationEPIC methylation array analysis of blood leukocytes DNA allowed the identification of 267 differentially methylated CpGs in the genomes of 4 women carriers of pathogenic SCMC gene variants with respect to age-matched controls (Fig. 1, upper panel). Also, the finding of accelerated epigenetic ageing in women with unsuccessful IVF (Li Piani et al. 2022) suggests that the blood whole-genome methylation profile may be used to predict the risk of generating progeny with methylation disturbances resulting in miscarriages or imprinting disorders. Specific tasks will be:

a) Determining the whole-genome methylation of blood leukocytes of the cohorts of women described in Aim 1 and comparing it with that of a similar number of sex- and age-matched controls;

b) Determining the whole-genome methylation of unfertilized oocytes derived from unsuccessful ART cycles of the same cohorts and comparing it with that of control oocytes (derived from either donation for research or public datasets).

DNA methylation will be determined in blood leukocytes by methylation array analysis and in unfertilized oocytes by single- cell BS-seq, as previously described (Demond et al., 2019). Concerning the bioinformatics analysis, particular care will be put to reduce batch-effects to a minimum, and variations due to blood cell composition and age differences will be taken into consideration, by using specific algorithms. Due to the complex distribution associated with whole-genome methylation profiles, dimensionality reduction techniques (e.g. PCA,MDS) and clustering methods (both hierarchical and centroid- based) will be used to assess the importance of specific regions (e.g. imprinted loci) to discriminate prioritized women from controls according to their methylation profiles. State-of-the art machine learning techniques (e.g. Hidden Markow models, Support vector machines, Deep learning) will be used to infer differentially methylated regions from methylation data and to classify women status based on methylation profiles. Methylation

profiles will be correlated with the results of the genetic analysis described in Aim 1 to see if specific episignatures are associated with MEGs. We expect the results obtained from this part of the project will allow to identify specific methylation changes in infertile women that can be used for diagnosis and patients' stratification.

Experimental design aim 3: The third aim of the project is to determine the molecular mechanisms underlying female infertility and imprinting disorder associated with damaging SCMC gene variants by employing a mouse model. SCMC gene knockout in the mouse generally causes female infertility and early developmental arrest, but DNA methylation and gene expression have not been investigated or have been done only in very limited manner so far (Mahadevan et al., 2017). In humans, PADI6 mutations have been associated with female infertility and imprinting disorders, and we recently identified a hypomorphic missense PADI6 variant (P632A) in compound heterozygosity with a truncating mutation in the mother of two siblings affected by Beckwith-Wiedemann syndrome and MLID (Cubellis et al. 2021). In order to generate a mouse model with a mutation that did not completely inactivate the Padi6 gene, we have introduced in the mouse the missense variant identified in the BWS family by homologous recombination. We found that the female mice carrying the Padi6 variant in homozygosity did not produce viable offspring after ten matings. However, when their MII oocytes were fertilized in vitro by wild-type sperm, 2-cell embryos were obtained with normal efficiency. At later stages, the embryos showed in vitro developmental delay and asymmetric cell divisions, but 10% of them reached the blastocyst stage after a delay of 2-3 days with respect to the wild- type. We hypothesize that epigenetic and genetic alterations resulting from defective maternal Padi6 cause developmental delay and female infertility in the mouse similarly to what happen in the human patients, suggesting that this mouse line is a good model to study the role of PADI6 in human infertility and imprinting disorders. Specific tasks will be:

a) Determining the whole-genome methylation and RNA profiles of the Padi6 mut/mut oocytes and pre-implantation embryos obtained after IVF with wild-type sperm;

b) Transfer of the blastocysts derived from the IVF described in a) into pseudopregnant females and analysis of whole- genome DNA methylation and RNA in the derived mid-gestation embryos by BS-seq and RNA-seq;

c) Mini-screening of epigenetics compounds on the 2-cell Padi6 mut/mut embryos described in a).

DNA methylation and RNA expression of the mouse oocytes and embryos will be determined by whole-genome scBS-seq and scRNA-seq, as previously described (Demond et al 2019). Analysis of DNA methylation and expression of imprinted genes will be prioritized, as previously described (Riso et al., 2016; Acurzio et al., 2021). For the drug screening, we will use compounds with defined activity against enzymes which carry out epigenetic modifications, such as HDACs, SIRTs, Lysine demethylases, HATs, Histone methyl transferases, DNA methyltransferases as well as SIRT activators (e.g. SCREEN-WELL® Epigenetics library). At least 40 compounds will be selected on the basis of the results of NGS analysis of mutant embryos. The compounds will be added to in vitro cultured 2-cell embryos, and time and efficiency of development to blastocyst will be determined. The results of this part of the project will pave the way for new therapeutic strategies to cure female infertility.

Hypothesis and significance: The hypotheses underlying this project arise from the anecdotal observations that some women who carry pathogenetic variants in genes coding for SCMC components display heterogeneous phenotypes suggestive of reproductive disturbances such as failures to conceive, multiple miscarriages and offspring affected by imprinting disorders. These reproductive disturbances have been

described in BWS families, where the proband had a genetic diagnosis of KCNQ1OT1:TSS-DMR hypomethylation associated with MLID. SCMC pathogenic variants were identified by U.O.2 and

U.O.3 (Sparago et al. 2019; Cubellis et al. 2020; Tannorella et al., 2022; see also Fig. 1, lower panel). The investigation of these variants in a population of infertile women with severe problems in achieving a pregnancy and in recurrent abortion patients will allow to evaluate the possibility that genetic issues underlying BWS may be involved in a wider spectrum of phenotypes responsible for reproductive dysfunctions. This possibility will have strong implications in the context of ART for the elucidation of the potential negative effects of the in vitro procedures on epigenetic traits and imprinting diseases. The safety of ART procedures represents an important matter of debate. Findings derived from this study may be crucial to clarify these safety issues and to change the current thinking on the association between ART and BWS.

Methods of data collection: Methylation analysis to disclose MLID: MS-MLPA (ME034, Multi-locus Imprinting, MRC-Holland) analysed by Coffalyzer Software; Targeted Methylation-sensitive single-nucleotide primer extension (MS-SNuPE) as described in (Tannorella et al, 2021).

Whole exome sequencing (WES): First, exome enriched libraries will be obtained in house starting from 500 ng of genomic DNA from each subject by using the SureSelect Human All Exon V7 Enrichment Kit (Agilent), which covers 35.7 Mb of content with a design size of 48.2 Mb that targets 99.7% of coding exons represented in four key databases (RefSeq, GENCODE v24, CCDS, and USCS Known Genes). Then, the obtained exome enriched libraries will be sequenced at Biodiversa s.r.l. on Illumina NovaSeq 6000 platform with flowcell S4 and using 2x150 bp paired-end reads. Sequencing will produce a total of 800Gbp of reads as output files format FastQ.

Variants validations: all variants considered of interest (with an ACMG classification 3,4 or 5) will be validated by Sanger sequencing or by Nextera XT if suggestive of mosaicism.

Splicing variants: cDNA of splicing variant will be sequenced by Sanger sequence to confirm the occurrence of an aberrant transcript, if the transcript is expressed in an achievable tissue (blood, fibroblast).

SNParray: Infinium CytoSNP-850K v1.2 BeadChip (Illumina) performed as recommended by Illumina protocols and analysed by Genomestudio/Blue Fuse.

Whole genome DNA methylation and RNA Profiling: Genomic DNA extracted from human peripheral blood leukocytes will be quantified and normalized and bisulfite conversion of DNA will be performed according to the manufacturer's recommendations for the Illumina Infinium Assay. Genome-wide DNA methylation assay will be performed using Epic 850K BeadChip, following Illumina protocol. After hybridization, staining and scanning of Illumina arrays, the methylation score for each CpG site will be represented as ß values according to the fluorescent intensity ratio between methylated and unmethylated probes. DNA and RNA extracted from unfertilized human and mouse oocytes and mouse 2-cell embryos will be analyzed by whole-genome single-cell bisulfite sequencing (scBS-seq) and single-cell RNA sequencing (scRNA-seq) as previously described (Demond et al., 2019; Angermueller et al., 2016 PMID: 26752769). DNA methylation and RNA will also be analyzed in E14 whole mouse embryos by whole-genome RRBS/BS-seq and RNA-seq, as previously described (Riso et al., 2016).

Mouse IVF: Heterologous IVF will be performed according to the method described by Taft (2017). 8-12 weeks old males and 8 weeks old females will be used. Female mice will be treated with pregnant mare serum gonadotropin (PMSG) and with human chorionic gonadotropin (hCG). Hyaluronidase will be used to

break down the oocyte-cumulus complex and collect MII oocyte. At the same time the cauda epididymis and vasa deferentia will be dissected and the spermatozoa collected. The oocytes will be incubated with sperm for 4-6 hours. Fertilized oocytes will be washed and incubated overnight. Embryo development to blastocyst will be evaluated at the microscope every 24 hours for 5 days.

Epigenetic drug screening: 2-cell embryos derived from homozygous Padi6-P632A/P632A and control heterozygous Padi6- P632A/+ females will be placed in 96 well plates containing EmbryoMax Advanced KSOM Embryo Medium (MR-101-D) and 40 epigenetics compounds of the SCREEN-WELL® Epigenetics library will be added at the concentrations indicated by the manufacturer. The compounds will be chosen on the basis of the results of NGS analysis of mutant embryos and tested in duplicate and different concentrations. The efficiency of development to blastocyst of the embryos will be determined after 5 day by microscope examination.

Statistic plan: Sample size has been estimated considering different aims and genetic and epigenetic analyses.

To AIM 1, for the sample size calculations, we claimed as clinically relevant the demonstration that these alterations could affect at least 20% (±10%) of women giving birth to a BWS newborn and 5% (±5%) of those with reproductive disorders. On these bases, setting a power at 80% and a significance at 5%, the number of women to be recruited in cohort 1, 2 and 3 is computed to 62, 73 and 73, respectively. To identify rare pathogenic and variants of uncertain significance (VUS), WES will be performed as the first approach in all the recruited patients. WES analysis will be carried out following various steps.

First, we will analyze different subsets of genes, belonging to: a) Maternal effect genes as SCMC components and other related genes; b) Genes essential in the maturation of the oocyte and zygote progression through the early phases of the embryogenesis or highly expressed at different stages of oocyte maturation; c) Genes with known and potential roles in the establishment and control of genomic imprinting and involved in DNA methylation reactions.

Subsequently, variants with a high pathogenicity score will be analyzed to identify any genes that may be associated with the phenomenon, but do not belong to the previously described categories of genes. Finally, we will conduct a WGS analysis on a selected subgroup of BWS mothers with peculiar clinical histories and negative WES analysis (n=30), to explore all the noncoding and regulatory regions not targeted by WES.

To AIM 2, epigenetic profiles from the enrolled n=208 women from the three patient cohorts will be obtained experimentally while age matched epigenetic profiles of an equivalent number of women from the general population will be retrieved from our data bank and public repositories. Sample size calculated using parameters from previous studies and assuming an effect size (methylation difference between two groups) of at least 5%, with a type I error of 10-8 and a power of 95% can be considered adequate.

Statistical analysis

Aim 1, WES: Following sequencing, the resulting reads will be first aligned with BWA algorithm while SNV and indel will be analyzed according to GATK best practices (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short- variant-discovery-SNPs-Indels-). Briefly, Fastq data will be aligned to the reference genome assembly GRCh37/hg19 using BWA-MEM (v.0.7.10). PCR duplicates will be removed employing Picard (v1.119) and the GATK4 will be used to recalibrate base quality scores while

HaplotypeCaller algorithm will be used for variant calling step. Finally, all variants will be annotated using wANNOVAR. A quality control analysis will be performed both on samples and genotypes obtained.

Variants with depth of coverage <10x and MAF >0.01 will be excluded. The American College of Medical Genetics and Genomics guidelines (ACMG) will be used for the variant Clinical Classification. Only genomic regions with a coverage > 10X in at least 95% of subjects will be considered for the analysis. For rare Pathogenic and VUS variants, the genotype phenotype association will be carried out using a Sequence Kernel Association Test (SKAT-O) approach.

Aim 2: Illumina EPIC BeadChip raw data will be analyzed using R and the RnBeads package (Assenov et al., 2014). A quality control analysis will be performed both on samples and probes. Probes and samples of highest impurity will be removed from the dataset, the background will be subtracted and the signal intensity values will be normalized. A particular attention will be dedicated to identifying potential confounding factors in particular way batch-effects, age and blood cell composition will be evaluated. Cell-type deconvolution algorithm proposed by Houseman et al. (2012) will be applied on DNA methylation data to estimate immune cell composition.

After quality control and normalization steps, the analysis on epigenetic data will evaluate four main points

a) Evaluation of distances in epigenetic profiles through dimensionality reduction techniques (e.g. PCA,MDS) and clustering methods (both hierarchical and centroid-based)

c) Identification of methylation signatures characterizing genetic background

d) Evaluation of rare epigenetic variations

e) Evaluation of epigenetic drift

Both cluster analysis and evaluation of epigenetic signatures as well as immune cell profiling will be performed using specific functions provided in the RnBeads package for DNA methylation analysis. Stochastic Epigenetic Mutation (SEMs) analysis will be instead used to detect important epigenomic alterations and to estimate epigenetic drift. This analysis proposed by Gentilini et al (2015) allows to detect rare epigenetic alteration potentially involved in etiopathogenetic mechanisms (Guida et al., 2021).

Aim 3: Concerning BS-seq and RNA-seq, sequencing quality of raw data will be analyzed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Low quality reads and adapter contamination will be removed with TrimGalore ((https://www.bioinformatics .babraham.ac.uk/projects/trim_galore/) and Cutadapt (Martin et al., 2011). For BSseq, Bismark (Krueger et al., 2011) will be used to map the data to the reference genome (GRCh38/GRCm38) followed by a deduplication process to remove the PCR duplicates. For single-cell experiments, a library of > 3M reads will be generated for each cell. About RNA-seq experiments, STAR (Dobin et al., 2013) will be used to align the reads to the reference genome and the gene count will be calculated with featureCount tool (Liao et al., 2014). The methylation data and the genes count will be imported in R software to perform the statistical analysis.

Timing of analysis data

Duration of the study: 20 months for the experimental part and 4 months to complete the analysis, disseminating the results and preparing the manuscript.

AIM 1: enrollment of the patient and control cohorts and sample collection are foreseen in 10 months. Methylation analysis on BWS KCNQ1OT1:TSS-DMR and other imprinted DMRs to widen cohort 1, WES at specific genes in all the cohorts and SNP-array to detect small CNVs and LOH will be performed following enrollment and will be finalized within month 17. The last part related to WGS on a selected subgroup of BWS mothers will be accomplished from month 16 to month 20.

AIM 2: blood leukocytes samples collected from the same cohorts and from controls enrolled in the first 10 months will be evaluated for whole-genome methylation within month 20. A similar timing can be foreseen for the evaluation of whole- genome methylation of unfertilized oocytes derived from unsuccessful ART cycles collected in the first 10 months.

AIM 3: for the experimental part based on the animal model, 10 months will be required for determining the whole-genome methylation and RNA profiles of the Padi6 mut/mut oocytes and pre-implantation embryos obtained after IVF. Following this procedure, blastocysts derived will be transferred into pseudopregnant females and analysed for whole-genome DNA methylation and RNA in the derived mid-gestation embryos from month 7 to 14. Almost simultaneously, epigenetic compounds will be tested within month 20.

Expected outcomes:

Aim 1.

-to enroll the three cohorts of women to be registered in a database with their clinical and reproductive histories and the control subjects

-to establish the frequency of pathogenic variants in the SCMC genes in each of the three cohorts and clarify whether the mutation type may impact the reproductive outcome (Imprinting defect, recurrent miscarriages, and infertility)

-to discover novel genes implicated in the pathomechanisms of BWS and able to explain the increased number of ID babies among children conceived with ART.

Aim 2.

-to identify DNA methylation changes in women with reproductive problems including offspring with BWS

-to correlate epigenetic profiles with genetic variants.

Aim 3.

-to determine the impact of maternal Padi6 hypomorphic variants on whole-genome DNA methylation and RNA profiles in female germ cells

-to determine the impact of maternal Padi6 hypomorphic variants on whole-genome DNA methylation and RNA profiles of pre- and post-implantation embryos.

-to identify epigenetic pathways that are affected in Padi6-mutant embryos and impact early development.

Risk analysis, possible problems and solutions:

Aim 1. Possible problem: For the goal of identifying novel rare pathogenic variants in SCMC genes or novel transacting genes, a wide number of women belonging to the three cohorts should be recruited in a

relatively short time. Solution: To overcome the possible problem of delay in recruitment, we count on the help of AIBWS, on the network of clinical geneticists, who have collaborated on the diagnostic activity with us for a long time. Moreover, for cohorts 2 and 3 we count on the support of the Italian Society of Human Reproduction, a scientific association dedicated to the study of infertile and abortion patients. The Society involves more than 20 Italian infertility centers that should guarantee support to identify these patients. The Co-PI is one of the Society's President. Possible problem: Another concern might be the finding of non-synonymous rare variants qualified as VUS or likely pathogenic variants by ACMG criteria, whose role is difficult to confirm without a functional role. The occurrence of variants with very low level of mosaicism may also be a challenge. Solution: We will use all possible in silico prediction tools including those based on sequence conservation and protein stability; the possible effect of splicing variants will be tested within mini-genes in transfected cells; a dedicated pipeline for mosaicism will also be applied.

Aim 2. Possible problems. Batch effects in methylome analysis Solution: Arrays will be processed each with randomized samples. Computational correction will be applied.

Aim 3. Possible problem: the Padi6 mouse model does not show imprinting defects similar to BWS patients. Solution: Focus on whole-genome methylation, copy-number and RNA expression profiles. This will reveal molecular mechanisms and pathways associated with Padi6 function and underlying female infertility that may pave the way to the identification of new therapeutic targets for human patients with SCMC mutations.

Significance and Innovation

The ART process mimics and also bypasses several physiological reproductive paths and it inherently involves a variable degree of invasiveness with unknown consequences. A recognized limitation in all studies is the confounding impact of the underlying cause of infertility of the biological parents in the obstetrical and perinatal complications. One of the suggested complications is the higher risk in MLID in children born following ART. The results of this project will have remarkable consequences in terms of safety of the ART procedures and will represent an important step forward in the understanding of the genetic causes underlying infertility and abortions. This will translate in the ability to prevent severe diseases such as BWS whose impact is extremely severe in terms of health, social and economic burden for the society and psychological consequences for the families.

Bibliography

Acurzio et al. Sci Rep. 2021;11(1):13802
Angermueller et al. Nat Methods. 2016;13(3):229-232
Assenov et al. Nat Methods. 2014;11(11):1138-1140
Brioude et al. Nat Rev Endocrinol. 2018;14(4):229-249
Cerrato et al. Genes (Basel). 2020; 11(4):355
Cubellis et al. Clin Epigenetics. 2020; 12(1):139
Demond et al. Genome Med. 2019;11(1):84
Dobin et al. Bioinformatics. 2013; 29(1):15-21
Eggermann et al. Clin Epigenetics. 2022;14(1):41
Gentilini et al. Aging (Albany NY). 2015;7(8):568-78
Guida et al. Int J Mol Sci. 2021; 22(3):1190

Henningsen et al. Hum Reprod. 2020;35(5):1178-1184
Houseman et al. BMC Bioinformatics. 2012;13:86
Krueger et al. Bioinformatics. 2011;27(11):1571-2
Liao et al. Bioinformatics. 2014;30(7):923-30
Li Piani et al. PLoS One. 2022;17(1):e0261591
Mahadevan et al. Sci Rep. 2017;7:44667
Martin. EMBnet.journal 17, 10, 2011
Pignata et al. Clin Epigenetics. 2022;14(1):71
Rehder et al. Genet Med. 2021;23(8):1399-1415
Richards et al. Genet Med. 2015;17(5):405-24.
Riso et al. Nucleic Acids Res. 2016;44(17):8165-78
Sparago et al, Clin Epigenetics. 2019;11(1):190
Taft. Cold Spring Harb Protoc. 2017;2017(11):pdb.prot094508
Tannorella et al. Clin Epigenetics. 2022;14(1):43

Timeline / Milestones

Duration of the study:

20 months for the experimental part

4 months to complete the analysis, disseminating the results and preparing the manuscript.

Milestones 12 months

- recruitment
- Whole-genome methylation and RNA profiles of the Padi6 mut/mut oocytes and derived pre-implantation embryos

Milestones 24 months

- Exome-seq profiles
- Whole-genome sequencing
- Whole-genome methylation profiles
- Analysis of the effects of 40 epigenetics compounds

Milestones 12 month

1. Recruitment of subjects from 1-3 cohorts and controls. Collection of samples.
2. Whole-genome methylation and RNA profiles of the Padi6 mut/mut oocytes and derived pre-implantation embryos

Milestones 24 month

1. Exome-seq profiles of all the cohorts of women
2. CNVs and LOH profiles of all the cohorts of women
3. Whole-genome sequencing of selected women

4. Whole-genome methylation profiles of blood leukocytes in all the cohorts of women
5. Whole-genome methylation profiles of unfertilized oocytes derived from unsuccessful ART cycles
6. Phenotypic and molecular analysis of mid-gestation embryos derived from Padi6 mutant blastocysts
7. Analysis of the effects of 40 epigenetics compounds on development of 2-cell Padi6 mutant embryos

Deliverables - 24 months

1. Identification of SNVs of MEGs in the recruited women with reproductive problems
2. Identification of CNVs and LOH in the recruited women with reproductive problems
3. Identification of DNA methylation changes in blood leukocytes and/or unfertilized oocytes of the recruited women with reproductive problems
4. Identification of DNA methylation and RNA changes in Padi6 mut/mut oocytes
5. Identification of DNA methylation and RNA changes in pre- and post-implantation embryos derived from Padi6 mut/mut oocytes
6. Identification of epigenetic compounds reverting the developmental delay of Padi6 mutant embryos

Equipment and resources available

Facilities Available: The study requires specific laboratory equipment. The experimental protocol is conducted in laboratories already accredited for research and makes use of routine quality controls implemented by the individual groups involved. NextSeq 550 (U.O.2 and U.O.3) Systems are available for NGS. Infinium MethylationEPIC BeadChip microarrays analyses will be performed using Illumina IScan System available at our institutions (U.O.2 and U.O.4). Tecan freedom evo platform will be used for wet lab steps. In addition, a QuantStudio 12k, Covaris E200, a BioAnalyzer 2100, a TapeStation, a Nanodrop (1 and 8 channels), a Qubit, an Illumina MiSeq, a PyroMark Q24, Vii A/ qPCR thermocycler and Hamilton and Eppendorf liquid handling workstations are available. Computational resources needed for processing of raw data from next generation sequencing include 2 server HP Proliant DL580 gen9 384 GB RAM 120 cores. A 250 TB storage system is built on top of a X9000 platform, which provides a network file system distributed on four HPE P2000 G3 MSA Array Systems, containing 6 separated enclosures each, equipped with more than 370 HardDrives. The different file systems are engineered to provide redundancy and prevent data loss.

Subcontract: For U.O.2 the subcontracts will be used for the Whole Genome Sequencing experiments, which will be run on a NovaSeq6000 sequencer. For U.O.3 and U.O.4 the subcontracts will be used for double Quality Control of NGS and Epigenomic analyses.

Translational relevance and impact for the national health system (SSN)

What is already know about this topic?

An association exists between ART and BWS in the newborns, although it is still unclear whether this can be ascribed to the ART protocol itself or the parental dysfunctions underlying infertility (Henningsen et al.,

2020). The SCMC components are exclusively expressed from the maternal genome in oocytes and early embryos and then degraded in embryonic development without compensation. Disruption of SCMC components in mice is associated with the developmental arrest of the embryo at the two-/four-cell stage. These reproductive disturbances have been described in BWS families, where the proband has a genetic diagnosis of MLID. SCMC pathogenic variants were identified by U.O.2 and U.O.3 (Sparago et al. 2019; Cubellis et al. 2020; Tannorella et al., 2022) and found to be present in women with troublesome reproductive histories such as failed pregnancy attempts and recurrent pregnancy loss (Eggerman et al, 2022).

Details on what is already know about this topic: Knowledge on the etiology and the clinical impact of MLID is still under study (Tannorella et al., 2022). Genomic sequencing of MLID patients' mothers highlighted the occurrence of pathogenic variants in genes transcribed by the maternal genome and deposited in the oocyte, where they persist until the first phases of embryogenesis. When the zygote genome is not yet transcriptionally active, these maternal genes contribute to the zygote genome activation (ZGA) and passage to early embryo. To date, pathogenic variants in maternal-effect genes in BWS patients with an IC2-LoM and MLID genotype have been disclosed in a limited number of studies. The reported cases include both biallelic and monoallelic variants, though the pathogenicity of the latter remains uncertain.

What this reasearch adds?

For the first time, a large cohort of mothers of BWS children will be inquired about their reproductive issues and/or recurrent miscarriages by experienced professionals. Previous studies point to a causative role for maternal effect damaging variants in genes deputed to the activation of zygote transcription, but the impact of these variants on the phenomenon remains to be investigated. This study faces also the crucial topic of ART role in imprinting disorders, specifically in BWS. Extensive (epi)genomic approaches (WES, WGS, and Whole methylation analyses) will be performed for the first time in this women's cohort. Knowledge of the underlying molecular cause of the reproductive problem is of fundamental importance in providing couples with personalized reproductive and genetic counseling. The mouse model and unfertilized oocytes derived from unsuccessful ART cycles in the women are both a precious and unique source to investigate the molecular link between BWS and MEGs.

Details on what this reasearch adds

For the first time:

1. Extensive (epi)genomic approaches (WES, WGS, and Whole methylation analyses) will be performed in mothers of BWS children.
2. Extensive (epi)genomic approaches will be performed in women with an unexplained and severe infertility difficult to solve with ART
3. Extensive (epi)genomic approaches will be performed in women with severe and recurrent abortions
4. The molecular link between BWS and MEGs will be investigated using a specific mouse model, allowing to explain the underlying mechanisms of infertility and abortion.
5. Mini-screening of epigenetics compounds will be tested on the 2-cell Padi6-P632A/P632A embryos in order to evaluate the possible deregulated pathways.

What are the implications for public health, clinical practice, patient care?

Even if rare, the increased incidence of BWS among ART babies is a clinically and epidemiological concern. The disease significantly impacts quality of life, and the economic burden per case is huge. Moreover, the awareness of this risk can affect couples' adherence to treatment. The proportion of infertile couples embarking in ART is still sub-optimal and treatments are negatively impacted by high dropouts (30-50%). The awareness of an increased risk of newborns with imprinting disorders such as BWS may play a role. Couples are informed about the possible causal relation between ART and BWS, a notion based on the existence of a biological rational (in vitro culture may disturb imprinting) and the strength of the epidemiological association (three folds higher risk). Demonstrating that, conversely, the association is not causal and couples at risk could be identified a priori would have the double benefits of preventing affected offspring and improve adherence.

Details on what are the implications for public health, clinical practice, patient care

1. defining couples at risk to have BWS babies following ART procedures (preventive measure)
2. to reduce important health costs to the society
3. to improve adherence to ART procedures
4. to improve fertility and abortion counselling