Use of hCG as prevention of breast cancer in BRCA1 carriers.

Principal Investigators

Jose Russo MD

Professor Director

Irma H Russo, MD-Breast Cancer Research Laboratory

Fox Chase Cancer Center, Philadelphia

Jose.Russo@fccc.edu

Jaak Janssens, MD, PhD

Professor ECP – Oncology ZOL

Jaak.janssens@zol.be

Herman Depypere MD, PhD

Professor Gynecology

University Ghent

Herman.Depypere@UGent.be

Eric de Jonge, MD, PhD

Professor Gynecology

Ziekenhuis Oost-Limburg

Eric.dejonge@zol.be

Content

Introduction

Three important recent models have shaped our current knowledge about breast cancer prevention: the accumulating evidence that the disease originates early in life, the impact of (epi-)genomic imprinting, and the recognition that breast cancer is a family of related but distinct diseases. The breast, the target organ, starting a unique intense growth after the first decade of life and involuting already during the third decade, is tremendously vulnerable to several endogenous and exogenous hormone disrupting molecules and other chemical and physical genotoxic factors. Lifestyle and toxins seem to generate long lasting (epi-)genomic marks especially in rapidly growing tissues such as the breast that can be reset for example by an early first full term pregnancy.

The preclinical interval between generation of susceptibility and appearance of the diseases offers opportunities for primary prevention and presumably has a period when genetic control is modifiable. Reversibility declines progressively when different premalignant or early malignant phenotypes appear. Early detection becomes now the priority and can be achieved through recognition of risk markers that reliably predict disease. Newer sophisticated imaging techniques detect the disease in phases where cure is expected. But these tools don't address the rapid mortal threat of breast cancer in the third world. This is now the prime concern for the next generations worldwide that probably are best served with affordable primary prevention. High priority must be given to lifestyle research on affordable reversion of (epi-) genomic alterations.

One event that restores effectively genomic alterations is an early full time pregnancy, that reduces breast cancer risk. High risk genomic signatures are converted to low risk signatures. The principal hormone responsible for this risk reduction is hCG. Consequently, treatment with hCG could have the same preventive capabilities compared to a full term pregnancy and could serve as a preventive treatment for high risk patients.

hCG is a safe and effective drug, commercialized under the name of Ovidrel. Earlier experience in treating women with breast cancer showed drastic reduction in Ki67 staining in cancer cells while no adverse reactions were noted. Similar experience is seen in infertility treatments where limited side effects are seen. The safety profile of this drug, together with the potential efficacy in reducing breast cancer risk, made the scientific community to believe that hCG treatment could be applied to healthy women at increased risk.

One of the best established risk factors is BRCA1 germ line mutation that confers a life time risk of up to 70%. In this proof-of-concept study, 25 patients will be selected for a 12 weeks treatment of Ovidrel. Before, immediately after and 1 year later the breast parenchyma will be examined by high-output molecular biology to find out if the "high-risk" genetic signature is effectively converted to "low-risk" signature in a similar way as early first full time pregnancy (FFTP) does.

If this project succeeds in its goal to show change is genetic risk signatures, not only an effective preventive hormonal therapy is provided for the woman at increased risk but also an intermediary risk factor (before the disease becomes clinical evident) is characterized. An intermediary risk factor identifies women at increased risk prior to the disease but also serves as monitor to guide preventive strategies.

Nederlandstalige samenvatting

Doelstelling en beschrijving van het studieprotocol

Dit is een klinische studie inzake een behandeling met studiegeneesmiddelen bij ca. 25 deelnemers, van wie zo'n 20 in België. Enkel gezonde vrouwen met een genetisch defect in BRCA1 gen zullen worden uitgenodigd om deel te nemen.

Het is een open studie dat het studiegeneesmiddel Ovidrel onderzoekt. Ovidrel zou de borstcellen beschermen tegen kankervorming doordat de kankerverwekkende eigenschap van het afwijkend gen BRCA1 gen door de veranderingen in het genetisch materiaal uitgeschakeld wordt.

Verloop van de studie

De studie neemt ongeveer *12* weken in beslag en omvat subcutane injecties van Ovidrel driemaal per week gedurende 12 weken. Maandelijks wordt de deelneemster gezien door de arts waarbij ook een bloedname gebeurt. Bovendien worden 3 lichte ingrepen uitgevoerd in de borst waarbij weefsel wordt weggenomen. Deze ingreep gebeurt ambulant (op de raadpleging) op een zo comfortabele manier als mogelijk. Hierdoor wordt de ingreep vrijwel pijnloos. De ingreep is voor de behandeling, onmiddellijk na de behandeling (week 12) en in week 36 vanaf de start van de behandeling.

Intake consultatie

Tijdens een eerste consultatie zal de arts-onderzoeker samen met de deelneemster nagaan of deze in aanmerking komt voor de studie (inclusiecriteria) en of er geen situaties zijn waarin deelname niet mogelijk is (exclusiecriteria).

Inclusiecriteria:

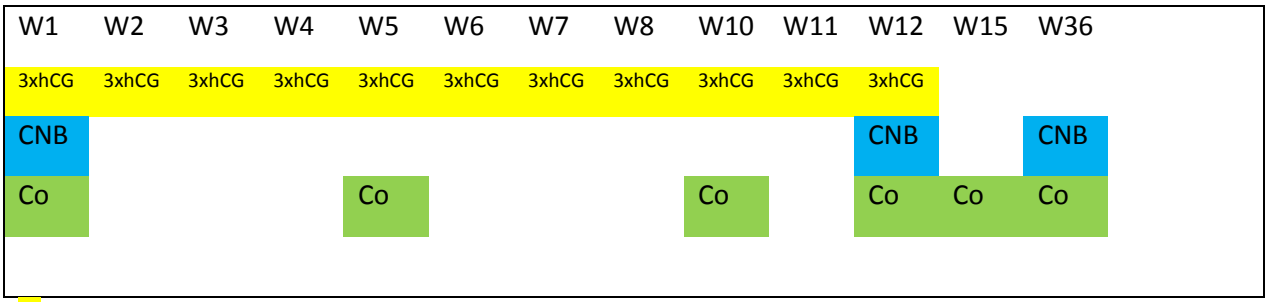
- Asymptomatische vrouw vanaf 18 jaar
- BRCA1 positief
- Karnofsky 100%
- Geen borstkanker
- Participeert niet in andere studie(s)
- Neemt geen tamoxifen
- Geen voorafgaande kanker gehad (uitz. Niet-melanoma huidkanker)
- Normale ovaria op ultrasound
- Begrijpt informatie en onderschrijft schriftelijk “informed consent”
- Akkoord om de klinische studie aan te gaan
- Niet zwanger
- Akkoord voor mechanische contraceptie (abstinentie, barrièremiddelen, tubaligatuur)

Exclusiecriteria

- Geschiedenis van allergie tov r-hCG
- Hormonale medicatie
- IVF behandeling in het verleden
- Actuele ziekten
- Hartafwijkingen
- Ernstige cognitieve stoornis
- Psychiatrische ziekte
- HIV-positief
- Hepatitis B of C

Onderzoeksfase.

Studie verloop kaart:



**) recombinant hCG 250 µg SC 3x/week op maandag, woensdag, vrijdag gedurende 12 weken*
**) CNB: core needle biopsy 1 sample van 8 Gauge x 2 cm onder echogeleide onder locale anesthesie.*
**) Consultatie voor anamnese en klinisch onderzoek en bloedname (hematologie & biochemie met beta-hCG bepaling)*

Vroegtijdige stopzetting:

Indien:

- de participant dat wenst
- intercurrente ziekte
- ernstige nevenwerkingen
- indien de onderzoeker dit nodig acht

Opvolgingsfase:

Na de laatste biopsie worden de participanten nog om de 3 maanden gezien op consultatie voor anamnese, klinisch onderzoek, bloedname (hematologie, biochemie, beta-hCG bepaling)

Belang van de studie:

Borstkanker is een levensbedreigende ziekte die alleen in België ongeveer 6000 nieuwe kankerpatienten per jaar betekent. Nog ongeveer 2000 overlijden jaarlijks eraan. Bovendien is het een mondiaal probleem met vooral expansie in de ontwikkelingslanden. Deze landen hebben niet de mogelijkheid om de (dure) screening en diagnosemogelijkheden aan te bieden. De mortaliteit is in deze landen nog steeds 50% en hoger. Een primaire preventie biedt zich dus aan en lijkt mogelijk door de bijdrage van jarenlang vorserswerk dat het belang van een eerste vroege volledig voldragen zwangerschap aantoont. Indien deze studie kan aantonen dat Ovidrel (of in een latere fase een deel van het zwangerschapshormoon hCG) hetzelfde beschermend effect kan nabootsen, lijkt de weg open om borstkanker effectief te bestrijden. In deze studie worden vrouwen uitgenodigd die ongeveer 70% risico hebben om de ziekte te krijgen. Later kan het principe aan een grotere groep aanboden worden.

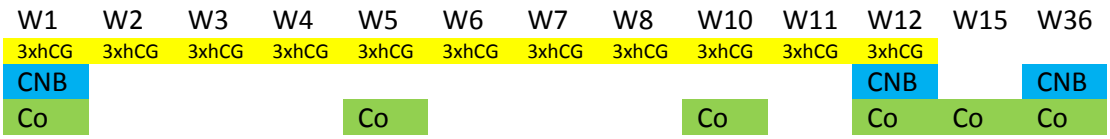
Specific aim:

To establish proof of principle in a phase 2 trial that the induction of differentiation by treatment with Ovidrel will revert a “high risk” to a “low risk” genetic signature, which would serve as a biomarker indicative of decreased breast cancer risk.

Study design

The study (Figure 1) is based on preclinical data that have demonstrated that hCG exerts a mammary cancer preventive effect, mediated by the induction of gland differentiation, which results in permanent changes in the genomic signature of this organ. This exploratory study will evaluate the genomic profile of breast epithelial cells obtained from Core Needle Biopsy (CNB) specimens performed in high risk women treated for 90 days with hCG.

Figure 1: Study Flow Chart:



**) hCG (Ovidrel) 250 µg SC 3x/week on Monday, Wednesday and Friday for 12 weeks*
**) CNB: core needle biopsy 1 sample of 8 Gauge x 2 cm under ultrasound guidance and local anesthesia.*
**) Consultation for anamnesis and physical exam with hematology, biochemistry and hCG measurement.*

Time line

Asymptomatic women of all racial and ethnic groups greater than 18 years old who are high risk for breast cancer (BRCA1 mutation); ECOG performance status 0 (Karnofski 100%); no clinical evidence of breast cancer; currently not participating in a chemoprevention trial for breast cancer or currently taking tamoxifen for chemoprevention; no history of prior cancer other than non-melanoma skin cancer; normal ovarian size report from pelvic ultrasound; ability to understand and willing to sign a written informed consent document; willing to self-administer or have the Ovidrel administered 3 X per week by a partner, (for 12 weeks, Subjects may not be pregnant because the effects of hCG on the developing human fetus are unknown. Subjects will be required to have a negative urine or serum pregnancy test within 72 hours of the 1st dose. Women of child-bearing potential must agree to use adequate contraception, including barrier method of birth control, abstinence or tubal ligation performed prior to study entry and for the duration of study participation. Should a woman become pregnant or suspect she is pregnant while participating in this study, she should inform her treating physician immediately.

Treatment plan

Women will receive three injections of 250 mcg hCG (Ovidrel) per week for a total of 12 weeks.

Recombinant human chorionic gonadotropin (r-hCG)(Ovidrel® Prefilled Syringe, Serono) will be purchased from Serono, Inc., Rockland, MA or Switzerland. Participants will receive a subcutaneous injection of 250 microgram-recombinant hCG three times a week (Monday, Wednesday and Friday) for 12 weeks. The first dose of study drug will be administered by a registered nurse. At that time, she will instruct each participant in the self-administration of the study drug in a skin fold of the upper abdominal region. Participants will return to receive doses 2 and 3, where they will be observed by the registered nurse in the self-administration of the drug for mastering of the skill and to answer any additional questions. The remainder of the drug doses will be self-administered by the participants or a partner trained in the procedure in their homes. All participants will be seen by a study physician once a month during the treatment phase for a history and physical exam and blood collection.

Protocol tissue collection

Normal breast tissue specimens will be collected by CNB at the beginning of treatment (0 day), at the end of treatment (90 days) and at six months post-treatment (270 days - 36 weeks). CNB specimens will be primarily utilized for analysis of genomic expression by cDNA microarray, RNA sequencing and epigenomic studies. In addition, a series of surrogate intermediate markers such as cytomorphologic evaluation and cell proliferation index will be analyzed. The primary objective is to compare the gene expression, and epigenomic profiles of sampled breast epithelial cells across the three time points (Fig. 1), before and after treatment, and identify differentially expressed or silenced genes (as determined by ≥ 2 fold change). Gene expression measurements will be obtained at baseline (time 0), after treatment with hCG at 90 days (time 1), and at 270 days from baseline (time 2). We are interested in comparing the expression profiles between all pairs of time points as well as across time. The comparison of profiles before and after treatment with either hCG, both at 90 and 270 days, are of particular interest (Figure 1).

Each breast sample consists of four core biopsies. One core will be used for histopathological analysis and the other three for genomic analysis. All breast core biopsies must be obtained by a qualified physician and specimens are obtained after an authorization for use and disclosure of protected health information for research has been approved by the Institutional Review Board (IRB).

From the four core biopsies collected, the first one is preserved in 70% ethanol for histopathological evaluation; and the other 3 are preserved in an RNA-preserving fluid (RNAlater®, Ambion) for genomic analysis.

For histopathological analysis the first core is expelled from the needle into a pre-labeled 15 mL tube containing 10 mL of 70% ethanol. The needle is washed in isotonic sodium chloride solution before next use. The specimen is identified with name, personal ID, date, core number and stored for up to 4 weeks at 4°C in the original collection tube until ready for shipment by express mail in wet ice.

For RNA preservation the second, third and fourth breast core biopsies are immersed in 2.0mL of RNAlater® (Ambion cat# AM7020), in an Eppendorf tube that is identified specifying the name, ID, date, and solution lot number. The tubes containing the biopsies and RNAlater® should be first incubated overnight at 4°C to allow thorough penetration of the solution into the tissue, then transfer to –20°C, where they should be kept until shipment. The shipment can be by express mail in wet ice (4°C).

Participant selection

Inclusion criteria

Asymptomatic women of all racial and ethnic groups who are carriers of a deleterious mutation in the BRCA1 gene, as determined by testing in a CLIA-certified clinical genetics laboratory; ECOG performance status 0 (Karnofski 100%); no clinical evidence of breast cancer; premenopausal women between the ages of 20-40; must have normal menstrual cycles and intact ovaries; nulliparous or parous women are eligible; not on oral contraceptives or hormone replacement therapy, and if taking them, stopped six weeks prior to the initiation of treatment and the performance of the Core Needle Biopsy (CNB) and blood drawing; currently not participating in a chemoprevention trial for breast cancer or currently taking tamoxifen for chemoprevention; no history of prior cancer other than non-melanoma skin cancer; normal ovarian size report from pelvic ultrasound; ability to understand and willing to sign a written informed consent document; willing to self-administer or have the hCG administered 3 X per week by a partner, over three months and to return for two repeat CNB. Subjects may not be pregnant because the effects of human chorionic gonadotropin on the developing human fetus are unknown. Subjects will be required to have a negative urine or serum pregnancy test within 72 hours of the 1st dose. Human chorionic gonadotropin is highly elevated in the serum of pregnant women, and is therefore not expected to be teratogenic. Nevertheless, women of child-bearing potential must agree to use adequate contraception, including barrier method of birth control, abstinence or tubal ligation performed prior to study entry and for the duration of study participation. Should a woman become pregnant or suspect she is pregnant while participating in this study, she should inform her treating physician immediately.

Exclusion criteria

Participants receiving any other agents, investigational or otherwise for the purpose of primary prevention; history of allergic reactions attributed to compounds of similar chemical or biologic composition to hCG preparations or one of its excipients; receiving medications that could interfere with the study protocol objectives such as hormonal contraceptives, androgens, prednisone, thyroid hormones, or insulin; previous treatment with follicle stimulating hormone for assisted reproduction; uncontrolled intercurrent illness including, but not limited to ovarian enlargement of undetermined origin, ongoing or active infection, NYHA ≥ class 1 congestive heart failure, unstable angina pectoris, cardiac arrhythmia, severe cognitive deficit or

psychiatric illness/social situations that could make her unable to give informed consent or would limit compliance with study requirements; HIV-positive, or infection with hepatitis B or C.

Criteria for Removal from Study:

Participants will be removed from study if they manifest any grade 2 or greater adverse reaction attributable to the recombinant hCG, non-compliance with treatment, or if the participant becomes pregnant. The reason for study removal and the date the participant was removed must be documented in the case report form. Subjects removed for reasons other than toxicity may be replaced.

Study assessment and procedures

A study physician will obtain written informed consent from the participants according to institutional and Federal guidelines. A questionnaire and a baseline physical exam will be performed along with a complete blood count, serum chemistries and a baseline beta hCG (which will also serve as a pregnancy test, when applicable). Baseline and repeat CNBs will be performed by the study surgeon.

Hematology and chemistries will be repeated monthly during the 90-day treatment period and 30 days after the final injection.

In the absence of treatment delays due to adverse events, treatment will continue for 12 weeks or until one of the following criteria applies: intercurrent illness that prevents further administration of treatment; unacceptable adverse event(s); participant decides to withdraw from the study, or; general or specific changes in the participant's condition render the participant unacceptable for further treatment in the judgment of the investigator.

Participants receiving *Ovidrel* will be followed for 30 days after completion of the study. Participants removed from study for unacceptable adverse events will be followed until resolution or stabilization of the adverse event.

Adverse events: List and reporting requirements

All side & Adverse events should be reported by the investigators.

Comprehensive Adverse Events and Potential Risks List (CAEPR)

for the Use of Recombinant Human Chorionic Gonadotropin (r-hCG) the NCI Guidelines: Adverse Event Reporting Requirements will be followed. Please see: http://ctep.cancer.gov/guidelines/

Treatment of Primary and Metastatic Breast Cancer.

Twenty-five postmenopausal women with newly diagnosed breast cancers of more than 1.5 cm were biopsied before randomization to receive either 500 micrograms recombinant chorionic gonadotropin (n=20) or placebo. No adverse events were reported for any of these participants

Treatment of Obesity with hCG in combination with a Very Low Calorie Diet (VLCD).

Human chorionic gonadotropin (hCG) alone has been orally administered to female obese volunteers participating in a double blind study. There are no reports in the literature regarding the oral administration route for comparison with our findings.

Ovarian Stimulation for Follicular recruitment with Gonadotropins for Assisted reproductive Technique (ART).:

The recombinant human chorionic gonadotropin (r-hCG) Ovidrel® PreFilled Syringe has been approved for its use as a subcutaneous injection for participant self-administration for the induction of final follicular maturation and early luteinization in infertile women who have been pretreated with follicle stimulating hormone (FSH) for *in vitro* fertilization and embryo transfer in Assisted Reproductive Technologies (ART). The side effects reported in the literature for the treatment with recombinant human chorionic gonadotropin are attributed to the pretreatment with FSH, but administration of recombinant human chorionic gonadotropin alone has not been used for the treatment of infertility.

Adverse Event Reporting

Adverse reactions must be reported to both the PI and Institutional Review Board of participating centers using rules listed below. This study will utilize the CTEP guidelines as well as the Adverse Event Expedited Reporting System accessed via the CTEP home page (http://ctep.cancer.gov/guidelines/). A written report must follow within 3 working days.

Transportation of Specimen(s) to laboratory:

All the tubes containing CNB samples collected as described need to be kept at 4°C before shipping. As an additional measure each sample tube should be wrapped with Parafilm to prevent possible leakage during shipment. The samples have to be shipped by overnight express mail to the Breast Cancer Research Laboratory at the Fox Chase Cancer Center with frozen ice packs in the Styrofoam box. The samples need to be packed tightly so they do not move during shipping. The shipments should be sent Monday through Thursday to ensure the samples are collected accordingly and not to ship on *Fridays to ensure that the samples do not sit for an extended period of time.*

Special Studies

The following special correlative studies will be performed:

#1, Histopathological evaluation of epithelial cell normality;
#2, evaluation of cell proliferation by Ki67 immunohistochemistry;
#3, serum analysis for hormone level and biomarker determinations;
4, genomic analysis of breast epithelial cells from CNB**.**

Genomic DNA and total RNA will be extracted using ALLPrep DNA/RNA Mini Kit (QIAGEN) according to manufacture instructions.

Special Correlative study #1 - Cell Morphology and Proliferation:

Histopathological analysis - Outcome measures: A minimum of two slides will be made one for histology and another one for Ki-67.

Special Correlative study #2 - Cell proliferation - Ki67 index:

Total epithelial cells will be delineated with Mayer's hematoxylin counter stain and proliferating cells with a MIB-1 (DAKO M7240) antibody for Ki-67.

Special Correlative study #3 - Serum analysis for hormone level and biomarker determination:

At the time of each CNB, whole blood will be collected by venipuncture and stored in the FCCC Breast Cancer Research Laboratory at -80^{0}C for batch determination of hormonal levels and aliquots will be stored for future studies. All serum assays for estradiol, progesterone and human chorionic gonadotropin will be conducted at the Breast Cancer Research Laboratory of the FCCC.

Special Correlative study #4 - Gene expression analysis.

Affymetrix U133plus2.0 microarrays will be used in this follow-up gene expression study involving 25 women each in the hCG treatment arm. Gene expression measurements will be obtained at baseline (before treatment) and at 3 months and 9 months after the respective treatment. We will identify genes that are

significantly up or down regulated at 3 months following treatment. In addition, we will compare the post- and pre-treatment differences and identify genes showing significantly altered expression. We will repeat the above analyses using the post-treatment expression at 9 months.

Analysis of biomarker data.

In addition to microarray data, we will obtain expression levels of Ki67, ER, PR, estradiol, progesterone and hCG using serum assays at baseline (before treatment) and at 3 months and 9 months after treatment. We will identify markers that are significantly up or down regulated at 3 months following treatment. In addition, we will compare the post- and pre-treatment differences and identify markers showing significantly altered expression. We will repeat the above analyses using the post-treatment expression at 9 months.

Statistical Considerations.

The primary efficacy analyses for the phase I/II study will concern baseline to 3-month follow-up changes in gene expression profiles measured by microarray methods as follows. For each woman *i* at biopsy time point *j* (0=pre-treatment, 1=3 month follow-up) we will compute the Mahalanobis distance between RMA normalized expression microarray data derived from the biopsy and the prototypic gene profile . We will denote this distance as *D(i,j)*. If *D(i,1)< D(i,0)* we will declare woman *i* a "success". This would indicate the post-treatment profile (genomic signature) is closer than the pre-treatment profile to the prototypic gene profile. At the conclusion of the study we will sum the number of successes for each arm. Whichever arm has a higher proportion of successes is declared the "preventive choice". With complete pre-treatment and post-treatment expression microarray data on 25 women we will correctly identify the changes in terms of "success" with at least 85% probability. Assuming a drop-out rate of 20%, we would need to enroll a total of 25 participants per arm.

In secondary analyses we will use an exact one-sided, one-sample test (α=0.05) to assess the null hypothesis of no drug induced change in *D(i,j)* (i.e., a null hypothesis that the probability of "success" is 0.5.). With data from 18 women we will have at least 80% power to detect the alternative hypothesis that the true probability of "success" is 0.80. We will also characterize the change in distance (i.e., $\Delta D(i)=D(i,1)-D(i,0)$) using standard methods (e.g., mean, median, standard deviations, etc.).

In exploratory analyses we will use LIMMA to identify genes whose expression changes significantly from pre- to post-treatment. Statistical significance will be measured by p-values adjusted for FDR using the method of Benjamini and Hochberg (108). Genes showing an FDR of less than 5% will be considered statistically significant. Biological significance will be measured by fold change. Genes showing more than 2-fold change in either direction (up and down regulated) will be considered biologically significant. Differentially expressed genes from each of the above filters will be combined and a list of common genes showing greater statistical and biological significance (lower q-values and up/down regulated by more than 2-fold) will be identified. Finally, we will use various bioinformatics methods (pathway analyses, enrichment analyses, etc.) to extract additional biological insights into the drug-induced changes in normal breast tissue. All of the above analyses will be repeated to examine baseline to 9-month follow-up differences. Assuming approximately 20000 genes of relevance in the study, a total sample size of 36 women/arrays (i.e., 18 each) will enable us to detect effect sizes of 1.39, 1.32 and 1.27 if the number of truly differentially expressed genes is 500, 1000 and 1500 respectively, with a power of 80% and a false discovery rate (FDR) of 5%. These calculations are based on the approach outlined in (105). In terms of $\log_2$ expression, an effect size of 1.32 corresponds to a 2.5-fold difference in gene expression between the control group and each treatment.

Mining for pathways and exploratory analyses:

In this study we are proposing two major high throughout genomic data platforms such as microarray and next generation sequencing platforms. The data mining and bioinformatics pipelines for both platforms are different. Following is a description of these data analyses pipelines.

Microarray data Mining:

First we will use various data mining methods to classify differentially expressed genes into functional groups based on Gene Ontology (GO) terms. Advanced statistical methodologies such as conditional hypergeometric test will be used to identify significantly enriched Gene Ontology terms such as molecular function, biological processes and sub-cellular localization. In order to identify underlying biological theme and pathways, pathway enrichment methods including Gene Set Enrichment Analysis (GSEA) (PMID: 16199517) and Q1-Q2 (PMID: 16174746) methods will be used. These methods essentially determine whether a priori set of genes such as pathways or protein interaction network module show significant association between phenotypes in a gene expression profiling experiment. We will also investigate if differentially expressed genes show significant enrichment for miRNA families, positional effect and transcription factor binding motifs. Furthermore we will investigate if profiles of co-regulated genes that are enriched for a process or pathway are similar across other studies by performing a meta-analysis. For this we will pool previous studies that are relavent to breast cancer from various public gene expression data repositories such as Gene Expression Omnibus (GEO) from NCBI and ArrayExpress from European Bioinformatics Institute (EBI).

Sequencing data mining:

The sequence data files from Illumina Genome Analyzer IIx will be processed using advanced sequencing data analysis pipelines. The analysis of sequence data will be divided into three phases (Figure 2).***Phase 1:*** Generation of primary sequence reads: First, image files from the GA IIX will be transformed using the Firecrest software module provided by the Illumina Genome Analyzer software suite. Next, the resulting intensity data will be converted to sequence data after a rigorous base-calling routine implemented by the Bustard module. The output of this phase is sequence data along with quality metrics. ***Phase 2:*** Sequence alignment and quality assessment: The raw sequence reads will be aligned to the reference genome in a multi-step process. The analytical procedures vary depending on the experiment. In case of Chip-seq, the raw sequence reads will be aligned to an unmasked reference genome (NCBI v37,hg19), whereas for RNA-seq the sequences will be aligned to both genome and RefSeq transcripts using Genomic Mapping and Alignment Program (GMAP) (http://www.gene.com/share/gmap/). We will use the Eland application module provided by

Illumina. Eland allows high throughput mapping of sequence reads by permitting no more than two mismatches per read sequence. This permits stringent mapping of reads to the genome. The sequence reads with partial mapping will be carefully evaluated for various genomic events such as fusions or alternative splicing. The sequences that cannot be mapped, even after relaxing stringent mapping criteria, will be discarded. ***Phase 3:*** Quantification and functional annotation of mapped sequence reads: Sequence reads obtained from Phase 2 will be assessed to compute statistical significance of the sequenced regions based on the abundance and length of the mapped regions of genome. Here we seek to identify locations on the genome where mapped sequence reads cluster and are observed as peaks. These peaks suggest regions of enrichment that represent in vivo locations where DNA modifications are associated. In the case of RNA-based experiments, these enriched regions on the genome suggest locations of transcribed exons of mRNA and miRNA. The significance of the enrichment-peak will be determined by calculating the probability of observing peaks with a particular height using monte-carlo simulation (PMID: 18599518). The significance of observed peaks will be assessed using the False Discover Rate (FDR) to account for multiple testing. This part of the analysis will be implemented using the FindPeaks 3.1 algorithm. The genomic locations of significantly enriched profiles thus obtained will then be combined with known transcription factor binding sites (TFBS), CpG islands, and other functional annotation data. To facilitate visualization and further processing of sequence data, unique aligned sequence reads will be transformed into UCSC browser-specific WIG files. These files can be easily uploaded onto UCSC browser. The UCSC browser allows visualization of the sequence reads as separate tracks (PMID: 19906737).

A list of genes related to the enriched profiles will be identified using UCSC annotation tables. The functional significance of the list of genes thus obtained will be examined using various bioinformatics pipelines that have been developed at FCCC. The specific methods applied in Phase 3 will depend on the type and scientific goals of the individual experiments. For example, in the case of RNA-seq based experiments designed to assess abundance of transcripts to identify digital gene expression, the analytic methods will be analogous to those used for expression microarray experiments.

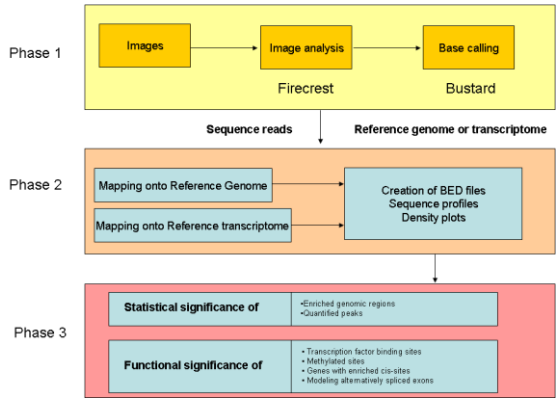


Figure 2. Workflow of data analysis for the Illumina GA IIX.Whole-Genome Chromatin Immunoprecipitation Sequencing experiments performed using the Genome Analyzer system and Pipeline Software can be parsed to the GenomeStudio ChIP Sequencing Module to create global binding site maps of DNA-associated proteins. Differential binding levels between experimental groups can be identified by comparing sequences, regions, and peaks in table or chromosome views.

In this setting, sequence read abundance will be transformed into expression measurements and normalized between data sets. Then statistical comparisons will be made to assess the differences among groups. In the case of studies aimed at identifying alternatively spliced transcripts, significant sequence reads obtained will

be modeled into an alternatively spliced exonic structure and compared with known isoforms. We will implement the ERANGE algorithm designed to quantify exonic read density from multi reads obtained from Phase 2 and Phase 3 (PMID: 18516045). The novel spliced forms will also be compared with ESTs available in dbEST.

Expected results and significance.

We are expecting that hCG is inducing genomic signature of protection in the breast. The signature must comprised of a distinctive set of markers indicating that chromatin remodeling had occurred, as indicated by the expression of genes associated with cell differentiation, cell proliferation, programmed cell death and DNA repair damage among others. It is also expected that specific non coding sequences and microRNAs will be expressed as part of the signature, as well as a complete epigenetic make up that will give us a more accurate and definitive profile of the mechanism of prevention. This knowledge will serve as the basis for establishing novel genomic signatures as intermediate biomarkers for larger preventive clinical trials at the completion of this project. The success of this project could significantly change the guidelines for risk reduction strategies, which currently recommend multi-modality screening and prophylactic mastectomy. Therefore, this application is highly innovative by the proposed approach of identifying how hCG activates differentiation pathways in the breast without affecting other organs and functions, thus facilitating the utilization in breast cancer prevention in the general population, going beyond the currently targeted high-risk groups

Further reading

J Ph Janssens, I Russo, J Russo, J Breitmeyer, M Lopez-Bresnahan.
Recombinant Human Chorionic Gonadotropin in postmenopausal primary breast cancer patients.
Monographs of the National Cancer Institute,, 2000.

J Ph Janssens, I Russo, J Russo, J Breitmeyer, M Lopez-Bresnahan.
Inhibitory effect of recombinant human chorionic gonadotropin (rhCG) in metastatic breast cancer in postmenopausal women.
Monographs of the National Cancer Institute, 2000.

Russo J, Janssens J, Russo IH.
Recombinant human chorionic gonadotropin (r-hCG) significantly reduces primary tumor cell proliferation in patients with breast cancer.
23rd Annual San Antonio Breast Cancer Symposium- Breast Cancer Research and Treatment, 2000.