# *Analysis of methylation status of CGB genes in common non-trophoblastic cancer*

*Thesis submitted to Middlesex University as partial fulfilment of the requirements for the degree of Master by Research*

Ivan Petkov Punev

M00353698

Director of study: Dr Beata Burczynska

Supervisors: Dr Helen Roberts and Dr Celia Bell

Department of Natural Sciences

Faculty of Science and Technology

Date of submission: 21st January 2020

# Abstract

hCG is a complex heterodimeric glycoprotein hormone essential during pregnancy. The hormone and its unique β-subunit are recognised tumour markers involved in enabling anti-apoptotic pathways, promotion of invasion and angiogenesis, and supporting tumour proliferation leading to more aggressive cancer and poor prognosis. The limiting factor for the hCG production is the presence of the β-subunit which is encoded by 4 non-allelic genes, *CGB3-9*. These genes are regulated via transcription factors AP2α, SP1, OCT3/4 and possibly DNA methylation. DNA methylation is a repressive mark found in cytosines in DNA. The aim of this study is to investigate the role of DNA methylation in relation to *CGB3-9* genes activation and hCG/hCGβ production in non-trophoblastic cancer cell lines. Bioinformatics methods were used to establish the region of interest and predict CpG islands (CGIs) within the promoter region of the target genes. Methylation-sensitive PCR and next generation sequencing were used to assess *CGB3*-9 promoter methylation in the cancer cell lines. The transcription level of *CGB3*-9 genes was investigated using qRT-PCR with TaqMan probes. Secretion in media of hCG/hCGβ was investigated with ELISA. The results demonstrated the presence of a CGI associated with *CGB7* which matches previous reports of the transcription of the gene in normal and cancer tissue. Notably, this study confirmed previously reported hypomethylation in trophoblastic cell lines. Hypomethylation of the genes was found in HeLa cell line which matches literature reported transcription studies. However, other results were discordant or inconclusive leading to no significant correlation between methylation and transcription levels. This first sequencing study of *CGB3*-9 methylation in non-trophoblastic cancer cell lines was inconclusive for the role of the epigenetic mark in *CGB3*-9 genes. Further investigations need to be performed to elucidate the role of methylation.

# Acknowledgments

Foremost, I would like to express my gratitude to my director of study Dr Beata Burczynska and my supervisors Dr Helen Roberts and Dr Celia Bell. Thank you for giving me this opportunity to complete this project and thank you for your patience, guidance and motivation on this learning journey.

I would also like to thank my colleagues and friends from the GAA team, research students and colleagues from the Dept. of Natural Sciences for the advice, stimulating discussions and support which helped me to arrive here.

Last but not least, I want to express my gratitude and appreciation for my family and friends who never stopped believing in me and were there for me through the difficult times.

# Table of Contents

# Abbreviations

| | |
|---|---|
| **(i)EVT** | (Invasive) Extravillous Cytotrophoblast |
| **5mC** | 5-Methyl Cytosine |
| **aa** | Amino Acid |
| **AKT** | Protein Kinase B |
| **ANOVA** | Analysis Of Variance |
| **AP2α** | Activator Protein 2 (Alpha) |
| **bp** | Base Pair |
| **BS** | Bisulphite Sequencing |
| **cAMP** | Cyclic Adenosine Monophosphate |
| **CD** | Cluster Of Differentiation |
| **cDNA** | Complementary DNA |
| ***CGA*** | Chorionic Gonadotropin Subunit Alpha |
| ***CGB*** | Chorionic Gonadotropin Subunit Beta |
| **CGI** | CpG Island |
| **c-Jun** | Jun Proto-Oncogene |
| **CpG** | Cytosine Guanine Dinucleotide |
| **CRE** | cAMP Response Element |
| ***DAPK*** | Death Associated Protein Kinase |
| **DNA** | Deoxyribonucleic Acid |
| **DNase** | Deoxyribonuclease |
| **DNMT** | DNA Methyltransferase |
| **EDTA** | Ethylenediaminetetraacetic Acid |
| **ELISA** | Enzyme Linked Immunosorbent Assay |
| **EMT** | Epithelial-To-Mesenchymal Transition |
| **ERK1/2** | Extracellular Signal-Regulated Kinases |
| **EZH2** | Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit |
| **F** | Forward |
| **FBS** | Foetal Bovine Serum |
| **FSH** | Follicle-Stimulating Hormone |
| **H3K27me** | Lysine 27 Histone 3 Methylation |
| **hCG** | Human Chorionic Gonadotropin |
| **hCG-H** | Hyperglycosylated hCG |
| **HDAC1/2** | Histone Deacetylase 1/2 |
| **kDa** | Kilo Dalton |
| ***(h)LH*** | (human)Luteinising Hormone |
| **LH/hCG-R** | LH and hCG Common Receptor |
| **M** | Methylated |
| **MSA** | Multiple Sequence Alignment |
| **MS-HRM** | Methylation Sensitive High Resolution Melting |
| **MSP** | Methylation Sensitive PCR |
| **MTA3** | Metastasis Associated Protein |
| **NGF** | Nerve Growth Factor |
| **NT** | Non-Trophoblastic |
| **NuRD** | Nucleosome Remodelling Deacetylase |
| **OCT3/4** | Octamer Binding Transcription Factor |
| **PBS** | Phosphate Buffered Saline |
| **PC** | Principal Component |
| **PCA** | Principal Component Analysis |
| **PCR** | Polymerase Chain Reaction |

| | |
|---|---|
| **PDGF** | Platelet-Derived Growth Factor |
| **PKA** | Protein Kinase A |
| **PPARγ** | Peroxisome Proliferator Activated Receptor Gamma |
| **PS** | Penicillin Streptomycin |
| **qRT-PCR** | Quantitative Reverse Transcription PCR |
| **R** | Reverse |
| **RE** | Restriction Enzymes |
| **RM-ANOVA** | Repeated Measures ANOVA |
| **RNA** | Ribonucleic Acid |
| **RNase** | Ribonuclease |
| **SAM** | S-Adenosyl Methionine |
| **SLUG** | Snail Family Transcriptional Repressor 2 |
| **SMAD2** | Mothers Against Decapentaplegic Homolog 2 |
| **SNAIL** | Snail Family Transcriptional Repressor 1 |
| **SNP** | Single Nucleotide Polymorphism |
| **SP1/3** | Specificity Protein 1/3 |
| **ST** | Syncytiotrophoblast |
| **T** | Trophoblastic |
| **TDG** | Thymine DNA Glycosylase |
| **TET** | Tet Methylcytosine Dioxygenase |
| **TF** | Transcription Factor |
| **TGFBR** | TGF-β Receptors |
| **TGFβ** | Transforming Growth Factor Beta |
| **Th** | T-Helper Cell |
| **TSE** | Trophoblastic Specific Element |
| **TSH** | Thyroid Stimulating Hormone |
| **TSS** | Transcription Start Site |
| **TWIST** | Twist Family BHLH Transcription Factor 1 |
| **U** | Unmethylated |
| **UTR** | Untranslated Region |
| **VCT** | Villous Cytotrophoblast |
| **VEGF** | Vascular Endothelial Growth Factor |
| **VEGF-R** | Vascular Endothelial Growth Factor Receptor |
| **WG** | Week Of Gestation |

# 1. Introduction

## 1.1 hCG function and structure

### 1.1.1 Normal physiology

Human chorionic gonadotropin (hCG) is a physiologically significant protein heterodimer highly synthesised in the first three months of pregnancy (Guo *et al.*, 2011). In normal physiology, hCG, or the pregnancy hormone, is released by the placenta. Its main role is to support pregnancy by stimulating progesterone secretion, promoting appropriate implantation, and preventing endometrial apoptosis (Rull *et al*., 2008).

hCG hormone is only found in humans and subhuman primate species. The hormone prevents corpus luteum regression and maintains its (ovarian) progesterone secretion until the shift in progesterone synthesis to placenta is completed by 9$^{th}$ week of gestation (WG) (Fournier, 2016**,** Rao, 2016). hCG is the first signal released by the conceptus after implantation and it is used to detect pregnancy. hCG and its beta subunit are detected in maternal blood from 1WG and their levels increase until 10-12WG then steadily drop until delivery (Fig. 1). The alpha subunit of hCG keeps increasing until delivery (Fournier, 2016).

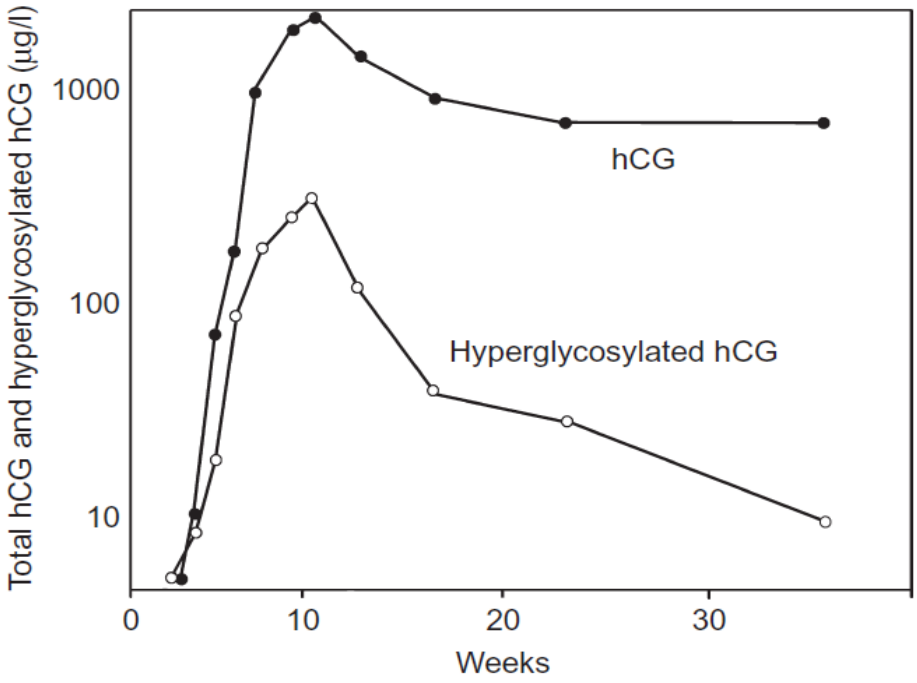



**Fig.1** Levels of hCG and its hyperglycosylated form in the serum during pregnancy. Adapted from (Cole, 2015)

The hCG molecule acts as a ligand activating the LH/hCG-R, a receptor that is shared between hCG and luteinising hormone (LH) (Guo *et al.*, 2011). This is due to high similarity between the 2 hormones. hCG binds to this receptor with higher affinity than LH which leads to hCG being more active *in vivo* (Casarini *et al.,* 2012). These receptors are widely distributed in the body including non-gonadal tissues. However, there is selectivity in the non-gonadal tissues: liver, lungs, kidney, spleen and skeletal smooth muscle in adults do not contain the receptor. This selectivity is not present in the human foetus which may indicate the possibility of a role of hCG in the promotion of tissue differentiation in embryo (Rao,2016).

In pregnancy, a complex interaction between the mother and conceptus occurs. The blastocyst needs to implant or invade the maternal endometrium and establish link between the developing baby and the mother for the purposes of its optimal growth and development *in utero*. This is achieved via the placenta that develops upon implantation from the trophoblast, the outer layer of the blastocyst. The trophoblast matures into the structures: villous cytotrophoblast (VCT), syncytiotrophoblast (ST), and invasive extravillous (cyto)trophoblast (iEVT) (Singh *et al.,* 2010).

The secretion of hCG in the formed placenta is done by the endocrine tissue known as syncytiotrophoblast (Fournier, 2016*,* Rao, 2016). The ST develops from the trophoblast via the autocrine action of hCG when the blastocyst invades the maternal decidua (endometrium) during implantation (Fig. 2) (Singh *et al.*, 2010). hCG promotes the formation of syncytiotrophoblasts by binding to the common LH/hCG-R which triggers secondary messengers like cAMP and inositol phosphates. This in turn starts the process of villous cytotrophoblast fusion into multinucleated ST (Fournier, 2016).

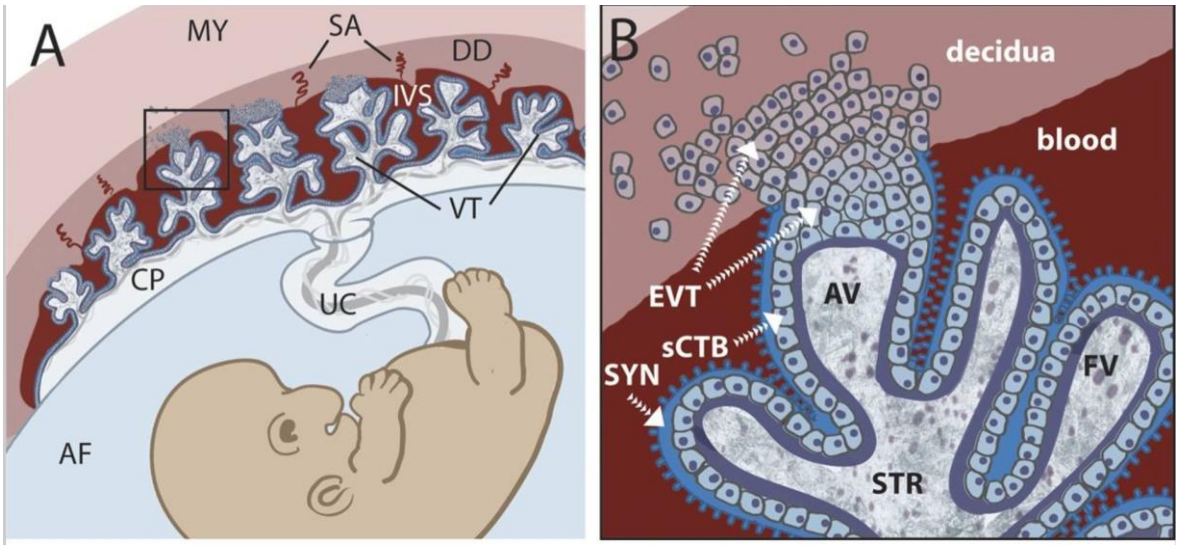



**Fig.2** Placental structure at approximately 6 weeks of gestation. A. Orientation of foetus and placenta. B. Closer view of the interface between mother and foetus. Abbreviations: MY: myometrium; SA: spiral arteries; DD: decidua; IVS: intervillous space filled with maternal blood; VT: villous tree; CP: chorionic plate; UC: umbilical cord; AF: amniotic fluid; AV: anchoring villi; FV: floating villi; SYN: syncytiotrophoblast; sCTB: subsyncytial cytotrophoblast; STR: villous stroma; EVT: extravillous cytotrophoblast. Adapted from Robbins *et al.*, 2012.

The invasion of the endometrium is essential to establish the foeto-maternal link needed for a successful pregnancy. Trophoblast cells cross into the maternal decidua to begin process of remodelling the maternal spiral arteries to establish the vasculature connecting the mother and the baby (Singh *et al.*, 2010). The trophoblast cells taking part in this process are the invasive extravillous trophoblast. Compared to the villous cytotrophoblast these cells express less cell adhesion molecules and more proteinases that degrade the extracellular matrix, allowing their invasive and migratory properties (Singh *et al.*, 2010).

*In vitro* studies have identified that the iEVT exhibits a different isoform of hCG – the hyperglycosylated hCG (hCG-H). Conditioned media from iEVT promotes trophoblast invasion suggesting that hCG-H isoform is responsible for the invasive properties of the trophoblast (Fournier, 2016). In ST conditioned media no hyperglycosylated protein was found and the media did not elicit invasion of the trophoblast. *In vivo* hCG-H concentration increases and peaks during 9WG, and steadily decrease thereafter reaching a plateau at the beginning of 2$^{nd}$ trimester corresponding to the end of the trophoblastic invasion. This further supports that hCG-H is responsible for the invasive properties of iEVT. It has also been proposed that the invasive properties may be a result from a pathway independent of LH/CG-R activation (Fournier, 2016).

hCG also promotes angiogenic activity in the placenta, maintains quiescence of myometrium and helps in development of embryonic immunotolerance in the mother (Fournier, 2016, Rull *et al*., 2008). The angiogenic activity is proposed to be due to hCG inducing upregulation of vascular endothelial growth factor (VEGF) and indirectly triggering angiogenesis. Also, LH/hCG-R has been found to be expressed in vascular endothelial cells. hCG acts on the endothelial cells by promoting their migration, permeability, and proliferation and increasing capillary sprout formation. Further to this direct effect hCG promotes vessel maturation by recruiting perivascular cells. hCG has also been reported to promote spiral artery angiogenesis (Połeć *et al.*, 2014, d'Hauterive *et al.*, 2011). There is some evidence to suggest that the pregnancy hormone may contribute towards the increase of the foeto-placental perfusion by dilation of the uterine blood vessels. This is thought to promote the acquisition of nutrition from the maternal circulation to meet the needs of the developing baby and to help with the foetal waste removal (Rao, 2016).

Upon implantation the hCG molecule contributes to the quiescence of the myometrium (Rao, 2016). Myometrial quiescence involves downregulation of the myometrium gap-junctions. hCG binds directly to myometrial LH/hCG-R to reduce their expression. Also, it has been shown that exogenous hCG inhibits oxytocin-induced labour (Ambrus and Rao, 1994). The decrease of hCG levels towards the end of pregnancy may allow for the activation of the myometrium to promote successful delivery of the baby (Rao,2016).Further to that, it has been shown that myometrium from term labour contains less LH/hCG-R, thus less chance of inhibition of labour (Ambrus and Rao, 1994).

Maternal immunotolerance is essential for successful pregnancy. The conceptus is a semi-allograft that needs to be protected from the maternal immune system to ensure there is no miscarriage due to immune rejection upon implantation. To ensure this the maternal leukocytes need to be signalled to tolerate the immunologically distinct blastocyst (Tsampalas *et al.*, 2010, Schumacher *et al.*, 2009). The main agents of this tolerance or rejection process are the T-cells. hCG has been shown to downregulate the Th1 cells, $CD8^+$ T-cells and macrophages which take part in the rejection of non-self-molecules. However, the pregnancy hormone increases the Th2 cells and the ratio of $CD4^+CD25^+/CD4^+$ T-cells in the spleen and pancreatic lymph nodes which helps tolerate the conceptus by the maternal immune system (Tsampalas *et al.*, 2010, Schumacher *et al.*, 2009).

Other roles of the hCG have also been reported. It has been documented that hCG hormone may play a role as anti-HIV agent *in utero* protecting the developing foetus from infection. hCG has also been proposed to reduce the manifestation of autoimmune disorders such as Rheumatoid arthritis and Diabetes Mellitus Type I due to the immunotolerance effects induced by the hormone during pregnancy. It has been suggested that hCG plays a role in the promotion of growth and differentiation of human foetal tissues. hCG has been shown to promote the growth and development of the breast tissue whereby the hormone triggers irreversible differentiation of the epithelial tissue to secretory cells which are immune to carcinogenic triggers (Rao, 2016).

### 1.1.2 hCG activation of LH/hCG-R and other alternative pathways

Activation of the LH/hCG-R by hLH or hCG can trigger several cell-signalling pathways: cAMP/PKA, ERK1/2 and AKT (Fig. 3) (Borisova *et al.*, 2017, Casarini *et al.*, 2012). hCG binding mainly activates the cAMP/PKA pathway by which the secondary messenger cAMP is released to modulate responses stimulating progesterone synthesis in the ovary and morphological changes such as angiogenesis in vascular endothelial cells (Połeć *et al.*, 2014, Casarini *et al.*, 2012, d'Hauterive *et al.*, 2011). ERK1/2 and AKT pathways are more often activated by hLH. The activation of the 2 pathways modulates proliferation, differentiation and survival (Casarini *et al.*, 2012).

The receptor is a glycoprotein of 675 amino acids encoded by a single gene. LH/hCG-R has an external domain rich in leucine residues that is responsible for the recognition and binding of hCG/LH coupled to 7 domains spanning through the membrane and an intracellular portion coupled with G-protein (Fig. 3). Once activated by hCG the receptor triggers a signal transduction via the secondary messenger cAMP part of the cAMP/PKA pathway to initiate progesterone secretion and fusion process of VCT (Tsampalas *et al.*, 2010).

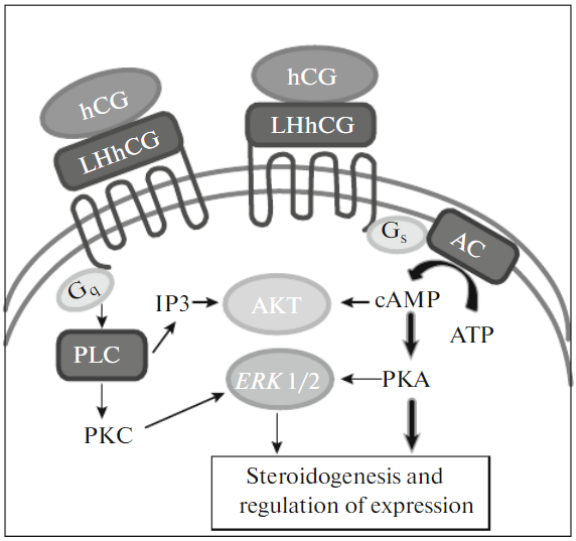



**Fig.3** LH/hCG-R receptor and pathways activated by hCG. Thicker arrows show characteristic pathway used. PKA - protein kinase A, PKC – protein kinase C, AKT—protein kinase B, ERK1/2—serine/threonine MAP kinases. Adapted from (Borisova *et al.*, 2017)

Research has shown that in some instances hCG may act independently of its native LH/hCG-R. Proliferation of uterine natural killer cells, mediated via hCG, has been shown to act via the mannose receptor CD206 (Tsampalas *et al.*, 2010). Murine LH/hCG-R knock out models have shown hyperglycosylated hCG mediated angiogenic activity suggesting alternative receptor activation. It was reported that angiogenic activity is mediated via TGF-β receptor (Fournier, 2016). Another study has shown empirically that hCG-β subunit is activating the TGF-β receptor to induce EMT (epithelial to mesenchymal transition) in colorectal cancer (Kawamata *et al.,* 2018). However, this remains a controversial topic as some of the preparations used in these studies may have been cross contaminated, with for instance, TGFβ**,** and other growth factors may have in fact activated the TGF-β receptor (Koistinen *et al.,* 2015).

All in all, the underlying processes in the developing conceptus are orchestrated in big part by the hCG hormone. Pregnancy is strictly regulated to ensure normal physiology is maintained (Singh *et al.*, 2010). An imbalance of the hormone can result in various pathologies. Lower hCG levels are associated with ectopic pregnancies or miscarriages. In cases of low hCG shallow implantation can occur and thus preeclampsia can arise (Fournier, 2016; Singh *et al.,* 2010). High levels of hCG in maternal blood can link with trisomy of the embryo or can be a marker of cancer such as choriocarcinoma (Keay *et al.,* 2004).

## 1.1.3 hCG Structure

Human chorionic gonadotropin is a heterodimeric molecule which shares similarities with luteinising hormone, follicle-stimulating hormone (FSH), and thyroid stimulating hormone (TSH). These hormones have a common α subunit and the β subunit determines their respective biological functions (Rull *et al*., 2008; Stenman *et al.,* 2004). hCG is also part of the diverse cysteine-knot growth factor superfamily together with TGF-β, NGF, PDGF (Cole, 2015, Iyer and Acharya, 2011). hCG is a glycosylated protein of about 37kDa that has 2 subunits which are non-covalently associated. The alpha subunit consists of 92 amino acids and the beta – consists of 145 amino acids (Fournier, 2016, Rao, 2016).

The tertiary structure of the hCG protein is maintained by a cysteine knot in the centre of each subunit. This cysteine knot is formed by 3 disulphide bridges linking the protein molecule of each subunit in a way that forms 3 loops. When the 2 subunits of hCG form the intact protein, they bind non-covalently in an antiparallel manner such that the first and third loop of one subunit and the second loop of the other molecule lie on the same side of the knot (Fig. 4) (Berger *et al.*, 2013).

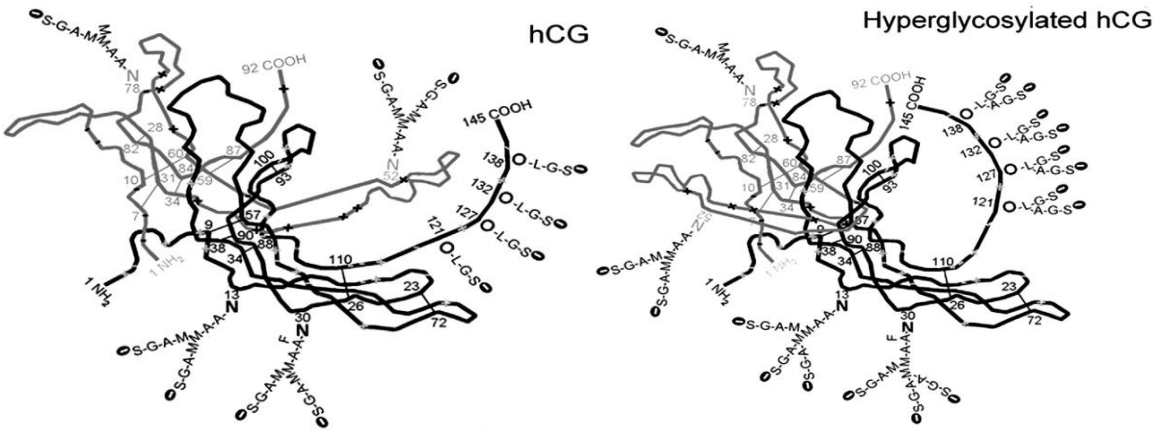



**Fig.4** Structures of hCG and hyperglycosylated hCG with proposed attached sugar moieties. (Cole, 2015)

Additionally, the protein structure is complicated by glycosylation which creates a diverse variety of hCG isoforms. Glycosylation is the covalent binding of a carbohydrate residue to a peptide. It determines the folding, receptor binding and half-life of the hCG molecule (Berger *et al.*, 2013).The sugar branches in hCG can be O-linked oligosaccharide containing N-acetylgalactosamine linked to a serine

residue or N-linked oligosaccharide containing N-acetylglucosamine residue linked to asparagine residue. The alpha subunit has 2 N-glycosylation sites, and the beta - 2 N-glycosylation and 4 O-glycosylation sites located in Carboxyl terminus of the polypeptide (Fournier, 2016).

Depending on the site of production, whether it is during pregnancy or in cancer, hCG has different sugar moieties, conferring different functions (Fournier, 2016; Cole, 2007). Studies investigating the difference in the hCG isoforms have reported important sites on the protein molecule that seem to distinguish between hCG and it highly glycosylated form hCG-H. The serine residues 127, 132, and 138 presenting with predominantly bi-antennary O-linked oligosaccharides are typical for the hCG-H whereas monoantennary O-linked oligosaccharides are more consistent with hCG (Fig. 4) (Cole, 2007).

hCG-H is a family of glycoproteins with 40-43 kDa molecular weight. hCG-H has tri-antennary N-linked oligosaccharide and double molecular size (hexasaccharide) O-linked oligosaccharides. As previously mentioned, it has been reported that hCG-H is important for the invasive properties of the iEVT in normal pregnancy and associated with successful embryo implantation (Fournier, 2016). The larger hCG-H isoform has been reported in choriocarcinoma, testicular germ cell malignancies and other invasive diseases supporting possible role of the hCG-H in promoting cellular invasion (Cole, 2007).

Due to high amount of glycosylation the protein dimer of hCG-H does not fold properly resulting in exposing the central cysteine knot. This knot is very similar to structures found in TGF-β and other members of the cysteine knot growth factor family (Berger *et al.*, 2013). The exposed cysteine knot in hCG-H may interact with TGF-β receptors to reduce trophoblast apoptosis during the 1$^{st}$ trimester and enhanced invasion associated with secretion of metalloproteinases (Fournier, 2016).

Given the involvement of hCG in pregnancy and pathological conditions such as miscarriages and cancer, it is important to be able to detect the hormone for prognostic and diagnostic purposes (Iles *et al.,* 2010, Uuskula *et al*., 2010). Discovering immunologically and biologically important epitopes to design antibodies for hCG is essential for the hormone's detection with methods such as the commonly employed ELISA. However, the variety of hCG isoforms due to glycosylation and thus folding poses a challenge in the epitope identification. A

further complication is the fact that the genes coding for hCG (*CGB3-9* genes) have evolved from the *LH* gene (*CGB4*). The hLH protein shares > 85% similarity with hCG. The main difference between the 2 different proteins is in the extended carboxyl terminus of hCG. This poses the challenge of designing and antibody that is specific only to hCG (Berger *et al.*, 2013).

hCG metabolism in the body generate different fragments or variants that have been applied in different assays in aiding diagnosis. There are 6 established variants which have been internationally defined: intact hCG, free β-subunit, free α-subunit, nicked hCG, nicked β-subunit, β-subunit core fragment. (Table 1) (Berger *et al.*, 2013, Stenman *et al.,* 2006). Assays based on one or multiple of these variants can give valuable information for the detection of pregnancy status and progression and cancer. hCG and its subunits can be detected in serum and urine. The nicked forms of hCG and its β-subunit are mainly found in urine but can be found in serum. The β-subunit core fragment comprises main hCG immunoreactivity in urine (Stenman *et al.*, 2006).

**Table 1** hCG and hCG-related variants. Adapted from (Berger *et al.*, 2013)

| Symbol | Molecular definition |
|---|---|
| hCG | Intact αβ heterodimer, bioactive |
| hCGn | Nicked αβ heterodimer, nicks in the region of aa hCGβ44-48 |
| hCGβ | Intact noncombined free hCGβ-subunit, aa hCGβ1-145 |
| hCGβn | Nicked hCGβ, nicks in the region of aa hCGβ44-48 |
| hCGβcf | Core fragment of hCGβ; aa hCGβ6-40 linked to hCGβ55-92 |
| hCGα | Noncombined free α-subunit of hCG; aa hCGα1-92 |
| Less well-defined hCG variants | |
| hCGβCTP | Carboxylterminal extension of hCGβ, aa hCGβ109/114-145 |
| -CTPhCG | hCGβ truncated core hCG, missing most of the hCGβCTP (aa hCGβ121-145) |
| -CTPhCGβ | hCGβ truncated core hCGβ (aa hCGβ1-120), missing most of the hCGβCTP |

Therefore, it is important to design monoclonal antibodies for the detection of the different molecules such as intact hCG, hCGβ, or metabolites such as the nicked hCG. These molecules share some epitopes but also have some unique ones due to their difference in folding. Extensive research has been performed to identify epitopes detecting one or combination of the hCG variants without cross-reactivity with similar molecules such as LH (Berger *et al.*, 2013).

### 1.1.4 Cancer and varied forms of hCG

Trophoblastic cancers, like choriocarcinomas and hydatidiform moles, and germ cell tumours, such as dysgerminomas and non-seminomatous germ cell tumours, are found to have elevated levels of intact hCG. Raised hCG and its β-subunit have also been confirmed in patients with non-trophoblastic tumours such as bladder cancer, gastrointestinal cancer and breast cancer (Guo *et al.*, 2011; Iles *et al.,* 2010; Stenman *et al.,* 2004). hCG elevation has been correlated with poor outcome in variety of different tumours from various origins including trophoblastic and non-trophoblastic. In the case of bladder cancer the elevation of the beta subunit is associated with a more aggressive cancer, with poor prognosis and resistance to available cancer therapies (Iles *et al.*, 1996). Intact hCG and also the free β subunit of hCG are used as a sensitive tumour marker for trophoblastic and testicular germ cell tumours (Rull *et al.,* 2008).

Some cancers express the hCG in small amounts thought to act as an autocrine growth factor. These small amounts are quickly eliminated and usually not detected in circulation. Some evidence of this role is found in non-trophoblastic lung cancers and trophoblastic neoplasms (Rao, 2016). hCGβ is also found to play a role as a cancer growth factor by blocking TGFβ-induced apoptosis rather than stimulation of mitosis and cancer proliferation (Iles *et al.*, 2010). This effect is thought to occur due to structural similarity between hCGβ and TGFβ as both can form homodimers and act as growth stimulating factors via their receptors (Cole, 2015). The autocrine mechanism behind this is thought to be due to the hCG-β molecule blocking the apoptosis induced by TGF-β pathways. Antibody against the β-subunit of hCG reverses the effect of cell growth. Study has confirmed this possibility of apoptosis blocking by showing that the β-subunit can bind to TGF-β receptors (TGFBR type 2) and act as antagonist. This further confirms previous research where the β-subunit has been proposed to competitively bind to TGFBR2 to prevent apoptosis activation (Cole, 2007).

A similar role in the inhibition of apoptosis due to hCG-β has been found *in vitro* in cervical cancers. In Jankowska *et al.*, (2008) reduction of the beta subunit increases programmed cell death rate confirming its action of growth factor by blocking of apoptosis. In ovarian cancer cell lines expressing hCG-β it has also been observed that when apoptosis-related proteins were investigated there was an upregulation of the pro-survival protein BCL-XL and decrease of the pro-

apoptotic protein Bad active form which suggests decreased apoptosis in the cells (Guo *et al.*, 2011).

Further proof of tumour proliferation role of hCG beta has been observed in ovarian cell lines (Guo *et al.*, 2011). Introduction of β-hCG vector into ovarian non-tumorigenic cell lines leads to overexpression inducing increased proliferation in those cells and potential *in vitro* tumour transformation. The overexpression of the beta subunit upregulates cyclins E and D1 together with their partner kinases Cdk-2 and Cdk-4 and cdk-6. This leads to progression through the G2 checkpoint of the cell cycle leading to increased cellular proliferation. Furthermore, the cells showed to be anchorage-independent suggesting increasing tumourigenicity and invasiveness (Guo *et al.*, 2011). Injecting xenografts of these cell lines in nude mice *(in vivo*) showed that the hCGβ overexpressing cells developed into tumours further supporting the *in vitro* data of the β-hCG mediated transformation of the surface epithelial cells into tumour cells (Guo *et al.*, 2011).

A study in colorectal carcinoma has also reported epithelial transformation due to hCGβ (Kawamata *et al.*, 2018). In the study transfected cell lines overexpressing hCGβ were tested for EMT markers SLUG, SNAIL, TWIST, phosphorylated SMAD2 and E-cadherin. Their results show that the markers change their expression as observed during epithelial transformation. They further propose that the EMT process occurs due to TGFBR activation since incubation with inhibitors for the type 1 and 2 TGFB-R reversed to normal the levels of EMT markers SNAIL and TWIST. In the study they also report the overexpression of the β-subunit of hCG increased invasiveness and migratory property of the cells (Kawamata *et al.*, 2018).

The hyperglycosylated form of hCG has also been reported to play a role in carcinogenesis. In trophoblastic disease, the hyperglycosylated hCG form has been associated with invasive properties. In cases of trophoblastic cancers hCG-H presence has been detected (Cole, 2007). *In vitro* studies with choriocarcinoma cell lines and testicular germ cell carcinoma cell lines have high levels of hCG-H and some amounts of hCG-β but no detectable regular hCG. The cell lines present with increased cellular growth, invasive properties, tumour formation. Similar trend was observed in *in vivo* models with nude mouse xenografts. These properties where lost after the cells were incubated with monoclonal antibodies against hCG-H confirming the invasive and metastasis-inducing properties of the

molecule. Furthermore, the hyperglycosylated form seems to have low interaction with the regular LH/CGR in comparison with the regular isoform suggesting activation of a different receptor to induce the invasive autocrine functions (Cole, 2007).

As mentioned above hCG-H and the beta subunit share a specific cysteine knot structure with a few other cytokines like TGF-β (Iyer and Acharya, 2011). This similarity has initiated research surrounding the possible interactions of hCG-H with the receptors of these cytokines and more specifically TGF-β. TGF-β and its pathway are involved in trophoblast invasion in normal and pathophysiology. hCG-H as well as the free β-subunit have both been proposed to interact with TGFBR to enhance cell proliferation and invasion through modulation of the apoptosis mechanism induced by TGF-β (Cole, 2015, Iles *et al.*, 2010). As to why both hCG-H and free β-subunit behave in such manner is that both present hyperglycosylation which in turn changes the protein structure to expose the cysteine knot which then interacts with the TGFBR. In the case of trophoblastic disease there is presence of α-subunit and this allows for hCG-H to be formed. However, non-trophoblastic malignancies tend to produce only the β-subunit in hyperglycosylated form which still has the cysteine knot. There has been research suggesting the existence of hCG-ββ homodimer, which is suggested to act as a growth factor similar to the homodimer of TGF-β (Iles *et al.*, 2010, Cole, 2007)

Further role of hCG in cancer relates to tumour angiogenesis. As mentioned previously hCG works in tandem with VEGF to establish the neovascularisation in the placenta (Połeć *et al.*, 2014, d'Hauterive *et al.*, 2011). As such it is thought similar process should occur in cancer. Arieta *et al.*, (2009) reports hCG as independent angiogenic factor in testicular germ cell tumours not related to VEGF expression. Further to that both VEGF and hCG are considered part of the cysteine knot growth factor family together with TGF-β. Therefore, it is possible that hCG may play a role in angiogenesis via the VEGF-R (Iles *et al.*, 2010).

## 1.2 *CGB* Gene Cluster

As established so far hCG has important roles both in normal and pathological states. However, understanding of the molecular mechanisms activating the gene and subsequently protein production, are still unclear. The complexity of the hCG protein structure is reflected in the genetics of the hormone. The common alpha subunit is encoded by single gene *CGA* on chromosome 6q21.1-23 and the gene

is expressed in large excess (Fournier, 2016, Cole, 2015). However, the beta subunit is encoded by a cluster of genes located on chromosome 19(19q13.32) and clustered together with the hLH-β gene (Fig.5). The beta subunit is unique for hCG; however, it is still showing high similarity with luteinising hormone (Berger *et al.*, 2013).

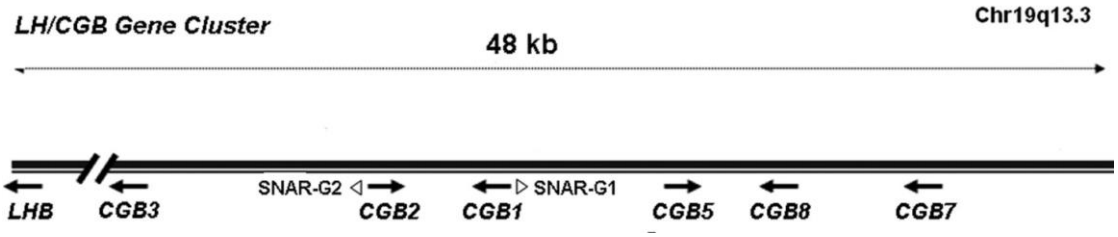



**Fig.5** *LHB/CGB* gene cluster. Diagram representing the relative positions of the genes Adapted from Burczynska *et al.,* 2014.

There are 6 number of genes referred to as *CGB1-9* divided into two types. The *CGB* genes are highly homologous but transcribed at different levels in trophoblastic tissues. Parrott *et al. (*2011) groups *CGB3-9* as type I genes and *CGB1/2* as type II. All *CGB* genes have 3 exons but type I and II genes produce different transcripts. Type I genes produce one transcript with the exception of *CGB*7 which has splice variants. Type I genes produce the mRNA needed to translate the hCGβ subunit 145aa. Type II genes have been found to have multiple splicing variants varying from the type I gene transcript (Rull *et. al*., 2008). This is because the *CGB*1-2 genes have a specific long insert of approx. 730 base pairs in the 5' untranslated region (UTR). This leads to Type II producing a transcript corresponding to predicted 132 aa sequence of unknown protein (Rull and Laan, 2005).

Type I genes are 4 genes and 2 alleles encoding for the biologically active hCGβ subunit. *CGB6* is an allele of *CGB7* and *CGB9* is an allele to *CGB3* (Stenman *et al*., 2004). *CGB3-9* genes can further be grouped in two types based on the translated beta subunit protein. *CGB*7 (1) has 3 aa difference from type 2 proteins (*CGB*3, 5, 8) (Fournier, 2016; Berger *et al.*, 2013). *CGB6/7* genes are expressed at a low level in non-trophoblastic tissues, where *CGB3/9, CGB5,* and *CGB8* are actively transcribed in placenta, testis and malignant tumours (Stenman *et al*., 2004)**.**

The most abundantly expressed in pregnancy are the *CGB*5, *3* and *8* genes (type I) (Glodek *et al.*, 2014). According to some studies *CGB*5 is with highest

expression in the placenta (Fournier, 2016). Rull and Laan (2005) reported *CGB*8 as the most transcriptionally active during pregnancy. However, these differences may be owed due to different polymorphic variant of tested genes or individual differences in trophoblastic differentiation. *CGB*7 has low expression profile in comparison to the other Type I genes (Rull and Laan, 2005).

*CGB1* and *2* (type II) were considered to be pseudogenes (Stenman *et al.*, 2004). However, recent research shows that these genes are transcribed in placental tissue and common epithelial cancers; but their protein products have not yet been identified (Burczynska *et al*., 2014; Glodek *et al.*, 2014). Type II genes are expressed in higher levels in the testes compared to other *CGB* genes (Rull *et al.*, 2007). Their proposed role is in the development of the male reproductive tract in the foetus (Parrott *et al.,* 2011).

The cluster of the beta subunit genes is proposed to have evolved from the *LHB* gene. This happened in recent evolutionary history whereby the *LHB* gene was duplicated and mutated to give rise to the *CGB*1-9 genes. In comparison with the *LHB* gene, the *CGB* genes have shifted transcription start site and extended transcription at 3' of the last exon corresponding to the protein's C-terminus **(**Fournier, 2016; Rao, 2016; Berger *et al.*, 2013).

### 1.2.1 *CGB* promoter and its regulation

Since the placental α subunit is available in excess, the limiting factor of hCG synthesis is the β subunit which production is influenced by the trophoblast via cAMP (Cole, 2015). The expression of genes coding for hCGβ (*CGB*) is proposed to be controlled by epigenetic modifications and availability of transcription factors AP2α, SP1, and SP3 mainly (Glodek *et al.*, 2014; Adams *et al.*, 2011).

As the *CGB* gene family has evolved from *LHB* via a proposed duplication event, the type I genes share a highly similar putative promoter region (Hallast *et al.,* 2007). The *CGB* genes' promoter does not have canonical regulatory elements or TATA box which has posed challenges to identify protein binding sites and how the gene is activated. Initial studies used the similarity with *LHB* gene to identify conserved elements. However, the promoter of *LHB* has been found out to be distinct from that of the *CGB* genes (Johnson and Jameson, 1999).

Empirical approaches were applied to elucidate the sequences required for the activity of *CGB* genes. These identified two enhancer regions upstream of the

transcription start site (TSS) that is responsive to cAMP (CREs) (Fig. 6). In these regions there are also sequences required for basal transcription. These have been found to interact with the ubiquitous Transcription Factors (TFs) AP2α, SP1 and SP3. The 5' CRE has 2 binding sites for AP2α and between them an SP1/SP3 binding site. The 3' CRE has 1 AP2α and one SP1 binding site (Cole, 2015; Johnson and Jameson, 1999; Pestell *et al.,* 1994).

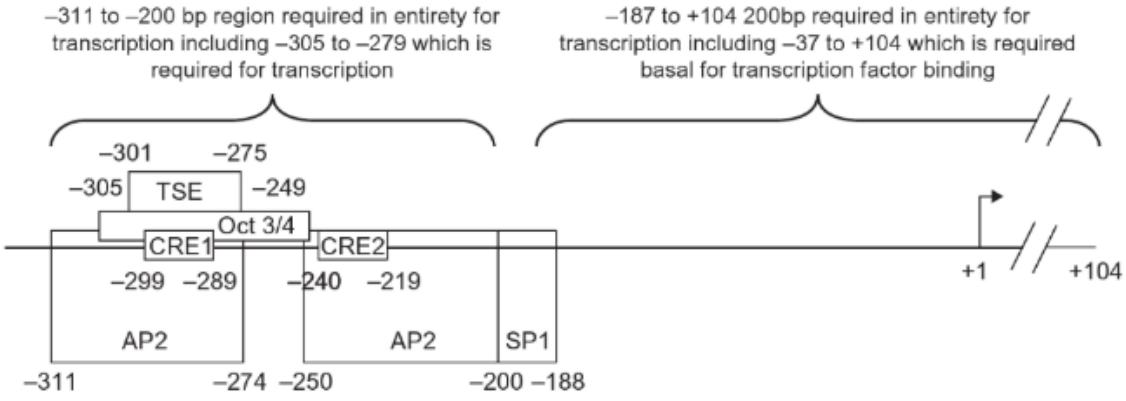



**Fig.6** Putative promoter region of *CGB5* indicating promoter elements and TF binding sites. Adapted from Cole, 2015.

The type II genes *CGB*1/2 appear to have a different promoter structure. They have an insertion of DNA of approx. 730bp that replaces the proximal 52 bp of the putative promoter and the 5' UTR in the *CGB*3-9 genes. This insert therefore creates a different putative promoter fragment just before the alternative 5'UTR and new exon 1 for the type II genes. This also creates a frameshift for the open reading frame by 1bp for exons 2 and 3 of the type II genes (Hallast *et al.,* 2007).

## 1.2.2 Regulation via Transcription Factors

AP2α and SP1 act to promote expression of the genes encoding for hCGβ in the placenta (Adams *et al.,* 2011). AP2α and SP1 are important in sustaining basal expression but AP2α has a further role of enhancing cAMP responsiveness. AP2α competes with SP1 in the 5' CRE. SP1 seems to have a stronger binding to the 5' CRE which competes away the AP2α binding. Even though AP2α and SP1 sequences seem to not overlap, both molecules cannot bind together at the 5' CRE. In the 3' CRE SP1 and AP2α can bind together without the competition seen in the upstream CRE (Fig. 6). (Cole, 2015; Johnson and Jameson, 1999).

AP2α function has been further proved in mutation studies. Disabling AP2α binding reduces cAMP responsiveness and decreases the transcription of *CGB* genes. Interestingly, when mutations are introduced in the SP1 sites the cAMP

responsiveness increases. However, both TFs are required for the basal expression of *CGB* genes (Johnson and Jameson, 1999).

Further confirmation of AP2α role in sustaining hCG level was done by Glodek *et al.* (2014). AP2α expression levels decrease significantly in pregnancies ending in miscarriage compared with normal pregnancy outcome. The same study shows that hCG serum levels also decrease in miscarriage; thus, a decrease in expression of both hCG and AP2α correlate positively with complications in pregnancy and can result in miscarriage (Glodek *et al.*, 2014).

The transcription factor SP3 has been shown to act as a repressor of the *CGB* genes (Glodek *et al.*, 2014). It is part of the same family as SP1 and competes for the same binding sites to reduce basal transcription (Johnson and Jameson, 1999). Another repressor of the *CGB* genes is c-Jun. c-Jun can actively bind to the CREs present in the *CGB* promoter to render transcription of the genes terminated (Cole, 2015; Pestell *et al.,* 1994). Oct3/4 has also been reported by Lui and Roberts, (1996) as suppressor of *CGB* gene transcription. Oct3/4 binds to a region between the 2 CRE elements uncompleted by another molecule showing potent repressing abilities (Fig. 6) (Johnson and Jameson, 1999; Lui and Roberts, 1996).

## 1.3 DNA Methylation and *CGB* genes

### 1.3.1 General function of methylation

Epigenetic modifications are a form of gene regulation where heritable changes to gene expression are introduced without changing the DNA sequence. These modifications can be grouped in distinct categories: chromatin remodelling, histone modification, DNA methylation, and microRNA silencing (Tammen *et al.,* 2013). DNA methylation is widespread across different organisms and is the most widely studied epigenetic mark (Cui and Xu, 2018).

Generally, DNA methylation refers to the addition of a methyl group to cytosine at the 5’ position which does not affect the DNA sequence (Fig. 7). Recent research has discovered methylation at 6’ of adenosine associated with transcriptional activation (Geer *et al.*, 2015). Here, DNA methylation refers to the methylated cytosine residues. It is a repressive mark inhibiting the transcriptional initiation.

The 5-methylcytosine followed by a guanine is termed CpG dinucleotide (Cui and Xu, 2018).

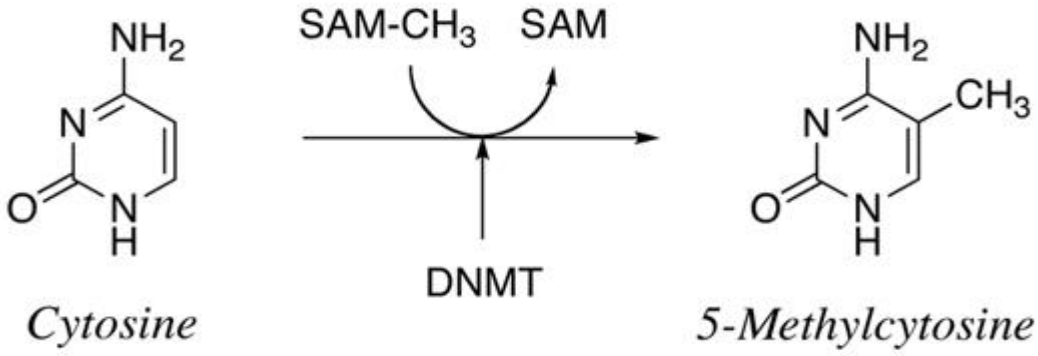



**Fig.7** Formation of 5mC molecule by addition of methyl group (-CH3) to cytosine. The reaction is catalysed by DNMT. Adapted from Gibney and Nolan, 2010.

Some regions of the DNA contain CpGs in large quantities known as CpG islands (CGIs), usually within the 5' untranslated region (5' UTR) of a gene promoter. More than half of vertebrate's genes are associated with such short (approx. 1kb) regions. Most CGIs associated with TSS are not methylated. Methylation of CGI at TSS is associated with long term silencing like germ cell expressed genes. They can also be found in intergenic regions and within gene bodies where CGIs are on occasion methylated. CGIs are thought to be more prevalent than CpG poor sites as they are never or shortly methylated leading to a decreased possibility of deamination. Methylated CpGs are mutagenic and can undergo spontaneous deamination converting the Cytosine to Uracil (Jones, 2012; Illingworth *et al.*, 2010).

DNA methylation is established and maintained by a class of enzymes called the DNA methyltransferases (DNMTs). DNMTs catalyse the addition of a methyl group to cytosine from the methyl donor *S*-Adenosyl methionine (SAM) (Varela-Rey *et al.*, 2014). There are 3 main enzymes identified: DNMT1, DNMT3a and 3b. DNMT1 maintains the established pattern of methylation as it has strong preference for hemimethylated DNA. DNMT3A and 3B are considered the *de novo* methyltransferases which establish the pattern of methylation in early development (Cui and Xu, 2018; Koh *et al.*, 2011).

Removing the methyl group from DNA can be achieved actively or passively. Passive demethylation is the termination of DNA methylation maintenance and allowing for cell division and DNA replication to decrease the amount of 5mC **(**Guo *et al.*, 2014). Active demethylation has recently been described. It involves

the enzymes TET methylcytosine dioxygenases and TDG. Active demethylation is not achieved by simply removing the methyl group. Most cases it involves a complex process of oxidation that will lead to replacement of the whole base via the Base Excision Repair mechanism or some form of DNA repair or cell division (Guo *et al.,* 2014; Koh *et al.,* 2011).

CpG rich regions in the vertebrate's genome are commonly associated with more than half of the protein producing genes. The other part of the genome is considered to be CpG depleted and about 70% of these CpGs are methylated. Comparing with CGI methylation, non-CGI CpGs have more dynamic methylation and are less uniform and rather tissue specific (Jones, 2012; Illingworth *et al.*, 2010).

Position of methylation is also important in determining the function of it. Methylation has mainly been studied when it is located at the TSS acting as a silencer. However, methylation can be found within the gene in intergenic regions and may have different function in splicing, or when associated with enhancers and insulators (Jones, 2012; Illingworth *et al.*, 2010).

As an epigenetic mark, DNA methylation fluctuates and can be reversed. Most dynamic changes in methylation are observed during embryogenesis (Guo *et al.*, 2015). The Tet enzymes and passive demethylation remove the mark to induce the totipotent state of the cell allowing proper embryogenesis. Lack of Tet enzymes can be detrimental in that period. If not present TET3 enzyme prevents the ability of demethylation of *Oct4* leading to delayed embryogenesis. Similarly, when the epigenetic reprogramming is complete and the establishment of the methylation towards the end of embryogenesis takes place DNMTs are essential. Lack of DNMTs and thus of methylation is incompatible with life (Guo *et al.*, 2015; Guo *et al.*, 2014; Li, 2002).

The reason why DNA methylation is important is when it comes to chromosome stability. Repeat regions are methylated like in centromeres to prevent alteration during processes like chromosomal segregation during cell division. Furthermore, methylation is thought to block transposable elements and thus promoting genomic stability. This silencing of the transposable elements however does not affect the transcriptional elongation. Lack of *de novo* methylation DNMT3A enzyme prevents from development of normal blood cells. This further proves that

5mC is essential for cellular differentiation and development. (Jones, 2012; Li, 2002).

DNA methylation also works with other epigenetic marks such as histone modifications, genomic imprinting and X-inactivation to carry out its functions. Upon methylation of CGI a histone modification follows to render this part of the DNA transcriptionally inactive (Grigoriu, *et al.*, 2011; Wojdacz and Dobrovic, 2007). Nucleosomal DNA is the substrate for *de novo* DNMTs is meaning the histone modifications can influence the introduction of methylation by these enzymes. During the process of X-inactivation, DNA methylation packages the chromosome tightly for long-term silencing. Also, methylation is part of Genomic imprinting, taking part in ensuring monoallelic expression of imprinted genes (Jones, 2012; Li, 2002).

### 1.3.2 Methylation and *CGB* genes

It has been reported that in the promoter sequence of the *CGB* genes, CpG dinucleotides are present and these are therefore sites for potential epigenetic gene regulation via methylation (Grigoriu *et al.*, 2011). Under normal physiological conditions (i.e. non-pregnant females and males), the *CGB* genes are highly hypermethylated, which prevents binding of transcription factors and consequent transcription (Grigoriu *et al.,* 2011). When the promoter is hypomethylated, as it occurs during pregnancy, access is granted to transcription machinery and consequently expression of genes coding for hCG is promoted. The placental *CGB* genes show very low methylation levels during the first trimester and this steadily increases until delivery (Grigoriu *et al.,* 2011; Campain *et al.,* 1993). This corresponds to the serum hCG level fluctuation in pregnancy which also increases until week 12 and then stabilises (Cole and Butler, 2015)

*CGB* aberrant promoter methylation has been linked with miscarriages and adverse pregnancy outcomes (Hanna *et al.*, 2013). Glodek *et al.,* (2014) show that chorionic tissues from miscarriages have higher promoter methylation compared to normal pregnancies. Therefore, hCGβ is expressed less in miscarriage leading to inability to reach full term pregnancy. The aberrant methylation may reflect problems with the embryo such as chromosomal abnormalities (Glodek *et al.*, 2014).

Additionally, a more specialised type of methylation, gain of imprinting, might be part of the regulation of *CGB*. It relates to the mechanism where only one allele of

a gene (paternal or maternal) is expressed and the other is silenced by methylation (Uuskula *et al.*, 2010). Uuskula *et al*. (2010) have investigated further the methylation of *CGB5* proposing methylation allelic polymorphism (gain of imprinting) that plays a role in pregnancy success. The study concluded that there is link between the methylation allelic polymorphism and miscarriages. Normal term pregnancy shows biallelic expression of the gene but when the paternal genes are silenced by methylation there is increased susceptibility to pregnancy termination (Uuskula *et al.*, 2010).

### 1.3.3 DNA methylation changes in cancer

DNA methylation has been shown to play a role in the development of cancer. Normally, intergenic regions containing transposable elements are highly methylated to maintain genome stability. However, one of the first signs of carcinogenesis is when these mobile regions (transposons) lose their methylation in an event known as global hypomethylation (Robertson, 2005). This allows them to move in the genome and cause genome instability, aberrant gene regulation, and generation of antisense transcripts. At the same time, CGIs associated with tumour suppressor genes become hypermethylated. This causes silencing leading to loss of their protective function. Gene specific hypomethylation may also occur, assisting in the adaptation of tumour cells (Jones, 2012; McCabe *et al.*, 2009; Robertson 2005).

During pregnancy it has been observed that there are changes in the methylation of the *CGB* promoter where it becomes hypomethylated (Grigoriu *et al.,* 2011). It stands to reason that similar changes are observed in cancers. Campain *et al.,* (1993) showed that choriocarcinoma cell lines present hypomethylation in the *CGB* promoter like that observed in normal pregnancy cells. In the same study by Campain *et al.,* (1993) non-trophoblastic cancerous tissues from glioblastoma multiform and lung cancer cell lines show decreased methylation of the *CGB* but the promoters are not as hypomethylated as in choriocarcinoma. Interestingly, only the glioblastoma cell line showed ectopic hCGβ release (Campain *et al.,* 1993). Further, based on the findings of Campain *et al.* (1993) it may be speculated that gene specific hypomethylation may contribute to the overexpression of *CGB* genes.

A more recent paper investigated DNA methylation of *CGB* in ovarian cancer (Śliwa *et al.*, 2019). The researchers found that the promoter of the *CGB* genes was demethylated when comparing between normal and cancerous ovarian tissue. In their study they also investigated the expression level of the *CGB* genes and associated transcription factors. It was observed that the *CGB*3-9 genes were expressed at higher levels in the cancer samples. However, no significant correlation was found between DNA methylation and the expression of the genes in this specific cancer type (Śliwa *et al.*, 2019).

## 1.4 Detection of DNA Methylation

A variety of methods can be utilised to establish the methylation pattern of *CGB* genes. The analysis in Campain *et al.* (1993) is based on selective cleavage of DNA at methylated or unmethylated cytosines by restriction enzymes (RE). However, RE analysis is restricted to specific sites and relies on high quality DNA as well as full digestion by the enzymes (Ammerpohl *et al.,* 2009; Fraga and Esteller, 2002). Currently, bisulphite sequencing (BS) is considered the gold standard in DNA methylation research. Bisulphite modification of DNA introduced more possibilities for methylation analysis since it converts unmethylated cytosines to uracils via deamination across the whole DNA molecule which allows wider gene analysis of all possible CpGs. The most widely applied method for specific locus investigation in research is Methylation Specific PCR (MSP) but this method is prone to false positives and has no quantitative output (Wojdacz and Dobrovic, 2007).

Methylation-sensitive high-resolution melting (MS-HRM) PCR, bisulphite pyrosequencing and MiSeq are examples of quantitative methods applied in DNA methylation analysis (Soto *et al.*, 2016; Colyer *et al.*, 2012; Wojdacz and Dobrovic, 2007). MS-HRM PCR is a relatively new method executed in a closed system that diminishes the possibility of errors. It is a modification of the HRM qPCR methodology developed for single nucleotide polymorphism (SNP) analysis which is able to differentiate 1 nucleotide change (Wojdacz and Dobrovic, 2007). MS-HRM is based on the different amplicon melting temperatures which are dependent on nucleotide content. Following PCR amplification the products are subjected to a steady temperature increase leading to their melting. The decrease of fluorescence upon disintegration of the products is recorded by the machine to produce graphical data. Using different ratios of negative and positive control a

standard curve is produced which is used to quantify the percentage of methylation (Wojdacz and Dobrovic, 2007).

Pyrosequencing is one of the established quantitative techniques providing fast and reliable results with the opportunity for high-throughput analysis (Ammerpohl *et al.*, 2009). The technique is considered as one of the first platforms for Next generation sequencing (Fakruddin *et al.*, 2013). It overcomes the need for bacterial cloning, an essential part of the labour intensive BS procedure. Pyrosequencing also provides more straightforward results for the methylation status of a CpG site compared to BS which relies on a lot of data conversion to convey this information due to the bacterial cloning (Reed *et al.*, 2009). It has been reported that pyrosequencing has a reasonably high sensitivity of detecting 5% methylation comparing to genomic sequencing which has considerably lower sensitivity of detecting 20% methylation (Wojdacz and Dobrovic, 2007).

Bisulphite pyrosequencing requires DNA to be converted with bisulphite as the initial step (Mikeska *et al.*, 2011). The target sequence is then amplified from the converted DNA via PCR with bisulphite sequencing primers. The key part of the amplification step is that one of the primers is biotinylated. The PCR product from the labelled primer is separated and subsequently incubated with sequencing primer. The sequencing primer is shorter than a regular PCR primer and is used as the starting point for the pyrosequencing assay. Pyrosequencing is described as a sequence by synthesis method. Upon the incorporation of a new nucleotide to the sequencing primer luminescence is emitted from the released pyrophosphate which is detected by the machine performing the method (Colyer *et al.*, 2012; Mikeska *et al.*, 2011). This allows for detection of methylation as an SNP at CpG dinucleotides where incorporation of a C denotes methylation and incorporation of a T denotes no methylation. In turn, this provides a quantitative analysis of all CpG dinucleotides covered in the PCR product (Mikeska *et al.*, 2011; Reed *et al*., 2009).

MiSeq is a newer Illumina-based platform for next generation sequencing. It is also sequencing by synthesis method but with the feature of “bridge” amplification (Soto *et al.*, 2016). Libraries fed to this system need to contain amplicons that are barcoded. These barcodes prime with adaptors on a solid phase to immobilise the amplicon while the sequencing is running creating a bridge. During amplification fluorescently labelled nucleotides are fed and machine detects their incorporation.

Like pyrosequencing, MiSeq is a high-throughput system but uses shorter reads in its analysis. However, Illumina based platforms are currently most used and pyrosequencing support for Roche platforms ended in 2016. MiSeq has moderate cost and comparted with other Illumina products has the fastest run times and provides the longest reads (Soto *et al.*, 2016). Methylation studies have employed MiSeq platform to reliably detect at a single nucleotide level the changes in DNA methylation in the context of disease (Roh *et al.*, 2018; Dukal *et al.*, 2017; Xiong *et al.*, 2017).

Currently, there is a gap in knowledge regarding changes in the methylation pattern of *CGB* genes in non-trophoblastic cancers. Glodek *et al.* (2014); Grigoriu *et al.* (2011) and Uskuula *et al.* (2010) have shown that methylation changes in the hCGβ genes are detrimental for successful pregnancy. However, limited studies have been performed to assess this in the context of non-trophoblastic cancers (Campain *et al.,* 1993; Śliwa *et al.*, 2019). Investigating the potential DNA methylation changes will shed light on possible novel molecular mechanisms involved in re-activation of the *CGB* genes in cancer.

## 1.5 Aims

This project aims to investigate the role of the repressive epigenetic mark (DNA methylation) on the production of hCG and its free β subunit in non-trophoblastic cancer cell lines. This will be achieved by the analysis of the methylation status in the promoter region of *CGB3*-9 genes via next generation sequencing. This will be followed by the quantification of the expression level of the *CGB3*-8 genes by qRT-PCR and establishing the level of the proteins hCG and its beta subunit secreted in culture media. The methylation status of the *CGB3*-8 promoter will be correlated to the expression and secretion of the human chorionic gonadotropin to establish the role of the suppressive mark in non-trophoblastic cancers cell lines.

# 2. Methods

## 2.1 Sample Collection

### 2.1.1 Tissue culture

Samples used in this study where acquired through cell culturing of 16 different cell lines. All of them are established cell lines already available within the laboratory at Middlesex University. Investigated samples included non-trophoblastic (NT) cancer cell lines, trophoblastic (T) cancer cell lines, normal colon cell line, and a mouse fibroblast cell line (Table 2). The trophoblastic cell lines were included as positive control, CRL-1790 - normal cell line control, and 3T3 – negative control. Cell lines were all grown with 10% Foetal bovine serum (FBS) and 1% penicillin-streptomycin (PS) at 37ºC and 5% $CO_2$. Table 2 shows culture media used for different cell lines. After the cell lines have become 80% confluent, the cells were collected using trypsinisation and then further processed for total DNA and RNA extraction. Cell pellets collected were stored at -80ºC

**Table 2** Investigated cell lines, their respective tissue of origin and used culture media

| # | **Cell Line (cancer type)** | **Tissue of origin** | **Culture media** |
|---|---|---|---|
| 1 | MDA-MB-468 (NT) | breast cancer | RPMI 10%FBS 1%PS |
| 2 | MCF-7 (NT) | breast cancer | RPMI 10%FBS 1%PS |
| 4 | HeLA (NT) | cervical cancer | DMEM 10%FBS 1%PS |
| 5 | HT-3 (NT) | cervical cancer | RPMI 10%FBS 1%PS |
| 6 | C-33a (NT) | cervical cancer | RPMI 10%FBS 1%PS |
| 8 | BeWo (T) | choriocarcinoma | DMEM 10%FBS 1%PS |
| 9 | JEG-3 (T) | choriocarcinoma | RPMI 10%FBS 1%PS |
| 10 | HCT116 (NT) | colon cancer | DMEM 10%FBS 1%PS |
| 11 | CRL-1790 | normal colon | EagleMEM 10% FBS 1%PS |
| 12 | 3T3 | mouse fibroblast | RPMI 10%FBS 1%PS |
| 13 | OAW42 (NT) | ovarian cancer | RPMI 10%FBS 1%PS |
| 14 | OVCAR-3 (NT) | ovarian cancer | RPMI 10%FBS 1%PS |
| 15 | HEY-T30 (NT) | ovarian cancer | RPMI 10%FBS 1%PS |
| 16 | SKOV-3 (NT) | ovarian cancer | RPMI 10%FBS 1%PS |

### 2.1.2 DNA samples

DNA was extracted from collected cell pellets. The DNA was extracted using column purification technology. The protocol supplied by the manufacturer was used for the extraction (DNA mini kit, Thermo Fisher Scientific). The cells suspended in 200µl PBS were lysed together with 20µl proteinase K and 20µl RNase-A. The lysate was incubated at 55ºC for 10 min and 200µl ethanol was added. The mixture was then spun through the columns at 10,000g. The bound

DNA was washed twice with 500µl Wash buffer and then eluted with 100µl EDTA-Tris by maximum speed centrifugation for 1.5min. DNA concentration and purity were checked with Qubit 3 and NanoDrop 2000 respectively (Table A1, Appendix A1). The DNA was then subsequently used for bisulphite conversion or sent for sequencing, DNA was stored at -20ºC.

### 2.1.3 RNA samples

RNA was extracted from the collected cell pellets using column purification technology (RNA mini kit, Thermo Fisher Scientific). The extraction followed the manufacturer's protocol. The cells were lysed with 1% beta-mercaptoethanol lysis solution (volume dependent on cell number in pellet) and homogenised mechanically using 20-gauge needle. The homogenate was mixed with equal volume 70% ethanol and was passed through the column. The bound RNA was washed twice with 700µl Wash buffer I and 500µl Wash buffer II. The column membrane was dried by centrifugation and RNA was eluted in RNase free water. RNA concentration and purity were then checked with Qubit 3 and NanoDrop respectively (Table A2, Appendix A1). The RNA samples were stored at -80ºC. Prior to further use in expression studies RNA was also treated with DNase I, Amplification grade (Thermo Fisher Scientific) to remove any possible DNA contaminant. 2µl DNase I in buffer was used to treat 1µg of RNA by incubating for 15min at RT. The enzyme was inactivated using 65ºC heat and then 1µl of 25mM EDTA was added to stop the reaction. The cleaned-up RNA was then used in expression studies.

### 2.1.4 Protein samples

Conditioned media from cell lines was collected prior to trypsinisation of confluent cells (Table 3). Trypsinised cells were counted using a haemocytometer. The media was used in ELISA to detect free beta hCG and intact hCG. The media was stored at -80ºC.

**Table 3** Growth characteristics of cell lines from which conditioned media was collected

| Cell Line | Time growing (days) | Volume of media (ml) | Cells counted ($10^6$ cells) |
|---|---|---|---|
| **HEY-T30 (HEY)** | 4 | 10 | 0.64 |
| **SKOV-3** | 4 | 10 | 4 |
| **Ovcar-3** | 4 | 9 | 0.688 |
| **CRL-1790** | 10 | 18 | 0.7 |
| **BEWO** | 4 | 20 | 8.3 |
| **OAW42** | 7 | 10 | 2.1 |
| **MCF-7** | 6 | 22 | 6.67 |
| **MDA-MB-468** | 4 | 19 | 8.5 |
| **3T3** | 3 | 20 | 6.8 |
| **HELA** | 3 | 18 | 9 |
| **C-33A** | 4 | 20 | 5 |
| **JEG-3** | 3 | 20 | 10.5 |

## 2.2 *CGB in silico*

### 2.2.1 Multiple sequence alignment

The sequences of the *CGB* genes and *LHB* gene were acquired from the ENSEMBL database and aligned with ClustalW software to identify the region of interest (Madeira *et al.*, 2019). The database used was Human genome version GRCh38.p10. For the Multiple sequence alignment (MSA) the 1st exon and 1000bp upstream of the Transcription start site (TSS) were used (Table 4). *CGB*6 and 9 were not used as they are allele forms of *CGB*7 and 3 respectively. As *CGB*7 has multiple splicing variants, the *CGB*7 transcript variant 2 was used. Based on previous reports in literature, the putative promoter elements were added on the MSA (Cole, 2015; Kerschgens *et al.,* 2011; Hallast *et al.*, 2007).

**Table 4** *CGB* and *LHB* genes and their location on chromosome 19 used for MSA

| Gene | ENSEMBL ID | Strand | Location |
|---|---|---|---|
| ***CGB1*** | ENSG00000267631 | Reverse | chr19:49,036,526-49,037,895 |
| ***CGB2*** | ENSG00000104818 | Forward | chr19:49,030,912-49,032,270 |
| ***CGB*** | ENSG00000104827 | Reverse | chr19:49,023,957-49,025,333 |
| ***CGB5*** | ENSG00000189052 | Forward | chr19:49,042,884-49,044,224 |
| ***CGB7*** | ENST00000377280 | Reverse | chr19:49,055,361-49,056,780 |
| ***CGB8*** | ENSG00000213030 | Reverse | chr19:49,048,725-49,050,107 |
| ***LHB*** | ENSG00000104826 | Reverse | chr19:49,017,067-49,018,081 |

### 2.2.2 CpG islands

The *CGB3-8* sequences in table 3 were also checked for possible CpG islands within the promoter region. The sequences used in the MSA were used in the Newcpgreport (Madeira *et al.*, 2019). The window was set to 200 bases, with minimal length of 200bp for CpG Island and 0.6 observed/expected ratio with minimum 50%CG content.

### 2.2.3 Primer design

Primers to be used in Next generation sequencing (NGS) were designed on MethPrimer to accommodate the region of interest within the *CGB*3-8 promoter (Li and Dahiya, 2002). Due to the length of region of interest and slightly different sequences for each gene different sets of primers were designed. The designs were based on the bisulphite converted *CGB*3 sequence. The preliminary primers were then run through ePCR software (BiSearch) where the expected products of each primer pairs were predicted (Arányi *et al.,* 2006).

## 2.3 *CGB* promoter methylation

### 2.3.1 Bisulphite conversion

Bisulphite conversion was done to prepare the DNA samples for the investigation of methylation level via either MSP or next generation sequencing. 500ng of DNA was used per conversion reaction as per manufacturer protocol (Diagenode, s.a.). The DNA was mixed with 130µl conversion reagent and placed in a thermocycler using the cycling program mentioned in table 5. Following this conversion column purification was used to collect the bisulphite converted DNA. The converted DNA was added to the column with 600µl of binding buffer. After the DNA was bound it was washed with 100µl buffer, desulphonated, and washed 2 more times with 200µl wash buffer. Then the converted DNA was eluted to be used in downstream reactions. Bisulphite converted DNA was stored in -20ºC.

**Table 5** Conditions used to bisulphite convert DNA

| **Stage** | **Temperature** | **Time** |
|---|---|---|
| **Denaturation** | 98ºC | 8min |
| **Conversion** | 54ºC | 60min |
| **Final hold** | 4ºC | ∞ (up to 20h) |

### 2.3.2 MSP

Bisulphite converted DNA was used in Methylation sensitive PCR (MSP) to establish the level of methylation in the *CGB* promoter region initially. The selected cell lines for MSP were HeLa, C-33a, HEY, SKOV-3, CRL-1790. The methylated (M) and unmethylated (U) primers used were from a previous publication (Glodek *et al.,* 2014). The housekeeping gene used was Death Associated Protein Kinase (DAPK). Primer sequences are mentioned in Table 6. The MSP was performed with HS polymerase mastermix as set in the instructions (PCR Biosystems). The cycling conditions used are detailed in table 7. Products were run on 2% agarose gel and stained with SafeView to visualise on Li-cor. Densitometry analysis was performed using the software ImageJ.

**Table 6** Primer sequences used in MSP

| **Primer ID** | **Binding strand** | **Sequence** |
|---|---|---|
| DAPK | Forward | 5'ATTGGGAAGGTTAAGGYGGAGGGAAATTTGGT3' |
| | Reverse | 5'CCCCAAACRAAACAATCCCCAAAACCACATTCCTA3' |
| *CGB*3-9_M | Forward | 5' TGTTTAGTTTGATGGTATCGC 3' |
| | Reverse | 5' ATACCCGAAACGATCCCC 3' |
| *CGB*3-9_U | Forward | 5'AATTGTTTAGTTTGATGGTATTGT 3' |
| | Reverse | 5' AAAATACCCAAAACAATCCCC 3' |

**Table 7** Cycling conditions for MSP

| **Step** | **Temperature** | **Duration** |
|---|---|---|
| **Initial denaturation** | 95ºC | 10min |
| **40 cycles** | | |
| **Denaturation** | 95ºC | 30s |
| **Annealing** | 58ºC(M) 55ºC(U) 59ºC(DAPK) | 30s |
| **Extension** | 72ºC | 30s |
| **Final extension** | 72ºC | 5min |
| **Final hold** | 4ºC | ∞ |

### 2.3.4 Sequencing

Sequencing was performed externally using the MiSeq Illumina platform (Bart's and the London Genome Centre - Queen Mary University, London). DNA samples from cultured cell lines were sent. In addition, DNA extracted from normal breast and cervix tissue was also sent to be sequenced (BioChain). DNA samples were bisulphite converted and then amplified with PCR to create the library for sequencing. Two sets of primers were selected from previously mentioned designs to accommodate the region of interest in the *CGB*3-9 genes (Table 8). The PCR products were then ligated with adaptors to perform the sequencing. The reads from the sequencing were mapped to the respective *CGB* genes.

Methylation of the CpG sites within the region of interest was given a Beta value using Bismark software. The value is the percent of methylated reads out of the total reads at each CpG.

**Table 8** Primers for sequencing

| Primer pair ID | Binding strand | Sequence |
|---|---|---|
| *CGB3-9*_1 (Pair 6) | Forward | 5' GGGGAAGGGATTAAGTTTAGA 3' |
| | Reverse | 5' ACTATACTACCAAAAAAACCACTTA 3' |
| *CGB*3-9_2 (Pair 7) | Forward | 5' GGGTATTTTGGTTTGAGGG 3' |
| | Reverse | 5' CCTCAACCCTCCTCTACTT 3' |

## 2.4 *CGB* expression and secretion

### 2.4.1 cDNA

1µg of DNase treated RNA was converted to cDNA using SSIV kit (Thermo Fisher Scientific). RNA samples were ligated with random hexamer for 10min at 65ºC. The reverse transcriptase was then added and placed in thermocycler for 10min at 25ºC, 10min at 55ºC and 10min at 80ºC to create cDNA.

### 2.4.2 qRT-PCR

qRT-PCR with TaqMan probes was employed to assess the transcription level of the *CGB*3-8 genes. 200ng of the cDNA was used during the reaction. FastStart Essential DNA Probes Master (Thermo Fisher Scientific) was used as instructed by manufacturer to amplify the cDNA. Due to the similarity of the genes 1 primer set and 1 probe from a previous publication was used to assess the *CGB*3-8 transcription (Śliwa *et al.,* 2019). Sequences are stated in Table 9. GAPDH TaqMan Assay from Thermo Fisher Scientificwas used as the housekeeping gene. Samples were plated in duplicate on a 96-well plate and run on the LightCycler 96 machine with the conditions mentioned in Table 10. Each run also included No Template Control and No Reverse Transcriptase Control. Relative *CGB*3-8 expression was calculated based on the $2^{-\Delta\Delta Ct}$ method whereby the expression was normalised to GAPDH expression (Livak and Schmittigen, 2001).

**Table 9** Primer and probe sequences used in qRT-PCR

| ID | Sequence |
|---|---|
| *CGB*3-9_RT F | 5'- GTGTCSAGCTCACYCCAGCATCCTA- 3' |
| *CGB*3-9_RT R | 5'- AGCAGCCCCTGGAACATCT -3' |
| TaqMan Probe | 6FAM-CCGAGGTYTAAAGCCAGGTACACSAGGC-BHQ |

**Table 10** Cycling conditions used to perform qRT-PCR

| **Step** | **Temperature** | **Duration** |
|---|---|---|
| **Pre-incubation** | 95ºC | 10min |
| **45 cycles of 3 step amplification:** | | |
| **Denaturation** | 95ºC | 10s |
| **Annealing** | 60ºC | 30s |
| **Extension** | 72ºC | 1s |
| | | |
| **Cooling** | 37ºC | 10s |

## 2.4.3 ELISA

Sandwich-based ELISA was employed to observe the secretion of hCG and hCG-beta in the cell line media. Intact hCG and free beta hCG kits were used from Demeditec Diagnostics (Germany). Both ELISA kits are pre-coated with the respective monoclonal antibody and standard protocol provided by the manufacturer was followed.

### 2.4.3.1 Free beta hCG

ELISA was performed as instructed by the manufacturer. Six standards ranging from 1.25ng/ml - 50ng/ml were used to create a standard curve. 50µl of the provided standards, controls, and conditioned media samples were loaded on the provided coated wells and diluted with 100µl zero buffer, then incubated at 37ºC for 30min. Plate was washed five times with wash solution and 150 µl enzyme conjugate containing the anti-beta-hCG ab conjugated to horseradish peroxidase was added to each well. After a further incubation at 37ºC for 30min the plate was washed 5 times and 100µl substrate was added. After 20min 100µl stop solution was employed to block the reaction and the wells were read at 460 nm. To account for background noise readings a second reading at 640nm was performed. The free beta hCG ELISA was run 2 times – once samples were in triplicate and once - in duplicate. Obtained values from standard were used to build a standard curve using the 4-parameter fit model. Sample concentration were calculated and then expressed as ng/ml.

### 2.4.3.2 Intact hCG

Similarly, the intact hCG ELISA was performed as per manufacturer instructions. Four standards ranging between 5-500 mIU/ml were used to build standard curve. 25 µl of the provided standards, controls and conditioned media samples were loaded on the provided wells. After 100µl of the enzyme conjugate containing the

monoclonal anti-hCG antibody was added to each well and the plate was incubated for 30min at room temperature and washed 5 times with distilled water. 100µl Substrate solution was added to develop colour for 10 min and the reaction was stopped.by adding 50µl stop solution. The samples were run once in duplicate. As with the previous ELISA plate was read at 460 and 640 nm. Standard curve was built based on the 4-parameter fit model and sample concentration was calculated.

## 2.5 Statistical Analysis

Methylation data was analysed using methylKit and R programming to create histograms, PCA, clustering analysis, and analysis of methylation association (Pearson) (Akalin *et al.*, 2012). Normally distributed data was analysed using parametric tests (Repeated Measures (RM) ANOVA, paired t-test), otherwise non-parametric tests (Mann Witney, Spearman's ranked correlation) were used. The nonparametric Mann Witney was used to analyse for difference in the MSP data. For the statistical analysis of methylation data Beta values of the sequencing were converted to M-values. Methylation data analysed for difference using paired t-test or RM-ANOVA paired with Tukey's test on Prism8. With smaller sample set the non-parametric Kruskal-Wallis test was used paired with Dunn's test on Prism8. Association analysis between the transcription and methylation data was done using Spearman's ranked correlation using Minitab18.

# 3. Results

## 3.1 Putative promoter Characteristics

### 3.1.1 MSA of target region

The aligned sequences of the *CGB*1-9 and *LHB* genes show the close identity of the genes. The region presented in Figure 9 shows the putative promoter region of the *CGB*3-8 genes and how it aligns with the *LHB* and *CGB*1-2 genes. All *CGB* genes have a conserved promoter sequence which starts to diverge towards the 3' end where the *CGB*1-2 specific insert begins. The putative promoter spans approximately 350bp upstream from the transcription start sight of the *CGB*3-8 genes. It contains 2 cAMP responsive elements (CREs), Trophoblast Specific element (TSE) and a CCAAT box. At the 5' end of the promoter region there is are 4 TF binding sites close together: an AP2α followed immediately by an SP1 and another AP2α transcription sites after which an Oct3/4 binding site. These 4 sites also coincide with the 5' CRE. The 3' CRE there are 2 more TF binding sites for AP2α followed by an SP1. Within the marked promoter region there are 57 CpG sites across all genes. The CpGs that are common between all *CGB*3-8 genes are labelled A1-A11, the one common between *CGB*3, 8 and 7 is B1, the one on *CGB*3, 5, and 7 – C1, the one on *CGB*3 – D1, the one on *CGB*5, 8, and 7 – E1, the one on *CGB*5 & 7 – F1, and the one on *CGB*7 – G1. MSA of the whole 1000bp 5'upstream region and the 1$^{st}$ exon can be found in Appendix A2.1.

### 3.1.2 CGI prediction

The Newcpgreport tool checked the *CGB3*-8 promoter genes for possible CGIs (Appendix A2.2) (Madeira *et al.*, 2019). It returned only one CpG island (Fig. 8). The island predicted by the tool was on the *CGB*7 gene with length of 720 bp. The software locates it -901 to -182bp from the TSS of the gene. The CG content within the region is 66% and an observed/expected ratio is 0.75. The other CpG genes did not present with a CpG island within the 1000bp region upstream from the TSS.

```
ID   CGB7  1419 BP.
XX
DE   CpG Island report.
XX
CC   Obs/Exp ratio > 0.60.
CC   % C + % G > 50.00.
CC   Length > 200.
XX
FH   Key              Location/Qualifiers
FT   CpG island       99..818
FT                    /size=720
FT                    /Sum C+G=474
FT                    /Percent CG=65.83
FT                    /ObsExp=0.75
FT   numislands       1
```

**Fig.8** Output from Newcpgreport tool for *CGB7* (Madeira *et al.*, 2019). ID row is the user defined ID and length of input sequence, DE- description of test, CC rows – conditions of the test, FH and FT rows present the predicted CpG island in relation to the input sequence.

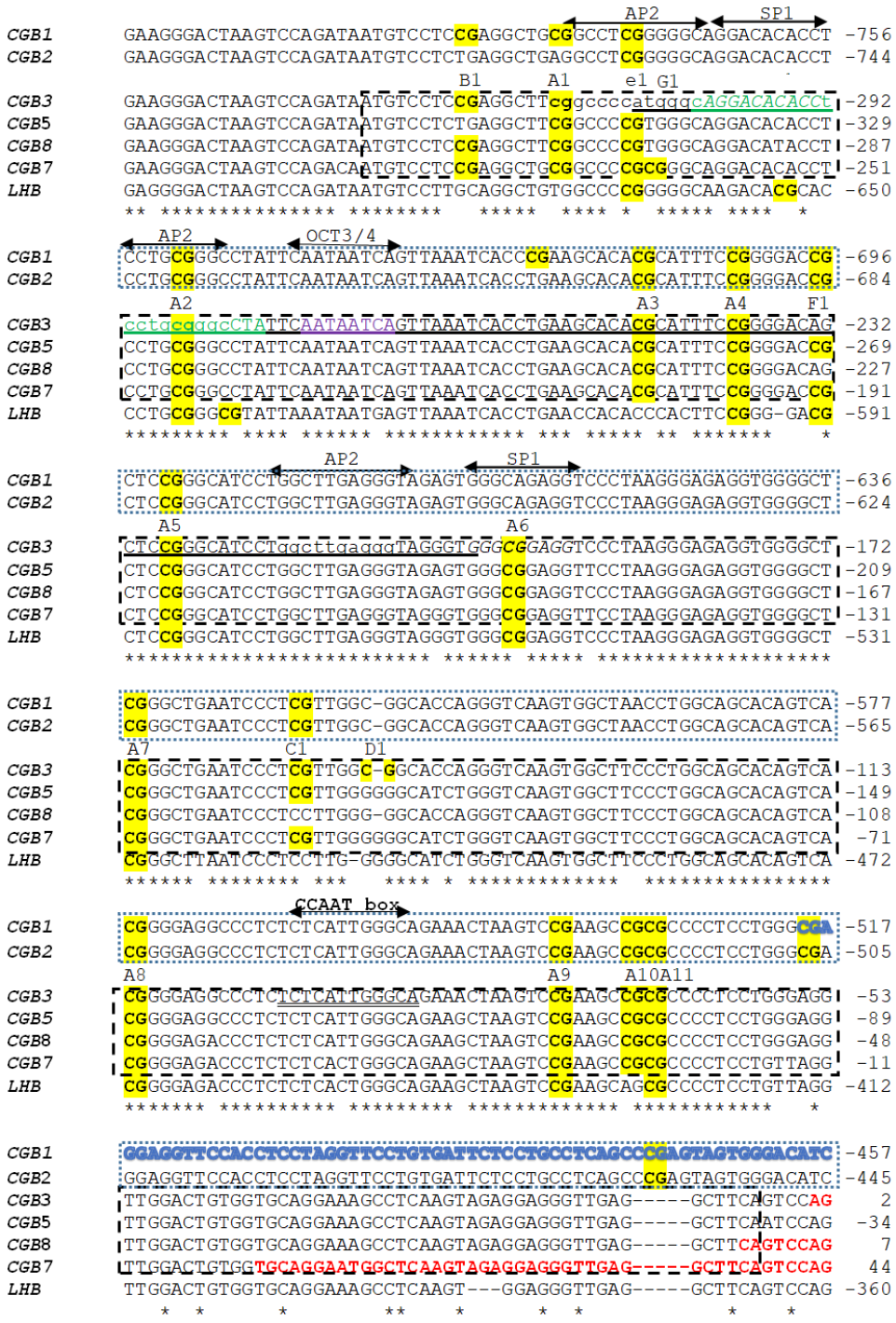



**Fig.9** Annotated MSA of the *CGB*3-8 putative promoter region together with *CGB*1-2 and *LHB*. Legend: CpG sites are bolded and highlighted in yellow*(***CG***)* and labelled above the *CGB* sequence, 5'UTR of exon 1 is labelled with red font and bolded *(***CGGCCCCATGG***)*, dash black line surrounds the putative promoter of the *CGB3*-8(- -)cAMP response elements (CREs) is the underlined region **(**CTGGCATCCTGGCTT*)*, TSE is written in green font*(*CCTGCGGGCCTA), CCAAT box is underlined with a double line(TCTCATTGGGCA*)*.AP2α binding is marked by lower case letters*(*cggccccatgggc*),* italicised letters denote SP1 binding *(CGGCCCCATGGGC),* purple font denotes the Oct3/4 binding site *(*TCTCATTGGGCA); AP2α, SP1 and OCT3/4 binding is also denoted by a double-headed arrow above *CGB1*, blue dotted line marks the *CGB1-2* putative promoter (⋯) and blue outlined letters denote *CGB1-2* specific insert **(**TCTCATTGGGCA*)*

## 3.2 Methylation profile

### 3.2.1 MSP

Initially the methylation profile was investigated using Methylation Sensitive PCR to check methylation at a CpG located at -550bp from TSS, F1 and A5. Methylated (M) primers showed consistent strong bands across all the cell lines. CRL-1790, HEY, and C-33a showed brighter bands compared to HeLa and SKOV-3. The obtained values from densitometry analysis were analysed for difference between the normal cell line (CRL-1790) and the non-trophoblastic cancer cell lines (HeLa, HEY, SKOV-3, C-33a) using Mann-Whitney (Fig. 10). No significant difference was found (p=0.687). Unmethylated (U) primers had weaker bands. HeLa showed brightest U product bands, and the other cell lines had weaker bands for the same product. Densitometry analysis and differential analysis was performed analogically. No significant difference was found (p=0.502) between the control and non-trophoblastic cell line. The house keeping gene showed consistent product of similar intensity in all the tested cell lines (Fig. 10). See Appendix A3 for densitometry data and statistical analysis.

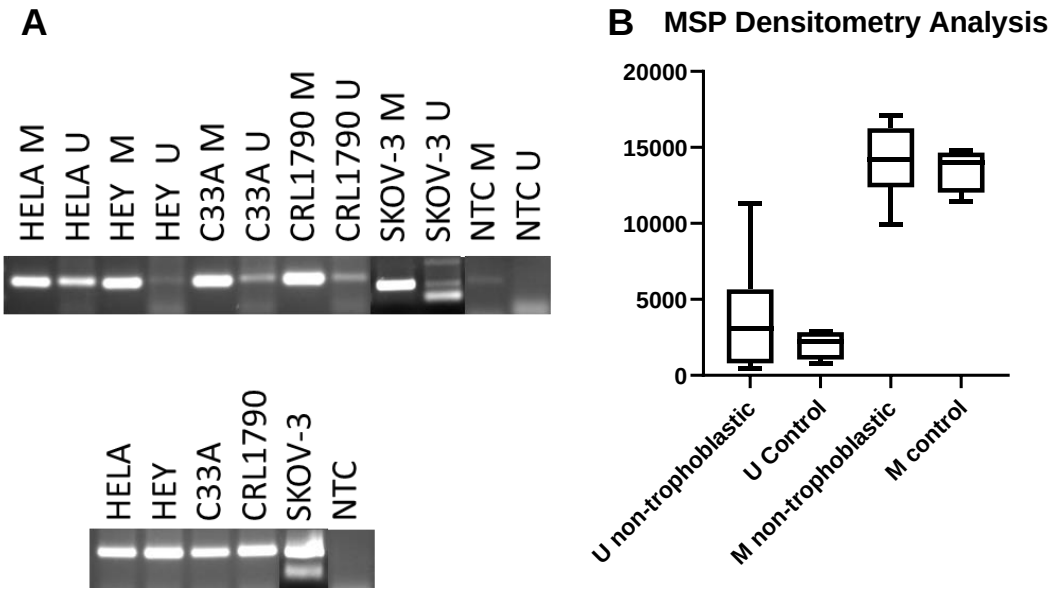



**Fig.10** MSP products and densitometry analysis. A. Methylated, unmethylated and housekeeping gene products run on 2% agarose gels and visualised on Li-cor. B. Boxplot showing the spread of the values obtained from densitometry analysis for each group – U (unmethylated) non-trophoblastic, U control, M (methylated) non-trophoblastic, M control. Line at the middle of box represents median. U - unmethylated, M – methylated, non-trophoblastic(C-33a, HeLa, HEY, SKOV-3), control (CRL-1790).

### 3.2.2 Sequencing primers

The next step to allow for more detailed analysis of methylation profile was to design sequencing primers. Seven primer pairs were selected from the output of MethPrimer as possible candidates (Table 10) (Li and Dahiya, 2002). The predicted product size of the selected primers varied between 200 and 290bp and the common CpGs covered within the region varied between 8 and 11. Pairs 1 and 6 have a predicted product covering the 5' of the putative promoter but not the 3' end of the region of interest. Pairs 2, 3, 5 and 7 are opposite – they cover the 3' part but not the 5' part of the promoter. Pair 4 covers most of the promoter region. As primer sequences were based on *CGB*3, mismatches after bisulphite conversion were observed (Table 11). Pairs 1, 3, 4, and 5 have 1 mismatch in the binding sites on each of the *CGB*5-8 genes for the reverse primer. Pair 2 has 2 mismatches on the reverse primer for *CGB*5 binding region. ePCR predicted products of only the target *CGB*3-8 genes for pairs 2 and 7 (Table 11). Pair 6 also has predicted product for all target genes but has predicted product for *LHB* as well. Pairs 3 and 5 have predicted product on all *CGB*1-8 genes same as pair 4 which has additional predicted product on *LHB*. Pair 1 has 4 predicted products only *CGB*3 and 8 from the target genes in addition to predicted product on *CGB*1-2. For default settings and output from ePCR see Appendix section A2.3.

**Table 10** Selected candidate sequencing primers generated from MethPrimer

| Pair | Primer Sequence | Start (bp from TSS) | Length (bp) | Product size (bp) | number of common CpGs |
|---|---|---|---|---|---|
| 1 | F: GGGAAGGGATTAAGTTTAGATAATGTT | -353 | 27 | 233 | 10 |
| | R: CTACCAAAAAAACCACTTAACCCTA | -121 | 25 | | |
| 2 | F: TTAATAATTAGTTAAATTATTTGAAGTATA | -278 | 30 | 290 | 11 |
| | R: AAAAAAATACTAAACTAAAACCTC | 11 | 24 | | |
| 3 | F: TTTAATAATTAGTTAAATTATTTGAAGTAT | -279 | 30 | 201 | 8 |
| | R: CTTAATTTCTACCCAATAAAAAAAA | -79 | 25 | | |
| 4 | F: GGGAAGGGATTAAGTTTAGATAATGTT | -353 | 27 | 275 | 11 |
| | R: CTTAATTTCTACCCAATAAAAAAAA | -79 | 25 | | |
| 5 | F: TTAATAATTAGTTAAATTATTTGAAGTATA | -278 | 30 | 200 | 8 |
| | R: CTTAATTTCTACCCAATAAAAAAAA | -79 | 25 | | |
| 6 | F: GGGGAAGGGATTAAGTTTAGA | -353 | 20 | 240 | 11 |
| | R: ACTATACTACCAAAAAAACCACTTA | -113 | 25 | | |
| 7 | F: GGGTATTTTGGTTTGAGGG | -208 | 19 | 219 | 8 |
| | R: CCTCAACCCTCCTCTACTT | -8 | 19 | | |

**Table 11** Mismatches of primers with other target genes and predicted products

| Pair | Mismatches in *CGBs* | | | ePCR predicted products | |
|---|---|---|---|---|---|
| | ***5*** | ***7*** | ***8*** | **sense strand** | **antisense strand** |
| **1** | 0 | 0 | 0 | 1.chr19:49031105-49031338 *(CGB2)* | 1.chr19:49024454-49024687*(CGB3)*<br>2.chr19:49037480-49037713(*CGB1*)<br>3.chr19:49049222-49049455*(CGB8)* |
| | 1 | 1 | 0 | | |
| **2** | 0 | 0 | 0 | 1.chr19:49043567-49043858(*CGB5*) | 1.chr.19:49024322-49024612(*CGB*3)<br>2.chr.19:49049090-49049380(*CGB*8)<br>3.chr.19:49055726-49056017(*CGB*7) |
| | 2 | 0 | 0 | | |
| **3** | 0 | 0 | 0 | 1.chr.19:49031179-49031380(*CGB*2)<br>2.chr.19:49043566-49043768(*CGB*5) | 1.chr19:49024412-49024613(*CGB*3)<br>2.chr19:49037438-49037639(*CGB*1)<br>3.chr19:49049180-49049381(*CGB*8)<br>4.chr19:49055816-49056018(*CGB*7) |
| | 1 | 1 | 1 | | |
| **4** | 0 | 0 | 0 | 1.chr19:49031105-49031380(*CGB*2)<br>2.chr19:49043492-49043768(*CGB*5) | 1.chr19:49017519-49017793(*LHB*)<br>2.chr19:49024412-49024687(*CGB*3)<br>3.chr19:49037438-49037713(*CGB*1)<br>4.chr19:49049180-49049455(*CGB*8)<br>5.chr19:49055816-49056092(*CGB*7) |
| | 1 | 1 | 1 | | |
| **5** | 0 | 0 | 0 | 1.chr19:49031180-49031380(*CGB*2)<br>2.chr19:49043567-49043768(*CGB*5) | 1.chr19:49024412-49024612(*CGB*3)<br>2.chr19:49037438-49037638(*CGB*1)<br>3.chr19:49049180-49049380(*CGB*8)<br>4.chr19:49055816-49056017(*CGB*7) |
| | 1 | 1 | 1 | | |
| **6** | 0 | 0 | 0 | 1.chr19:49043492-49043732(*CGB*5) | 1.chr19:49017556-49017794(*LHB*)<br>2.chr19:49024449-49024688(*CGB*3)<br>3.chr19:49049217-49049456(*CGB*8)<br>4.chr19:49055853-49056093(*CGB*7) |
| | 0 | 0 | 0 | | |
| **7** | 0 | 0 | 0 | 1.chr19:49043619-49043838(*CGB*5) | 1.chr19:49024343-49024561(*CGB*3)<br>2.chr19:49049111-49049329(*CGB*8)<br>3.chr19:49055747-49055966(*CGB*7) |
| | 0 | 0 | 0 | | |

### 3.2.3 Global profile of *CGB*3-8 methylation

The sequences generated from the region of interest were mapped back to the promoters of the *CGB*3-8 genes and assigned a Beta value (Appendix A4, Table A4). The Beta value corresponds to the percentage methylation – a beta value of 10 denotes 10% methylation at that site. Fig. 11 presents the percentage of the reads mapped to each gene by cell type. *CGB*5 and 8 contribute most to the data set, and *CGB*7 – the least. *CGB*3 and 7 have some CpGs with low or no reads which were excluded from the data. 3T3 negative control mouse cell line did not return any sequencing result.

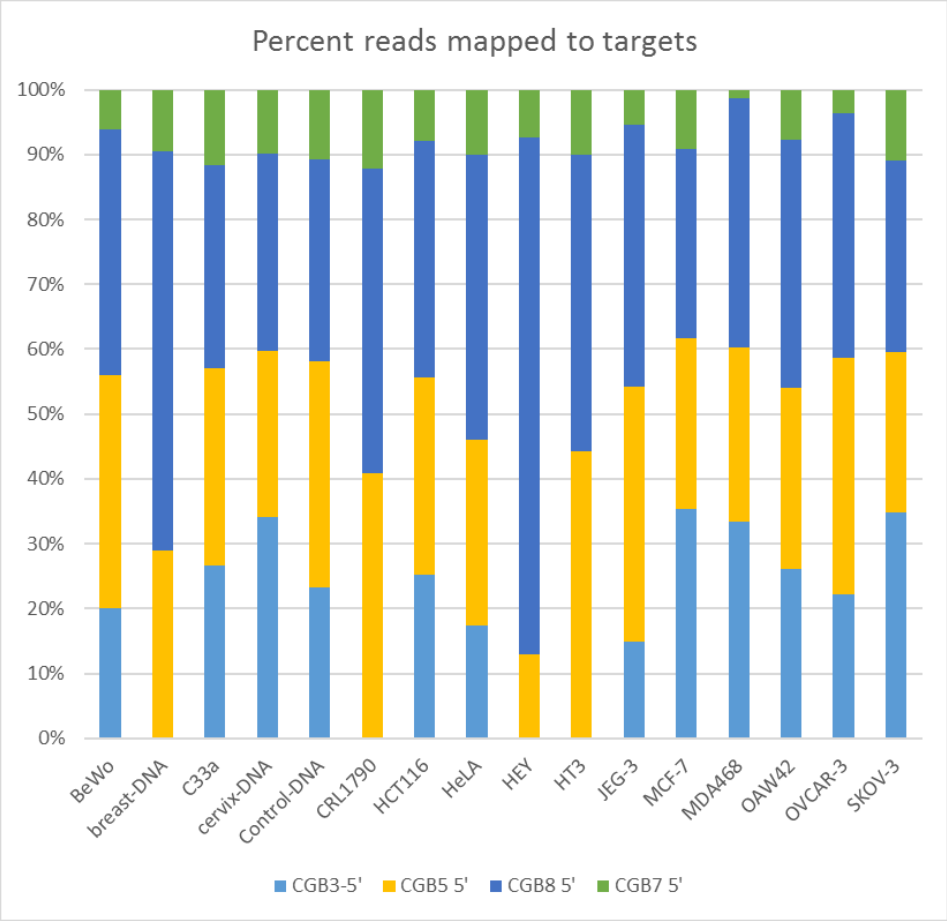



**Fig.11** Percent mapped reads to targets within the *CGB*3-8 genes.

At a first glance the methylation reads within the promoter CpGs seem consistent per gene (Fig. 12, Table 12). *CGB*3, 5 and 8 show higher methylation values across the investigated cell lines. The steady high methylation drops between A7 and A8 across most of the cell lines. *CGB*7 presents with lower Beta values across the CpGs. When looking at individual cell lines JEG-3 and BeWo stand out as having lower methylation reads for *CGB3* but BeWo has beta value for the other genes as other samples, and JEG-3 maintains lower methylation in the other 3 genes. HeLa seems to have lower beta values at the *CGB*3 sites. Observing the histograms in Fig. 13, it shows that predominantly CpG reads range between 90-100% methylation. JEG-3 presents with similar percent methylation reads spanning beta values 20-80% with a peak at 60%. More varied reads of methylation are present in the MCF-7, MDA-MB-468, breast tissue, CRL-1790, and OVCAR-3.

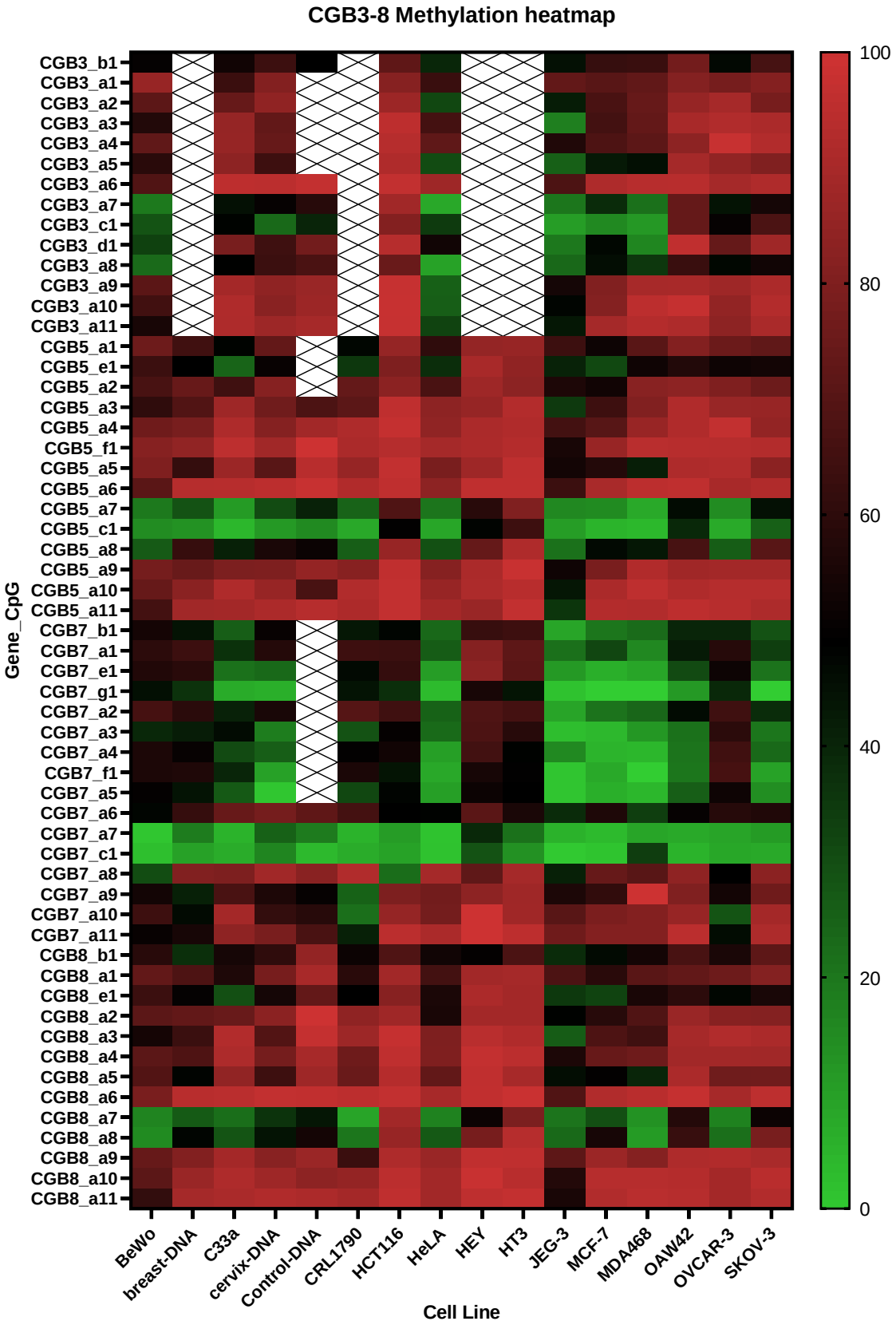



**Fig.12** CpG methylation heat map. The x-axis denotes the cell lines, and the y - denotes the CpG location. Within a gene CpGs are ordered 5'->3'. Green-black-red colour scheme is used to represent the Beta value range (0-50-100%)

**Table 12** Average methylation of the investigated promoter region in the CGB3-8 genes per cell line

| Cell line | Tissue origin | *CGB3* | *CGB5* | *CGB8* | *CGB7* |
|---|---|---|---|---|---|
| **MCF-7** | breast cancer | 63% | 60% | 65% | 30% |
| **MDA-MB-468** | breast cancer | 62% | 68% | 64% | 32% |
| **breast-DNA** | breast tissue | NA | 67% | 65% | 49% |
| **C-33a** | cervical cancer | 75% | 66% | 70% | 44% |
| **HeLA** | cervical cancer | 42% | 65% | 67% | 34% |
| **HT-3** | cervical cancer | NA | 90% | 91% | 62% |
| **cervix-DNA** | cervix tissue | 72% | 70% | 72% | 42% |
| **BeWo** | choriocarcinoma | 55% | 62% | 61% | 47% |
| **JEG-3** | choriocarcinoma | 40% | 45% | 48% | 23% |
| **CRL-1790** | colon | NA | 66% | 66% | 44% |
| **HCT116** | colon cancer | 90% | 88% | 90% | 52% |
| **HEY** | ovarian cancer | NA | 84% | 87% | 68% |
| **OAW42** | ovarian cancer | 85% | 80% | 82% | 43% |
| **OVCAR-3** | ovarian cancer | 76% | 72% | 71% | 48% |
| **SKOV-3** | ovarian cancer | 81% | 76% | 81% | 39% |
| **Control-DNA** | | 73% | 73% | 84% | 51% |

To observe the similarity between the methylation profiles of the cell lines *CGB*3-8 promoter Pearson correlation in package methylKit was used (see code in Appendix A5.1). The correlation study (Fig. 13) between the methylation profiles of the different cell lines showed strong correlation between most of them. The correlation was strong within most tissue groups and between them. For instance, OAW42 shows strong correlations with the other ovarian cell lines – SKOV-3, OVCAR-3, HEY, but also with the cervical group cell lines – HeLa, C-33a, and HT-3. It was also observed that some cell lines had weaker correlations with the others. Both HCT116 and HT-3 seem to correlate strongly with a few cell lines, but otherwise correlations are moderate.

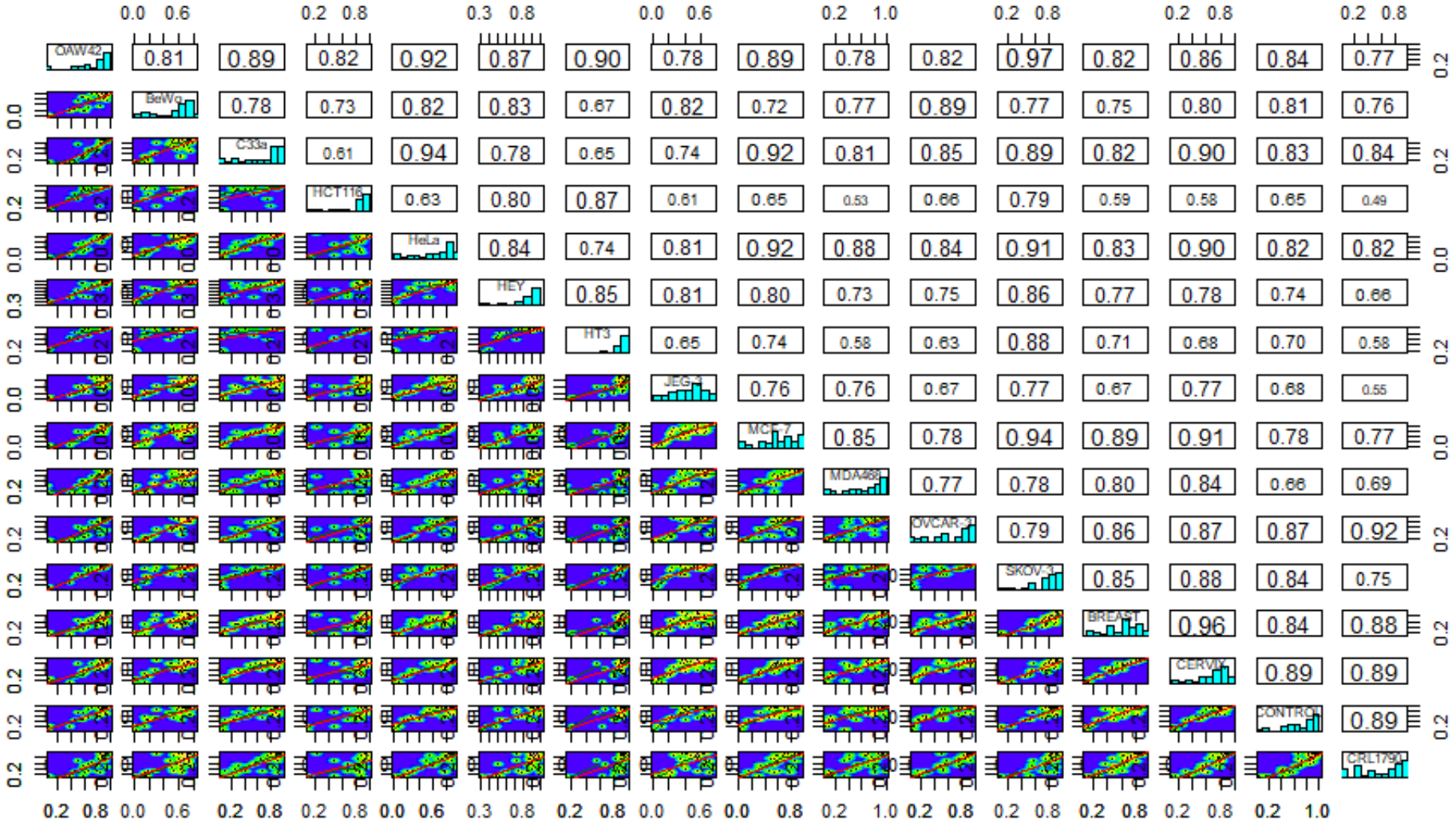



**Fig.13** CpG methylation correlation between samples. Diagonally histograms of CpG percentage reads per cell line are plotted. Right side of diagonal shows the Pearson coefficients between the different samples and the left side – scatterplots of the correlation. Created in methylKit

To further investigate the clustering of the methylation signatures the correlation data was used to build a dendrogram (Fig. 14) (see code in Appendix A5.1). The clustering shows that majority of cell lines have similar methylation profiles. HTC116 shows least correlation with the other cell lines. Together with JEG-3, BeWo, and HEY the 4 cell lines form the cluster furthest away from the other cell lines. OVCAR-3, Control-DNA, and CRL-1790 from a cluster and the rest of the cell lines and tissue DNA form the last cluster which show closer correlation based on methylation footprint.

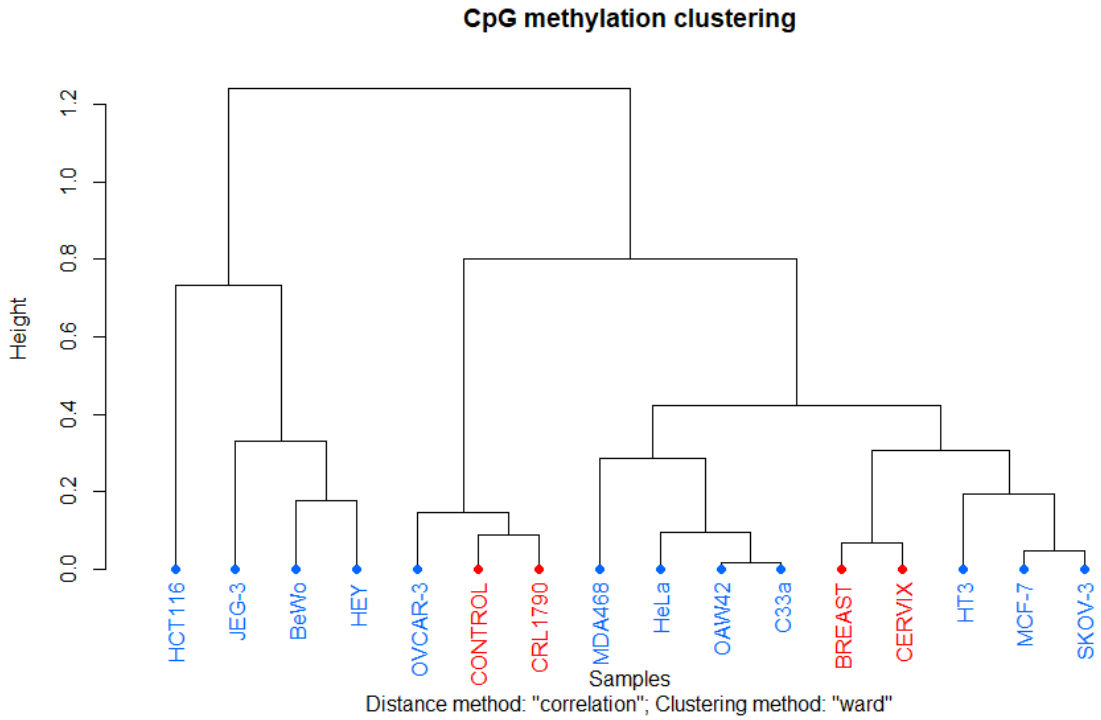



**Fig.14** CpG methylation clustering dendrogram based on correlation data. Produced via methylKit

As the methylation data has high dimensionality Principal Component Analysis (PCA) was applied to observe patterns within the obtained data (see code in Appendix 5.2). PCA reduces the dimensions of the data to 2 principal components (PCs) and retains the maximum variability. In the case of the *CGB*3-8 promoter methylation data the PCA retains 63.7% of the data variability (Fig. 15). Most samples stay fairly close together but 4 distinct clusters can be observed. BeWo and JEG-3 do not cluster with any of the other samples.

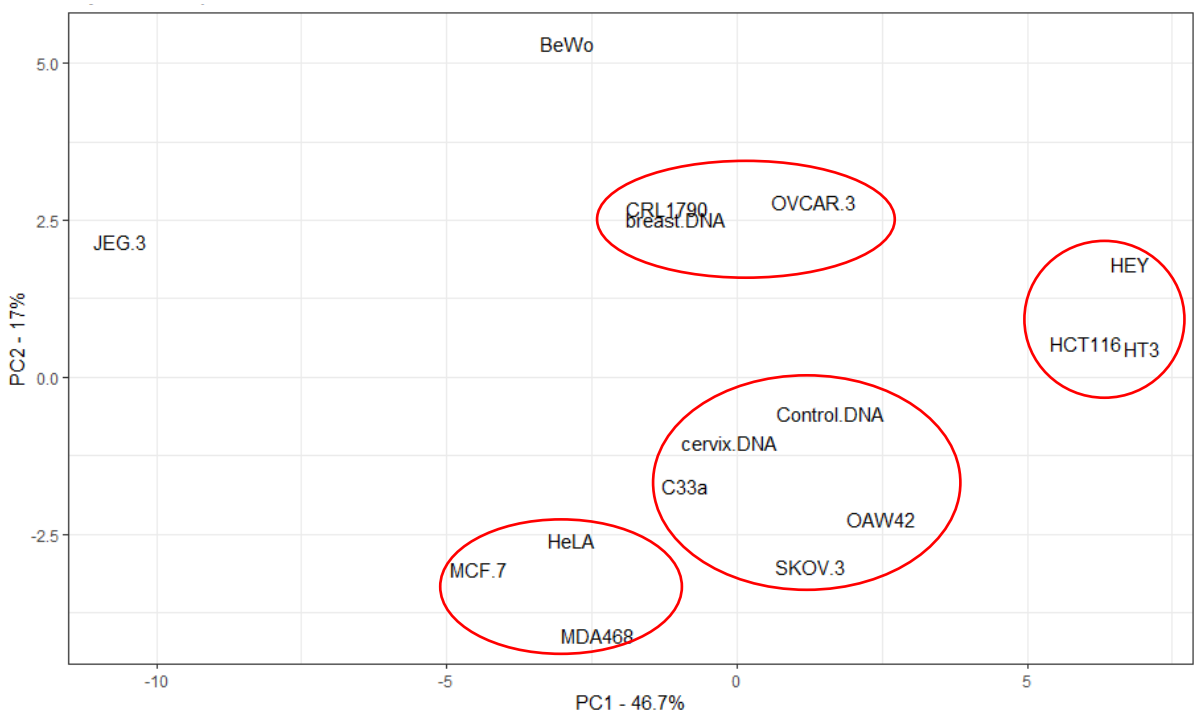



**Fig.15** Principal component analysis of the CpG methylation. Produced using R statistics

Cell line average methylation of region of interest by tissue type was also plotted to observe possible trends in the motif (Fig. 16). As observed previously cell lines maintain their respective higher methylation beta value throughout the promoter region but these values drop between the A7 and A8 CpG. Site D1 which is between A7 and A8 presents with an increase in the percentage methylation for the A7-8 region in the cervical samples (excluding HT-3), ovarian samples (excluding HEY), MCF-7, and Control-DNA. Trophoblastic cell lines present possible lower methylation compared to the control DNA. The 2 cell lines also seem to have distinct motif as beta-values for CpGs at the 5' end do not overlap (Fig. 16A). In the colon group the cancer cell line (HTC116) tends to have higher methylation than the normal colon one (CRL-1790) (Fig. 16B). In the cervical group HT-3 tends to have higher methylation than the other samples within the group which seem to follow similar pattern (Fig. 16C). In ovarian and breast groups samples seem to be following similar pattern of methylation (Fig. 16D).

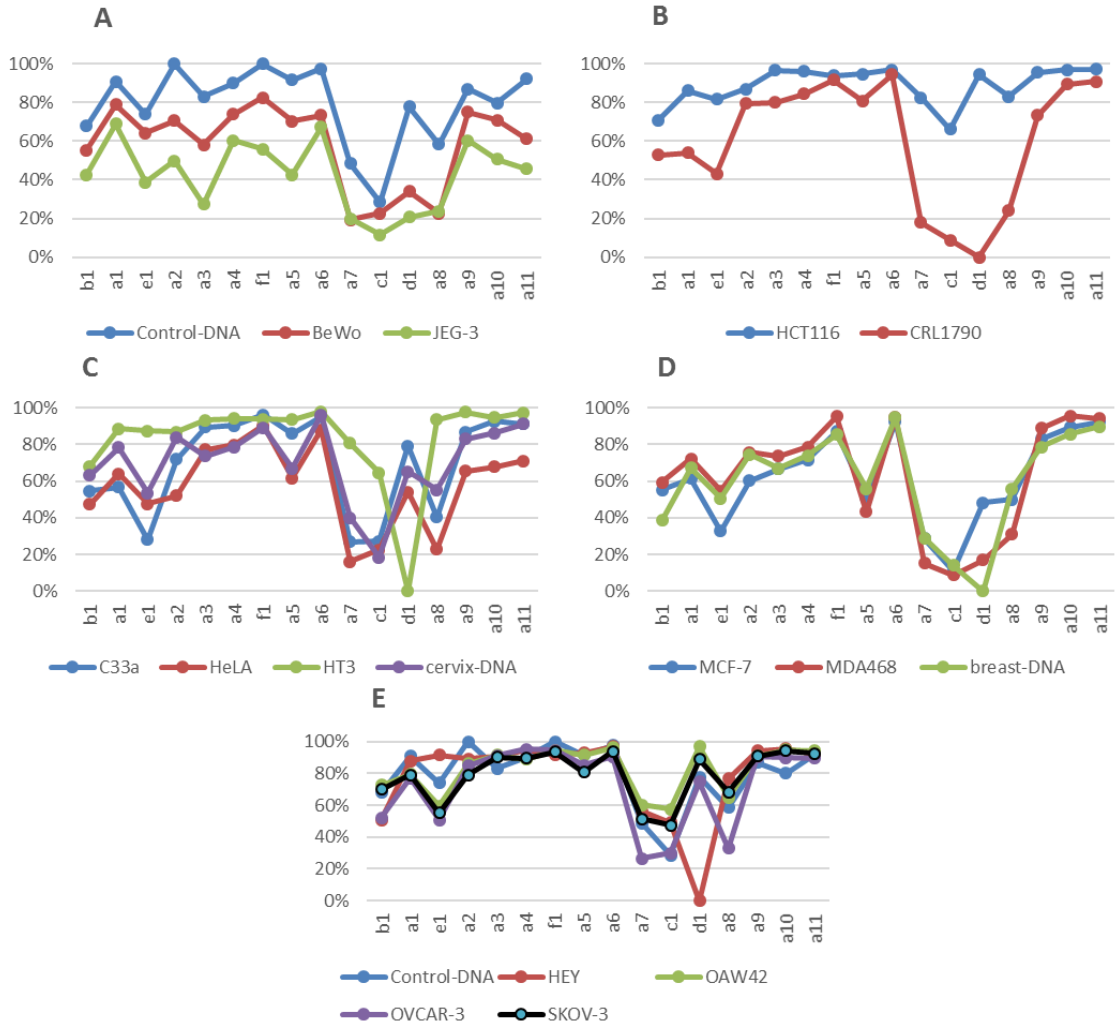



**Fig.16** Samples promoter region averaged methylation grouped by tissue of origin. The x-axis contains the investigated CpGs ordered 5'->3' direction. The y-axis denotes the Beta-value (in %). A. Trophoblastic cell lines and control-DNA B. Colon cell lines. C. Cervical cell lines and cervical tissue. D. Breast cell lines and breast tissue. E. Ovarian cell lines and control-DNA

### 3.2.4 Differential analyses of methylation

To further understand the methylation profile of the region of interest analyses of difference were performed. For these analyses the Beta-value was converted to M-value using the formula $M = log_2(\frac{beta\ value}{1-beta\ value})$ to allow parametric statistical analyses as mentioned in Du *et al.* (2010) (Table A5, Appendix A4). Firstly, the promoter methylation per gene was checked for difference. The average promoter M-value per cell line per gene was compared using ANOVA (Appendix A6.1). The results show strong significance of difference ($p<0.0001$) in the methylation of the genes. Tukey's pairwise comparison test was performed to identify where the difference is. The results show that *CGB7* has significantly lower methylation in

comparison with the other 3 genes ($p<0.0001$ for each of the differences with *CGB*3, 5, and 8).

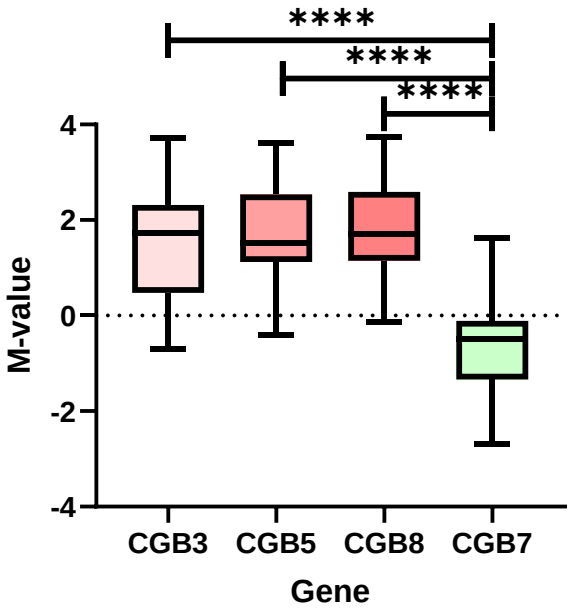



**Fig.17** Box plot of the averaged cell line M-value per investigated *CGB* gene. Lines above the box denote the significance in difference based on Tukey's test.

The next step in the differential analysis was to assess if there is a difference in the methylation profile between trophoblastic and non-trophoblastic cancers and control sample DNAs. Firstly, the averaged M-values of *CGB*3, 5 and 8 genes whole investigated region per cell line were compared (Figs. A23-24, Appendix A6.2). In Fig. 18A it can be seen that the trophoblastic and control samples have narrower spread of the M-value. Trophoblastic samples seem to have lower methylation compared to the other 2 groups. The analysis of difference used was the non-parametric Kruskal-Wallis due to small sample size in the trophoblastic group. A difference of weak significance was returned as result ($p=0.0404$); however, Dunn's pairwise test showed no significant differences between the groups.

The average methylation of *CGB*3, 5, and 8 was also compared in two specific regions within the investigated promoter (Figs. A25-26, Appendix A6.2). Fig. 18B shows the float bar plot for the 5' TF binding region binding AP2α-SP1-AP2α-OCT3/4 transcription factors, which covers the B1, A1-3, and E1 CpG sites. This 5' TF region has more widely spread averaged M-values within the control samples compared to the whole promoter region. The spread of the trophoblastic and non-trophoblastic samples were similar to the whole promoter region. The 5'

TF region was analysed analogically to the whole investigated region. No significant difference was observed.

The second specific region is within the investigated promoter's 3' TF binding site binding AP2α-SP1 transcription factors and covering A4-6 and F1 CpGs (Fig. 18C). Averaged M-values are widely spread within the normal and non-trophoblastic regions. Trophoblastic samples seem to be having lower methylation as observed in Fig. 18C. The same approach was used to analyse for difference as with the other 2 regions. No significant difference was found between the 3 groups.

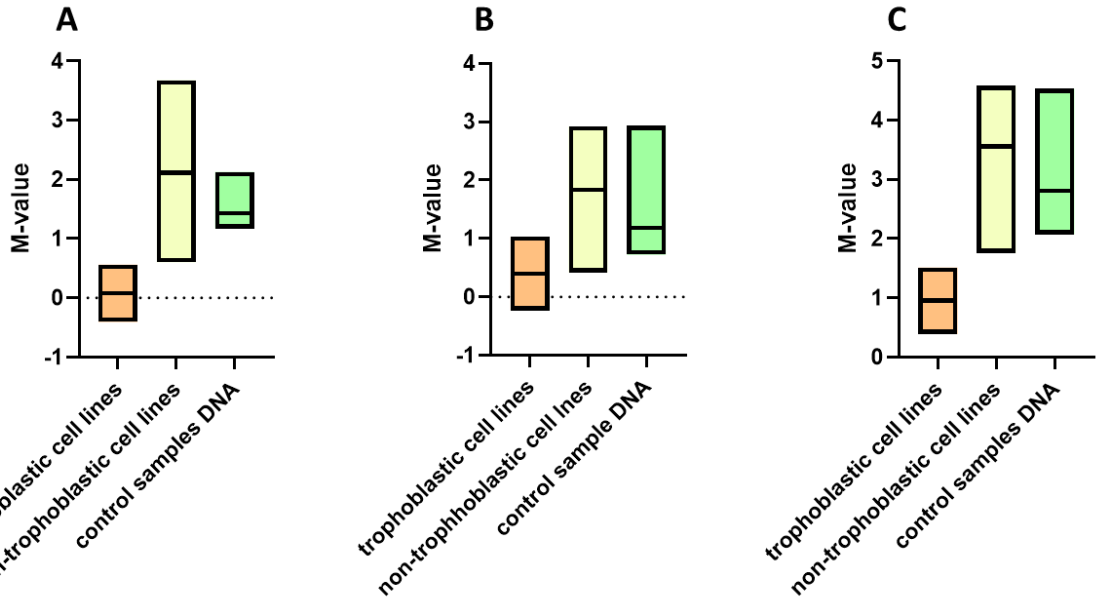



**Fig.18** Float bar plot representing the averaged *CGB*3, 5, and 8 methylation per cell line grouped by cancer type. A. Whole region of interest. B. AP2α-SP1-AP2α-OCT3/4 region. C. AP2α-SP1 region

The same analysis was repeated for the *CGB*7 averaged methylation (Figs. A27-29, Appendix A6.2). As seen in Fig. 19 the data is spread wider in the trophoblastic and non-trophoblastic cell lines in comparison to the control samples. Methylation seems to be similar between the sample groups in the whole interest region and in the 5' and 3' TF biding region. As previously, Kruskal-Wallis was applied to test the difference in methylation in the above mentioned regions which yielded no significant result.

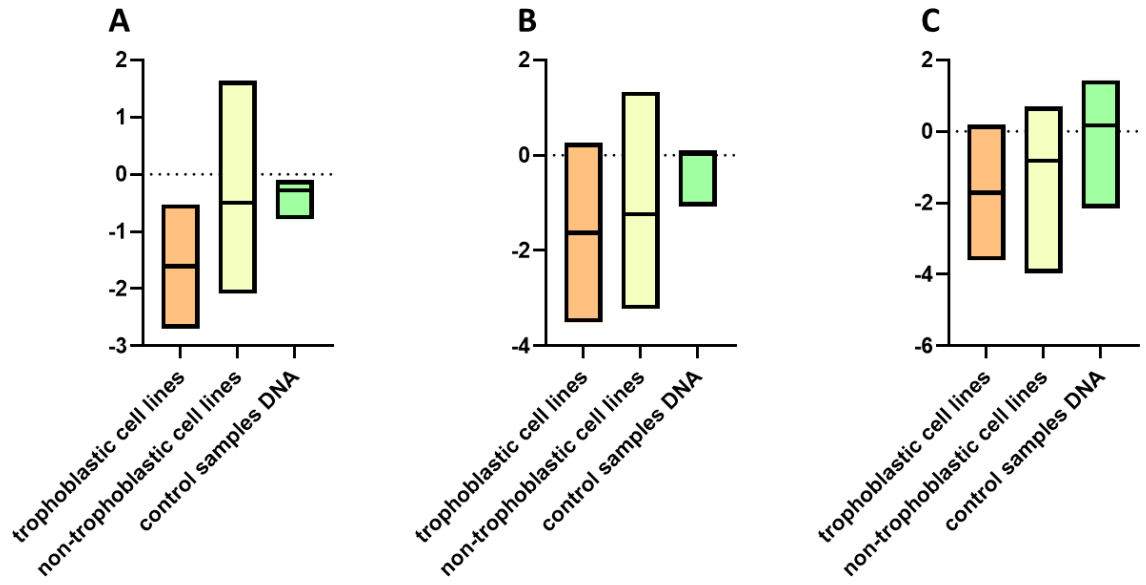



**Fig.19** Float bar plot representing the averaged *CGB*7 methylation per cell line grouped by cancer type. A. Whole region of interest. B. AP2α-SP1-AP2α-OCT3/4 region. C. AP2α-SP1 region

The last set of tests to see a difference in the methylation profile analysed the difference in M-values within tissue of origin groups. The M-values for *CGB*3, 5, and 8 were averaged per CpG and samples were grouped by tissue of origin. Also the difference in methylation within the control samples was assessed. Repeated measure ANOVA (RM ANOVA) was to assess the difference within groups of 3 or mores samples. This was paired with Tukey's pairwise comparison to establish where the significant difference lies. For sets of 2 samples paired t-test was used (Appendix A6.3.1-6). No significant difference in the methylation profile was found in the breast and ovarian tissue group (Fig. 20B and E).

In the trophoblastic cell line group, it can be observed that both the cell lines have lower M-values in comparison to the control-DNA (Fig. 20A). The performed RM ANOVA confirms that there is a very strong significant difference ($p<0.0001$). Tukey's test reveals that control-DNA is significantly more methylated than JEG-3 and BeWo (both with $p<0.0001$). It also revealed that BeWo is significantly more methylated than JEG-3 ($p<0.0001$).

In the control samples group the control-DNA seems to have slightly higher methylation (Fig. 20C). The RM ANOVA confirms moderate significant difference between the M-values ($p=0.0014$). Tukey's test revealed that control-DNA is more methylated with moderate significance than CRL-1790($p=0.0096$) and breast-DNA ($p=0.0079$) and weak significance from cervix-DNA ($p=0.0114$).

Observing the cervix group no difference stands out between the samples at a first glance (Fig. 20 D). RM ANOVA shows that there is a difference of weak significance (p= 0.0125). Tukey's test shows that the difference comes from lower methylation in the HeLa cell line in comparison with C-33a (p=0.0014), cervix-DNA (p=0.0041), and HT-3 (p=0.0209).

In the colon group there were only 2 samples to compare (Fig. 20F). HCT116 seems to be more methylated than the normal colon cell line CRL-1790. Paired t-test confirms that with strong significance (p=0.0003).

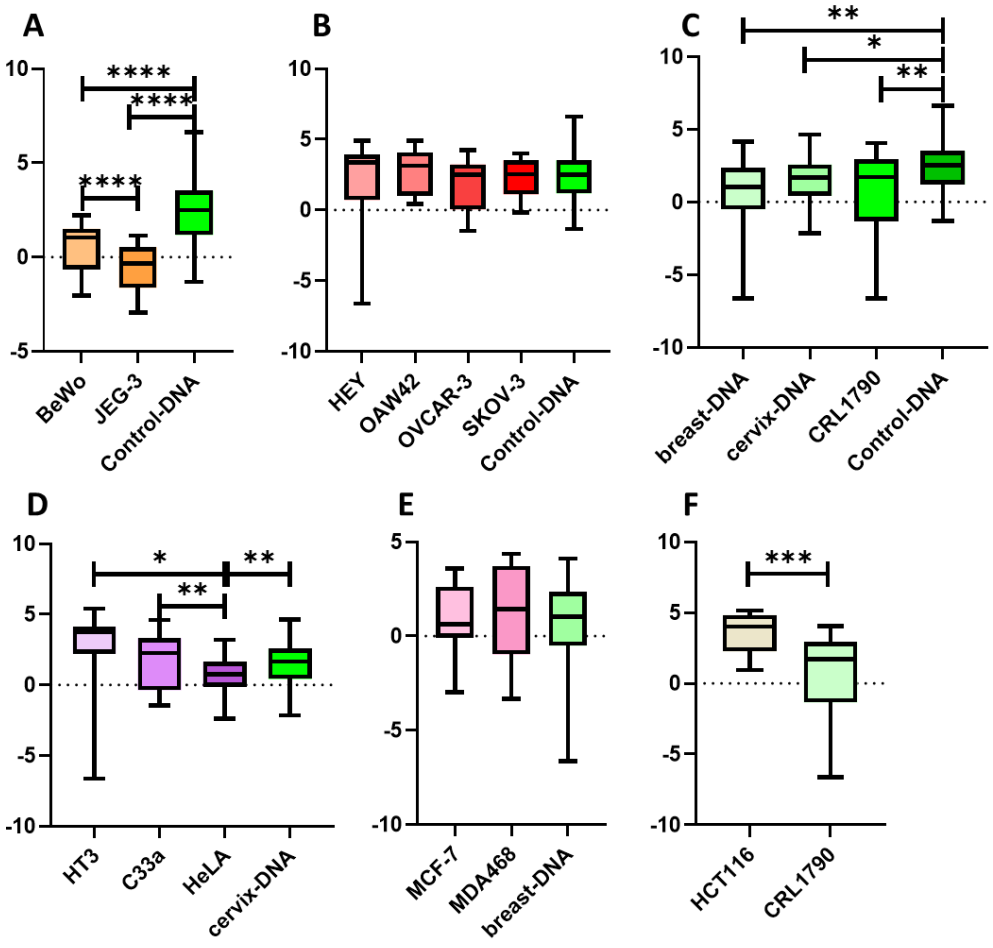



**Fig.20** Boxplot of the average M-values for *CGB*3, 5, and 8 of each sample compared within their respective tissue of origin group. A. Trophoblastic cell lines and control DNA. B. Ovarian cancer cell lines and control DNA. C. Control samples. D. Cervical cancer cell lines and normal cervix tissue. E. Breast cancer cell lines and normal breast tissue DNA. F. Colon samples. Lines above boxes denote significance of difference between groups based on Tukey's test (A-E) or paired t-test (F)

Analogically, the M-values of the *CGB*7 gene were analysed for differences in methylation within the samples' tissue of origin group (Appendix A6.3.7-12). Control-DNA sample was excluded from the analysis due to missing data points

preventing the RM ANOVA test. No significant difference was observed in the control DNA samples and colon groups (Fig. 21C and F).

JEG-3 tends to have lower methylation in the trophoblastic group for *CGB*7 (Fig. 21A). RM ANOVA confirms difference of moderate significance ($p=0.0020$) within the trophoblastic group. Tukey's test shows that the difference is due to lower methylation of JEG-3 in comparison to CRL-1790 ($p=0.0120$) and BeWo ($p=0.0126$).

In the ovarian group HEY tends to have higher methylation values than the other samples for *CGB*7 (Fig. 21B). Statistical test presents moderate difference within the group ($p=0.0047$). Pairwise comparisons reveal that HEY has higher methylation than OAW42 ($p<0.0001$), OVCAR-3 ($p=0.0002$) and OVCAR-3($p=0.0396$).

HT-3 cell line form the cervical group appears with slightly higher methylation when observing Fig. 21D. RM ANOVA presents that there is a difference in the samples with strong significance ($p<0.0001$). HT-3 is more methylated than HeLa ($p<0.0001$) and C-33a ($p=0.0009$) with strong significance and cervix-DNA ($p=0.0106$) with weak significance. C-33a is also more methylated than HeLa ($p=0.0187$) according to Tukey's.

Finally, in the breast group breast-DNA seems to have higher methylation than the breast cancer cell lines (Fig 21E). RM ANOVA confirms a weak significance in the difference of $p=0.0333$ in the group. Tukey's test shows that the difference lies between MCF-7 and breast-DNA ($p=0.0100$) whereby MCF-7 has lower methylation than breast-DNA.

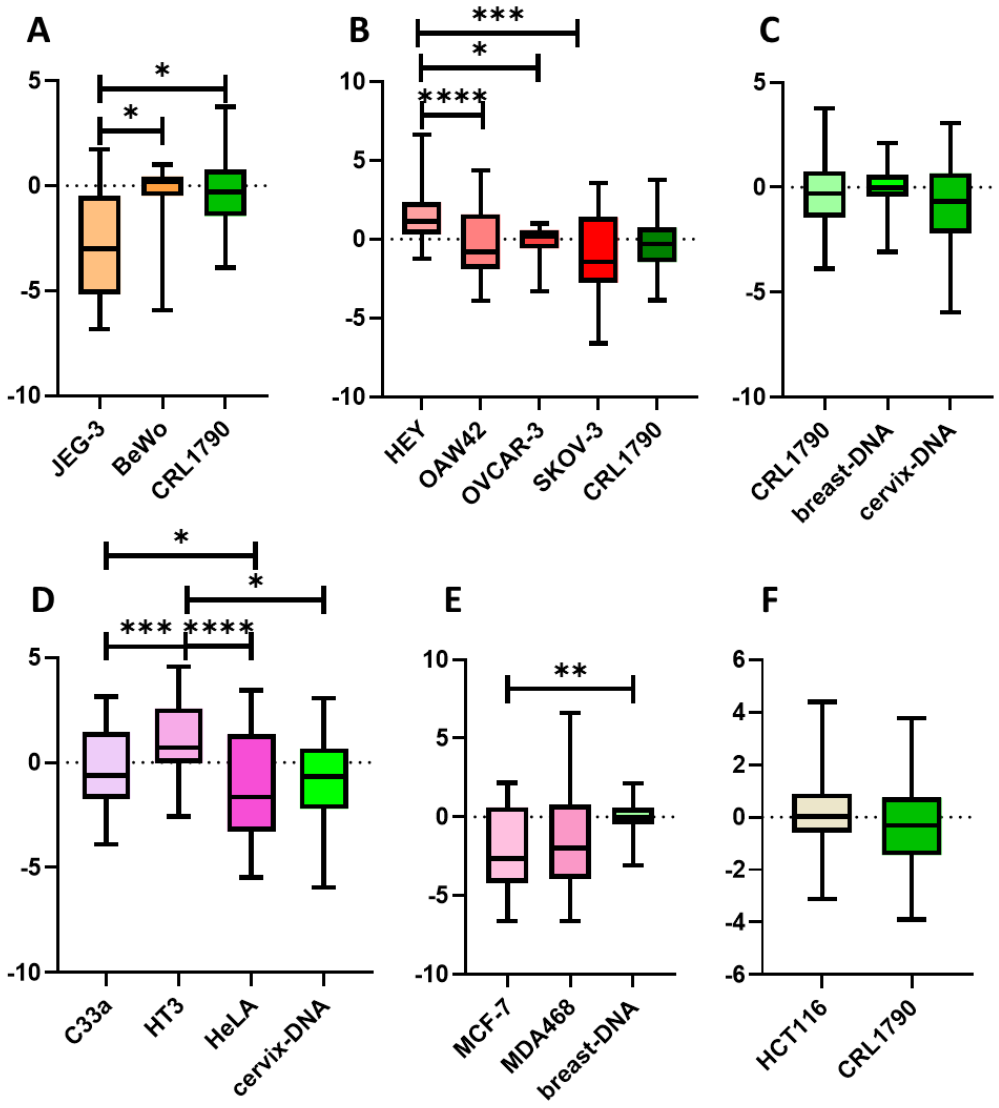



**Fig.21** Boxplot of the M-values for *CGB*7 of each sample compared within their respective tissue of origin group. A. Trophoblastic cell lines and control DNA. B. Ovarian cancer cell lines and control DNA. C. Control samples. D. Cervical cancer cell lines and normal cervix tissue. E. Breast cancer cell lines and normal breast tissue DNA. F. Colon samples. Lines above boxes denote significance of difference between groups based on Tukey's test (A-E)

## 3.3 Transcription level of *CGB*3-8

Relative quantification expressed as fold difference for the *CGB*3-8 mRNA transcript was calculated by comparing the cancer sample $C_t$ ($C_q$) values to the normal cell line CRL-1790 (Appendix A7). The trophoblastic cell lines present with very high fold difference in mRNA levels in comparison to the normal cell line (Fig. 22A; BeWo: 1767-fold; JEG-3: 588-fold). From the non-trophoblastic cell lines (Fig. 22B) HTC116 and SKOV-3 have highest mRNA transcript levels – 127- and 121-fold difference, respectively. OAW42, HT-3, and C-33a also tend to have moderately higher transcription level of the *CGB*3-8 genes – 65-, 57-, and 30-fold difference respectively. MDA-MB-468, OVCAR-3, MCF-7, and HEY tend to show

slightly higher transcription of the beta-subunit genes – 10-, 8-, 7-, and 2-fold difference respectively. HeLa shows lower transcription level of the *CGB*3-8 genes with a 0.29-fold difference from the normal CRL-1790. 3T3 did not produce any product as expected.

The data from the transcription study was compared for association with the methylation level (Appendix A7.1). Spearman's ranked correlation was used to perform the test. The transcription fold difference was compared to the average whole promoter methylation, 5' TF binding, and 3' TF binding regions. No significant association was discovered between the methylation profiles and transcription level.

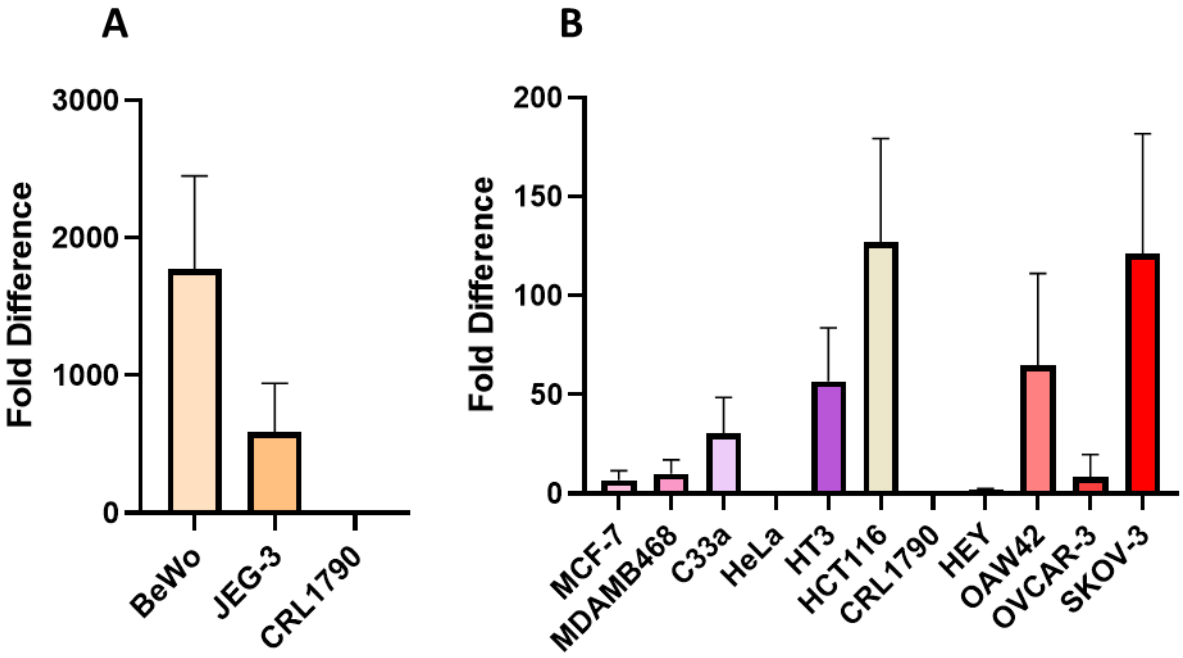



**Fig.22** Bar chart showing fold difference in mRNA transcript in the investigated cell lines. A. trophoblastic cell lines and CRL-1790. B. Non-trophoblastic cell lines and CRL-1790

## 3.4 Secretion of hCG and hCG-beta

The concentration of intact hCG and free hCG-beta was calculated using the standard curves produced by each ELISA and the samples' optical density. Each ELISA had internal controls provided with the kit (Appendix A8). The calculated concentration for those in each assay matched the expected range provided by manufacturer. In the ELISA for intact hCG, the target molecule was only detected in the 2 trophoblastic samples – BeWo and JEG-3. BeWo media had hCG concentration of 117mIU/ml and JEG-3 -121 mIU/ml. The values were converted

to ng/$10^6$cells/24h to adjust for the number of cells and days the media was conditioned. JEG-3 had higher secretion of intact hCG (8.39 ng/$10^6$cells/24h) than BeWo (7.7 ng/$10^6$cells/24h).

For the free beta hCG ELISA 4 samples were positive for the free beta subunit. The molecule was found in the media of the 2 trophoblastic cell lines (BeWo – 3.67ng/ml and JEG-3 – 1.78ng/ml) and 2 non-trophoblastic cell lines (HEY – 1.37ng/ml SKOV-3 – 4.6ng/ml). After adjusting for cells and days in culture HEY seems to secrete the most free beta hCG (5.35 ng/10^6cells/24h), followed by SKOV-3 (2.88 ng/10^6cells/24h), BeWo (2.21 ng/10^6cells/24h), and JEG-3 (1.13 ng/10^6cells/24h). No statistical analysis was performed on the data from ELISA due to the small sample size.

# 4. Discussion

hCGβ has been studied extensively in relation to its ectopic production in cancers. Presence of hCGβ is indicative of worsened prognosis and increased metastasis. However, there is still missing knowledge in the molecular mechanism of re-activation in cancer **(**Zhong *et al.*, 2019; Schüler-Toprak *et al.*, 2017; Szczerba *et al.*, 2016; Kubiczak *et al.*, 2013; Jankowska *et al.*, 2008).

In previous reports *(*Grigoriu *et al.,* 2011; Campain *et al., 1993) CGB* genes demonstrate hypomethylated profile as the genes are important during pregnancy and actively transcribed. In normal physiology as the genes are not needed, higher methylation of the genes is observed in comparison to the state during pregnancy. This confers the silent state of these genes (Grigoriu *et al.,* 2011). This suggests that reactivation in cancer should also present with decrease of the methylation.

In the present study the main focus was investigation of the methylation profile of the *CGB*3-8 genes promoter in non-trophoblastic cancer from epidermal origin. Previous research has found ectopic hCGβ in bladder, colon, lung, ovarian, cervical and breast cancers (Zhong *et al.*, 2019; Schüler-Toprak *et al.*, 2017; Szczerba *et al.*, 2016; Kubiczak *et al.*, 2013; Jankowska *et al.*, 2008). However, only 3 other studies have been found to investigate methylation in non-trophoblastic cancer tissues and cell lines (Śliwa *et al.*, 2019; Campain *et al.*, 1993; Whitfield and Kourides, 1985).

## 4.1 MSP analysis did not find differential methylation

Initially, studies via MSP were applied to investigate the *CGB3*-8 promoter region methylation on a few selected non-trophoblastic cell lines. The selected cell lines are representative of previously reported hCG-β positive non-trophoblastic cancers (Sinnappan, 2015; Acevedo *et al.*, 1992). Thus, it was suspected that the cancer cell lines could possess hypomethylated profile in comparison to a normal cell line (CRL-1790). Densitometry analysis of the MSP products did not yield significant difference in the methylation profile of *CGB* genes in the investigated cancer cell lines versus the normal cell line.

However, the applied MSP approach has its caveats that may have not represented the full picture of the *CGB* promoter methylation. Firstly, comes the caveat of sample size and heterogeneity. DNA methylation profile is

heterogeneous between different tissues and within cancer tumours (Wen *et al.*, 2017). The samples investigated via MSP were from 2 different tissues of origin – ovarian and cervical cancer. Therefore, their methylation profile could be quite varied. Combined with the small sample, size it may not be a good representation of the wider non-trophoblastic cancer population.

Other limitations of the MSP approach are that the obtained data is semi-quantitative and relative only to the CpGs contained in the primers. The primers for MSP are designed in such a way that they cover the same CpGs – one set detecting their methylated version and one – detecting the unmethylated version (Fraga and Esteller, 2011; Ammerpohl *et al.,* 2009). Therefore, information on any CpG site within the product is not represented, unless the product is sequenced. In this study the MSP targets only 3 CpG sites within the promoter – one located -550bp upstream from TSS and 2 between the Oct3/4 and AP2 binding site – A5 and F1 (Fig. 9). Furthermore, the data obtained cannot be reliably quantitated and MSP does not provide nucleotide level resolution of the sample methylation (Fraga and Esteller, 2011).

## 4.2 MiSeq based bisulphite sequencing

In order to achieve more detailed picture of the DNA methylation within the *CGB3-8* promoter MiSeq (Illumina) based targeted bisulphite sequencing was applied. The first step was to design primers amplifying the region of interest. Bisulphite conversion of DNA changes the original sequences by converting non-CpG cytosines to uracils. CpG cytosines remain ambiguous due to their possible methylation (Li and Tollefsbol, 2011). Bisulphite sequencing primers should be able to amplify methylated and unmethylated sequences with equal efficiency. That is why the primers should not contain CpG sites (Correa *et al.*, 2012; Warnecke *et al.*, 2002). Another criterion is that the length of the product should not be longer than 300bp which is the maximum read length of MiSeq (Soto *et al.*, 2016). Further to that, it is recommended that 2 primer sets are used for the same locus in a way that they overlap. This is to prevent mispriming and amplification of non-desired regions (Correa *et al.*, 2012, Warnecke *et al.*, 2002).

Seven primer pairs based on *CGB*3 were selected as possible candidates that fit the above mentioned criteria. A further complication to the primer design was that the oligonucleotides need to amplify 4 different genes simultaneously. To resolve this, the primer sequences were checked for mismatches with the *CGB*5-8 genes

and then run in the BiSearch ePCR tool to check for predicted products (Table 11) (Li and Dahiya, 2002). Due to mismatches in the primer sequence and possible amplification of more than the target genes (primer pairs 3-5) or no predicted amplification in all genes of interest (primer pair 1), primer pairs 1-5 designs were not used for the downstream sequencing. Primer pairs 6 and 7 were selected as they had no mismatches and were predicted to amplify all genes of interest. Furthermore, the product sequences also overlap to ensure specific target amplification. However, a possible issue with these pairs was the possible *LHB* product in primer pair 6. Nonetheless, the primer products overlapping and/or mapping of the library post-sequencing should resolve this issue based on single nucleotide differences (Correa *et al.*, 2012; Warnecke *et al.*, 2002).

The selected primers were run in optimised PCR assay to create the library based on bisulphite converted DNA from the studied cell lines and tissue samples which was consecutively sequenced using next generation sequencing MiSeq Illumina platform. The MiSeq platform has been previously employed with methylation studies to reliably obtain methylation status and differences in various genes in conditions ranging from cancers to psychiatric disorders (Dukal *et al.*, 2016; Roeh *et al.*, 2016; Masser *et al.*, 2013; Xiong *et al.*, 2012). It has been shown that this technology is cost-effective, high-throughput, and also sensitive even in samples where target DNA sequence is in low amount (Ward *et al.*, 2016; Luthra *et al.*, 2013). MiSeq has been reported as a tool for accurate absolute 5-methyl cytosine quantification in low diversity samples such as the bisulphite-converted cytosine-poor DNA (Masser *et al.*, 2013).

## 4.3 Trends in DNA methylation of *CGB3*-8 promoter

Sample sequencing was performed on total DNA from selected cell lines and internal control DNA sample (labelled control-DNA). The sequencing data was aligned to *CGB3-8* reference sequences using commonly employed Bismark software. The software is a reliable tool for alignment and methylation calling of sequencing data obtained from varied bisulphite-based sequencing methods (Xiong *et al.*, 2012; Krueger *et al.*, 2011). Based on the methylation calling, a beta value is assigned at each site covered by the sequencing which is the percentage of methylated reads from the total reads at a particular CpG site (Du *et al.*, 2011).

The sequencing interrogated the methylation status of a total of 57 CpG sites across the promoter of the *CGB*3-8 genes. This included 14 sites in *CGB*3, 5; 13 -

in *CGB*8; and 16 - in *CGB*7. Most mapped reads come from *CGB*5 and *CGB*8. The generally observed trend was that *CGB*3, 5 and 8 were more methylated compared to *CGB*7 which showed lower DNA methylation across the 13 of the studied cell lines and 2 normal tissue samples.

Looking at the trends of the methylation pattern within samples, it is observed that the positive control trophoblastic cells JEG-3 show varied methylation reads at individual sites (Fig. 13 histograms) but tend to have lower methylation. The other positive control, BeWo, has lower average methylation especially at the *CGB3* gene but higher at *CGB7*. These findings match results from Campain *et al.,* (1993) that claim choriocarcinoma is hypomethylated. Both positive control cell lines are used as models to study human trophoblast *in vitro*, therefore they share some phenotypic features with the placenta (Serranoa *et al.*, 2007; Wolfe, 2006). Grigoriu *et al.,* (2011) performing studies on human placenta DNA methylation, have confirmed hypomethylation of the *CGB*3-9 genes in trophoblast cells during gestation.

The negative control cell line 3T3 did not have any reads as expected. 3T3 is a mouse fibroblast cell line which does not have the *CGB* genes as these are only found in human and some primate species (Tuncay *et al.*, 2018; Fournier, 2016; Rao, 2016). In the other non-trophoblastic cell lines interestingly, HeLa presented with lower methylation reads at the *CGB3* gene averaging at 42%. Cell lines and tissues from breast origin (MCF-7, MDA-MB-468, normal breast tissue), the normal colon (CRL-1790) and ovarian OVCAR-3 showed varied methylation reads at investigated CpG sites (Fig. 13). Another interesting feature in the data was that HEY, HT-3 and HCT116 cell lines tend to have higher average methylation in the *CGB7* gene in comparison to the relatively low reads in the other samples (Table 12).

The trends of the averaged methylation for *CGB*3-8 genes per CpG site in the different tissue groups present similar as the ones stated above (Fig. 16). Compared to the control-DNA, JEG-3 and BeWo tend to have lower methylation as reported previously (Campain *et al.* 1993). The variability in methylation reads per site is visualised for the breast tissue group, JEG-3, and CRL-1790. Some samples seem to have distinct patterns hinting differences in methylation. In Fig. 16 the lines of JEG-3, BeWo, and control in trophoblast group; the lines of CRL-1790 and HCT116 in colon group; and the HT-3 line in cervical group separate distinctly from the other samples within the respective groups.

Only limited other studies (Campain *et al.*, 1993; Whitfield and Kourides, 1985) have been found from literature to perform some degree of research into the *CGB* genes DNA methylation of non-trophoblastic cancer cell lines. Campain *et al.,* (1993) studies 2 non-trophoblastic cell lines – normal transformed fibroblast from lung (2RA) and ectopically hCG producing glioblastoma (CBT). The research discovered that both present with some degree of demethylation. In Whitfield and Kourides, (1985), D-98 (cervical cancer), Lu-65, and Sand (both lung tumour) cell lines were studied which showed some degree of methylation changes. However, these findings are from more than two decades ago and are based on restriction enzyme analyses of methylation. RE analyses on their own are outdated as they require high quality DNA and are restricted to enzyme recognised sites (Fraga and Esteller, 2011; Ammerpohl *et al.,* 2009). A more recent study investigated methylation of *CGB*3-9 genes in ovarian cancer tissue via MSP (Śliwa *et al.*, 2019). It was reported demethylation of the *CGB* genes in cancer tissue corresponds to the increase of mRNA transcript in the diseased samples. However, the methodology used by Śliwa *et al.*, (2019) limits the investigation to only a few sites covered by the primers of the MSP (Fraga and Esteller, 2011; Ammerpohl *et al.,* 2009). In the present research a more reliable method is applied based on bisulphite conversion that can interrogate all CpG sites in the region of interest spanning the *CGB* promoter region of approx. 350bp (Masser *et al.*, 2013).

## 4.4 Similarities in DNA methylation of cancer cell lines

To further elucidate the methylation profile in the studied cell lines and tissues the data from the sequencing was analysed with the software package methylKit (Akalin *et al.*, 2012). This approach of analysis is based on the R programming language. The package is flexible in its data input and allows for rapid analyses in the realm of DNA methylation from high-throughput methylation sequencing. methylKit can be used to summarise and cluster data from sequencing and visualise patterns from the supplied methylation calls. This allows discerning outliers in the data set and finding similarly methylated samples (Wreczycka *et al.*, 2017; Akalin *et al.*, 2012).

Based on the data in this research Pearson correlation analysis was used initially to understand similarity of the investigated samples methylation profile in the *CGB3-8* genes promoter. The methylation profiles showed high Pearson

coefficient between samples within and between different tissue of origin groups. HCT116 and HT-3 cell lines stand out as they present lower correlation values to the other samples (Fig. 13)

To further clarify the similarity of the samples, two different approaches were applied in methylKit – plotting a dendrogram based on correlation results and principal component analysis (Figs. 14 & 15). In these approaches CpG sites that have missing data have been excluded (Akalin *et al.*, 2012). The dendrogram clusters the majority of the samples relatively close together. HCT116 diverges furthest from the other samples based on its correlation. JEG-3, BeWo and HEY show closer correlation but still their correlation values are different from the rest of the samples. Together the four cell lines cluster furthest away from the rest of the investigated samples. Another more defined cluster is formed between OVCAR-3, Control-DNA, and CRL-1790. These samples' correlation is closer to the majority of the rest of the samples in comparison to the above mentioned cell lines.

The above discussed clustering approach by correlation, however, may not be the most appropriate for this data set. The distance method applied in Fig. 14 to build the dendrogram is solely based on the Pearson coefficient (Akalin *et al.*, 2012). Comparing the averaged methylation data per gene and the dendrogram reveals a few discrepancies. JEG-3 and BeWo cell lines showed lower methylation values compared to HCT116 and HEY with which they are clustered. CRL-1790 stands out in its cluster as the cell line with low methylation values. OAW42 and HT-3 cell lines in the big cluster show higher average methylation per investigated gene compared to the other samples. Therefore, a different approach for clustering might be more appropriate.

Principal component analysis (PCA) is a technique in statistics that can be applied to large data sets with multiple variables attached. It aims to reduce the complexity of data by reducing the number of variables and retaining as much as possible from the original data (Akalin *et al.*, 2012; Joliffe and Morgan, 1992). This can help distinguish outliers in a data set and show more clearly patterns which may have been omitted by simple observation. The PCA creates new variables, which are linear functions of original variables called principal components (PC). The PCs are ordered so that the first PC hast the highest variance among the rest of the PCs, the second PC - has second most variance, and so forth. PCs are

built from a correlation matrix of the original data which aims to standardise it (Joliffe and Morgan, 1992).

In the present study the data set has a high number of variables which means PCA can be a good approach to cluster the methylation profile for the investigated samples. The PCA plot on Fig. 15 uses PC1 and PC2 derived from the methylation percentage to build a scatter to observe the grouping in the samples (Akalin *et al.*, 2012; Joliffe and Morgan, 1992). The plot built from these 2 components considers 63.7% of the variability in the data. The cell lines JEG-3 and BeWo do not seem to cluster with any other form the investigated samples. This is comparable to what was observed above (Tale 12, Fig. 12) where JEG-3 shows lower methylation across all investigated *CGB* genes, and BeWo presenting with lower average *CGB3* methylation but relatively higher *CGB7* methylation. HEY, HT-3, and HCT116 cluster together which reflects their higher average methylation values in comparison to the other samples. Observing the next cluster CRL-1790, breast-DNA, and OVCAR-3, shows a profile of high *CGB*5, lower *CGB*8, and lowest *CGB*7 methylation relative to the average values of the interested genes. The last two clusters of the rest of the samples are in a relatively close proximity on the plot. In the last two clusters the profile of methylation based on the averaged methylation percentage is high *CGB*8, low *CGB*5, and lowest *CGB*7 averaged methylation call. What separates HeLa, MCF-7, and MDA-MB-468 in a defined cluster is the fact that their *CGB*7 methylation is lower to the other samples within the two clusters. It is worth noting that the PCA analysis did not take into account *CGB3* reads due to missing or low values in the investigated samples.

## 4.5 *CGB7* has a CGI within promoter

The next step in investigating the *CGB*3-8 promoter profile was to establish whether the differences in methylation percentages between the four studied *CGB* genes bared any statistical significance. To achieve this aim the reported beta-values were converted to M-values as previous reports suggested (Weinhold *et al.*, 2016; Du *et al.*, 2011). The M-value is logarithmic conversion of the reported percentage methylation. This removes the statistical limitations posed by using percentages in subsequent differential analyses. The beta-value is more intuitive in interpretation but limits the available methods to distinguish statistically differences in samples (Weinhold *et al.*, 2016; Du *et al.*, 2011).

Comparison of the average methylation of the investigated promoter expressed in M-value revealed a significant difference ($p<0.0001$) between the *CGB*3-8 genes. Tukey's test confirmed that the difference comes from the lower methylation of the *CGB*7 compared to the other 3 genes across all samples. This matches the previous results from Fig. 12 and Table 12 where this trend was observed.

Furthermore, the *CGB3*-8 genes' promoter sequences were tested *in silico* for presence of CGIs. CGIs are usually defined as regions of at least 200bp length with at least 50% cytosine and guanine content that has observed over expected ratio of minimum 0.6 CpG dinucleotides. These regions are generally unmethylated and found in about 60% promoter regions of human genes (Jones, 2012; Straussman *et al.*, 2009). The returned results show that *CGB*7 has a CGI of 720bp length associated with the promoter region and no other from the investigated *CGB* genes. Thus, the results from Newcpgreport tool and the differential analysis of promoter methylation between the genes of interest strongly suggest *CGB*7 indeed has a CGI associated with its promoter (Madeira *et al.*, 2019).

As mentioned above CGIs are commonly associated with promoter regions and have relatively uniform low level of methylation. The DNA structure within this site is poor for nucleosome assembly which allows for maintenance of a more relaxed chromatin state to induce transcription (Illingworth *et al.,* 2011). This is further supported by the fact that the CGI are commonly associated with histone 3 lysine 4 trimethylation (H3K4me), an active transcription mark (Thomson *et al.,* 2010). Further to this, CGIs can also be found remotely from promoter regions and termed as "orphan" CGI. They also maintain a mostly unmethylated state with H3K4me. They are proposed to be associated with transcription start site of regulatory molecules like *HOTAIR* and *Xist* – non protein coding RNA transcripts, which take part in regulating chromatin state and X-inactivation respectively (Illingworth *et al.,* 2011).

On occasion CGIs can be methylated in normal tissues which is usually associated with long term silencing (Jones, 2012). This methylation usually occurs in mono-allelic fashion to silence only one allele as seen in genomic imprinting. Another example of CGI methylation in normal tissue is the inactivation of X-chromosome whereby adequate gene dosage is achieved (Portela and Esteller, 2010). However, research of CGI methylation is focused on the aberrant hypermethylation in cancer. This is considered as one of the general changes in

the methylome that occurs during carcinogenesis (Robertson, 2005). CGI hypermethylation in cancer is linked with a wide range of genes associated with DNA repair (*MGMT*), cell cycle (*Rb),* cell adherence (*E-cadherin)*, and apoptosis (*DAPK1)* (Portela and Esteller, 2010; Esteller, 2007). These are all pathways which when dysregulated help the establishment of a neoplastic tumour (Portela and Esteller, 2010; Esteller, 2007).

However, in the current study *CGB*7 demonstrates low methylation state across all samples. This suggests activity of the gene in both normal and cancerous cells. Stenman *et al.*, 2004 reports transcrptionally active *CGB7* at low levels in non-trophoblastic tissues. Zimmermann *et al.*, (2012) also claims *CGB*7 is exressed in breast, lung, bladder, and colon. Their study investigates *CGB* expression in normal endomethrium. The findings in Zimmermann *et al.*, (2012) confirm hCG expression in the normal endomethrium in secretory phase derived from the *CGB7* and its allele form *CGB6.* Another study by Giovangrandi *et al.*, (2001) investigates the transcriptional activity of the *CGB* genes in normal and cancerous breast tissue. They confirm the detection of *CGB7* transcripts in both tissue types; however, there is no change in the transcription level of the *CGB*7 gene between normal and cancerous state (Giovangrandi *et al.*, 2001). Therefore, it could be stated that the detected low level expression of *CGB7* in prevoiusly mentioned studies possibly is due to the CGI within the *CGB7* promoter maintaining the transcriptionally available state of DNA.

## 4.6 Methylation differences within tissue groups

### 4.6.1 Trophoblastic versus non-trophoblastic cell lines

After assessing the differences between *CGB*3-8 genes promoter methylation, the differences between normal samples, trophoblastic and non-trophoblastic cancers were considered. As mentioned above M-value was used in the statistical tests. Two sets of tests for difference were used – one aimed at the *CGB*3, 5, and 8 results, and one aimed at the *CGB*7 results to account for their different methylation. Each set of tests tested 3 different groups of averaged M-values – one for the whole region of interest, one for the TF binding sites at the 5’ part of the region of interest, and one for the 3’ TF binding site of the region of interest. The 5’ part corresponds to CpGs found in the vicinity of 4 TF sites ordered AP2α-SP1-AP2α-OCT3/4, and the 3’ – to the 2 TF binding sites AP2α-SP1 (Fig. 9). This was done to assess if the methylation of the whole promoter or a specific TF

binding region contributes to the possible reactivation of the beta subunit in non-trophoblastic caners.

These tests of differences found weakly significant differences (p=0.404) only in the average methylation of *CGB*3, 5, and 8 in the whole promoter. However, Dunn's test could not point to significant difference between the control, trophoblastic, and non-trophoblastic samples. The limitation in these sets of statistical tests was the small sample size of trophoblastic and normal samples. Even though no differences were established, an interesting trend is seen in the spread of the data in the different groups (Figs. 18 & 19). In the *CGB*3, 5, 8 averaged M-values the trophoblastic group in all 3 regions tends to lower average methylation in comparison to normal and non-trophoblastic groups, and the non-trophoblastic group maintains wider spread in the 3 regions compared to the other 2 groups (Fig. 18). The trophoblastic tendency of lower methylation matches what has been found by Campain *et al.*, 1993 as discussed previously.

In the *CGB7* group, the tendency for lower methylation of trophoblastic samples is not obvious, but the wider spread of the non-trophoblastic group is maintained (Fig. 19). The previously mentioned lower methylation of the *CGB*7 genes and the fact that there is no significant difference is concordant with the findings in Zimmermann *et al.*, (2012) which suggest that *CGB*7 consistent transcript is present in low levels in normal and non-trophoblastic tissue. The *CGB*7 results also further support the presence previously mentioned CGI associated with the gene as islands tend to keep more uniform methylation across tissues (Illingworth *et al.*, 2010).

### 4.6.2 Differences in tissue of origin groups

The PCA analysis demonstrated that samples from same tissue of origin cluster with samples from different origin. Furthermore, in the set of difference tests done just prior it was noticed how widely spread the values of the 5mC content within the non-trophoblastic group are in comparison to the other 2 groups. Therefore, the significance of these differences between samples within tissue of origin groups needs to be considered. Again two sets of difference tests were used – one set aimed at *CGB*3, 5, and 8 averaged values, and one – at *CGB*7 values. In this instance the M-value per site in each sample was used. The selected test of difference was parametric RM-ANOVA which allows comparing the difference in repeated measures of a variable (CpG site) at different conditions (cell line)

(Singh *et al.*, 2013). Normality assumption was violated in some cases of the test; however, previous research has shown ANOVA is robust enough to handle this (Harwell *et al.*, 1992).

In the positive control group (JEG-3 and BeWo) methylation differences were as expected for *CGB*3, 5 and 8. Both choriocarcinoma cell lines were compared to the control-DNA for the *CGB*3, 5 and 8, and CRL-1790 was used as normal reference for *CGB*7 RM-ANOVA analysis. The control DNA was found to be more methylated at *CGB*3, 5 and 8, with a strong significant difference ($p<0.0001$) for both choriocarcinoma cell lines. These findings are concordant with the reports by Campain *et al.* (1993) and Grigoriu *et al.* (2011) that show hypomethylation in the *CGB* genes for choriocarcinoma and trophoblast tissue respectively.

Interestingly, JEG-3 showed also significantly lower methylation ($p<0.0001$) than BeWo. This is also observed in the PCA plot where the two cell lines have been plotted at different locations, not clustering with other samples. This could be due to the difference in the two cell lines' characteristics. Serranoa *et al.* (2007) reports that even though the two cell lines are active producers of intact hCG, they differ in the degree of differentiation and proliferation. JEG-3 is more differentiated than BeWo, but proliferation of the cells is at higher rate in BeWo (Serranoa *et al.*, 2007). This difference in phenotype could be reflected also in the methylation of the cells as this mark is fluctuating (Guo *et al.,* 2014). Furthermore, JEG-3 was found to be less methylated than CRL-1790 and BeWo for *CGB7*. This matches with previous observations - *CGB*7 appears to have higher average methylation in BeWo than JEG-3 which could be contributed to the reported phenotypic differences in the selected choriocarcinoma cell lines (Guo *et al.,* 2014; Serranoa *et al.*, 2007).

In the normal samples or control group, including the cervix tissue, breast tissue, CRL-1790 and Control-DNA, differences were observed as well. The control-DNA used as reference from the sequencing facility was more methylated than all other normal samples. All normal samples send for sequencing – CRL-1790, breast-DNA, and cervix-DNA are from epidermal origin. However, the tissue of origin for the control-DNA was not reported back which would help in understanding the reasons for the observed difference. As mentioned previously DNA methylation fluctuates in its pattern between different tissues (Wen *et al.*, 2017). Therefore, it is likely that the control-DNA is from a different tissue type which would contribute to the observed difference in methylation. The normal samples group analysis of

difference for *CGB7* did not contain control-DNA due to missing reads. There was no significant difference in *CGB7* methylation between the 3 samples in the normal samples group.

Observing the tests of difference in the non-trophoblastic cervical tissue group, a few differences can be observed. HeLa presents with a hypomethylated profile when compared to the other cervix samples in *CGB3*, 5 and 8. This is also observed in the clustering from the PCA where HeLa is separately grouped with MCF-7 and MDA-MB-468. The average methylation values (Table 12) also show that the cell line exhibits lower *CGB*3 methylation, comparable with JEG-3 and BeWo. This is suggesting that HeLa should have active transcription of the *CGB* genes mentioned above. Articles Jankowska *et al.*, (2008); Chen *et al.*, (1996); and Goldstein *et al.*, (1990) report the presence of *CGB* mRNA transcript in HeLa cells. However, these transcripts are relatively low compared to the choriocarcinoma cells (Chen *et al.*, 1996).

Observing the results in the same tissue group for *CGB7* showed that HT-3 is hypermethylated in comparison to the normal cervical tissue and the other two cervical cancer cell lines. A similar trend was observed in the *CGB*3, 5, and 8 genes but was not significant. As mentioned before HT-3 is also part of the cluster with HEY and HCT116 in the PCA, that tend to have higher average methylation values. Thus, it is expected that there should be reduction of the transcript of *CGB* genes and in particular less transcripts from *CGB7*. However, Acevedo *et al.*, (1992) reported that free hCG-beta has been found associated with the membrane of the HT-3 cells in low amounts suggesting active transcription of the genes. The other notable result was that *CGB*7 appears to be significantly more methylated on C-33a in comparison with HeLa but no difference found with the normal cervical tissue. Furthermore, C-33a and the cervical tissue are clustering together on the PCA. This is further confirmation that HeLa is hypomethylated in the cervical group.

In the colon tissue group, a significant difference between the cancerous and normal tissue was only found for the *CGB*3, 5 and 8 averaged M-values. The cancerous cell line HCT116 presented with significantly higher methylation. This matches what was observed in the PCA where HTC116 is grouped with more methylated on average HT-3 and HEY cell lines. This would indicate that the investigated *CGB3, 5* and *8* genes are likely silenced in this colorectal cancer cell line. However, HCT116 has been reported to produce low levels of CGB mRNA

transcript when compared to the choriocarcinoma JAR (Li *et al.*, 2018; Sohr and Engeland, 2011). No significant differences were found in the colon group for *CGB*7.

Interestingly, no significant differences were observed in the ovarian group for the methylation status of *CGB*3, 5, and 8 methylation averages. This contradicts the PCA plot which groups the 4 ovarian cancer cell lines in 3 different clusters. One possible reason for that is that the PCA does not include data from *CGB*3 methylation. The only significant difference in the ovarian group comes in the *CGB*7 gene methylation, where the HEY cell line shows hypermethylation in comparison to the other cancer cell lines. This is concordant with the PCA grouping where HEY clusters together with HCT116 and HT-3 that show higher average methylation which was confirmed statistically for the *CGB3*,5 and 8 (HCT116) and *CGB7* (HT-3). The higher methylation of HEY in *CGB*7 suggests that the gene expression would be silenced. Sinnappan (2015) reports that HEY is a potent producer of hCGβ mRNA. However, the study does not discern from which gene the transcripts come from so the effect of methylation on *CGB*7 transcript is not confirmed.

One of the few studies done on *CGB* methylation in cancer investigates the methylation changes in ovarian tissue (Śliwa *et al.*, 2019). The study is suggesting that DNA methylation plays a role in the *CGB* expression in ovaries (Śliwa *et al.*, 2019). Śliwa *et al.*, (2019) analysed methylation within *CGB* promoter via MSP. Their results find significant difference between the unmethylated product of cancer and normal ovary, whereby the cancer tissue is significantly demethylated. These findings were not confirmed by the current study. This is possibly due to the different approach in analysis – here the all CpGs of the sequenced promoter are considered, and the MSP in Śliwa *et al.*, (2019) focusses on a few CpG sites found only in their designed primers (Fraga and Esteller, 2011; Ammerpohl *et al.,* 2009). Furthermore, ovarian tissue from both healthy and cancer samples expresses *CGB*3-9 transcripts but the level of these transcripts is significantly increased in cancer tissues (Śliwa *et al.*, 2019). Therefore, it could be speculated that possibly other CpG sites in the *CGB* promoter in normal tissue have lower methylation enabling the *CGB* activation (Grigouriu *et al.*, 2011). Thus, the observed difference in methylation by Śliwa *et al.* (2019) cannot be accounted for when considering the whole promoter as is done in this study (Śliwa *et al.*,

2019).The applied approach here should provide a more accurate picture due to the number of methylation sites analysed.

The last tissue group to discuss for the methylation sequencing differences is the breast tissue. No significant differences in methylation were found in the samples for *CGB*3, 5, and 8 genes. Interestingly, the PCA plot groups the cancer cell lines in different clusters from the normal breast-DNA. However, looking into the average methylation values the most difference is observed in the *CGB*7 gene. There was only significant difference in the pattern of MCF-7 which has presented with lower methylation than the control breast-DNA for *CGB*7. As mentioned above Giovangrandi *et al.*, (2001) has already reported that *CGB*7 is active in both normal and cancerous tissues. Therefore, the lower methylation in MCF-7 seems to have no impact on *CGB7* expression.

## 4.7 Transcription and translation level of *CGB*3-8 genes

In order to understand whether the DNA methylation has direct effect on *CGB*3-8 genes, the transcription and translation of the hCG-beta subunit was assessed via qRT-PCR and ELISA. mRNA transcript for the *CGB3*-9 gene was detected in all samples with the exception of the negative control, 3T3. The positive control samples from choriocarcinoma had the highest relative transcription level (BeWo- 1767 and JEG-3- 588 fold difference). These findings are concordant with the observed methylation pattern found in this study and previous reports (Grigoriu *et al.*, 2011; Campain *et al.*, 1993). The transcription pattern is as expected since the cell lines are from trophoblast origin and behave in a similar fashion to what is observed in pregnancy, i.e. active *CGB* transcription (Serranoa *et al.*, 2007; Acevedo *et al.*, 1995; Whitfield and Kourides, 1985).

The observed transcription level for the non-trophoblastic cancer was quite diverse. Previous studies agree that all non-trophoblastic cancers show *CGB*3-9 gene activity either by the detection of the mRNA transcript or the beta subunit itself (Śliwa *et al.*, 2019; Li *et al.*, 2018; Sinnappan, 2015; Sohr and Engeland, 2011; Jankowska *et al.*, 2008; Chen *et al.*, 1996; Acevedo *et al.*, 1992; Goldstein *et al.*, 1990). One study in particular uses a fairly different approach to assess the level of hCGβ in cancer cell lines – flow cytometry (Acevedo *et al.*, 1992). In this particular study polyclonal and monoclonal antibodies target different fragments of the intact hCG molecule or its subunits that are associated with the cellular membrane. Due to the differences in applied methods and the employment of

relative quantification of mRNA transcripts, accurate comparisons with this project cannot be achieved.

Correlation using Spearman's ranked test was used to asses any association between the observed methylation and level of *CGB* expression. The test did not present any significant correlation between the transcription level and average methylation of the cell lines in the promoter region, 3' TF binding and 5' TF binding region. Studies by Whitfield and Kourides (1985) and Śliwa *et al.*, (2019) which explored DNA methylation in *CGB3-9* genes also report no association between the methylation and expression level of the hCGβ genes. This agrees with the correlation results in this project. A possible reason for the lack of association in this study is the contradictory results of averaged methylation and transcription level data. For instance, HCT116 cell line shows high average methylation of ~90% (hypermethylation for *CGB3, 5, 8*) suggesting no transcription. However, in the transcription study HCT116 has the highest fold difference in non-trophoblastic tissues (Fig. 22)

Comparing data between methylation and transcription at tissue of origin level may show some associations which may not be detected by the statistical tests. Unexpectedly, HeLa cells show transcription level of *CGB*3-9 genes lower than the normal CRL-1790, even though methylation analyses reveal HeLa hypomethylation. As previously mentioned, HeLa is transcriptionally active (Jankowska *et al.*, 2008; Chen *et al.*, 1996; Goldstein *et al.*, 1990). Acevedo *et al.* (1995) further reports that hCGβ is produced in HeLa cells and the main contributor to the product was identified as *CGB*3. This finding supports the methylation data obtained from the sequencing in this study. The unexpected transcriptional result may be due to issues in contamination of HeLa cells in our lab, as recently colleagues have reported changed phenotype of the cell line. Proposed reasons for this change is cross-contamination with another cell line.

The other cell lines in the cervical group C-33a and HT-3 do not show significant methylation changes from the normal cervix DNA. However, Acevedo *et al.*, (1992) reports free beta subunit association with the membrane of C-33a and HT-3 corresponding to 47.5% and 15.4% of the studied populations confirming the activity of the genes. This correlates with the transcription data in this study where HT-3 and C-33a have moderately high level of transcription compared to the normal CRL-1790 with the difference that HT-3 has higher transcription than C-33a. HT-3 did present hypermethylation in the *CGB*7 sites; however, the

transcriptional assay does not discern between different gene transcripts and its role cannot be assessed.

Further to that, Jankowska *et al.* (2008) presents a study on gynaecological cancers where hCGβ and LH/hCG-R transcription level are evaluated in ovarian, endometrial, and cervical cancer. Their study found that all cancer samples were positive for the *CGB3*-9 transcript. The normal tissue used in the study did not detect any activity of the *CGB* genes suggesting that expression of the hormone subunit is typical for cancer tissue. In the same study the LH/hCG-R is co-expressed in cancers suggesting possible autocrine/paracrine role of the free β hormone subunit. Furthermore a consecutive study by the same group report that in cervical carcinoma U1 snRNA blocking of hCGβ expression results in increased apoptosis (Jankowska *et al.*, 2008).

In the other investigated gynaecological cancer, ovarian cell lines SKOV-3 is showing highest transcriptional activity followed by OAW42, OVCAR-3 and HEY has lowest relative expression in the group. Śliwa *et al.* (2019) and Sinnappan (2015) confirm expression of the *CGB*3-9 genes at a low level in normal and higher level in cancerous ovarian cells. The fact that *CGB* genes are expressed in both normal and cancerous could be a reason why methylation data that shows no significant difference in ovarian cells. Sinnappan (2015) further shows that HEY and SKOV-3 have the most active *CGB*3-9 genes in their sample, and OVCAR-3 has low level of expression. This contradicts the results in the present study where HEY has lowest transcription level. This could be due to differences in applied qRT-PCR assays.

Tissues from ovarian cancer patients have revealed increased transcript levels of hCGβ mRNA in comparison to healthy tissue (Zhong *et al.*, 2019). In the same study it was also established that this increase matched with the protein level increase in the cancer tissue samples. The increase of hCGβ level in these samples was found to be associated with worse patient prognosis denoted by advanced tumour stages and increased metastasis. However, exact mechanism of how hCG acts on ovarian cancer progression needs to be elucidated (Zhong *et al.*, 2019).

The studies Szczerba *et al. (*2016) and Kubiczak *et al.* (2013) have also confirmed samples of ovarian cancer tissues positive for hCGβ expression. In the studies transcript of the *CGB3*-9 genes were found both in normal and cancerous tissue.

The cancerous tissue presented with higher level of *CGB* gene activity. However, they also report that the level of activity in the analysed samples was highly varied due to cellular heterogeneity and genetic instability typical for cancerous cells. Szczerba *et al.* (2016) and Junker and Oudenaarden (2014) further suggest that induced overexpression of *CGB5* modulates apoptosis regulated genes *BCL2, BAX* and *BRIC5* so that apoptosis in ovarian cancer is suppressed.

In the breast cancer group the 2 cell lines MCF-7 and MDA-MB-468 are transcribed similarly at low levels compared to most other investigated cell lines. Acevedo *et al.* (1992) reports presence of membrane associated free hCGβ confirming that the genes are transcriptionally active in the cell line. Giovangrandi *et al.* (2001) investigates the expression of the genes in breast cancers and confirms that elevated *CGB3*, 5 and 8 transcription is associated only with malignant tumours. Methylation data did not confirm hypomethylation in the *CGB3*, 5 and 8 genes to suggest the activation of these genes.

Furthermore, Giovangrandi *et al.* (2001) proposes that the beta subunit of hCG is acting as a tumour growth factor independently from the classic receptor in breast cancer. A different report states that beta-hCG presence in breast cancer is associated with apoptosis inhibition and down regulation of epithelial cell adhesion to allow tumour migration and metastasis. However, intact hCG has a protective role on breast tissue against malignancy. Early completion of full term pregnancy changes the breast epithelium protecting it from malignant changes. The paradoxical nature of hCG and its beta subunit on breast cancer remains controversial (Schüler-Toprak *et al.*, 2017).

In the colon group the transcription level of HCT116 has the highest relative transcription level of all non-trophoblastic cell lines used in this study. Data from previous studies confirm the presence of mRNA transcripts in the cell line but due to the relative approaches applied no comparison of the values can be done (Li *et al.*, 2018; Sohr and Engeland, 2011). This is in contrast with the methylation data which show HCT116 is hypermethylated in the *CGB*3, 5 and 8 genes.

Li *et al.*, (2018) further comments on the role of hCGβ in colorectal cancer which is mainly in tumour invasion and migration but no effect on tumour proliferation. This is thought to be due to triggering of epithelial-to-mesenchymal transition (EMT), a process in which the cell loses its epithelial characteristics such as the cuboidal cell shape and adhesiveness to become more spindle shaped and able

to migrate. Kawamata *et al.* (2018) is another report confirming findings that hCGβ activates EMT via TGFβ receptor in colorectal carcinoma. The study tests this *in vitro* using western blot and qPCR to show that overexpressing cells display EMT associated changes. The EMT associated changes were reversed in the studied cell line by addition of TGFβ receptor inhibitor (Kawamata *et al.*, 2018).

The last piece of information collected for this study was the protein level of hCGβ in the cell line conditioned media. The positive control trophoblastic cell lines behaved as expected and showed positive results for intact and free beta hCG. This matches previous reports of the cell lines as potent hCG producers (Serranoa *et al.*, 2007). The negative control 3T3 had no protein detected. ELISA data showed positive results for free hCGβ in only two of the non-trophoblastic samples – SKOV-3 and HEY. All other cell lines were negative and below detection limit for hCGβ of 0.2 ng/ml. Transcription data for SKOV-3 from this study is relatively high in the tested samples and matches the detection of the beta subunit in the media. However, HEY showed little transcriptional activity in this study but displayed the highest level of free hCGβ per million cells over 24h. Wu *et al.* (2019) and Sinnappan (2015) show that hCGβ protein is produced by both of the cell lines. Findings in (Sinnappan, 2015) report that SKOV-3 produces more of the hCGβ protein as detected in culture media.

As the other samples did not show positivity for the protein ELISA data could not be used for statistical analysis to evaluate translation associations with transcription and methylation. The transcription data suggests that at least HT-3, HCT116, and possibly C-33a should express some level of the protein. Possible reasons for this discrepancy could be an artefact from repeated freezing-thawing. As previously mentioned tumours tend to be heterogeneous and only select few cells secrete the hCGβ leading to too low of a concentration to be detected in the media (Szczerba *et al.*, 2016; Junker and Oudenaarden, 2014). Rao (2016) further reports that small amounts of secreted hCG can be quickly eliminated/absorbed from circulation leading to no detection. It is possible that this could have happened in the cultured cell lines.

## 4.8 Limitations of study

One limitation of this study is the lack of replicates of the sequenced samples. The study used a single set of cell line and tissue samples for the sequencing. Yet, tumour cells are heterogeneous and unstable in nature (Szczerba *et al.*, 2016;

Junker and Oudenaarden, 2014). DNA methylation fluctuates and the heterogeneity of tumour cells may be reflected in the DNA methylation (Jones, 2012). Replicate samples of each cell line could address the issue of heterogeneous methylation by increasing the chance to capture the variety in the profiles and establish significant methylation motifs that may have a biological role. Furthermore, no validation study has been coupled with the sequencing performed here. Combining the sequencing with a second approach can show whether the data is reproducible and sensitive (Roeh *et al.*, 2016; Luthra *et al.*, 2013).

Further to this, each possible methylation site at a given cell can be either methylated or not. However, there are multiple reads per site which are reported as beta-value. A beta value of 60% indicates that 60% of the total reads showed methylation but the other 40% are not methylated (Du *et al.*, 2011). That could be a possible reason why some samples with higher average methylation percentage or beta value still show transcription. To address that some sequencing approaches apply cut-off ranges based on beta value to define methylated, unmethylated and heterogeneous samples (Warden *et al.*, 2013).

Additionally, DNA methylation is not well studied in regards to non-CGI associated promoters such as *CGB3*, 5 and 8. CGIs show clear inverse correlation between transcription initiation and DNA methylation (Jones, 2012). However, Weber *et al.* (2007) stated that methylated non-CGI sites still possess transcriptional activity. This statement is not in agreement with other research where it is presented that the methylation of DNA behaves similarly between CGI and non-CGI associated TSS (Han *et al.*, 2011).

The recommended approach for DNA sequencing analysis is using R programming based packages (Wreczycka *et al.*, 2017; Weinhold *et al.*, 2016; Akalin *et al.*, 2012). However, limited expertise in R programming prevented the full utilisation of the available functions of methylKit or other similar tools. As reported these can provide comprehensive analysis that identifies differentially methylated regions and differentially methylated cytosines. This may recognise other motifs which may have been omitted by conventional statistics (Wreczycka *et al.*, 2017; Akalin *et al.*, 2012).

In the transcription study the major limitation is the use of a universal primer pair to amplify multiple genes. This does not allow checking for the transcript levels of

specific CGB genes. Therefore, the methylation data obtained per gene needed to be averaged and possibly masking some associations between the specific gene methylation and transcription. This is especially important with the *CGB*7 gene which as observed and reported in Giovangrandi *et al.*, (2001) behaves differently than the rest of the *CGB* genes.

## 4.9 Further research

This is the first study of methylation profile in the promoter of the *CGB*3-9 genes in non-trophoblastic cancer cell lines employing next generation sequencing method. The study was coupled with qRT-PCR analysis and ELISA to establish how methylation of the promoter affects downstream gene expression. As such the data from the methylation study could only be inferred from the obtained transcription levels and the ones reported in literature. Śliwa *et al.* (2019); Uuskula *et al. (*2010); Glodek *et al.* (2014); Campain *et al.* (1993); and Whitfield and Kourides (1985) are the only found previous reports of the *CGB*3-8 genes methylation specifically. However, the studies apply varied approaches in analysis of methylation status and utilise different cancer and non-cancer samples making comparison between the methylation data difficult.

Free hCG beta is a tumour marker which is associated with poor prognosis and overall lower survival time (Rull *et al.,* 2008, Iles *et al.,* 2010; Stenman *et al.,* 2004). hCGβ can be produced by numerous common non-trophoblastic cancers as already established (Schüler-Toprak *et al.*, 2017; Sinnappan, 2015; Kubiczak *et al.*, 2013; Jankowska *et al.*, 2008; Iles *et al.,* 1996). Understanding the mechanisms which are involved in the transcriptional regulation can shed a light on the development on both diagnostic tools and possible therapeutic targets to combat aggressive cancers (Śliwa *et al.*, 2019). As previously reported methylation pattern of the *CGB* genes changes between the cancerous and normal tissues (Śliwa *et al.*, 2019; Campain *et al.* 1993); however, no association is found between the transcription and methylation data (Śliwa *et al.*, 2019; Whitfield and Kourides, 1985). The obtained data in this project showed some discrepancies where hypermethylated regions have active transcription. Other methylation results could not be compared to see if there is significance due to the approach of the transcription study where a universal primer set was used for all *CGB3-8* genes and thus, no discrepancy between the individual gene transcripts. Therefore, future research needs to be carried out to validate the information

obtained here and elucidate the role of DNA methylation in non-trophoblastic cancers.

DNA methylation does not exist as an isolated mark – it is involved with the other epigenetic mechanisms such as histone modifications and non-coding RNAs (Grigoriu, *et al.*, 2011; Thomsonal, 2010; Wojdacz & Dobrovic, 2007). An example is the lysine 4 methylation at histone 3 which is the active mark associated with CGI overlaying the already hypomethylated (active) CGI (Thomsonal, 2010). Therefore, it could be speculated that the methylation observed in *CGB3*-8 genes may be a consequence of other epigenetic marks that dictate the methylation pattern. For instance in cancers, histones, as with DNA methylation, have global changes – these are less defined but typically loss of the active acetylation marks is observed. The methyltransferase EZH2 is overexpressed in several cancers which alters the H3K27me profile in the genome. In turn, this histone methyltransferase interacts with DNMTs and by extension controls DNA methylation (Portela and Esteller, 2010).

Another aspect of *CGB* transcriptional control is the TFs involved in its expression. The studies by Glodek *et al.* (2014) and Śliwa *et al.* (2019) use combined approach to assess the methylation level and compare it with TFs levels in the samples. AP2α levels have been correlated positively with transcriptional activation of the *CGB* genes. Pairing methylation studies with transcriptional factor level may provide insights into the relationship of the two. Furthermore, not all TFs are directly influenced by DNA methylation directly – establishing how DNA methylation and TFs influence each other could further elucidate the transcription mechanism of hCGβ genes (Portela and Esteller, 2010).

Other TFs that have been found to interact with *CGB* genes are PPARγ and MTA-3. PPARγ is a nuclear receptor involved in trophoblast differentiation and invasion (Handschuh *et al.*, 2009). Handschuh *et al.* (2009) showed different behaviour depending on the site of the trophoblast: in VCT activation of PPARγ leads to higher amount of free hCGβ and secretion of hCG. In iEVT PPARγ activation decreased the transcript and hCG secretion. Fournier *et al.*, 2011 showed similar findings to Handschuh *et al.*, (2009) with regards to PPARγ differential regulation of trophoblast cell subtypes. Further to the nuclear receptor association with increased hCG expression and secretion in VCT, it was reported PPARγ plays a role in villous trophoblast differentiation to ST (Fournier *et al.*, 2011).

MTA3 has been shown to repress hCG gene expression mediated via HDAC1/2 component of NuRD in BeWo cell line. Forskolin treatment in BeWo decreased MTA3 and increased hCG expression (*CGB5*) suggesting inverse relationship between MTA3 and hCG expression in trophoblast (Chen *et al.*, 2013). Cytotrophoblasts show higher staining for MTA3 than syncytiotrophoblasts, which are usually associated with higher amount of hCG secretion between the 2 sites, further suggesting inverse correlation between MTA3 and hCG expression in placenta (Chen *et al.*, 2013). MTA3 may have a role in proliferation and differentiation in cytotrophoblast as it shows stable concentration until CTs are fully differentiated to EVT or ST (Horii *et al.*, 2015). MTA3 is required for terminal differentiation as its knockdown leads to decrease in hCG secretion and reduction of mRNA transcripts (Horii *et al.*, 2015). Understanding where these 2 factors (MTA3 and PPARγ) bind in the promoter and their interaction with DNA methylation can present novel insights in the transcriptional regulation of the *CGB* genes.

# 5. Conclusions

This study focused on the role of DNA methylation in the activation of the *CGB3*-8 genes whose protein products have functions as anti-apoptotic, angiogenic, growth and invasion stimulating factors in cancers. The results from this study showed that the *CGB7* gene is hypomethylated compared to *CGB3*-9. This, coupled with *in silico* predictions strongly suggests *CGB7* has a CGI associated with its promoter which is further confirmed in the literature where it was reported low levels of *CGB7* mRNA were present in normal tissues and in breast cancer.

Comparing the methylation profiles of the studied non-trophoblastic cell lines and tissues revealed 4 distinct groups, separate from the trophoblastic cancer cell lines, which did not assemble with each other. Other notable finds show hypomethylation of JEG-3 and BeWo cell lines which matches the expression and secretion data obtained by this study and is further supported by literature. Hypomethylation of HeLa was also detected which has confirmed previously reported transcription studies but did not match the expression data in this study. The results from the methylation sequencing showed no significant correlation with the data from expression studies.

This is the first study investigating the *CGB3*-8 promoter methylation in non-trophoblastic cancer cell lines. The results did not show conclusive methylation changes associated with non-trophoblastic cancer. Further studies should be completed to fully understand the role of *CGB* gene family DNA methylation in non-trophoblastic cancer.

# 6. References


Acevedo, H. F., Krichevsky, A., Campbell-Acevedo, E. A., Galyon, J. C., Buffo, M. J., and Hartsock, R. J. (1992) 'Expression of membrane-associated human chorionic gonadotropin, its subunits, and fragments by cultured human cancer cells', 69(7), pp. 1829–1842, doi: 10.1002/1097-0142(19920401)69:7<1829::aid-cncr2820690727>3.0.co;2-0

Acevedo, H.F., Tong, J.Y. and Hartsock, R.J. (1995) 'Human chorionic gonadotropin-beta subunit gene expression in cultured human fetal and cancer cells of different types and origins' *Cancer*, 76, pp.1467-1475. doi:10.1002/1097-0142(19951015)76:8<1467::aid-cncr2820760826>3.0.co;2-a.

Adams, C., Henke, A., and Gromoll, J. (2011) 'A novel two-promoter-one-gene system of the chorionic gonadotropin beta gene enables tissue-specific expression', *Journal of Molecular Endocrinology*, 47(3), pp. 285-298. doi:10.1530/JME-11-0026.

Akalin, A., Kormaksson, M., Li, S. *,* Garrett-Bakelman, F.E., Figueroa, M. E., Melnick, A. and Mason, C.E.(2012) 'methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome Biology,* 13, doi:10.1186/gb-2012-13-10-r87.

Ambrus, G. and Rao, C. V.,(1994) 'Novel regulation of pregnant human myometrial smooth muscle cell gap junctions by human chorionic gonadotropin', *Endocrinology*, 135(6), pp. 2772-2779, doi:10.1210/endo.135.6.7988470

Ammerpohl, O., Martin-Subero, J. I., Ritcher, J., Vater, I., and Siebert, R. (2009) 'Hunting for the 5th base: Techniques for analyzing DNA methylation', *Biochimica et Biophysica Acta- General Subjects*, 1790(9), pp. 847-862, doi:10.1016/j.bbagen.2009.02.001.

Arányi, T., Váradi, A., Simon, I. and Tusnády, G.E. (2006) 'The BiSearch web server', *BMC Bioinformatics,* 7, doi:10.1186/1471-2105-7-431

Arrieta, O., Michel Ortega, R.M., Ángeles-Sánchez, J., Villarreal-Garza, C., Avilés-Salas, A., Chanona-Vilchis, J.G., Aréchaga-Ocampo, E., Luévano-González, A., Jiménez, M.A. and Aguilar, J.L.(2009) 'Serum human chorionic gonadotropin is associated with angiogenesis in germ cell testicular tumors', *Journal of Experimental & Clinical Cancer Research,* 28(1), doi:10.1186/1756-9966-28-120.

Berger, P., Paus, E., Hemken, P.M., Sturgeon, C., Stewart, W.W., Skinner, J.P., Harwick, L.C., Saldana, S.C., Ramsay, C.S., Rupprecht, K.R., Olsen, K.H., Bidart, J.M. and Stenman, U.H.(2013) 'Candidate epitopes for measurement of hCG and related molecules: the second ISOBM TD-7 workshop', *Tumor Biology,* 34(6), pp. 4033-57, doi:10.1007/s13277-013-0994-6

Borisova, M., Moiseenko, D., and Smirnova, O. (2017) 'Human Chorionic Gonadotropin: Unknown about Known', *Human Physiology,* 43, pp. 93-104, doi:10.1134/S0362119716060050.

Burczynska, B., Kobrouly, L., Butler, S. A., Naase, M., and Iles, R. K. (2014) 'Novel insights into the expression of *CGB*1 & 2 genes by epithelial cancer

cell lines secreting ectopic free hCGβ', *Anticancer Res*earch, 34(5), pp.2239-2248, Available at: http://ar.iiarjournals.org/content/34/5/2239.long (Accessed: 16 December 2019)

Campain, J. A., Gutkin, D. W., and Cox, G. S. (1993) 'Differential DNA Methylation of the Chorionic Gonadotropin beta-subunit Multigene Family', *Molecular Endocrinology*, 7(10), pp.1331-1346, doi:10.1210/mend.7.10.7505395

Casarini, L., Lispi, M., Longobardi, S., Milosa, F., La Marca, A., Tagliasacchi, D., Pignatti, E. and Simoni, M.(2012) 'LH and hCG action on the same receptor results in quantitatively and qualitatively different intracellular signalling', *PLoS One,* 7(10), pp. e46682, doi:10.1371/journal.pone.0046682.

Chen, Y. H., Chen, T.M. and Hsieh, C.Y. (1996) 'The expression of the human chorionic gonadotropin beta subunit gene depends on negative control'. *Cell Biochem. Funct.*, 14, pp. 297-301, doi:10.1002/cbf.686.

Chen, Y., Miyazaki, J., Nishizawa, H., Kurahashi, H., Leach, R., and Wang, K. (2013) 'MTA3 regulates *CGB*5 and Snail genes in trophoblast', *Biochemical and biophysical research communications*, 433(4), pp.379-384, doi:10.1016/j.bbrc.2013.02.102.

Cole, L. (2015). *Human Chorionic Gonadotropin (hCG).* 2nd edn. Edited by L. Cole and S. Butler. San Diego: Elsevier.

Cole, L.A. (2007) 'Hyperglycosylated hCG', *Placenta*, 28(10), pp. 977-986 doi:10.1016/j.placenta.2007.01.011.

Colyer, H., Armstrong, R., Sharpe, D., and Mils, K. (2012*) Detection and Analysis of DNA Methylation by Pyrosequencing.* In R. Dumitrescu, and M. Verma (Eds.), Cancer Epigenetics (pp. 281-292). Humana Press. doi:10.1007/978-1-61779-612-8_17.

Correa, P. A., Pujol-Borrell, R. and Colobrana, R.(2012) 'Bisulfite genomic sequencing to uncover variability in DNA methylation: Optimized protocol applied to human T cell differentiation genes', *Inmunología,* 31(4), pp. 97-105 doi:10.1016/j.inmuno.2012.09.002.

Cui, D., and Xu, X. (2018) 'DNA Methyltransferases, DNA Methylation, and Age-Associated Cognitive Function', *International journal of molecular sciences*, *19*(5), doi:10.3390/ijms19051315.

d'Hauterive, S.P., Huge, F., Polese, B., Berndt, S., Gaspard, O., Munaut, C., Foidart, J.M., Geenen, V. and Gridelet V.(2011) 'Immune tolerance and angiogenesis during embryo implantation: the role of hCG', *Journal of Reproductive Immunology*, 90(2), pp. 161-162, doi:10.1016/j.jri.2011.06.057.

Du, P., Zhang, X., Huang, C. , Jafari, N., Kibbe, W.A., Hou L. and Lin, S.M. (2010) 'Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis', *BMC Bioinformatics,* 11**,** doi:10.1186/1471-2105-11-587.

Dukal, H., Schendel, D., Lang, M., Mossner, M., Jann, J.C., Frank, J., Treutlein, J., Witt, S. and la Rietschel, M. (2017) 'T22 - Next-Generation Bisulfite Sequencing Of Cacna1c With Illumina Miseq', *European*

*Neuropsychopharmacology*, 27, pp. S445, doi:10.1016/j.euroneuro.2016.09.510.

Esteller, M.(2007) 'Epigenetic gene silencing in cancer: the DNA hypermethylome', *Human Molecular Genetics*, 16(R1) pp. R50–R59, doi:10.1093/hmg/ddm018

Fakruddin, Mazumdar, R., Chowdhury, A., Hossain, N., Mohajan, S., and Islam, S. (2013) 'Pyrosequencing-A Next Generation Sequencing Technology', *World Applied Sciences Journal*, 24(12), pp. 1558-1571, doi:10.5829/idosi.wasj.2013.24.12.2972.

Fournier, T. (2016) 'Human chorionic gonadotropin: different glycoforms and biological activity depending on its source of production', *Annales d'Endocrinologie*, 77(2), pp. 75-81 doi: 10.1016/j.ando.2016.04.012.

Fournier, T., Guibourdenche, J., Handschuh K., Tsatsaris V., Rauwel B., Davrinche C. and Evain-Brion D. (2011) 'PPARγ and human trophoblast differentiation', *Journal of Reproductive Immunology*, 90(1), pp. 41-49 doi: 10.1016/j.jri.2011.05.003.

Fraga, M. F., and Esteller, M. (2002) 'DNA methylation: a profile of methods and application', *BioTechniques*, 33(3), pp. 632-649 doi:10.2144/02333rv01.

Gibney, E. and Nolan, C.M. (2010) 'Epigenetics and gene expression', *Heredity*, 105, pp. 4-13, doi:10.1038/hdy.2010.54.

Giovangrandi, Y., Parfait, B., Asheuer, M., Olivi, M., Lidereau, R., Vidaud, M. and Bièche, I (2001) 'Analysis of the human *CGB*/LHB gene cluster in breast tumors by real-time quantitative RT-PCR assays' , *Cancer Letters*, 168 (1), pp. 93-100, doi:10.1016/S0304-3835(01)00496-7.

Glodek, A., Kubiczak, M. J., Walkowiak, G. P., Nowak-Markwitz, E., and Jankowska, A. (2014) 'Methylation status of human chorionic gonadotropin beta subunit promoter and TFAP2A expression as factors regulating *CGB* expression in placenta', *Fertility and Sterility*, 102, pp. 1175-1182 doi:10.1016/j.fertnstert.2014.06.016.

Goldstein, S., Jones, R., Hardin, J., Glenn D. Braunstein, and Reis, R.J.S. (1990) 'Expression of α- and β-Human Chorionic Gonadotropin Subunits in Cultured Human Cells'. *In Vitro Cellular & Developmental Biology, 26*(9), pp. 857-864. Available at: www.jstor.org/stable/4296528 (Accessed date 10.01.2020)

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.H., Aravind, L., He, C. and Shi, Y.(2015) 'DNA methylation on N6-adenine in *C. elegans*', *Cell*, 161(4), pp 868–878, doi:10.1016/j.cell.2015.04.005.

Grigoriu, A., Ferreira, J. C., Choufani, S., Baczyk, D., Kingdom, J., and Weksberg, R. (2011) 'Cell specific patterns of methylation in the human placenta', *Epigenetics*, 6(3), pp. 368-379. doi:10.4161/epi.6.3.14196.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B., Walsh, C.P., Li, J., Tang, F. and Xu G. L.(2014) **'**Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote', *Cell Stem Cell*, 15(4), pp. 447-459, doi:10.1016/j.stem.2014.08.003.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., Wang, W., Li, R., Yan, J., Zhi, X., Zhang, Y., Jin, H., Zhang, W., Hou, Y., Zhu P, Li, J., Zhang, L., Liu, S., Ren, Y., Zhu, X., Wen, L., Gao, Y. Q., Tang, F. and Qiao, J. (2015) 'The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells', *Cell*, 161(6), pp. 1437-1452, doi:10.1016/j.cell.2015.05.015.

Guo, X., Liu, G., Schauer, I.G., Yang, G., Mercado-Uribe, I., Yang, F., Zhang, S., He, Y. and Liu, J. (2011) 'Overexpression of the beta subunit of hCG promotes the transformation of human ovarian epithelial cells and ovarian tumorigenesis', *The American journal of pathobiology,* 179(3) , pp. 1385-1393 doi:10.1016/j.ajpath.2011.05.018.

Hallast, P., Rull, K. and Laan, M. (2007) 'The evolution and genomic landscape of CGB1 and CGB2 genes', *Molecular and cellular endocrinology*, *260-262*(9), pp. 2–11, doi:10.1016/j.mce.2005.11.049.

Han, H., Cortez, C. C., Yang, X., Nichols, P. W., Jones, P. A., and Liang, G. (2011) 'DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter', *Human molecular genetics*, 20(22), pp. 4299–4310, doi:10.1093/hmg/ddr356.

Handschuh, K., Guibourdenche, J., Cocquebert, M., Tsatsaris, V., Vidaud, M., Evain-Brion, D. and Fournier, T. (2009) 'Expression and Regulation by PPARγ of hCG alpha and beta subunits: Comparison between villous and invasive extravillous trophoblastic cells', *Placenta*, 30(12), pp. 1016-1022, doi:10.1016/j.placenta.2009.09.006.

Hanna, C.W., McFadden, D.E. and Robinson, W.P. (2013) 'DNA Methylation Profiling of Placental Villi from Karyotypically Normal Miscarriage and Recurrent Miscarriage', *The American journal of pathology*,182(6), pp. 2276-2284 doi: 10.1016/j.ajpath.2013.02.021.

Harwell, M.R., Rubinstein, E.N., Hayes, W.S. and Olds, C.C.(1992) 'Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases' *J. Educ. Stat*. 17, pp. 315-339, doi:10.2307/1165127.

Horii, M., Moretto-Zita, M., Nelson, K. K., Li, Y., and Parast, M. M. (2015) 'MTA3 regulates differentiation of human Cytotrophoblast stem cells', *Placenta*, 36(9), pp. 974-980, doi: 10.1016/j.placenta.2015.07.122.

Iles, R., Delves, P., and Butler, S. A. (2010) 'Does hCG or hCG-beta play a role in cancer cell biology?', *Molecular and Cellular Endocrinology*, 329(1-2), pp. 67-70. doi:10.1016/j.mce.2010.07.014.

Iles, R., Persad, R., Trivedi, M., Sharma, K., Dickinson, A., Smith, P. and Chard, T. (1996) 'Urinary concentration of human chorionic gonadotrophin and its fragments as a prognostic marker in bladder cancer', *British Journal of Urology*, 77(1), pp. 61-69, doi:10.1046/j.1464-410X.1996.82910.x.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D.J., Andrews, R. and Bird, A. P. (2010) 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS genetics*, *6*(9), e1001134, doi:10.1371/journal.pgen.1001134.

Iyer, S. and Acharya, K.R. (2011) 'Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines', *The FEBS Journal,* 278(22), pp. 4304-4322. doi:10.1111/j.1742-4658.2011.08350.x.

Jankowska, A., Gunderson, S. I., Andrusiewicz, M., Burczynska, B., Szczerba, A., Jarmolowski, A., Nowak-Markwitz, E. and Warchol, J. B. (2008) 'Reduction of human chorionic gonadotropin beta subunit expression by modified U1 snRNA caused apoptosis in cervical cancer cells', *Molecular cancer*, 7, doi:10.1186/1476-4598-7-26.

Jankowska, A., Andrusiewicz, M., Grabowski, J., Nowak-Markwitz, E., and Warchol, J.b. (2008) 'Coexpression of human chorionic gonadotropin beta subunit and its receptor in nontrophoblastic gynecological cancer', *International Journal of Gynecologic Cancer,* 18(5), pp. 1102-1107 doi:10.1111/j.1525-1438.2007.01151.x.

Johnson, W. and Jameson, J. L.(1999) 'AP-2 (Activating Protein 2) and Sp1 (Selective Promoter Factor 1) Regulatory Elements Play Distinct Roles in the Control of Basal Activity and Cyclic Adenosine 3′,5′-Monophosphate Responsiveness of the Human Chorionic Gonadotropin-β Promoter', *Molecular Endocrinology*, 13(11), pp. 1963–1975, doi:10.1210/mend.13.11.0386.

Joliffe, I., and Morgan, B. (1992). Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research*, *1*(1), pp. 69-95, doi:10.1177/096228029200100105.

Jones, P.A.(2012) 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nature Reviews Genetics,* 29(17), pp.484-492 doi:10.1038/nrg3230.

Junker, J.P. and Oudenaarden, A. (2014) 'Every Cell Is Special: Genome-wide Studies Add a New Dimension to Single-Cell Biology', *Cell*, 157(1), pp8-11, doi:10.1016/j.cell.2014.02.010

Kawamata, F., Nishihara, H., Homma, S., Kato, Y. Tsuda, M., Konishi, Y., Wang, L., Kohsaka, S., Liu, C., Yoshida, T., Tanino, M., Tanaka, S., Kawamura, H., Kamiyama, T. and Taketomi, A., (2018) 'Chorionic Gonadotropin-β Modulates Epithelial-Mesenchymal Transition in Colorectal Carcinoma Metastasis', *The American Journal of Pathology*, 188(1), pp. 204-215, doi:10.1016/j.ajpath.2017.08.034.

Keay, S.D., Vatish, M., Karteris, E., Hillhouse, E.W. and Randeva, H.S. (2004) 'REVIEW: The role of hCG in reproductive medicine', *British Journal of Obstetrics & Gynaecology*, 111(11), pp. 1218-1228, doi:10.1111/j.1471-0528.2004.00412.x.

Kerschgens, J., Renaud, S., Schütz, F., Grasso, L., Egener-Kuhn, T., Delaloye, J.F., Lehr, H.A., Vogel, H. and Mermod, N. (2011) 'Protein-Binding Microarray Analysis of Tumor Suppressor AP2α Target Gene Specificity', *Plos One,* 6(8), e22895. doi: 10.1371/journal.pone.0022895.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., Lahesmaa, R., Orkin, S.H., Rodig, S.J., Daley, G.Q. and Rao, A. (2011) 'Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in

Mouse Embryonic Stem Cells', *Cell Stem Cell*, 8(2), pp. 200-213, doi:10.1016/j.stem.2011.01.008.

Koistinen, H., Hautala, L., Koli, K. and Stenman, U.H.(2015) 'Absence of TGF-β Receptor Activation by Highly Purified hCG Preparations', *Molecular Endocrinology,* 29(12), pp. 1787-91, doi:10.1210/me.2015-1187.

Krueger, F. and Andrews, S.R.,(2011) 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics,* 27(11) pp.1571-1572, doi:10.1093/bioinformatics/btr167.

Kubiczak, M., Walkowiak, G. P., Nowak-Markwitz, E. and Jankowska, A. (2013) 'Human Chorionic Gonadotropin Beta Subunit Genes *CGB*1 and *CGB*2 are Transcriptionally Active in Ovarian Cancer', *International Journal of Molecular Sciences,* 14(6), pp. 12650-12660, doi10.3390/ijms140612650

Li, E. (2002) 'Chromatin modification and epigenetic reprogramming in mammalian development', *Nature Reviews Genetics,* 3**,** pp**.** 662–673, doi:10.1038/nrg887.

Li, J., Yin, M., Song, W., Cui, F., Wang, W., Wang, S. and Zhu, H. (2018) 'B Subunit of Human Chorionic Gonadotropin Promotes Tumor Invasion and Predicts Poor Prognosis of Early-Stage Colorectal Cancer', *Cell Physiol Biochem*, 45, pp. 237-249, doi: 10.1159/000486770.

Li, L.C. and Dahiya R. (2002) 'MethPrimer: designing primers for methylation PCRs', *Bioinformatics*, 18(11), pp.1427-1431, doi:10.1093/bioinformatics/18.11.1427

Li, Y. and Tollefsbol, T.O. (2011) 'DNA methylation detection: bisulfite genomic sequencing analysis', *Methods in Molecular Biology,* 791, pp11-21, doi:10.1007/978-1-61779-316-5_2.

Liu, L. and Roberts, R.M. (1996) 'Silencing of the gene for the β subunit of human chorionic gonadotropin by the embryonic transcription factor Oct3/4'. *The Journal of Biological Chemistry,* 271(28), pp. 16683–16689, Available at: http://www.jbc.org/content/271/28/16683.full.pdf (Accessed: 16 December 2019).

Livak, K.J. and Schmittgen, T.D.(2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods,* 25(4), pp. 402-408, doi:10.1006/meth.2001.1262

Luthra, R., Patel, K.P., Reddy, N.G., Haghshenas, V., Routbort, M.J., Harmon, M.A., Barkoh, B.A., Kanagal-Shamanna, R., Ravandi, F., Cortes, J.E., Kantarjian, H.M., Medeiros, L.J. and Singh, R.R (2014) 'Next-generation sequencing-based multigene mutational screening for acute myeloid leukemia using MiSeq: applicability for diagnostics and disease monitoring', *Haematologica*, 99(3), pp. 465-473, doi:10.3324/haematol.2013.093765.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D. and Lopez, R.I.(2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Research,* 47(W1), pp. W636-W641, doi:10.1093/nar/gkz268.

Masser, D.R., Berg, A.S. and Freeman, W.M. (2013) 'Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-

generation sequencing', *Epigenetics Chromatin,* 6(1), doi:10.1186/1756-8935-6-33.

McCabe, M. T., Brandes, J. C., and Vertino, P. M. (2009) 'Cancer DNA methylation: molecular mechanisms and clinical implications', *Clinical cancer research : an official journal of the American Association for Cancer Research*, *15*(12), pp. 3927–3937, doi:10.1158/1078-0432.CCR-08-2784.

Mikeska, T., Felsberg, J., Hewitt, C., and Dobrovic, A. (2011) *Analysing DNA Methylation Using Bisulphite Pyrosequencing*. In T. Tollefsbol (Ed.), Epigenetics protocols (pp. 33-53). Humana Press. doi:10.1007/978-1-61779-316-5_4.

Parrott, A. M., Sriram, G., Liu, Y., and Mathews, M. B. (2011) 'Expression of Type II Chorionic Gonadotropin Genes Supports a Role in the Male Reproductive System', *Molecular and Cellular Biology*, 31(2), pp. 287-299. doi:10.1128/MCB.00603-10.

Pestell, R.G., Hollenberg, A.N., Albanese, C. and Jameson, J.L.(1994) 'c-Jun represses transcription of the human chorionic gonadotropin α and β genes through distinct types of CREs', *The Journal of Biological Chemistry*, 269(49), pp. 31090–31096, Available at: http://www.jbc.org/content/269/49/31090.full.pdf (Accessed: 16 December 2019).

Połeć, A., Fedorcsák, P., Eskild, A. and Tanbo, T.G.(2014) 'The interplay of human chorionic gonadotropin (hCG) with basic fibroblast growth factor and adipokines on angiogenesis *in vitro'*, *Placenta*, 35(4), pp. 249-253, doi:10.1016/j.placenta.2014.02.002**.**

Portela, A. and Esteller, M.(2010) 'Epigenetic modifications and human disease', *Nature Biotechnology,* 28, pp. 1057–1068, doi:10.1038/nbt.1685.

Qui, C., Zhi, Y., Shen, Y., Gong, J., Li, Y., and Li, X. (2015) 'High-Resolution melting analysis of HPV-16 *L1* gene methylation: a promising method for prognosing cervical cancer', *Clinical Biochemistry*, 48(13-14), pp. 855-859 doi:10.1016/j.clinbiochem.2015.05.006.

Rao, C.V. (2016) 'There is no turning back on the paradigm shift on the actions of hCG and LH', *Journal of Reproductive Health and Medicine*, 2(1), pp.4-10 doi:10.1016/j.jrhm.2015.07.001.

Reed, K., Poulin, M., Yan, L., and Parisseenti, A. (2009) 'Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation', *Analytical Biochemistry,* 397(1) pp.96-106, doi:10.1016/j.ab.2009.10.021.

Robbins, J. R., Zeldovich, V. B., Poukchanski, A., Boothroyd, J. C., and Bakardjiev, A. I. (2012) 'Tissue barriers of the human placenta to infection with Toxoplasma gondii', *Infection and immunity*, *80*(1), pp. 418–428, doi:10.1128/IAI.05899-11.

Robertson, K. D. (2005) 'DNA methylation and human disease', *Nature Reviews Genetics*, 6, pp. 597-610, doi:10.1038/nrg1655.

Roeh, S., Wiechmann, T., Sauer, S., Ködel, M., Binder, E.B. and Provençal, N.(2018) 'HAM-TBS: high-accuracy methylation measurements via

targeted bisulfite sequencing', *Epigenetics Chromatin*, 11(1) doi:10.1186/s13072-018-0209-x.

Rull, K. and Laan, M. (2005) 'Expression of β-subunit of HCG genes during normal and failed pregnancy', *Human Reproduction*, 20(12), pp.3360-3368, doi:10.1093/humrep/dei261

Rull, K., Hallast, P., Uuskula, L., Jackson, J., Punab, M., Salumets, A., Campbell, R.K. and Laan, M. (2008) 'Fine-scale quantification of HCG beta gene transcription in human trophoblastic and non-malignant non-trophoblastic tissues', *Molecular Human Reproduction*, 14(1), pp. 23-31, doi:10.1093/molehr/gam082.

Schüler-Toprak, S., Treeck, O. and Ortmann, O. (2017) 'Human Chorionic Gonadotropin and Breast Cancer', *International Journal of Molecular Sciences*., 18(7), doi:10.3390/ijms18071587

Schumacher, A., Brachwitz, N., Sohr, S., Engeland, K., Langwisch, S., Dolaptchieva, M., Alexander, T., Taran, A., Malfertheiner, S.F., Costa, S., Zimmermann, G., Nitschke, C., Volk, H., Alexander, H., Gunzer, M. and Zenclussen, A. C. (2009) 'Human Chorionic Gonadotropin Attracts Regulatory T Cells into the Fetal-Maternal Interface during Early Human Pregnancy', *The Journal of Immunology*, 182(9), pp. 5488-5497, doi:10.4049/jimmunol.0803177.

Serranoa, M.A., Maciasb, R.I.R., Brizb, O., Monteb, M.J., Blazquezb, A.G., Williamsonc, C., Kubitzd, R. and Marinb, J.J.G. (2007) 'Expression in Human Trophoblast and Choriocarcinoma Cell Lines, BeWo, Jeg-3 and JAr of Genes Involved in the Hepatobiliary-like Excretory Function of the Placenta', *Placenta*, 28(2–3), pp. 107-117, doi:10.1016/j.placenta.2006.03.009.

Singh, H., Makino, S.I., Endo, Y. and Nie, G.(2010) 'Inhibition of HTRA3 stimulates trophoblast invasion during human placental development.', *Placenta*, 31(12), pp. 1085–1092. doi:10.1016/j.placenta.2010.10.003.

Singh, V., Rana, R. K., and Singhal, R. (2013) 'Analysis of repeated measurement data in the clinical trials', *Journal of Ayurveda and integrative medicine*, *4*(2), pp. 77–81, doi:10.4103/0975-9476.113872

Sinnappan, S. M. (2015) *The role of free beta subunit of human chorionic gonadotropin in high-grade serous cancer*. PhD thesis. The University of Sydney. Available at: http://hdl.handle.net/2123/14706 (Accessed 10.01.2020)

Śliwa, A., Kubiczak, M., Szczerba, A., Walkowiak, G., Nowak-Markwitz, E., Burczyńska, B., Butler, S., Iles, R., Białas, P. and Jankowska, A. (2019) 'Regulation of human chorionic gonadotropin beta subunit expression in ovarian cancer', *BMC cancer*, *19*(1), doi:10.1186/s12885-019-5960-2.

Sohr, S., and Engeland, K. (2011) 'The tumor suppressor p53 induces expression of the pregnancy-supporting human chorionic gonadotropin (hCG) *CGB*7 gene', *Cell cycle*, 10(21), pp. 3758–3767, doi:10.4161/cc.10.21.17946

Soto, J.,Rodriguez-Antolin, C., Vallespín, E., Carpeño, J.C. and Caceres, I.I. (2016) 'The impact of next-generation sequencing on the DNA

methylation–based translational cancer research', *Translational Research,* 169, pp. 1-18.e1, doi:10.1016/j.trsl.2015.11.003.

Stenman, U. H., Tiitinen, A., Alfthan, H. and Valmu, L. (2006) 'The classification, functions and clinical use of different isoforms of HCG', *Human Reproduction Update*, 12(6) , pp. 769–784, doi:10.1093/humupd/dml029

Stenman, U.H., Alfthan, H., and Hotakanien, K. (2004) 'Human chorionic gonadotropin in cancer', *Clinical Biochemistry*, 37(7), pp. 549-561, doi:101016/j.clinbiochem.2004.05.008.

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H.(2009) 'Developmental programming of CpG island methylation profiles in the human genome'. *Nature Structural & Molecular Biology,* 16, pp. 564–571, doi:10.1038/nsmb.1594.

Szczerba, A., Śliwa, A., Kubiczak, M., Nowak-Markwitz, E., and Jankowska, A. (2016) 'Human chorionic gonadotropin β subunit affects the expression of apoptosis-regulating factors in ovarian cancer', *Oncology Reports*, 35, pp. 538-545. doi:10.3892/or.2015.4386

Tammen, S., Frisco, S., and Choi, S. (2012) 'Epigenetics: The link Between Nature and Nurture', *Molecular Aspects of Medicine*, 34(4), pp.753-764, doi:10.1016/j.mam.2012.07.018.

Thomson, J., Skene, P., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., Turner, D.J., Illingworth, R. and Bird, A.(2010) 'CpG islands influence chromatin structure via the CpG-binding protein Cfp1', *Nature,* 464, pp. 1082–1086, doi:10.1038/nature08924.

Tsampalas, M., Gridelet, V., Berndt, S., Foidart, J.M., Geenen, V. and d'Hauterive, S.P. (2010) 'Human chorionic gonadotropin: A hormone with immunological and angiogenic properties', *Journal of Reproductive Immunology,* 85(1), pp. 93-98, doi:10.1016/j.jri.2009.11.008.

Tuncay, S., Senol, H., Guler, E.M., Ocal, N., Secen, H., Kocyigit, A. and Topcu, G., (2018) 'Synthesis of Oleanolic Acid Analogues and Their Cytotoxic Effects on 3T3 Cell Line', *Medicinal Chemistry***,** 14(6), pp. 617-625(9), doi:10.2174/1573406414666180222094544.

Uuskula, L., Rull, K., Nagirnaja, L., and Laan, M. (2010) 'Methylation Allelic Polymorphism (MAP) in Chorionic Gonadotropin beta5 (*CGB*5) and Its Association with Preganancy Success', *Clinical Endocrinology Metabolism*, 96(1), pp. e199-e207, doi:10.1210/jc.2010-1647.

Varela-Rey, M., Iruarrizaga-Lejarreta, M., Lozano, J. J., Aransay, A. M., Fernandez, A. F., Lavin, J. L., Mósen-Ansorena, D., Berdasco, M., Turmaine, M., Luka, Z., Wagner, C., Lu, S.C., Esteller, M., Mirsky, R., Jessen, K.R., Fraga, M.F., Martínez-Chantar, M.L., Mato, J.M. and Woodhoo, A. (2014) 'S-adenosylmethionine levels regulate the Schwann cell DNA methylome', *Neuron*, *81*(5), pp. 1024–1039, doi:10.1016/j.neuron.2014.01.037.

Ward, D. G., Baxter, L., Gordon, N.S., Ott, S., Savage, R.S., Beggs, A.D., James, J.D., Lickiss, J., Green, S., Wallis, Y., Wei, W., James, N.D., Zeegers, M.P., Cheng, K.K., Mathews, G.M., Patel, P., Griffiths, M. and Bryan, R.T

(2016) ‘Multiplex PCR and Next Generation Sequencing for the Non-Invasive Detection of Bladder Cancer’, *PLoS One,* 11(2), pp. e0149756, doi:10.1371/journal.pone.0149756.

Warden, C.D., Lee, H., Tompkins, J.D., Li, X., Wang, C., Riggs, A.D., Yu, H., Jove, R. and Yuan, Y.C.(2013) ‘COHCAP: an integrative genomic pipeline for single-nucleotide resolution DNA methylation analysis’, *Nucleic Acids Research*, 41(11), pp. e117 doi:10.1093/nar/gkt242

Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R. and Clark, S.J. (2002) ‘Identification and resolution of artifacts in bisulfite sequencing’, *Methods*, 27(2), pp. 101-107, doi:10.1016/S1046-2023(02)00060-9.

Weber, M., Hellmann, I., Stadler, M., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007) ‘Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome’, *Nat Genet,* 39, pp.457–466, doi:10.1038/ng1990.

Weinhold, L., Wahl, S., Pechlivanis, S., Hoffmann, P. and Scmid, M. (2016) ‘A statistical model for the analysis of beta values in DNA methylation studies’, *BMC Bioinformatics,* 17, doi:10.1186/s12859-016-1347-4.

Wen, Y., Wei, Y., Zhang, S., Li, S., Liu, H., Wang, F., Zhao, Y., Zhang, D. and Zhang, Y. (2017) ‘Cell subpopulation deconvolution reveals breast cancer heterogeneity based on DNA methylation signature’, *Briefings in Bioinformatics*, 18(3), pp 426–440, https://doi.org/10.1093/bib/bbw028

Whitfield, G. K. and Kourides, I.A. (1985) ‘Expression of Chorionic Gonadotropin *α*- and *β*-Genes in Normal and Neoplastic Human Tissues: Relationship to Deoxyribonucleic Acid Structure’, *Endocrinology*, 117(1), pp. 231–236, doi:10.1210/endo-117-1-231

Wojdacz, T. K., and Dobrovic, A. (2007) ‘Methylation-sensitive high-resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation’, *Nucleic Acid Research*, 6, pp. e41, doi:10.1093/nar/gkm013.

Wojdacz, T. K., Borgbo, T., and Hansen, L. L. (2009) ‘Primer design versus PCR bias in methylation independent PCR amplifications’, *Epigenetics*, 4(4), pp. 231-234, doi:10.4161/epi.9020.

Wolfe, M.W. (2006) *Culture and Transfection of Human Choriocarcinoma Cells*. In: Soares M.J., Hunt J.S. (eds) Placenta and Trophoblast. Methods in Molecular Medicine™, vol 121. Humana Press doi:10.1385/1-59259-983-4:227

Wreczycka, K., Gosdschan, A., Yusuf, D., Grüning, B., Assenov, Y. and Akalin, A. (2017) ‘Strategies for analyzing bisulfite sequencing data’, *Journal of Biotechnology*, 261, pp. 105-115, doi:10.1016/j.jbiotec.2017.08.007.

Wu, W., Gao, H., Li, X., Peng, S., Yu, J., Liu, N., Zhan, G., Zhu, Y., Wang, K., and Guo, X. (2019) β-hCG promotes epithelial ovarian cancer metastasis through ERK/MMP2 signalling pathway, *Cell Cycle,* 18(1), pp. 46-59, doi:10.1080/15384101.2018.1558869.

Xiao, Z., Li, B., Wang, G., Zhu, W., Wang, Z., Lin, J., Xu, A. and Wang, X. (2014) ‘Validation of methylation-sensitive high-resolution melting (MS-HRM) for

the detection of stool DNA methylation in colorectal neoplasms', *Clinica Chimica Acta*, 431, pp. 154-163, doi:10.1016/j.caa.2014.01.044.

Xiong, T., Meister, G.E., Workman, R.E., Kato, N.C, Spellberg, M.J., Turker, F., Timp, W., Ostermeier, M. and Novina, C.D (2017) 'Targeted DNA methylation in human cells using engineered dCas9-methyltransferases', *Scientific Reports,* 7, doi:10.1038/s41598-017-06757-0.

Zhong, Y., Wang, Y., Huang, J., Xu, X., Pan, W., Gao, S., Zhang, Y. and Su, M. (2019) 'Association of hCG and LHCGR expression patterns with clinicopathological parameters in ovarian cancer', *Pathology - Research and Practice*, 215(4) , pp. 748-754, doi:10.1016/j.prp.2019.01.001.

Zimmermann, G., Ackermann, W. and Alexander, H. (2012) 'Expression and Production of Human Chorionic Gonadotropin (hCG) in the Normal Secretory Endometrium: Evidence of *CGB7* and/or *CGB6* Beta hCG Subunit Gene Expression', *Biology of Reproduction*, 86(3), pp. 1-14, doi:10.1095/biolreprod.111.092429

# Appendix

# A1. DNA and RNA samples

**Table A1.** DNA samples with Quantity and Quality values from Nanodrop and Qbit

| Cell Line | Tissue of origin | Qbit | Nanodrop | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Conc. | Concentration | | a260 | a280 | 260/280 | 260/230 |
| MDA-MB-468 | breast cancer | 110 ng/µl | 89.65 | ng/µl | 1.793 | 0.9515 | 1.885 | 2.49 |
| MCF-7 | breast cancer | 804 ng/µl | 200.9 | ng/µl | 4.019 | 2.152 | 1.87 | 2.23 |
| Breast-DNA | Normal breast | N/A | | | | | | |
| HeLA | cervical cancer | 93.6ng/µl | 110.8 | ng/µl | 2.2155 | 1.1705 | 1.89 | 2.23 |
| HT-3 | cervical cancer | 95.6ng/µl | 170.55 | ng/µl | 3.411 | 1.797 | 1.9 | 2.39 |
| C-33a | cervical cancer | 82.2ng/µl | 55.25 | ng/µl | 1.105 | 0.59 | 1.875 | 2.215 |
| Cervix-DNA | Normal cervix | N/A | | | | | | |
| BeWo | choriocarcinoma | 93.2ng/µl | 265.35 | ng/µl | 5.308 | 2.755 | 1.93 | 2.04 |
| JEG-3 | choriocarcinoma | 106ng/µl | 172.85 | ng/µl | 3.457 | 1.823 | 1.895 | 2.315 |
| HCT116 | colon cancer | 94.4ng/µl | 241.6 | ng/µl | 4.832 | 2.5345 | 1.905 | 2.445 |
| CRL-1790 | colon normal | 104 ng/µl | 29.3 | ng/µl | 0.586 | 0.313 | 1.86 | 4.61 |
| 3T3 | mousefibroblast | 104 ng/µl | 115.6 | ng/µl | 2.3115 | 1.1955 | 1.935 | 2.3 |
| OAW42 | ovarian cancer | 76 ng/µl | 59.75 | ng/µl | 1.1955 | 0.635 | 1.88 | 2.695 |
| OVCAR-3 | ovarian cancer | 95.4ng/µl | 124.1 | ng/µl | 2.4815 | 1.309 | 1.895 | 2.62 |
| HEY | ovarian cancer | 96.8ng/µl | 94.55 | ng/µl | 1.8905 | 0.987 | 1.915 | 2.575 |
| SKOV-3 | ovarian cancer | 288 ng/µl | 212.3 | ng/µl | 4.246 | 2.235 | 1.9 | 2.525 |

**Table A2**. RNA samples with Quantity and Quality values from Nanodrop and Qbit

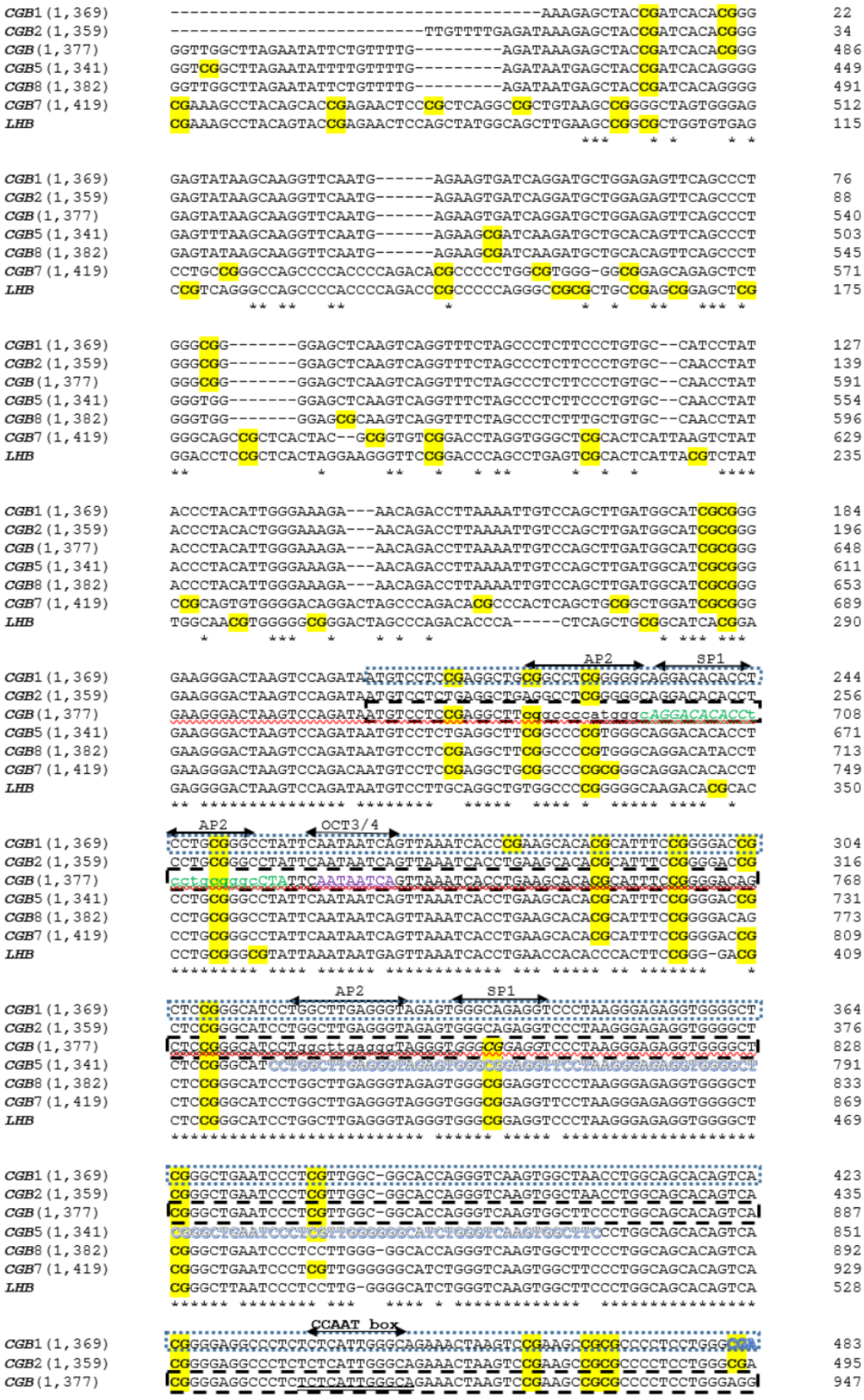
| Cell Line | Tissue of origin | Qbit | Nanodrop | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Conc. | Concentration | | a260 | a280 | 260/280 | 260/230 |
| 3T3 | mouse fibroblast | 166 ng/µl | 348.8 | ng/µl | 8.72 | 4.049 | 2.15 | 0.98 |
| BeWo | choriocarcinoma | 400 ng/µl | 622.7 | ng/µl | 15.569 | 7.301 | 2.13 | 2.12 |
| C-33a | cervical cancer | 348 ng/µl | 400.7 | ng/µl | 10.017 | 4.728 | 2.12 | 1.86 |
| CRL-1790 | colon normal | 190 ng/µl | 155.55 | ng/µl | 3.89 | 1.84 | 2.12 | 1.55 |
| HCT116 | colon cancer | 573 ng/µl | 864.3 | ng/µl | 21.607 | 10.179 | 2.12 | 1.9 |
| HeLA | cervical cancer | 300 ng/µl | 419.6 | ng/µl | 10.491 | 4.916 | 2.13 | 1.16 |
| HEY | ovarian cancer | 808 ng/µl | 781.7 | ng/µl | 19.543 | 9.21 | 2.12 | 2.07 |
| HT-3 | cervical cancer | 578 ng/µl | 584.6 | ng/µl | 14.616 | 6.821 | 2.14 | 1.5 |
| JEG-3 | choriocarcinoma | 799 ng/µl | 832.9 | ng/µl | 20.822 | 9.776 | 2.13 | 2.16 |
| MCF-7 | breast cancer | 126 ng/µl | 258 | ng/µl | 6.451 | 3.002 | 2.15 | 1.17 |
| MDA-MB-468 | breast cancer | 1100ng/µl | 993.5 | ng/µl | 24.839 | 11.682 | 2.13 | 1.86 |
| OAW42 | ovarian cancer | 178 ng/µl | 274.2 | ng/µl | 6.856 | 3.223 | 2.13 | 1.7 |
| OVCAR-3 | ovarian cancer | 356 ng/µl | 554.8 | ng/µl | 13.869 | 6.447 | 2.15 | 1.74 |
| SKOV-3 | ovarian cancer | 132 ng/µl | 164.6 | ng/µl | 4.115 | 1.939 | 2.12 | 1.43 |

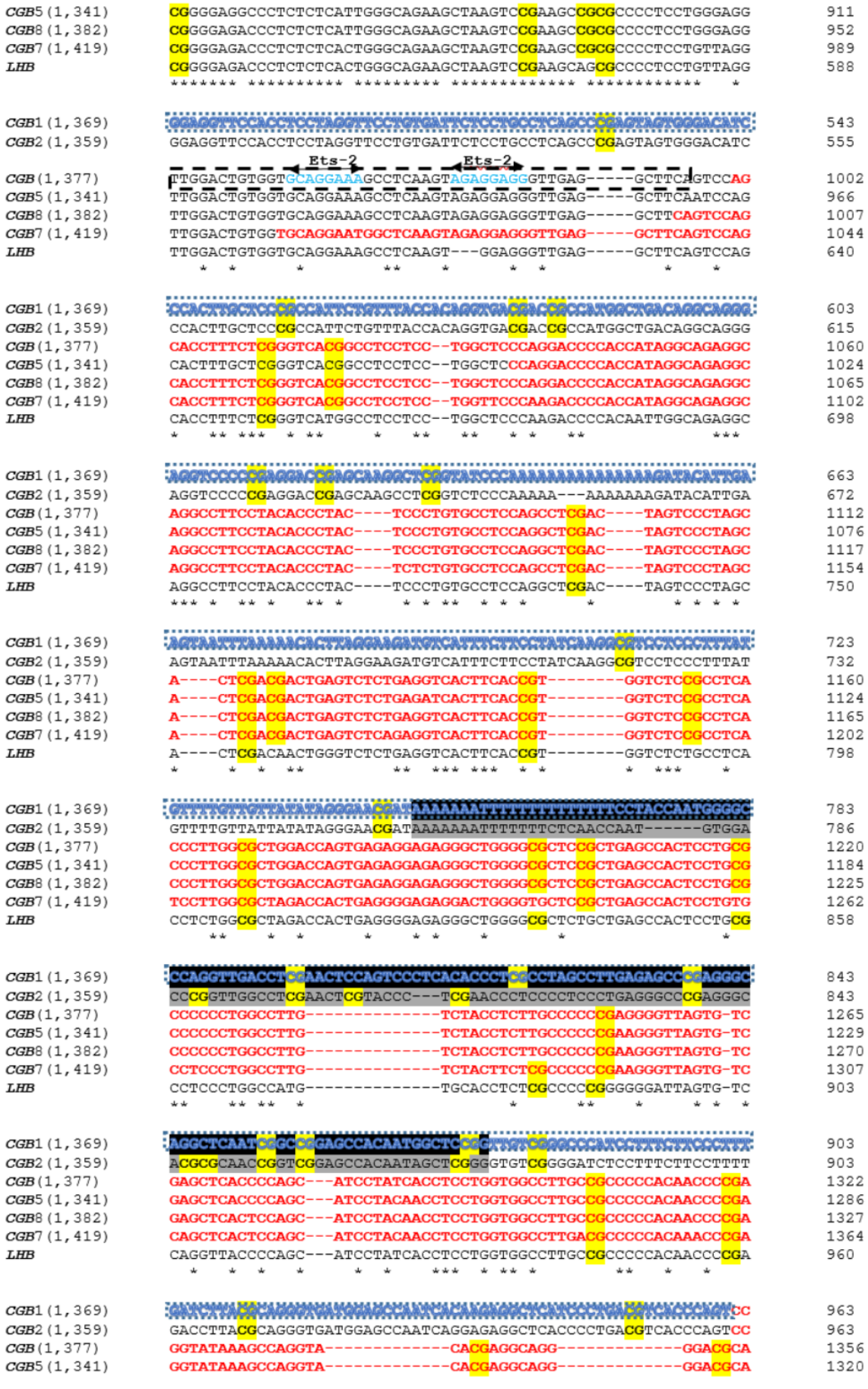
# A2. *In silico* analyses

## A2.1 MSA with promoter elements

Below is the Multiple Sequence Alignment of the *CGB* genes. The sequences used are the 1 exon and 1000bp upstream for the *CGB1*-9 genes and the *LHB* genes TSSs.

```
CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       --------------ACTCTCTGAATTTGAGCTCAATCTCTGCAGGATGGGTGCCACCACA  46
CGB5(1,341)      ------------------------------------AGGATGGGTGCCACCACA  18
CGB8(1,382)      ATCACAAAATCACAACTCTCTGAATTTGAGCTCAGTCTCTGCAGGATGGGTGCCACCACG  60
CGB7(1,419)      --TCCAGGACGTTAAGTACCTCGA------------GGACCCAGGAACGCAGGCAACGCG  46
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       TGTGGTTTTGAAGGTTGAATAGGAGCTCTCCCAGGGGAACTTGAGGGTAATCATGATGAT  106
CGB5(1,341)      TGGGGTTTTGAAGGTTGAATAGGAGTTCTCCTGGGGGAACTTGAGGATAATCATGATG--  76
CGB8(1,382)      TGGGGTTTTGAAGGTTGAATAGGAGTTCTCCAGGGGGAAATTGAGAGTAATCATGATG--  118
CGB7(1,419)      TGCCCGGTCGCGTCCTTCCTAAGGGGTGCTCGCCTGGAAATGCAGGAGTTTGAATGTGCA  106
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       GATG---ATGATGA----TGATAATAATAATAG---------CCACTATTTACTGAGTGT  150
CGB5(1,341)      ----------ATGA----TGATAATAATAATAG---------GCACTATTTACTGAGTGT  113
CGB8(1,382)      ----------ATGA----TGATGATAATAATAG---------CCACTATTTACTGAGTGT  155
CGB7(1,419)      GACTCCTTAGATCCCGAGGTCTGAAACTCGTCCCCAAGGATCCAGCTGTATCCTCACCAT  166
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       TTACTCTTTCGTAGCCCTAATACATAACTCCTCGGATCAACTCTCATGGATTTGATCAT-  209
CGB5(1,341)      TTACTCTTTCTTAGCCATAATACATAACTCCTCGGATCAACTCTCATGGATTTGATCAT-  172
CGB8(1,382)      TTACTCTTTCGTAGCCCTAATACATAACTCCTCAGATCAACTCTCATGGATTTGATCAT-  214
CGB7(1,419)      CCCGCCCTGCTCC--CACAGGACGGAGCCGTCCGGATACCCCCGCAAGGACTACCTGACC  224
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       TGGTGACCTTTGGTGTTAAGTTGCTGACTGCTCAGTCAC-------AGAGGACACCAC--  260
CGB5(1,341)      TGGTGACCTTTGGTGTTAAGTTGCTGACTGCTCAGTCAC-------AGAGGACACCAC--  223
CGB8(1,382)      TGGTGACCTTTGGTGTTAAGTTGCTGACTGCTCAGTCAC-------AGAGGACACCAC--  265
CGB7(1,419)      TACTGGAGCTTAGAAGAGACGCCCCAACTCTGGAGTCACGACCCGCAGCGGCAGCCCTGT  284
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       -CTTGCTCATCCTGGGGAGTGGGAGG---GCACATTTCACGATGTGCATGGGGGAGGAGG  316
CGB5(1,341)      -CTTCCTCATCCTGGGGAGTGGGAGG---GCACATTTCACGATGTGCATGGGGGAGGAGA  279
CGB8(1,382)      -CTTGCTCATCCTGGGGAGTGGGAGG---GCACATTTCACGATGTGGATGGGGGTGGAGA  321
CGB7(1,419)      CTCTGCTCTTCCAGGCCGCCTCGAGCGCCGCACGGCGGGCCATGTCTCCGGCTCAGGAAA  344
LHB              ------------------------------------------------------------  0

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       GAAACTGGAACATGCAAGCAGATGGCCAGGGGACCTTGAGGACAT-GGTCTACAGAAGGC  375
CGB5(1,341)      GAAACTGGAACATGCAAGCAGATGGCCAGGGGACCTTGAGAACAT-GGTCTACAGAAGGC  338
CGB8(1,382)      GAAACTGGAGCATGCAAGCAGATGGCCAGGGGACCTTGAGGACAT-GGTCTACAGAAGGC  380
CGB7(1,419)      CCTACAGCTCCTCGCC-GCACCCGCTCCGCC-GGCTCGACCGCTTCCGCCCGCTGGAGGC  402
LHB              ----------------------------------------------------GAGGC  5

CGB1(1,369)      ------------------------------------------------------------  0
CGB2(1,359)      ------------------------------------------------------------  0
CGB(1,377)       CTTTAAGTATCTGGGAGCTGGGGTTCAAATGAGAAATCTTACTTGGTGAGAGCAGGCAGG  435
CGB5(1,341)      CTTTCAGTATCTGGGAGCTGGGGTTCAAATGAGAAATCTTACTTGGTGAGAGTGGGCAGG  398
CGB8(1,382)      CTTTAAGTATCTGGGAGCTGGGGTTCAAATGAGAAATCTTACTTGGTGAGAGCAGGCAGG  440
CGB7(1,419)      GAC---------TTGGGGTGGC-CCCCACTGGAAGCCCCTGCCGGGCATACACAGCGTCC  452
LHB              GCC---------CTGGGGCGGC-CCGCACTGGAAGCCCCTGCGGGGCATCTACAGCGTGC  55
```


AP2
SP1
AP2
OCT3/4
AP2
SP1
CCAAT box


Ets-2
Ets-2

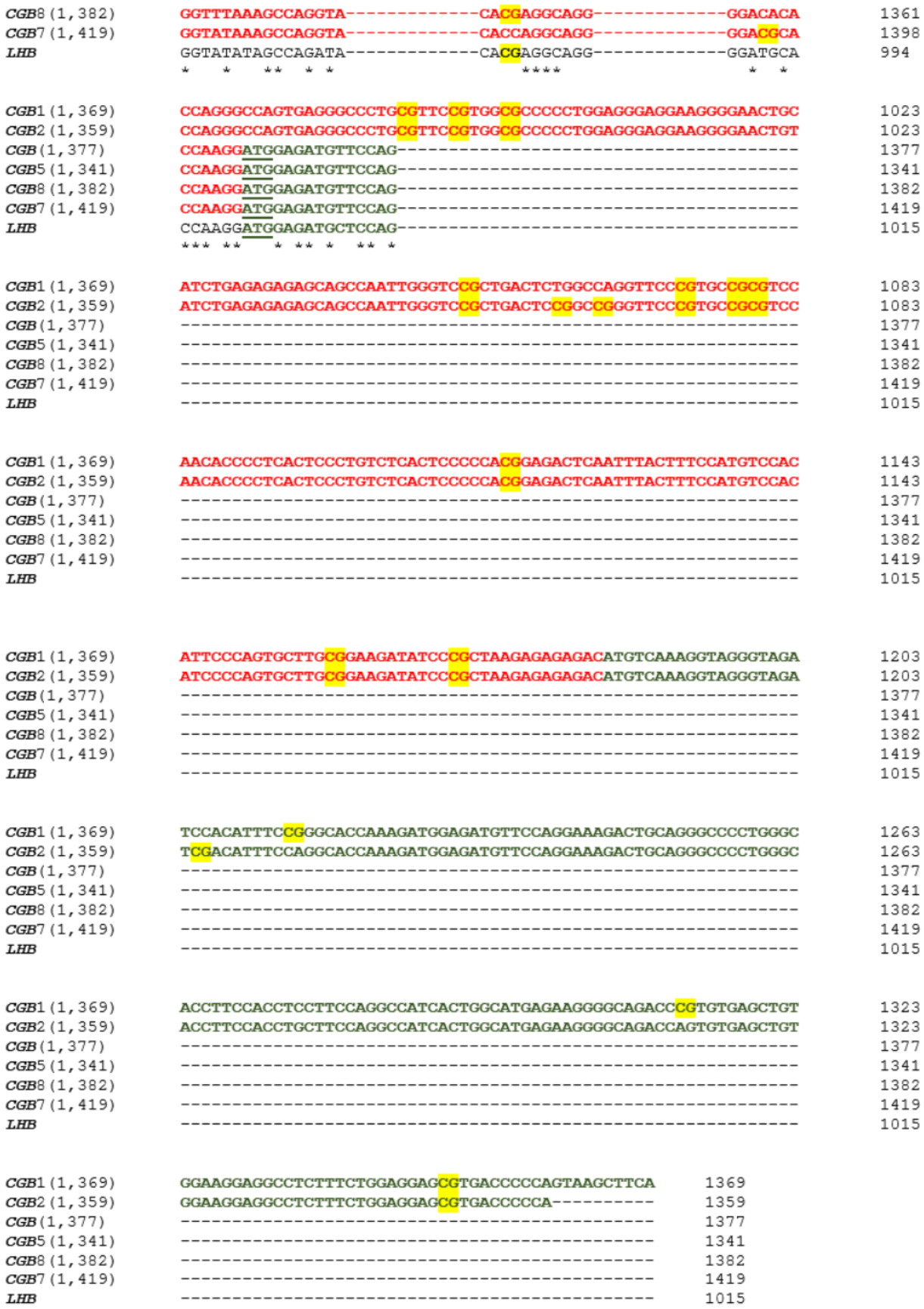



**Fig.A1** Multiple Sequence alignment of *CGB1*-9 and *LHB*. Legend for MSA: **CG** – CpG site; cggcccc **–** AP2 binding; *CGGCCCC* **–** SP1 binding; Putative promoter *CGB*3-9; Putative promoter *CGB*1-2;**CGGCCCC - 1st** exon 5' UTR;**ATGCGG** – 1st exon after ATG;CTC**CG**GGC- cAMP response elements; cctg**cg**gg-TSE ; TCTCATT-CCAAT box; TCTCATT **–** Ets-2; TCTCATT-Oct3/4; **TCTCATT** -*CGB*1-2 insert **;** snaR-G1 reverse complement; **TGGC** snaR-G2 reverse complement;

# A2.2 CpG Island predictions

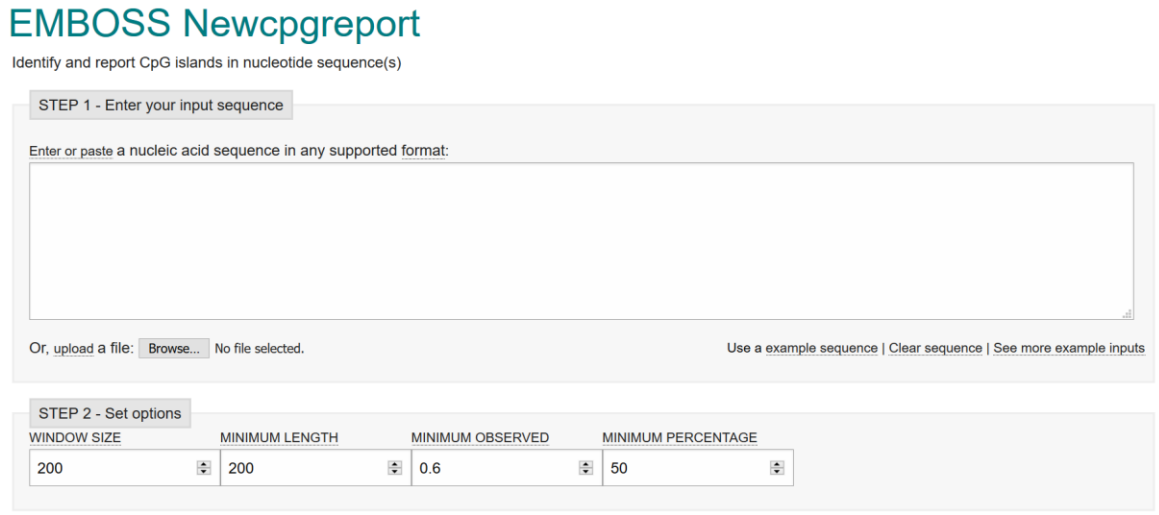



**Fig.A2** Screen capture of the Newcpgreport tool and its settings used to predict the CGIs for CGB3-8

```
ID   CGB3  1377 BP.
XX
DE   CpG Island report.
XX
CC   Obs/Exp ratio > 0.60.
CC   % C + % G > 50.00.
CC   Length > 200.
XX
FH   Key             Location/Qualifiers
FT   no islands detected

ID   CGB5  1341 BP.
XX
DE   CpG Island report.
XX
CC   Obs/Exp ratio > 0.60.
CC   % C + % G > 50.00.
CC   Length > 200.
XX
FH   Key             Location/Qualifiers
FT   no islands detected


ID   CGB8  1382 BP.
XX
DE   CpG Island report.
XX
CC   Obs/Exp ratio > 0.60.
CC   % C + % G > 50.00.
CC   Length > 200.
XX
FH   Key             Location/Qualifiers
FT   no islands detected
```

**Fig.A3** Output from Newcpgreport tool for CGB3, 5 and 8. ID row is the user defined ID and length of input sequence, DE- description of test, CC rows – conditions of the test, FH and FT rows present the predicted CpG island in relation to the input sequence.

# A2.3 Primer design and ePCR

MethPrimer

NEW! Invitation to test MethPrimer 2.0

Paste an ORIGINAL source sequence. Try this Sample sequence
You don't need to modify your sequence (e.g. convert 'C' to 'T') before pasting.

Pick primers for bisulfite sequencing PCR or restriction PCR .

Pick MSP primers.

| Use CpG island prediction for primer selection? | Window | Shift | Obs/Exp | GC% |
|---|---|---|---|---|
| ☐ | 100 | 1 | 0.6 | 50 |

Submit Reset

| General Parameters for Primer Selection | | |
|---|---|---|
| Sequence name (optional): | | |
| Target (optional): | | "start, size", such as (560, 30) |
| Excluded Regions (optional): | | "start, size", such as (160, 50 1100, 50) |
| Number of output pairs (optional): | 5 | |

| | | | |
|---|---|---|---|
| Product Size: | Min: 100 | Opt: 200 | Max: 300 |
| Primer Tm: | Min: 50 | Opt: 55 | Max: 60 |
| Primer Size: | Min: 20 | Opt: 25 | Max: 30 |
| Product CpGs: | 4 | Primer Poly X: | 5 |
| Primer non-CpG 'C's: | 4 | Primer Poly T: | 8 |

| Parameters for MSP primers | |
|---|---|
| 3'CpG constraint: | 3 |
| CpG in primer: | 1 |
| Max Tm difference: | 5 |

Submit Reset

**Fig.A4** Screen capture of the MethPrimer tool and its settings used to design primers for *CGB3-8* putative promoter

```
      Primer              Start Size  Tm      GC%    'C's  Sequence

1. Left  primer            647   27   59.54   48.15   4
   GGGAAGGGATTAAGTTTAGATAATGTT
   Right primer            879   25   59.03   64.00   7
   CTACCAAAAAAACCACTTAACCCTA
                  Product size: 233, Tm: 67.2, CpGs in product: 10
```

**Fig.A5** Sample output of primer pair with primer characteristics

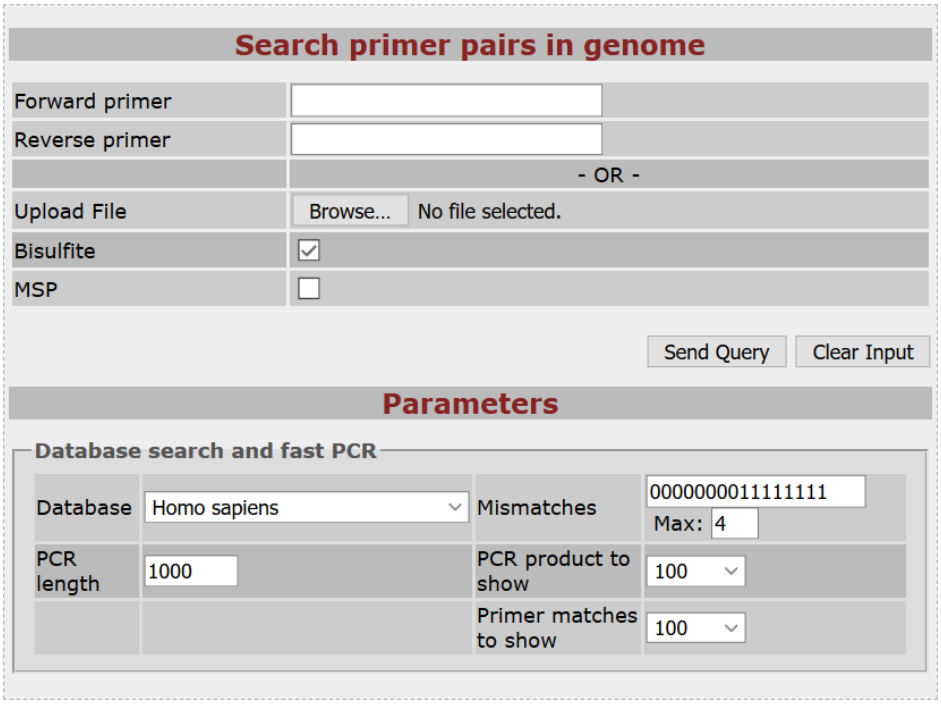



**Fig.A6** Screen capture of the BiSearch ePCR page and its settings used to predict PCR products for proposed primer pairs

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: GGGAAGGGATTAAGTTTAGATAATGTT
      Reverse primer: CTACCAAAAAAACCACTTAACCCTA
1. Chromosome 19 (CGB2)(len: 233)
  49031105                                  GGG AAGGGATTAA GTTTAGATAA
           TGTTTTTTGA GGTTGAGGTT TTGGGGGTAG GATATATTTT TTGTGGGTTT
           ATTTAATAAT TAGTTAAATT ATTTGAAGTA TATGTATTTT TGGGGATTGT
           TTTGGGTATT TTGGTTTGAG GGTAGAGTGG GTAGAGGTTT TTAAGGGAGA
           GGTGGGGTTT GGGTTGAATT TTTTGTTGGT GGTATTAGGG TTAAGTGGTT
           AATTTGGTAG                                          49031338
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: GGGAAGGGATTAAGTTTAGATAATGTT
      Reverse primer: CTACCAAAAAAACCACTTAACCCTA
```

3 PCR products should be generated.

```
1. Chromosome 19(CGB3) (len: 233)
  49024454                                  CTAC CAAAAAAACC ACTTAACCCT
           AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
           CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
           TACATATACT TCAAATAATT TAACTAATTA TTAAATAAAC CCACAAAAAA
           TATATCCTAC CCATAAAACC AAAACCTCAA AAAACATTAT CTAAACTTAA
           TCCCTTCCC                                           49024687
2. Chromosome 19(CGB1)(len: 233)
  49037480    CTACCAAA TTAACCACTT AACCCTAATA CCACCAACAA AAAATTCAAC
           CCAAACCCCA CCTCTCCCTT AAAAACCTCT ACCCACTCTA CCCTCAAACC
           AAAATACCCA AAACAATCCC CAAAAATACA TATACTTCAA ATAATTTAAC
           TAATTATTAA ATAAACCCAC AAAAAATATA TCCTACCCCC AAAACCACAA
           CCTCAAAAAA CATTATCTAA ACTTAATCCC TTCCC               49037713
3. Chromosome 19 (CG8)(len: 233)
  49049222                                                 CTACCA
           AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
           CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
           CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
           AAATAAACCC ACAAAAAATA TATCCTACCC ACAAAACCAA AACCTCAAAA
           AACATTATCT AAACTTAATC CCTTCCC                       49049455
```

**Fig.A7** ePCR Predicted PCR products for primer pair 1.

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: TTAATAATTAGTTAAATTATTTGAAGTATA
      Reverse primer: AAAAAAATACTAAACTAAAACCTC
1. Chromosome 19 (CGB5)(len: 291)
  49043567                                                  T TAATAATTAG
            TTAAATTATT TGAAGTATAT GTATTTTTGG GGATTGTTTT GGGTATTTTG
            GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
            TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
            TAGTTATGGG GAGGTTTTTT TTTATTGGGT AGAAGTTAAG TTTGAAGTTG
            TGTTTTTTTT GGGAGGTTGG ATTGTGGTGT AGGAAAGTTT TAAGTAGAGG
            AGGGTTGAGG TTTTAATTTA GTATTTTGTT                  49043858
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: TTAATAATTAGTTAAATTATTTGAAGTATA
      Reverse primer: AAAAAAATACTAAACTAAAACCTC
```

3 PCR products should be generated.

```
1. Chromosome 19 (CGB3)(len: 290)
  49024322                                                  AAAAAA
            ATACTAAACT AAAACCTCAA CCCTCCTCTA CTTAAAACTT TCCTACACCA
            CAATCCAACC TCCCAAAAAA AACACAACTT CAAACTTAAT TTCTACCCAA
            TAAAAAAAAA CCTCCCCATA ACTATACTAC CAAAAAAACC ACTTAACCCT
            AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
            CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
            TACATATACT TCAAATAATT TAACTAATTA TTAA              49024612
2. Chromosome 19 (CGB8)(len: 290)
  49049090               AAAAAAAT ACTAAACTAA AACCTCAACC CTCCTCTACT
            TAAAACTTTC CTACACCACA ATCCAACCTC CCAAAAAAAA CACAACTTCA
            AACTTAACTT CTACCCAATA AAAAAAAATC TCCCCATAAC TATACTACCA
            AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
            CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
            CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
            AA                                                49049380
3. Chromosome 19 (CGB7)(len: 291)
  49055726                                                  AA
            AAAAATACTA AACTAAAACC TCAACCCTCC TCTACTTAAA CCATTCCTAC
            ACCACAATCC AACCTAACAA AAAAAACACA ACTTCAAACT TAACTTCTAC
            CCAATAAAAA AAAATCTCCC CATAACTATA CTACCAAAAA AACCACTTAA
            CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
            AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCCAA AACAATCCCC
            AAAAATACAT ATACTTCAAA TAATTTAACT AATTATTAA          49056017
```

**Fig.A8** ePCR Predicted PCR products for primer pair 2.

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: TTTAATAATTAGTTAAATTATTTGAAGTAT
      Reverse primer: CTTAATTTCTACCCAATAAAAAAAA
```

2 PCR products should be generated.

```
1. Chromosome 19 (CGB2)(len: 201)cgb2
  49031179   TTTAATAAT TAGTTAAATT ATTTGAAGTA TATGTATTTT TGGGGATTGT
             TTTGGGTATT TTGGTTTGAG GGTAGAGTGG GTAGAGGTTT TTAAGGGAGA
             GGTGGGGTTT GGGTTGAATT TTTTGTTGGT GGTATTAGGG TTAAGTGGTT
             AATTTGGTAG TATAGTTATG GGGAGGTTTT TTTTTATTGG GTAGAAATTA
             AG                                                  49031380
2. Chromosome 19 (CGB5)(len: 202)
  49043566                                              TT TAATAATTAG
             TTAAATTATT TGAAGTATAT GTATTTTTGG GGATTGTTTT GGGTATTTTG
             GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
             TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
             TAGTTATGGG GAGGTTTTTT TTTATTGGGT AGAAGTTAAG         49043768
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: TTTAATAATTAGTTAAATTATTTGAAGTAT
      Reverse primer: CTTAATTTCTACCCAATAAAAAAAA
```

4 PCR products should be generated.

```
1. Chromosome 19 (CGB3)(len: 201)
  49024412                                          CTTAAT TTCTACCCAA
             TAAAAAAAAA CCTCCCCATA ACTATACTAC CAAAAAAACC ACTTAACCCT
             AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
             CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
             TACATATACT TCAAATAATT TAACTAATTA TTAAA              49024613
2. Chromosome 19 (CGB1)(len: 201)
  49037438             CTTAATTTCT ACCCAATAAA AAAAAACCTC CCCATAACTA
             TACTACCAAA TTAACCACTT AACCCTAATA CCACCAACAA AAAATTCAAC
             CCAAACCCCA CCTCTCCCTT AAAAACCTCT ACCCACTCTA CCCTCAAACC
             AAAATACCCA AAACAATCCC CAAAAATACA TATACTTCAA ATAATTTAAC
             TAATTATTAA A                                        49037639
3. Chromosome 19 (CGB8)(len: 201)     cgb8
  49049180    CTTAACTT CTACCCAATA AAAAAAAATC TCCCCATAAC TATACTACCA
             AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
             CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
             CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
             AAA                                                 49049381
4. Chromosome 19 (CGB7)(len: 202)
  49055816                                              CT TAACTTCTAC
             CCAATAAAAA AAAATCTCCC CATAACTATA CTACCAAAAA AACCACTTAA
             CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
             AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCCAA AACAATCCCC
             AAAAATACAT ATACTTCAAA TAATTTAACT AATTATTAAA         49056018
```

**Fig.A9** ePCR Predicted PCR products for primer pair 3.

PCR product(s) on the bisulfite transformed sense chain

Forward primer: GGGAAGGGATTAAGTTTAGATAATGTT

Reverse primer: CTTAATTTCTACCCAATAAAAAAAA

2 PCR products should be generated.

```
1. Chromosome 19 (CGB2)(len: 275)
  49031105                                  GGG AAGGGATTAA GTTTAGATAA
           TGTTTTTTGA GGTTGAGGTT TTGGGGGTAG GATATATTTT TTGTGGGTTT
           ATTTAATAAT TAGTTAAATT ATTTGAAGTA TATGTATTTT TGGGGATTGT
           TTTGGGTATT TTGGTTTGAG GGTAGAGTGG GTAGAGGTTT TTAAGGGAGA
           GGTGGGGTTT GGGTTGAATT TTTTGTTGGT GGTATTAGGG TTAAGTGGTT
           AATTTGGTAG TATAGTTATG GGGAGGTTTT TTTTTATTGG GTAGAAATTA
           AG                                                    49031380
2. Chromosome 19 (CGB5)(len: 276)
  49043492                  GGGAAG GGATTAAGTT TAGATAATGT TTTTTGAGGT
           TTTGGTTTTG TGGGTAGGAT ATATTTTTTG TGGGTTTATT TAATAATTAG
           TTAAATTATT TGAAGTATAT GTATTTTTGG GGATTGTTTT GGGTATTTTG
           GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
           TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
           TAGTTATGGG GAGGTTTTTT TTTATTGGGT AGAAGTTAAG          49043768
```

PCR product(s) on the bisulfite transformed antisense chain

Forward primer: GGGAAGGGATTAAGTTTAGATAATGTT

Reverse primer: CTTAATTTCTACCCAATAAAAAAAA

5 PCR products should be generated.

```
1. Chromosome 19 (LHB)(len: 274)
  49017519                                                 CTTAACTTC
           TACCCAATAA AAAAAAATCT CCCCATAACT ATACTACCAA AAAAACCACT
           TAACCCAAAT ACCCCCAAAA AAAAATTAAA CCCAAACCCC ACCTCTCCCT
           TAAAAACCTC CACCCACCCT ACCCTCAAAC CAAAATACCC AAAACATCCC
           CAAAAATAAA TATAATTCAA ATAATTTAAC TCATTATTTA ATACACCCAC
           AAAATACATA TCTTACCCCC AAAACCACAA CCTACAAAAA CATTATCTAA
           ACTTAATCCC CTCTC                                    49017793
2. Chromosome 19 (CGB3)(len: 275)
  49024412                                        CTTAAT TTCTACCCAA
           TAAAAAAAAA CCTCCCCATA ACTATACTAC CAAAAAAACC ACTTAACCCT
           AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
           CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
           TACATATACT TCAAATAATT TAACTAATTA TTAAATAAAC CCACAAAAAA
           TATATCCTAC CCATAAAACC AAAACCTCAA AAAACATTAT CTAAACTTAA
           TCCCTTCCC                                           49024687
3. Chromosome 19 (CGB1)(len: 275)
  49037438              CTTAATTTCT ACCCAATAAA AAAAAACCTC CCCATAACTA
           TACTACCAAA TTAACCACTT AACCCTAATA CCACCAACAA AAAATTCAAC
           CCAAACCCCA CCTCTCCCTT AAAAACCTCT ACCCACTCTA CCCTCAAACC
           AAAATACCCA AAACAATCCC CAAAAATACA TATACTTCAA ATAATTTAAC
           TAATTATTAA ATAAACCCAC AAAAAATATA TCCTACCCCC AAAACCACAA
           CCTCAAAAAA CATTATCTAA ACTTAATCCC TTCCC               49037713
4. Chromosome 19 (CGB8)(len: 275)
  49049180    CTTAACTT CTACCCAATA AAAAAAAATC TCCCCATAAC TATACTACCA
           AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
           CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
           CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
           AAATAAACCC ACAAAAAATA TATCCTACCC ACAAAACCAA AACCTCAAAA
           AACATTATCT AAACTTAATC CCTTCCC                       49049455
5. Chromosome 19 (CGB7)(len: 276)
  49055816                                              CT TAACTTCTAC
           CCAATAAAAA AAAATCTCCC CATAACTATA CTACCAAAAA AACCACTTAA
           CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
           AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCCAA AACAATCCCC
           AAAAATACAT ATACTTCAAA TAATTTAACT AATTATTAAA TAAACCCACA
           AAAAATATAT CCTACCCACA AAACCACAAC CTCAAAAAAC ATTATCTAAA
           CTTAATCCCT TCCC                                     49056092
```

**Fig.A10** ePCR Predicted PCR products for primer pair 4.

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: TTAATAATTAGTTAAATTATTTGAAGTATA
      Reverse primer: CTTAATTTCTACCCAATAAAAAAAA
```

2 PCR products should be generated.

```
1. Chromosome 19  (CGB2)(len: 200)
  49031180    TTAATAAT TAGTTAAATT ATTTGAAGTA TATGTATTTT TGGGGATTGT
            TTTGGGTATT TTGGTTTGAG GGTAGAGTGG GTAGAGGTTT TTAAGGGAGA
            GGTGGGGTTT GGGTTGAATT TTTTGTTGGT GGTATTAGGG TTAAGTGGTT
            AATTTGGTAG TATAGTTATG GGGAGGTTTT TTTTTATTGG GTAGAAATTA
            AG                                                  49031380
2. Chromosome 19 (CGB5)(len: 201)
  49043567                                                T TAATAATTAG
            TTAAATTATT TGAAGTATAT GTATTTTTGG GGATTGTTTT GGGTATTTTG
            GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
            TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
            TAGTTATGGG GAGGTTTTTT TTTATTGGGT AGAAGTTAAG         49043768
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: TTAATAATTAGTTAAATTATTTGAAGTATA
      Reverse primer: CTTAATTTCTACCCAATAAAAAAAA
```

4 PCR products should be generated.

```
1. Chromosome 19 (CGB3)(len: 200)
  49024412                                        CTTAAT TTCTACCCAA
            TAAAAAAAAA CCTCCCCATA ACTATACTAC CAAAAAAACC ACTTAACCCT
            AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
            CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
            TACATATACT TCAAATAATT TAACTAATTA TTAA               49024612
2. Chromosome 19 (CGB1)(len: 200)
  49037438             CTTAATTTCT ACCCAATAAA AAAAAACCTC CCCATAACTA
            TACTACCAAA TTAACCACTT AACCCTAATA CCACCAACAA AAAATTCAAC
            CCAAACCCCA CCTCTCCCTT AAAAACCTCT ACCCACTCTA CCCTCAAACC
            AAAATACCCA AAACAATCCC CAAAAATACA TATACTTCAA ATAATTTAAC
            TAATTATTAA                                        49037638
3. Chromosome 19 (CGB8)(len: 200)
  49049180    CTTAACTT CTACCCAATA AAAAAAAATC TCCCCATAAC TATACTACCA
            AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
            CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
            CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
            AA                                                  49049380

4. Chromosome 19 (CGB7)(len: 201)
  49055816                                              CT TAACTTCTAC
            CCAATAAAAA AAAATCTCCC CATAACTATA CTACCAAAAA AACCACTTAA
            CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
            AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCCAA AACAATCCCC
            AAAAATACAT ATACTTCAAA TAATTTAACT AATTATTAA          49056017
```

**Fig.A11** ePCR Predicted PCR products for primer pair 5.

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: GGGGAAGGGATTAAGTTTAGA
      Reverse primer: ACTATACTACCAAAAAAACCACTTA
1. Chromosome 19 (CGB5)(len: 241)
  49043492                GGGGAAG GGATTAAGTT TAGATAATGT TTTTTGAGGT
            TTTGGTTTTG TGGGTAGGAT ATATTTTTTG TGGGTTTATT TAATAATTAG
            TTAAATTATT TGAAGTATAT GTATTTTTGG GGATTGTTTT GGGTATTTTG
            GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
            TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
            TAGT                                                49043732
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: GGGGAAGGGATTAAGTTTAGA
      Reverse primer: ACTATACTACCAAAAAAACCACTTA
```

4 PCR products should be generated.

```
1. Chromosome 19 (LHB)(len: 239)
  49017556                              ACT ATACTACCAA AAAAACCACT
            TAACCCAAAT ACCCCCAAAA AAAAATTAAA CCCAAACCCC ACCTCTCCCT
            TAAAAACCTC CACCCACCCT ACCCTCAAAC CAAAATACCC AAAACATCCC
            CAAAAATAAA TATAATTCAA ATAATTTAAC TCATTATTTA ATACACCCAC
            AAAATACATA TCTTACCCCC AAAACCACAA CCTACAAAAA CATTATCTAA
            ACTTAATCCC CTCTCC                                   49017794

2. Chromosome 19 (CGB3) (len: 240)
  49024449                        ACTATACTAC CAAAAAAACC ACTTAACCCT
            AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
            CTCCACCCAC CCTACCCTCA AACCAAAATA CCCAAAACTA TCCCCAAAAA
            TACATATACT TCAAATAATT TAACTAATTA TTAAATAAAC CCACAAAAAA
            TATATCCTAC CCATAAAACC AAAACCTCAA AAAACATTAT CTAAACTTAA
            TCCCTTCCCC                                          49024688
3. Chromosome 19 (CGB8)(len: 240)
  49049217                                          AC TATACTACCA
            AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
            CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
            CAAAACTATC CCCAAAAATA CATATACTTC AAATAATTTA ACTAATTATT
            AAATAAACCC ACAAAAAATA TATCCTACCC ACAAAACCAA AACCTCAAAA
            AACATTATCT AAACTTAATC CCTTCCCC                      49049456
4. Chromosome 19  (CGB7) (len: 241)
  49055853                              ACTATA CTACCAAAAA AACCACTTAA
            CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
            AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCCAA AACAATCCCC
            AAAAATACAT ATACTTCAAA TAATTTAACT AATTATTAAA TAAACCCACA
            AAAAATATAT CCTACCCACA AAACCACAAC CTCAAAAAAC ATTATCTAAA
            CTTAATCCCT TCCCC                                    49056093
```

**Fig.A12** ePCR Predicted PCR products for primer pair 6.

PCR product(s) on the bisulfite transformed sense chain

```
      Forward primer: GGGTATTTTGGTTTGAGGG
      Reverse primer: CCTCAACCCTCCTCTACTT
1. Chromosome 19 (CGB5)(len: 220)
  49043619                                                  GGGTATTTTG
            GTTTGAGGGT AGAGTGGGTG GAGGTTTTTA AGGGAGAGGT GGGGTTTGGG
            TTGAATTTTT TGTTGGGGGG TATTTGGGTT AAGTGGTTTT TTTGGTAGTA
            TAGTTATGGG GAGGTTTTTT TTTATTGGGT AGAAGTTAAG TTTGAAGTTG
            TGTTTTTTTT GGGAGGTTGG ATTGTGGTGT AGGAAAGTTT TAAGTAGAGG
            AGGGTTGAGG                                          49043838
```

PCR product(s) on the bisulfite transformed antisense chain

```
      Forward primer: GGGTATTTTGGTTTGAGGG
      Reverse primer: CCTCAACCCTCCTCTACTT
```

3 PCR products should be generated.

```
1. Chromosome 19 (CGB3)(len: 219)
  49024343                 CCTCAA CCCTCCTCTA CTTAAAACTT TCCTACACCA
            CAATCCAACC TCCCAAAAAA AACACAACTT CAAACTTAAT TTCTACCCAA
            TAAAAAAAAA CCTCCCCATA ACTATACTAC CAAAAAAACC ACTTAACCCT
            AATACCACCA ACAAAAAATT CAACCCAAAC CCCACCTCTC CCTTAAAAAC
            CTCCACCCAC CCTACCCTCA AACCAAAATA CCC                49024561

2. Chromosome 19 (CGB8)(len: 219)
  49049111                                      CCTCAACC CTCCTCTACT
            TAAAACTTTC CTACACCACA ATCCAACCTC CCAAAAAAAA CACAACTTCA
            AACTTAACTT CTACCCAATA AAAAAAAATC TCCCCATAAC TATACTACCA
            AAAAAACCAC TTAACCCTAA TACCCCCAAA AAAAAATTCA ACCCAAACCC
            CACCTCTCCC TTAAAAACCT CCACCCACTC TACCCTCAAA CCAAAATACC
            C                                                   49049329
3. Chromosome 19 (CGB7) (len: 220)
  49055747                     CC TCAACCCTCC TCTACTTAAA CCATTCCTAC
            ACCACAATCC AACCTAACAA AAAAAACACA ACTTCAAACT TAACTTCTAC
            CCAATAAAAA AAAATCTCCC CATAACTATA CTACCAAAAA AACCACTTAA
            CCCAAATACC CCCCAACAAA AAATTCAACC CAAACCCCAC CTCTCCCTTA
            AAAACCTCCA CCCACCCTAC CCTCAAACCA AAATACCC            49055966
```

**Fig.A13** ePCR Predicted PCR products for primer pair 7.

# A3. MSP Densitometry

**Table A3** MSP Densitometry data from ImageJ

| **Cell Line** | **HeLa (NT cancer)** | | **HEY (NT cancer)** | | **SKOV-3 (NT cancer)** | |
|---|---|---|---|---|---|---|
| **Primer** | u | m | u | m | u | m |
| **1** | 11299.96 | 16703.4 | 805.385 | 17103.69 | 1370.477 | 16540.45 |
| **2** | 9506.861 | 13705.45 | 917.234 | 14228.57 | 473.92 | 12383.86 |
| **3** | 8492.154 | 12351.86 | 755.406 | 13649.45 | 3093.205 | 16031.69 |
| **4** | 6208.548 | 9947.912 | 447.335 | 11001.15 | - | - |
| **5** | 5105.376 | 16496.05 | 787.213 | 14401.64 | - | - |
| **Cell Line** | **C-33a (NT cancer)** | | **CRL-1790 (control)** | | | |
| **Primer** | u | m | u | m | | |
| 1 | - | - | - | - | | |
| 2 | 3244.426 | 14456.33 | 2913.841 | 14332.91 | | |
| 3 | 3573.134 | 13804.98 | 2603.548 | 13724.98 | | |
| 4 | 2019.77 | 11796.28 | 1848.184 | 11460.15 | | |
| 5 | 3228.497 | 15594.93 | 775.749 | 14774.59 | | |

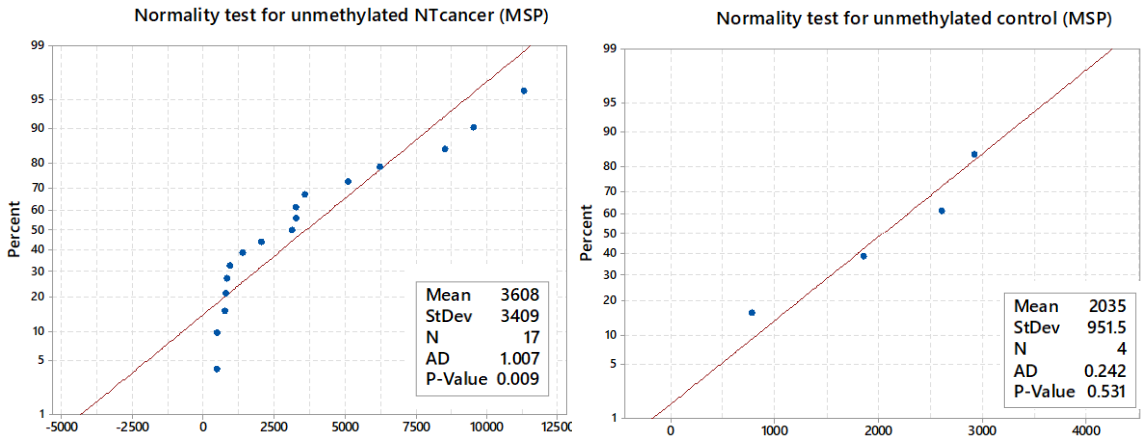



**Fig.A14** Plot for unmethylated densitometry data normality test (Anderson-Darling). Output from Minitab. (Left: non-trophoblastic cancer right: control)

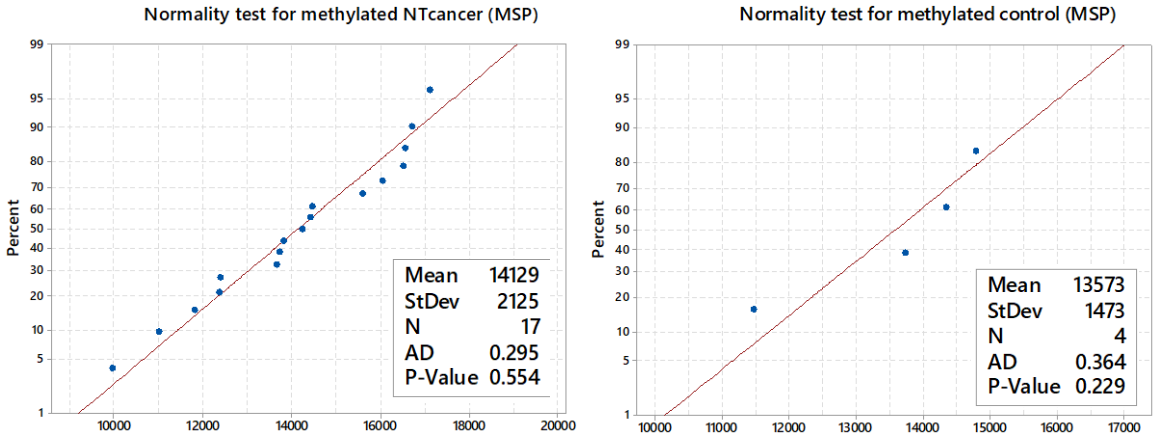



**Fig.A15** Plot for methylated densitometry data normality test (Anderson-Darling). Output from Minitab. (Left: non-trophoblastic cancer right: control)

## Mann-Whitney: unme c, unmectrl

### Method

$\eta_1$: median of unme c

$\eta_2$: median of unmectrl

Difference: $\eta_1 - \eta_2$

### Descriptive Statistics

| Sample | N | Median |
|---|---|---|
| unme c | 17 | 3093.20 |
| unmectrl | 4 | 2225.87 |

### Estimation for Difference

| Difference | CI for Difference | Achieved Confidence |
|---|---|---|
| 542.193 | (-1686.31, 5578.31) | 95.61% |

### Test

| | |
|---|---|
| Null hypothesis | $H_0: \eta_1 - \eta_2 = 0$ |
| Alternative hypothesis | $H_1: \eta_1 - \eta_2 \neq 0$ |

| W-Value | P-Value |
|---|---|
| 195.00 | 0.502 |

## Mann-Whitney: met c, met ctrl

### Method

$\eta_1$: median of met c

$\eta_2$: median of met ctrl

Difference: $\eta_1 - \eta_2$

### Descriptive Statistics

| Sample | N | Median |
|---|---|---|
| met c | 17 | 14228.6 |
| met ctrl | 4 | 14028.9 |

### Estimation for Difference

| Difference | CI for Difference | Achieved Confidence |
|---|---|---|
| 590.122 | (-1949.05, 2771.07) | 95.61% |

### Test

| | |
|---|---|
| Null hypothesis | $H_0: \eta_1 - \eta_2 = 0$ |
| Alternative hypothesis | $H_1: \eta_1 - \eta_2 \neq 0$ |

| W-Value | P-Value |
|---|---|
| 192.00 | 0.687 |

**Fig.A16** Output of Mann-Whitney test of difference from Minitab. (Left: unmethylated MSP data right: methylated MSP data) Legend: unme c: unmethylated NT cancer; unmectrl: unmethylated control; met c: methylated NT cancer; met ctrl: methylated control

# A4. Sequencing data

**Table A4** Received Beta-values Output from the MiSeq sequencing

| | | 3T3-mouse | BeWo | breast-DNA | C-33a | cervix DNA | Control-DNA | CRL-1790 | HCT116 | HeLA | HEY | HT-3 | JEG-3 | MCF-7 | MDMBA 468 | OAW42 | OVCAR-3 | SKOV-3 | Annotation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr12 | 52538990 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0 | NA | NA | NA | |
| chr12 | 52539026 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0 | NA | NA | NA | |
| chr19 | 49520764 | NA | NA | NA | NA | NA | NA | NA | 100 | NA | NA | 100 | NA | NA | NA | 0 | NA | NA | |
| chr19 | 49520772 | NA | NA | NA | NA | NA | NA | NA | 0 | NA | NA | 100 | NA | NA | NA | 100 | NA | NA | |
| chr19 | 49520808 | NA | NA | NA | NA | NA | NA | NA | 0 | NA | NA | 100 | NA | NA | NA | 0 | NA | NA | |
| chr19 | 49520867 | NA | NA | 0 | NA | 0 | NA | NA | 100 | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520895 | NA | NA | 100 | NA | NA | NA | NA | 100 | NA | NA | 100 | NA | NA | NA | 0 | NA | NA | |
| chr19 | 49520924 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520929 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520935 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520978 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520982 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49520991 | NA | NA | 0 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49521004 | NA | NA | 100 | NA | 0 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | 0 | NA | 100 | |
| chr19 | 49527657 | NA | 56 | 88 | 92 | 88 | 90 | 100 | 98 | 34 | 75 | 100 | 44 | 90 | 94 | 92 | 84 | 91 | CGB3-5' |
| chr19 | 49527659 | NA | 66 | 88 | 93 | 83 | 88 | 100 | 98 | 27 | 100 | 100 | 49 | 82 | 96 | 98 | 86 | 93 | CGB3-5' |
| chr19 | 49527665 | NA | 72 | 100 | 90 | 85 | 87 | 100 | 98 | 26 | 100 | 100 | 55 | 81 | 91 | 91 | 88 | 92 | CGB3-5' |
| chr19 | 49527701 | NA | 24 | 79 | 50 | 64 | 68 | 0 | 75 | 10 | 43 | 100 | 25 | 47 | 37 | 64 | 48 | 54 | CGB3-5' |
| chr19 | 49527740 | NA | 34 | 100 | 79 | 65 | 78 | 0 | 94 | 54 | 100 | NA | 21 | 48 | 17 | 97 | 74 | 89 | CGB3-5' |
| chr19 | 49527746 | NA | 30 | 33 | 49 | 24 | 41 | 17 | 82 | 36 | 57 | 0 | 12 | 16 | 13 | 75 | 51 | 68 | CGB3-5' |
| chr19 | 49527760 | NA | 20 | 56 | 46 | 51 | 59 | 50 | 89 | 9 | 100 | 100 | 21 | 39 | 23 | 74 | 45 | 55 | CGB3-5' |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr19 | 49527788 | NA | 69 | 89 | 96 | 96 | 98 | 75 | 97 | 88 | 100 | 0 | 68 | 92 | 94 | 95 | 90 | 93 | CGB3-5' |
| chr19 | 49527817 | NA | 60 | 100 | 84 | 65 | 78 | 50 | 93 | 32 | 100 | NA | 27 | 44 | 46 | 90 | 85 | 81 | CGB3-5' |
| chr19 | 49527829 | NA | 73 | 100 | 86 | 75 | 78 | 50 | 94 | 73 | 100 | NA | 58 | 69 | 72 | 84 | 98 | 93 | CGB3-5' |
| chr19 | 49527837 | NA | 58 | 100 | 86 | 74 | 100 | 100 | 96 | 66 | 100 | NA | 19 | 66 | 74 | 91 | 93 | 92 | CGB3-5' |
| chr19 | 49527876 | NA | 72 | 100 | 75 | 85 | 100 | 100 | 88 | 33 | 100 | NA | 43 | 68 | 75 | 86 | 90 | 79 | CGB3-5' |
| chr19 | 49527904 | NA | 87 | 0 | 64 | 82 | 100 | 50 | 82 | 64 | 100 | NA | 74 | 71 | 74 | 82 | 79 | 82 | CGB3-5' |
| chr19 | 49527912 | NA | 51 | NA | 54 | 65 | 50 | 100 | 73 | 41 | 100 | NA | 46 | 63 | 64 | 78 | 48 | 67 | CGB3-5' |
| chr19 | 49546787 | NA | 76 | 66 | 49 | 74 | 64 | 48 | 86 | 61 | 86 | 87 | 65 | 53 | 72 | 82 | 76 | 73 | CGB5 5' |
| chr19 | 49546793 | NA | 64 | 50 | 26 | 52 | 43 | 36 | 81 | 39 | 91 | 85 | 42 | 32 | 54 | 58 | 53 | 54 | CGB5 5' |
| chr19 | 49546815 | NA | 67 | 75 | 65 | 82 | 93 | 74 | 84 | 68 | 89 | 84 | 57 | 54 | 83 | 84 | 81 | 76 | CGB5 5' |
| chr19 | 49546854 | NA | 61 | 70 | 88 | 77 | 68 | 72 | 96 | 84 | 87 | 93 | 36 | 65 | 81 | 93 | 87 | 87 | CGB5 5' |
| chr19 | 49546862 | NA | 77 | 79 | 92 | 82 | 89 | 92 | 98 | 85 | 92 | 93 | 66 | 71 | 87 | 93 | 97 | 86 | CGB5 5' |
| chr19 | 49546869 | NA | 82 | 85 | 96 | 89 | 100 | 92 | 94 | 90 | 92 | 94 | 56 | 87 | 95 | 94 | 94 | 94 | CGB5 5' |
| chr19 | 49546874 | NA | 81 | 63 | 88 | 71 | 95 | 86 | 97 | 79 | 89 | 96 | 55 | 58 | 43 | 92 | 93 | 84 | CGB5 5' |
| chr19 | 49546903 | NA | 72 | 94 | 95 | 96 | 98 | 93 | 96 | 84 | 97 | 96 | 64 | 91 | 96 | 97 | 91 | 93 | CGB5 5' |
| chr19 | 49546931 | NA | 20 | 30 | 12 | 32 | 42 | 26 | 69 | 21 | 60 | 81 | 17 | 16 | 9 | 47 | 16 | 46 | CGB5 5' |
| chr19 | 49546945 | NA | 16 | 14 | 5 | 12 | 16 | 9 | 50 | 9 | 49 | 65 | 11 | 6 | 5 | 40 | 8 | 26 | CGB5 5' |
| chr19 | 49546991 | NA | 28 | 63 | 42 | 56 | 53 | 27 | 87 | 30 | 75 | 92 | 22 | 48 | 44 | 67 | 27 | 71 | CGB5 5' |
| chr19 | 49547027 | NA | 79 | 75 | 80 | 81 | 86 | 83 | 97 | 83 | 92 | 99 | 53 | 79 | 93 | 89 | 90 | 90 | CGB5 5' |
| chr19 | 49547033 | NA | 75 | 84 | 93 | 87 | 67 | 93 | 97 | 87 | 92 | 95 | 45 | 92 | 96 | 93 | 94 | 94 | CGB5 5' |
| chr19 | 49547035 | NA | 66 | 89 | 90 | 92 | 94 | 92 | 97 | 90 | 87 | 97 | 37 | 94 | 93 | 96 | 94 | 92 | CGB5 5' |
| chr19 | 49552425 | 100 | 62 | 90 | 91 | 93 | 92 | 90 | 96 | 90 | 96 | 98 | 56 | 93 | 95 | 94 | 90 | 93 | CGB8 5' |
| chr19 | 49552427 | 100 | 72 | 87 | 92 | 88 | 84 | 86 | 95 | 89 | 99 | 95 | 59 | 94 | 94 | 94 | 90 | 95 | CGB8 5' |
| chr19 | 49552433 | 100 | 75 | 82 | 90 | 83 | 88 | 64 | 92 | 87 | 97 | 97 | 72 | 88 | 82 | 92 | 93 | 91 | CGB8 5' |
| chr19 | 49552469 | 0 | 16 | 49 | 29 | 45 | 55 | 21 | 87 | 28 | 79 | 94 | 24 | 55 | 12 | 63 | 23 | 79 | CGB8 5' |
| chr19 | 49552528 | 100 | 18 | 28 | 23 | 38 | 44 | 10 | 89 | 18 | 53 | 80 | 22 | 30 | 14 | 59 | 18 | 53 | CGB8 5' |
| chr19 | 49552556 | 100 | 79 | 95 | 95 | 97 | 97 | 96 | 97 | 91 | 97 | 99 | 69 | 93 | 95 | 98 | 90 | 96 | CGB8 5' |
| chr19 | 49552585 | NA | 70 | 49 | 85 | 65 | 88 | 75 | 94 | 74 | 97 | 91 | 47 | 51 | 41 | 91 | 77 | 77 | CGB8 5' |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr19 | 49552597 | NA | 72 | 69 | 92 | 78 | 91 | 77 | 96 | 81 | 98 | 95 | 57 | 75 | 77 | 89 | 89 | 89 | CGB8 5' |
| chr19 | 49552605 | NA | 55 | 64 | 93 | 70 | 98 | 88 | 98 | 81 | 95 | 93 | 28 | 69 | 65 | 90 | 93 | 92 | CGB8 5' |
| chr19 | 49552644 | NA | 72 | 74 | 75 | 84 | 100 | 85 | 89 | 56 | 90 | 90 | 49 | 59 | 69 | 87 | 83 | 82 | CGB8 5' |
| chr19 | 49552666 | NA | 64 | 51 | 31 | 55 | 74 | 50 | 83 | 56 | 92 | 90 | 36 | 33 | 56 | 61 | 48 | 56 | CGB8 5' |
| chr19 | 49552672 | NA | 74 | 69 | 57 | 79 | 91 | 60 | 89 | 66 | 89 | 90 | 68 | 60 | 71 | 74 | 76 | 82 | CGB8 5' |
| chr19 | 49552680 | NA | 60 | 39 | 55 | 61 | 86 | 53 | 69 | 54 | 51 | 68 | 39 | 47 | 55 | 67 | 56 | 73 | CGB8 5' |
| chr19 | 49559061 | NA | 52 | 56 | 85 | 79 | 68 | 42 | 95 | 92 | 100 | 96 | 77 | 82 | 82 | 95 | 47 | 92 | CGB7 5' |
| chr19 | 49559063 | NA | 65 | 47 | 90 | 62 | 59 | 23 | 86 | 78 | 100 | 89 | 71 | 80 | 82 | 87 | 29 | 90 | CGB7 5' |
| chr19 | 49559069 | NA | 55 | 42 | 68 | 57 | 51 | 26 | 80 | 78 | 84 | 89 | 57 | 62 | 100 | 81 | 55 | 77 | CGB7 5' |
| chr19 | 49559105 | NA | 31 | 81 | 80 | 89 | 83 | 93 | 23 | 90 | 73 | 90 | 42 | 75 | 71 | 85 | 50 | 84 | CGB7 5' |
| chr19 | 49559151 | NA | 3 | 11 | 8 | 18 | 5 | 8 | 10 | 3 | 30 | 14 | 1 | 2 | 35 | 6 | 9 | 9 | CGB7 5' |
| chr19 | 49559165 | NA | 2 | 20 | 6 | 26 | 20 | 6 | 12 | 2 | 40 | 22 | 7 | 5 | 10 | 9 | 10 | 12 | CGB7 5' |
| chr19 | 49559193 | NA | 49 | 63 | 75 | 79 | 73 | 67 | 50 | 49 | 72 | 56 | 39 | 57 | 35 | 51 | 59 | 58 | CGB7 5' |
| chr19 | 49559222 | NA | 51 | 45 | 28 | 2 | 25 | 33 | 49 | 11 | 53 | 50 | 2 | 7 | 5 | 27 | 53 | 15 | CGB7 5' |
| chr19 | 49559227 | NA | 57 | 57 | 41 | 11 | 25 | 57 | 45 | 9 | 55 | 50 | 2 | 9 | 0 | 21 | 67 | 10 | CGB7 5' |
| chr19 | 49559234 | NA | 57 | 51 | 32 | 27 | 25 | 51 | 54 | 11 | 66 | 49 | 16 | 6 | 5 | 21 | 65 | 24 | CGB7 5' |
| chr19 | 49559242 | NA | 40 | 43 | 47 | 19 | 25 | 30 | 51 | 24 | 68 | 59 | 4 | 5 | 13 | 22 | 60 | 21 | CGB7 5' |
| chr19 | 49559281 | NA | 67 | 60 | 42 | 56 | 67 | 71 | 65 | 26 | 69 | 66 | 10 | 22 | 25 | 47 | 65 | 39 | CGB7 5' |
| chr19 | 49559301 | NA | 46 | 38 | 8 | 7 | 33 | 45 | 39 | 4 | 56 | 45 | 3 | 0 | 0 | 13 | 40 | 0 | CGB7 5' |
| chr19 | 49559303 | NA | 58 | 60 | 22 | 24 | 33 | 48 | 63 | 12 | 84 | 72 | 13 | 7 | 10 | 32 | 53 | 22 | CGB7 5' |
| chr19 | 49559309 | NA | 61 | 65 | 38 | 59 | 33 | 65 | 64 | 28 | 82 | 72 | 23 | 33 | 17 | 43 | 59 | 35 | CGB7 5' |
| chr19 | 49559317 | NA | 55 | 45 | 27 | 52 | 0 | 44 | 49 | 24 | 63 | 65 | 10 | 21 | 24 | 41 | 41 | 30 | CGB7 5' |
| chrM | 2333 | NA | NA | NA | NA | NA | 100 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | |
| chrM | 2345 | NA | NA | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | |
| chrM | 2477 | NA | NA | NA | NA | NA | 100 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | |
| chrM | 2492 | NA | NA | NA | NA | NA | 100 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | |

**Table A5** Calculated M-values based on the Beta-values obtained from sequencing

| gene | CpG | BeWo | breast-DNA | C-33a | cervix-DNA | Control-DNA | CRL-1790 | HCT116 | HeLA | HEY | HT-3 | JEG-3 | MCF-7 | MDA-MB468 | OAW42 | OVCAR-3 | SKOV-3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CGB3** | **a11** | 0.332575 | | 3.603341 | 2.904966 | 3.246074 | | 5.745955 | -0.98682 | | | -0.3322 | 3.190864 | 3.926882 | 3.59856 | 2.414046 | 3.415655 |
| **CGB3** | **a10** | 0.934297 | | 3.662206 | 2.311732 | 2.861803 | | 5.915879 | -1.44144 | | | -0.07985 | 2.205749 | 4.476814 | 5.559695 | 2.567475 | 3.831081 |
| **CGB3** | **a9** | 1.36257 | | 3.148183 | 2.478631 | 2.75619 | | 5.666351 | -1.47349 | | | 0.296393 | 2.1165 | 3.288125 | 3.29799 | 2.929064 | 3.510461 |
| **CGB3** | **a8** | -1.66647 | | -0.02223 | 0.839823 | 1.087463 | | 1.613817 | -3.10949 | | | -1.6172 | -0.19605 | -0.79558 | 0.810428 | -0.10516 | 0.235813 |
| **CGB3** | **d1** | -0.95553 | | 1.923152 | 0.896765 | 1.807355 | | 4.07227 | 0.232817 | | | -1.94662 | -0.10039 | -2.27599 | 4.897043 | 1.546374 | 2.97837 |
| **CGB3** | **c1** | -1.25511 | | -0.04879 | -1.6443 | -0.52607 | | 2.180121 | -0.84643 | | | -2.9169 | -2.34845 | -2.76129 | 1.551942 | 0.070598 | 1.097429 |
| **CGB3** | **a7** | -1.9622 | | -0.23461 | 0.079999 | 0.554762 | | 3.030697 | -3.35614 | | | -1.90314 | -0.62758 | -1.77423 | 1.544645 | -0.2766 | 0.291531 |
| **CGB3** | **a6** | 1.173648 | | 4.533432 | 4.411631 | 5.300123 | | 5.110614 | 2.925269 | | | 1.109229 | 3.606658 | 4.087463 | 4.145488 | 3.172213 | 3.630354 |
| **CGB3** | **a5** | 0.573735 | | 2.423682 | 0.870787 | | | 3.658544 | -1.11171 | | | -1.46778 | -0.36463 | -0.22125 | 3.197217 | 2.525462 | 2.106622 |
| **CGB3** | **a4** | 1.450033 | | 2.668192 | 1.57414 | | | 4.058894 | 1.439006 | | | 0.454032 | 1.131438 | 1.384664 | 2.426597 | 5.930738 | 3.790638 |
| **CGB3** | **a3** | 0.478047 | | 2.660251 | 1.485427 | | | 4.542527 | 0.987289 | | | -2.10109 | 0.973926 | 1.506107 | 3.294304 | 3.741932 | 3.441598 |
| **CGB3** | **a2** | 1.378512 | | 1.566872 | 2.498548 | | | 2.873043 | -1.04495 | | | -0.40054 | 1.065802 | 1.569366 | 2.6734 | 3.215013 | 1.900867 |
| **CGB3** | **a1** | 2.717857 | | 0.828888 | 2.160647 | | | 2.233119 | 0.821663 | | | 1.485427 | 1.310223 | 1.475733 | 2.203284 | 1.876618 | 2.187627 |
| **CGB3** | **b1** | 0.068713 | | 0.227667 | 0.86755 | 0 | | 1.43024 | -0.54749 | | | -0.23704 | 0.769516 | 0.814444 | 1.815575 | -0.12102 | 1.025091 |
| **CGB5** | **a1** | 1.69883 | 0.926502 | -0.04891 | 1.501881 | | -0.09398 | 2.65966 | 0.672296 | 2.625835 | 2.70955 | 0.869744 | 0.177709 | 1.330308 | 2.160891 | 1.655352 | 1.44859 |
| **CGB5** | **e1** | 0.841935 | -0.013 | -1.54027 | 0.108934 | | -0.80735 | 2.050123 | -0.64618 | 3.321928 | 2.492844 | -0.48794 | -1.06782 | 0.202817 | 0.473371 | 0.187527 | 0.254719 |
| **CGB5** | **a2** | 1.043943 | 1.584963 | 0.910139 | 2.232331 | | 1.520572 | 2.38054 | 1.061076 | 2.946419 | 2.401716 | 0.397964 | 0.23012 | 2.302882 | 2.430181 | 2.088056 | 1.690045 |
| **CGB5** | **a3** | 0.666263 | 1.198451 | 2.936903 | 1.756804 | 1.115477 | 1.366881 | 4.756729 | 2.423542 | 2.686501 | 3.82905 | -0.83446 | 0.879231 | 2.133856 | 3.679237 | 2.746068 | 2.720477 |
| **CGB5** | **a4** | 1.73954 | 1.9245 | 3.59614 | 2.22037 | 3.087463 | 3.571867 | 5.394124 | 2.514321 | 3.455195 | 3.67013 | 0.972906 | 1.294488 | 2.735066 | 3.696608 | 5.08625 | 2.594947 |
| **CGB5** | **f1** | 2.233797 | 2.525462 | 4.61471 | 3.036423 | 6.629357 | 3.448053 | 3.97429 | 3.201634 | 3.503963 | 3.923379 | 0.339052 | 2.739472 | 4.264703 | 4.089435 | 4.059781 | 3.880111 |
| **CGB5** | **a5** | 2.08655 | 0.740168 | 2.8391 | 1.316122 | 4.169925 | 2.67184 | 5.118489 | 1.942691 | 2.944423 | 4.654177 | 0.266336 | 0.484909 | -0.42951 | 3.5869 | 3.736541 | 2.36257 |
| **CGB5** | **a6** | 1.349727 | 4.046142 | 4.115947 | 4.449561 | 5.892796 | 3.731281 | 4.753217 | 2.430634 | 4.807355 | 4.72792 | 0.840588 | 3.348728 | 4.448081 | 4.785329 | 3.315107 | 3.675565 |
| **CGB5** | **a7** | -1.97199 | -1.2387 | -2.86262 | -1.11038 | -0.48294 | -1.52202 | 1.167868 | -1.87883 | 0.565854 | 2.103138 | -2.32729 | -2.38355 | -3.38702 | -0.16381 | -2.41504 | -0.23792 |
| **CGB5** | **c1** | -2.42068 | -2.59205 | -4.15482 | -2.81493 | -2.37693 | -3.37541 | 0.018958 | -3.28113 | -0.06319 | 0.870549 | -2.94642 | -3.95675 | -4.21214 | -0.58187 | -3.42909 | -1.47926 |
| **CGB5** | **a8** | -1.36507 | 0.766887 | -0.4637 | 0.359972 | 0.149344 | -1.43624 | 2.718004 | -1.20224 | 1.5479 | 3.614092 | -1.80812 | -0.13618 | -0.33673 | 1.039813 | -1.41219 | 1.318271 |
| **CGB5** | **a9** | 1.868562 | 1.603341 | 2.008364 | 2.050912 | 2.616572 | 2.262495 | 4.863118 | 2.242201 | 3.453957 | 6.191903 | 0.198183 | 1.944633 | 3.729104 | 2.996593 | 3.101917 | 3.098989 |
| **CGB5** | **a10** | 1.5667 | 2.352714 | 3.63743 | 2.716991 | 1.052568 | 3.757513 | 5.150942 | 2.695808 | 3.54432 | 4.145524 | -0.31586 | 3.462343 | 4.624491 | 3.630843 | 3.987363 | 4.005294 |
| **CGB5** | **a11** | 0.966597 | 3.055444 | 3.126532 | 3.504231 | 4.043422 | 3.520909 | 5.148552 | 3.165465 | 2.77259 | 5.101037 | -0.76447 | 3.859402 | 3.751544 | 4.539733 | 4.100137 | 3.542527 |
| **CGB8** | **a11** | 0.709704 | 3.184104 | 3.40387 | 3.665336 | 3.489966 | 3.160177 | 4.729104 | 3.091922 | 4.672425 | 5.296617 | 0.345028 | 3.754887 | 4.258734 | 4.025535 | 3.14505 | 3.815415 |
| **CGB8** | **a10** | 1.375509 | 2.767728 | 3.600787 | 2.941701 | 2.408876 | 2.617694 | 4.380424 | 3.089694 | 6.295642 | 4.16501 | 0.496153 | 4.046693 | 4.053111 | 3.921246 | 3.14505 | 4.220409 |
| **CGB8** | **a9** | 1.568488 | 2.158189 | 3.162939 | 2.301626 | 2.812563 | 0.829751 | 3.570655 | 2.776767 | 4.833574 | 4.816984 | 1.397631 | 2.836732 | 2.223454 | 3.59547 | 3.727095 | 3.367425 |
| **CGB8** | **a8** | -2.39169 | -0.07326 | -1.25899 | -0.27669 | 0.274251 | -1.89707 | 2.706554 | -1.3505 | 1.906562 | 4.017922 | -1.6429 | 0.316857 | -2.85137 | 0.787173 | -1.74974 | 1.920128 |
| **CGB8** | **a7** | -2.22447 | -1.36234 | -1.74291 | -0.73647 | -0.32269 | -3.17734 | 3.086316 | -2.18237 | 0.162615 | 2.007016 | -1.85114 | -1.18903 | -2.58496 | 0.506369 | -2.17688 | 0.162166 |
| **CGB8** | **a6** | 1.937587 | 4.205114 | 4.34685 | 5.147607 | 4.942983 | 4.470027 | 5.061549 | 3.277766 | 5.013984 | 6.576095 | 1.181256 | 3.767554 | 4.219814 | 5.501252 | 3.224205 | 4.599913 |
| **CGB8** | **a5** | 1.222392 | -0.07013 | 2.522887 | 0.882867 | 2.925999 | 1.621735 | 3.968951 | 1.487468 | 4.925885 | 3.265345 | -0.19071 | 0.031896 | -0.53219 | 3.426797 | 1.747019 | 1.754325 |
| **CGB8** | **a4** | 1.368445 | 1.139823 | 3.546422 | 1.852193 | 3.285402 | 1.724597 | 4.729404 | 2.083017 | 5.29749 | 4.366782 | 0.389771 | 1.577831 | 1.719794 | 3.08801 | 3.087463 | 3.039966 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CGB8** | **a3** | 0.283426 | 0.835597 | 3.793465 | 1.215013 | 5.392318 | 2.885252 | 5.590961 | 2.049244 | 4.258136 | 3.673634 | -1.39818 | 1.134961 | 0.902878 | 3.232661 | 3.758276 | 3.449611 |
| **CGB8** | **a2** | 1.369695 | 1.487977 | 1.603857 | 2.359081 | 6.629357 | 2.488101 | 2.97537 | 0.343853 | 3.097697 | 3.095157 | -0.04108 | 0.541709 | 1.174874 | 2.767091 | 2.274941 | 2.203872 |
| **CGB8** | **e1** | 0.841818 | 0.061001 | -1.17284 | 0.292409 | 1.5025 | -0.01132 | 2.271655 | 0.374589 | 3.485846 | 3.111508 | -0.8225 | -0.99165 | 0.345135 | 0.616111 | -0.09558 | 0.350907 |
| **CGB8** | **a1** | 1.49526 | 1.135437 | 0.41622 | 1.901682 | 3.321928 | 0.571842 | 3.070389 | 0.97777 | 3.07298 | 3.200952 | 1.114216 | 0.578703 | 1.326756 | 1.495186 | 1.693684 | 2.20014 |
| **CGB8** | **b1** | 0.560715 | -0.67479 | 0.304087 | 0.663475 | 2.584963 | 0.1676 | 1.162271 | 0.238288 | 0.046867 | 1.092849 | -0.62149 | -0.15347 | 0.279617 | 1.045229 | 0.36759 | 1.402098 |
| **CGB7** | **a11** | 0.106915 | 0.318822 | 2.459432 | 1.944657 | 1.061776 | -0.46467 | 4.386059 | 3.449307 | 6.629357 | 4.566815 | 1.731183 | 2.174498 | 2.169925 | 4.380822 | -0.16992 | 3.551796 |
| **CGB7** | **a10** | 0.881356 | -0.152 | 3.147307 | 0.719892 | 0.545968 | -1.76867 | 2.656046 | 1.8314 | 6.629357 | 3.013297 | 1.31259 | 1.986061 | 2.169925 | 2.697265 | -1.26303 | 3.132156 |
| **CGB7** | **a9** | 0.263034 | -0.46042 | 1.06114 | 0.39752 | 0.08017 | -1.50546 | 2.034646 | 1.805272 | 2.432959 | 2.972693 | 0.393342 | 0.693424 | 6.629357 | 2.09818 | 0.263034 | 1.733756 |
| **CGB7** | **a8** | -1.12801 | 2.124328 | 2.02526 | 3.053298 | 2.307428 | 3.767339 | -1.70974 | 3.206029 | 1.447459 | 3.187244 | -0.455 | 1.57447 | 1.321928 | 2.476586 | 0 | 2.357279 |
| **CGB7** | **c1** | -4.88753 | -3.07621 | -3.53325 | -2.23704 | -4.30743 | -3.52788 | -3.12338 | -5.18587 | -1.22239 | -2.56906 | -6.84549 | -5.50885 | -0.89308 | -3.92539 | -3.33298 | -3.41504 |
| **CGB7** | **a7** | -5.94642 | -2.04439 | -3.90138 | -1.49513 | -2.01436 | -3.89308 | -2.88408 | -5.47168 | -0.58496 | -1.80453 | -3.83008 | -4.37504 | -3.22239 | -3.39232 | -3.16993 | -2.84874 |
| **CGB7** | **a6** | -0.0813 | 0.754888 | 1.584963 | 1.86994 | 1.431951 | 0.993367 | -0.01355 | -0.03395 | 1.353637 | 0.354843 | -0.62344 | 0.424885 | -0.90689 | 0.071553 | 0.530515 | 0.454032 |
| **CGB7** | **a5** | 0.054448 | -0.27357 | -1.34958 | -5.9542 | | -1.04975 | -0.05772 | -3 | 0.167295 | 0 | -5.78136 | -3.66297 | -4.16992 | -1.4458 | 0.195016 | -2.48543 |
| **CGB7** | **f1** | 0.383329 | 0.431716 | -0.52793 | -3.05889 | | 0.380031 | -0.28951 | -3.36923 | 0.311944 | 0.015942 | -5.67243 | -3.38247 | -6.62936 | -1.89308 | 1.013806 | -3.11548 |
| **CGB7** | **a4** | 0.383329 | 0.08573 | -1.09954 | -1.4361 | | 0.033553 | 0.226771 | -3.01495 | 0.958421 | -0.03206 | -2.35364 | -4 | -4.16992 | -1.86876 | 0.920566 | -1.64386 |
| **CGB7** | **a3** | -0.56188 | -0.40663 | -0.16046 | -2.05889 | | -1.23447 | 0.057715 | -1.65711 | 1.110054 | 0.551642 | -4.75489 | -4.2854 | -2.72247 | -1.79837 | 0.606989 | -1.89308 |
| **CGB7** | **a2** | 1 | 0.593375 | -0.47207 | 0.353637 | | 1.26628 | 0.91427 | -1.5025 | 1.164387 | 0.983512 | -3.16993 | -1.8413 | -1.58496 | -0.16046 | 0.906891 | -0.63743 |
| **CGB7** | **g1** | -0.22239 | -0.72085 | -3.5025 | -3.66297 | | -0.27563 | -0.66558 | -4.45943 | 0.3254 | -0.28728 | -5.2854 | -6.62936 | -6.62936 | -2.80735 | -0.58496 | -6.62936 |
| **CGB7** | **e1** | 0.459432 | 0.576501 | -1.80735 | -1.63227 | | -0.1375 | 0.747234 | -2.93074 | 2.39689 | 1.345135 | -2.80735 | -3.67243 | -3.24793 | -1.09115 | 0.17585 | -1.848 |
| **CGB7** | **a1** | 0.616671 | 0.863938 | -0.70454 | 0.4975 | | 0.880418 | 0.854149 | -1.39593 | 2.212994 | 1.394279 | -1.78427 | -1.03953 | -2.32193 | -0.37851 | 0.532874 | -0.90689 |
| **CGB7** | **b1** | 0.289507 | -0.27009 | -1.44746 | 0.106915 | | -0.3388 | -0.074 | -1.62803 | 0.792195 | 0.869939 | -3.24793 | -1.89308 | -1.67807 | -0.53051 | -0.53051 | -1.22239 |

# A5. R statistics

## A5.1 MethylKit analysis

Based on the provided beta values, a text file was created for each cell line to be fed in the program for analysis. Each file contained the following columns: chrBase chr base strand coverage freqC freqT. The first 4 column provided information on location, coverage was the amount of reads and the freqC and freqT provide information on the methylation at that location.

```
library(methylKit)
file.list=list( system.file("dnaproject","OAW42.txt", package = "methylKit"),
                system.file("dnaproject","bewo.txt", package = "methylKit"),
                system.file("dnaproject","c33a.txt", package = "methylKit"),
                system.file("dnaproject","HCT116.txt", package = "methylKit"),
                system.file("dnaproject","HeLa.txt", package = "methylKit"),
                system.file("dnaproject","hey.txt", package = "methylKit"),
                system.file("dnaproject","HT3.txt", package = "methylKit"),
                system.file("dnaproject","JEG-3.txt", package = "methylKit"),
                system.file("dnaproject","MCF-7.txt", package = "methylKit"),
                system.file("dnaproject","MDA468.txt", package = "methylKit"),
                system.file("dnaproject","OVCAR-3.txt", package = "methylKit"),
                system.file("dnaproject","SKOV-3.txt", package = "methylKit"),
                system.file("dnaproject","breast.txt", package = "methylKit"),
                system.file("dnaproject","cervixDNA.txt", package = "methylKit"),
               system.file("dnaproject","ControlDNA.txt", package = "methylKit"),
               system.file("dnaproject","CRL1790.txt", package = "methylKit") )

myobj=methRead(file.list,
           sample.id=list(
                           "OAW42",
                           "BeWo",
                           "C33a",
                           "HCT116",
                           "HeLa",
                           "HEY",
                           "HT3",
                           "JEG-3",
                           "MCF-7",
                           "MDA468",
                           "OVCAR-3",
                           "SKOV-3",
                           "BREAST-DNA",
                           "CERVIX-DNA",
                           "CONTROL-DNA",
                           "CRL1790"
                           ),
           assembly="hg18",
           treatment=c(
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   1,
                                   0,
                                   0,
                                   0,
                                   0
                           ),
           context="CpG"
                           )
>meth=unite(myobj, destrand=FALSE)

>getCorrelation(meth,plot=TRUE)

>clusterSamples(meth, dist="correlation", method="ward", plot=TRUE)
```

**Fig.A17** Code in R to be fed in methylKit to produce the cluster dendrogram and CpG correlation figures.

# A5.2 PCA

```
 db<-read.csv("~/b-values3.csv")
 pca <- prcomp(t(db<- as.matrix), scale=TRUE)
 plot(pca$x[,1], pca$x[,2])
 pca.var <- pca$sdev^2
 pca.var.per <- round(pca.var/sum(pca.var)*100, 1)
 barplot(pca.var.per, main="Scree Plot", xlab="Principal Component", ylab="Percent Variati
on")

 library(ggplot2)
  pca.data <- data.frame(Sample=rownames(pca$x),
                                                          X=pca$x[,1],
                                                          Y=pca$x[,2])
pca.data
ggplot(data=pca.data, aes(x=X, y=Y, label=Sample)) +
           geom_text() +
            xlab(paste("PC1 - ", pca.var.per[1], "%", sep="")) +
            ylab(paste("PC2 - ", pca.var.per[2], "%", sep="")) +
            theme_bw() +
            ggtitle("My PCA Graph")
```

**Fig.A18** Code in R to produce the Scree pot and PCA plot. Code adapted from https://github.com/StatQuest/pca_demo/blob/master/pca_demo.R

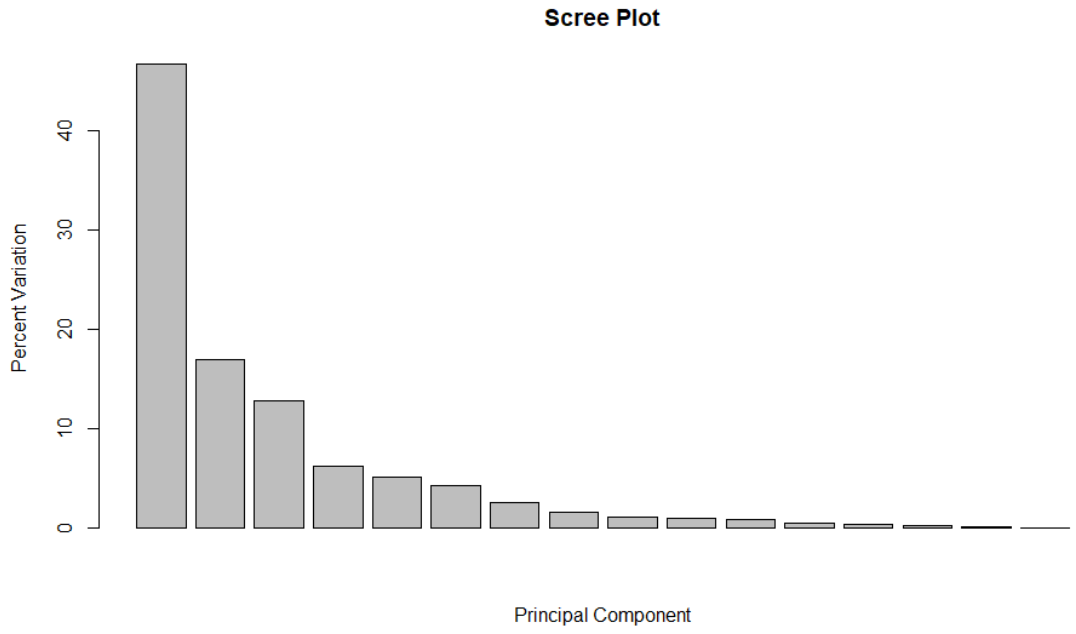



**Fig.A19** Scree plot based on the methylation data used in PCA. Produced with R

# A6. Analyses of difference

## A6.1 Average Promoter methylation

| Normality and Lognormality Tests<br>Tabular results | A<br>CGB3 | B<br>CGB5 | C<br>CGB8 | D<br>CGB7 |
|---|---|---|---|---|
| **Test for normal distribution** | | | | |
| **Anderson-Darling test** | | | | |
| A2* | 0.1859 | 0.2728 | 0.4926 | 0.3080 |
| P value | 0.8833 | 0.6188 | 0.1863 | 0.5227 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes |
| P value summary | ns | ns | ns | ns |
| | | | | |
| **Number of values** | 12 | 16 | 16 | 16 |

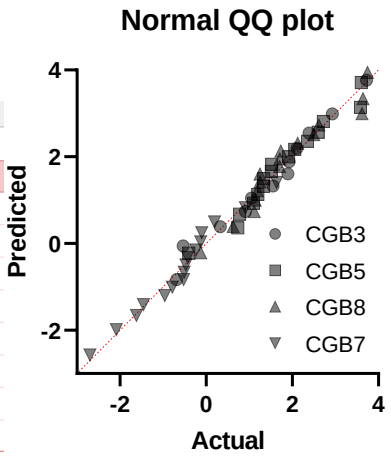



**Fig.A20** Normality test of average promoter methylation. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| ANOVA table | SS | DF | MS | F (DFn, DFd) | P value |
|---|---|---|---|---|---|
| Treatment (between columns) | 62.88 | 3 | 20.96 | F (3, 56) = 16 | P<0.0001 |
| Residual (within columns) | 71.83 | 56 | 1.283 | | |
| Total | 134.7 | 59 | | | |

**Fig.A21** Ordinary one-way ANOVA output of the average promoter methylation by gene. Produced by Prism8

ANOVA results | Multiple comparisons

| Ordinary one-way ANOVA<br>Multiple comparisons | | | | | | |
|---|---|---|---|---|---|---|
| Number of families | 1 | | | | | |
| Number of comparisons per family | 6 | | | | | |
| Alpha | 0.05 | | | | | |
| | | | | | | |
| **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** | |
| CGB3 vs. CGB5 | -0.2813 | -1.427 to 0.8639 | No | ns | 0.9149 | A-B |
| CGB3 vs. CGB8 | -0.3990 | -1.544 to 0.7462 | No | ns | 0.7929 | A-C |
| CGB3 vs. CGB7 | 2.047 | 0.9016 to 3.192 | Yes | **** | <0.0001 | A-D |
| CGB5 vs. CGB8 | -0.1176 | -1.178 to 0.9426 | No | ns | 0.9911 | B-C |
| CGB5 vs. CGB7 | 2.328 | 1.268 to 3.388 | Yes | **** | <0.0001 | B-D |
| CGB8 vs. CGB7 | 2.446 | 1.386 to 3.506 | Yes | **** | <0.0001 | C-D |

**Fig.A22** Ordinary one-way ANOVA Multiple comparisons output of the average promoter methylation by gene (Tukey's test). Produced by Prism8

# A6.2 Average promoter methylation vs cancer type

| | Kruskal-Wallis test<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | Average methylation of the promoter |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.0404 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | * |
| 7 | Do the medians vary signif. (P < 0.05)? | Yes |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 5.698 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 16 |

**Fig.A23** Analysis of difference between the average methylation of *CGB3, 5 &8* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

| | Kruskal-Wallis test<br>Multiple comparisons | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | | |
| 2 | Number of comparisons per family | 3 | | | | | |
| 3 | Alpha | 0.05 | | | | | |
| 4 | | | | | | | |
| 5 | **Dunn's multiple comparisons test** | **Mean rank diff.** | **Significant?** | **Summary** | **Adjusted P Value** | | |
| 6 | trophoblastic cell lines vs. non-trophoblastic cell lines | -8.700 | No | ns | 0.0550 | A-B | |
| 7 | trophoblastic cell lines vs. control samples DNA | -6.250 | No | ns | 0.3887 | A-C | |
| 8 | non-trophoblastic cell lines vs. control samples DNA | 2.450 | No | ns | >0.9999 | B-C | |
| 9 | | | | | | | |
| 10 | **Test details** | **Mean rank 1** | **Mean rank 2** | **Mean rank diff.** | **n1** | **n2** | **Z** |
| 11 | trophoblastic cell lines vs. non-trophoblastic cell lines | 1.500 | 10.20 | -8.700 | 2 | 10 | 2.359 |
| 12 | trophoblastic cell lines vs. control samples DNA | 1.500 | 7.750 | -6.250 | 2 | 4 | 1.516 |
| 13 | non-trophoblastic cell lines vs. control samples DNA | 10.20 | 7.750 | 2.450 | 10 | 4 | 0.8698 |

**Fig.A24** Multiple comparisons between the average methylation of *CGB3, 5 &8* in the 3 groups of cell line cancer type (T, NT, and Control). Dunn's test (Kruskal-Wallis test) on Prism8

| | Kruskal-Wallis test<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | AP2-SP-AP2-Oct3/4 region |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.3759 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | ns |
| 7 | Do the medians vary signif. (P < 0.05)? | No |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 2.144 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 16 |

**Fig.A25** Analysis of difference between the average 5' promoter region methylation of *CGB3, 5 &8* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

| | Kruskal-Wallis test<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | AP2-SP1 region |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.0634 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | ns |
| 7 | Do the medians vary signif. (P < 0.05)? | No |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 5.080 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 16 |

**Fig.A26** Analysis of difference between the average 3' promoter region methylation of *CGB3, 5 &8* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

| | Kruskal-Wallis test<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | average promoter CGB7 |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.3428 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | ns |
| 7 | Do the medians vary signif. (P < 0.05)? | No |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 2.345 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 16 |

**Fig.A27** Analysis of difference between the average promoter methylation of *CGB7* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

| | Kruskal-Wallis test<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | AP2-SP1-AP"-OCT CGB |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.8521 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | ns |
| 7 | Do the medians vary signif. (P < 0.05)? | No |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 0.3750 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 15 |

**Fig.A28** Analysis of difference between the average 5' promoter region methylation of *CGB7* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

| | Kruskal-Wallis test | |
|---|---|---|
| 1 | Table Analyzed | AP2-SP1 cgb7 |
| 2 | | |
| 3 | **Kruskal-Wallis test** | |
| 4 | P value | 0.4372 |
| 5 | Exact or approximate P value? | Exact |
| 6 | P value summary | ns |
| 7 | Do the medians vary signif. (P < 0.05)? | No |
| 8 | Number of groups | 3 |
| 9 | Kruskal-Wallis statistic | 1.815 |
| 10 | | |
| 11 | **Data summary** | |
| 12 | Number of treatments (columns) | 3 |
| 13 | Number of values (total) | 16 |

**Fig.A29** Analysis of difference between the average 3' promoter region methylation of *CGB7* in the 3 groups of cell line cancer type (T, NT, and Control). Kruskal-Wallis Test on Prism8

# A6.3 Average promoter methylation vs cell line origin

## A6.3.1 Choriocarcinoma cell lines

| Normality and Lognormality Tests<br>Tabular results | A<br>BeWo | B<br>JEG-3 | C<br>Control-DNA |
|---|---|---|---|
| | | | |
| **Test for normal distribution** | | | |
| **Anderson-Darling test** | | | |
| A2* | 0.9541 | 0.3285 | 0.2604 |
| P value | 0.0118 | 0.4810 | 0.6615 |
| Passed normality test (alpha=0.05)? | No | Yes | Yes |
| P value summary | * | ns | ns |
| | | | |
| **Number of values** | 16 | 16 | 16 |

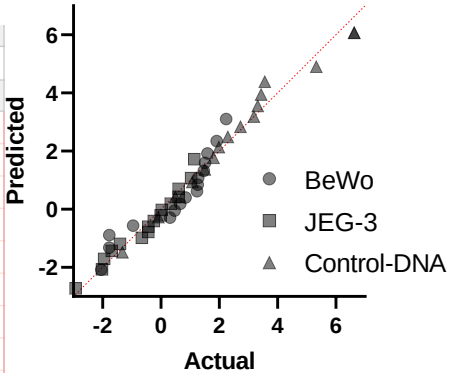



**Fig.A30** Normality test of average promoter methylation for *CGB3, 5 & 8* in choriocarcinoma group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | | |
|---|---|---|
| 1 | Table Analyzed | chorio |
| 2 | | |
| 3 | **Repeated measures ANOVA summary** | |
| 4 | Assume sphericity? | No |
| 5 | F | 51.56 |
| 6 | P value | <0.0001 |
| 7 | P value summary | **** |
| 8 | Statistically significant (P < 0.05)? | Yes |
| 9 | Geisser-Greenhouse's epsilon | 0.6067 |
| 10 | R squared | 0.7746 |

**Fig.A31** RM one-way ANOVA output of the average promoter methylation for *CGB3, 5 & 8* in choriocarcinoma group. Produced by Prism8

| RM one-way ANOVA<br>Multiple comparisons | | | | | | |
|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | |
| 2 | Number of comparisons per family | 3 | | | | |
| 3 | Alpha | 0.05 | | | | |
| 4 | | | | | | |
| 5 | **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| 6 | BeWo vs. JEG-3 | 1.015 | 0.6352 to 1.395 | Yes | **** | <0.0001 |
| 7 | BeWo vs. Control-DNA | -2.145 | -3.069 to -1.222 | Yes | **** | <0.0001 |
| 8 | JEG-3 vs. Control-DNA | -3.160 | -4.183 to -2.137 | Yes | **** | <0.0001 |

**Fig.A32** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB3, 5 & 8* in choriocarcinoma group (Tukey's test). Produced by Prism8

## A6.3.2 Ovarian cell lines

| Normality and Lognormality Tests<br>Tabular results | A<br>HEY | B<br>OAW42 | C<br>OVCAR-3 | D<br>SKOV-3 | E<br>Control-DNA |
|---|---|---|---|---|---|
| **Test for normal distribution** | | | | | |
| **Anderson-Darling test** | | | | | |
| A2* | 1.438 | 0.4895 | 0.7273 | 0.4963 | 0.2604 |
| P value | 0.0007 | 0.1898 | 0.0460 | 0.1822 | 0.6615 |
| Passed normality test (alpha=0.05)? | No | Yes | No | Yes | Yes |
| P value summary | *** | ns | * | ns | ns |
| **Number of values** | 16 | 16 | 16 | 16 | 16 |

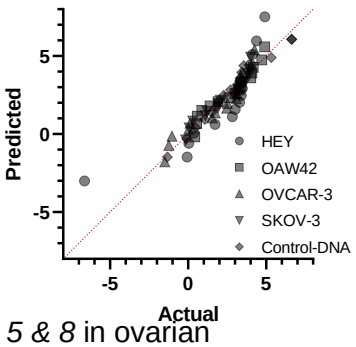



**Fig.A33** Normality test of average promoter methylation for *CGB3, 5 & 8* in ovarian group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | ovarian |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 1.030 |
| P value | 0.3620 |
| P value summary | ns |
| Statistically significant (P < 0.05)? | No |
| Geisser-Greenhouse's epsilon | 0.4378 |
| R squared | 0.06427 |

**Fig.A34** RM one-way ANOVA output of the average promoter methylation for *CGB3, 5 & 8* in choriocarcinoma group. Produced by Prism8

## A6.3.3 Control group

| Normality and Lognormality Tests<br>Tabular results | A<br>breast-DNA | B<br>cervix-DNA | C<br>CRL1790 | D<br>Control-DNA |
|---|---|---|---|---|
| **Test for normal distribution** | | | | |
| **Anderson-Darling test** | | | | |
| A2* | 0.5651 | 0.1468 | 0.5447 | 0.2604 |
| P value | 0.1193 | 0.9565 | 0.1355 | 0.6615 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes |
| P value summary | ns | ns | ns | ns |
| **Number of values** | 16 | 16 | 16 | 16 |

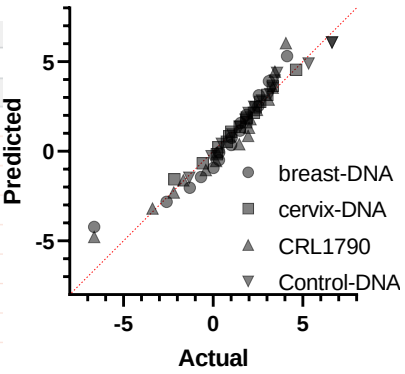



**Fig.A35** Normality test of average promoter methylation for *CGB3, 5 & 8* in control group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

RM one-way ANOVA
ANOVA results

| | | |
|---|---|---|
| 1 | Table Analyzed | control |
| 2 | | |
| 3 | **Repeated measures ANOVA summary** | |
| 4 | Assume sphericity? | No |
| 5 | F | 9.634 |
| 6 | P value | 0.0014 |
| 7 | P value summary | ** |
| 8 | Statistically significant (P < 0.05)? | Yes |
| 9 | Geisser-Greenhouse's epsilon | 0.5525 |
| 10 | R squared | 0.3911 |

**Fig.A36** RM one-way ANOVA output of the average promoter methylation for *CGB3, 5 & 8* in control group. Produced by Prism8

RM one-way ANOVA
Multiple comparisons

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | |
| 2 | Number of comparisons per family | 6 | | | | |
| 3 | Alpha | 0.05 | | | | |
| 4 | | | | | | |
| 5 | **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| 6 | breast-DNA vs. cervix-DNA | -0.9464 | -2.236 to 0.3429 | No | ns | 0.1928 |
| 7 | breast-DNA vs. CRL1790 | -0.07971 | -0.7591 to 0.5997 | No | ns | 0.9862 |
| 8 | breast-DNA vs. Control-DNA | -2.113 | -3.702 to -0.5236 | Yes | ** | 0.0079 |
| 9 | cervix-DNA vs. CRL1790 | 0.8667 | -0.5643 to 2.298 | No | ns | 0.3360 |
| 10 | cervix-DNA vs. Control-DNA | -1.167 | -2.089 to -0.2444 | Yes | * | 0.0114 |
| 11 | CRL1790 vs. Control-DNA | -2.033 | -3.603 to -0.4632 | Yes | ** | 0.0096 |

**Fig.A37** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB3, 5 & 8* in control group (Tukey's test). Produced by Prism8

## A6.3.4 Cervical group

Normality and Lognormality Tests
Tabular results

| | | A | B | C | D |
|---|---|---|---|---|---|
| | | HT3 | C33a | HeLA | cervix-DNA |
| 1 | **Test for normal distribution** | | | | |
| 2 | **Anderson-Darling test** | | | | |
| 3 | A2* | 1.467 | 0.4873 | 0.2728 | 0.1468 |
| 4 | P value | 0.0006 | 0.1923 | 0.6188 | 0.9565 |
| 5 | Passed normality test (alpha=0.05)? | No | Yes | Yes | Yes |
| 6 | P value summary | *** | ns | ns | ns |
| 7 | | | | | |
| 8 | **Number of values** | 16 | 16 | 16 | 16 |

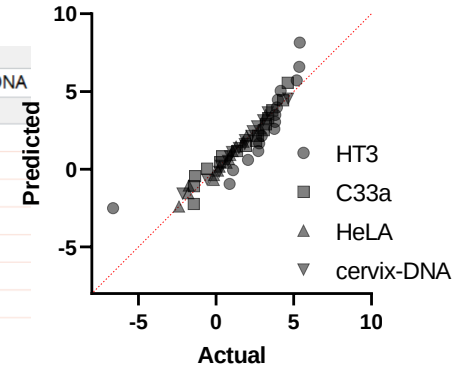



**Fig.A38** Normality test of average promoter methylation for *CGB3, 5 & 8* in cervical group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | cervical |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 6.544 |
| P value | 0.0125 |
| P value summary | * |
| Statistically significant (P < 0.05)? | Yes |
| Geisser-Greenhouse's epsilon | 0.4485 |
| R squared | 0.3037 |

**Fig.A39** RM one-way ANOVA output of the average promoter methylation for *CGB3, 5 & 8* in cervical group. Produced by Prism8

| RM one-way ANOVA<br>Multiple comparisons | | | | | |
|---|---|---|---|---|---|
| Number of families | 1 | | | | |
| Number of comparisons per family | 6 | | | | |
| Alpha | 0.05 | | | | |
| | | | | | |
| **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| HT3 vs. C33a | 1.157 | -0.9566 to 3.271 | No | ns | 0.4195 |
| HT3 vs. HeLA | 2.280 | 0.3113 to 4.249 | Yes | * | 0.0209 |
| HT3 vs. cervix-DNA | 1.330 | -0.5197 to 3.180 | No | ns | 0.2066 |
| C33a vs. HeLA | 1.122 | 0.4354 to 1.810 | Yes | ** | 0.0014 |
| C33a vs. cervix-DNA | 0.1727 | -0.6375 to 0.9829 | No | ns | 0.9259 |
| HeLA vs. cervix-DNA | -0.9498 | -1.607 to -0.2925 | Yes | ** | 0.0041 |

**Fig.A40** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB3, 5 & 8* in cervical group (Tukey's test). Produced by Prism8

## A6.3.5 Breast group

| Normality and Lognormality Tests<br>Tabular results | A<br>MCF-7 | B<br>MDA468 | C<br>breast-DNA |
|---|---|---|---|
| **Test for normal distribution** | | | |
| **Anderson-Darling test** | | | |
| A2* | 0.2635 | 0.3135 | 0.5651 |
| P value | 0.6506 | 0.5131 | 0.1193 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes |
| P value summary | ns | ns | ns |
| | | | |
| **Number of values** | 16 | 16 | 16 |

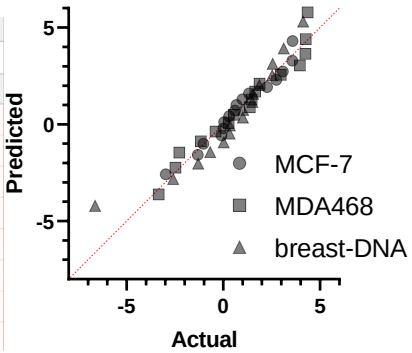



**Fig.A41** Normality test of average promoter methylation for *CGB3, 5 & 8* in breast group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| | RM one-way ANOVA<br>ANOVA results | |
|---|---|---|
| 1 | Table Analyzed | breast |
| 2 | | |
| 3 | **Repeated measures ANOVA summary** | |
| 4 | Assume sphericity? | No |
| 5 | F | 1.154 |
| 6 | P value | 0.3194 |
| 7 | P value summary | ns |
| 8 | Statistically significant (P < 0.05)? | No |
| 9 | Geisser-Greenhouse's epsilon | 0.7571 |
| 10 | R squared | 0.07142 |

**Fig.A42** RM one-way ANOVA output of the average promoter methylation for *CGB3, 5 & 8* in breast group. Produced by Prism8

## A6.3.6 Colon group

| | Normality and Lognormality Tests<br>Tabular results | A<br>CRL1790 | B<br>HCT116 |
|---|---|---|---|
| 1 | **Test for normal distribution** | | |
| 2 | **Anderson-Darling test** | | |
| 3 | A2* | 0.5289 | 0.6057 |
| 4 | P value | 0.1611 | 0.1081 |
| 5 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 6 | P value summary | ns | ns |
| 7 | | | |
| 8 | **Number of values** | 27 | 41 |

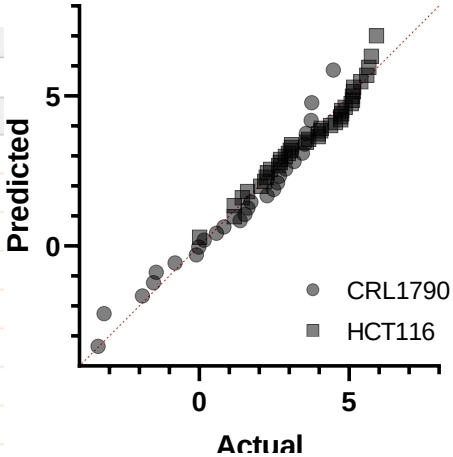



**Fig.A43** Normality test of average promoter methylation for *CGB3, 5 & 8* in colon group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| | | |
|---|---|---|
| 1 | Table Analyzed | colon |
| 2 | | |
| 3 | Column B | HCT116 |
| 4 | vs. | vs. |
| 5 | Column A | CRL1790 |
| 6 | | |
| 7 | **Paired t test** | |
| 8 | P value | <0.0001 |
| 9 | P value summary | **** |
| 10 | Significantly different (P < 0.05)? | Yes |
| 11 | One- or two-tailed P value? | Two-tailed |
| 12 | t, df | t=9.217, df=26 |
| 13 | Number of pairs | 27 |
| 14 | | |
| 15 | **How big is the difference?** | |
| 16 | Mean of differences (B - A) | 2.348 |
| 17 | SD of differences | 1.324 |
| 18 | SEM of differences | 0.2547 |
| 19 | 95% confidence interval | 1.824 to 2.871 |
| 20 | R squared (partial eta squared) | 0.7657 |

**Fig.A44** Paired t test output of the average promoter methylation for *CGB3, 5 & 8* in breast group. Produced by Prism8

## A6.3.7 Choriocarcinoma group (*CGB7*)

| Normality and Lognormality Tests<br>Tabular results | A<br>JEG-3 | B<br>BeWo | C<br>CRL1790 |
|---|---|---|---|
| **Test for normal distribution** | | | |
| **Anderson-Darling test** | | | |
| A2* | 0.2082 | 2.410 | 0.3403 |
| P value | 0.8358 | <0.0001 | 0.4504 |
| Passed normality test (alpha=0.05)? | Yes | No | Yes |
| P value summary | ns | **** | ns |
| | | | |
| **Number of values** | 16 | 16 | 16 |

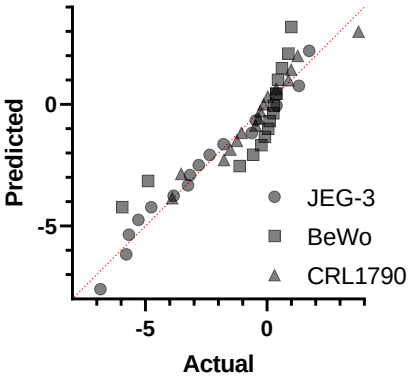



**Fig.A45** Normality test of average promoter methylation for *CGB7* in choriocarcinoma group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | chorio cgb7 |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 9.127 |
| P value | 0.0020 |
| P value summary | ** |
| Statistically significant (P < 0.05)? | Yes |
| Geisser-Greenhouse's epsilon | 0.8050 |
| R squared | 0.3783 |

**Fig.A46** RM one-way ANOVA output of the average promoter methylation for *CGB7* in choriocarcinoma group. Produced by Prism8

| RM one-way ANOVA<br>Multiple comparisons | | | | | |
|---|---|---|---|---|---|
| Number of families | 1 | | | | |
| Number of comparisons per family | 3 | | | | |
| Alpha | 0.05 | | | | |
| | | | | | |
| **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| JEG-3 vs. BeWo | -2.174 | -3.884 to -0.4638 | Yes | * | 0.0126 |
| JEG-3 vs. CRL1790 | -2.269 | -4.040 to -0.4974 | Yes | * | 0.0120 |
| BeWo vs. CRL1790 | -0.09466 | -1.209 to 1.020 | No | ns | 0.9736 |

**Fig.A47** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB7* in choriocarcinoma group (Tukey's test). Produced by Prism8

## A6.3.8 Ovarian group (*CGB7*)

| Normality and Lognormality Tests<br>Tabular results | A<br>HEY | B<br>OAW42 | C<br>OVCAR-3 | D<br>SKOV-3 | E<br>CRL1790 |
|---|---|---|---|---|---|
| **Test for normal distribution** | | | | | |
| **Anderson-Darling test** | | | | | |
| A2* | 1.143 | 0.3843 | 1.240 | 0.3352 | 0.3403 |
| P value | 0.0038 | 0.3516 | 0.0021 | 0.4633 | 0.4504 |
| Passed normality test (alpha=0.05)? | No | Yes | No | Yes | Yes |
| P value summary | ** | ns | ** | ns | ns |
| **Number of values** | 16 | 16 | 16 | 16 | 16 |

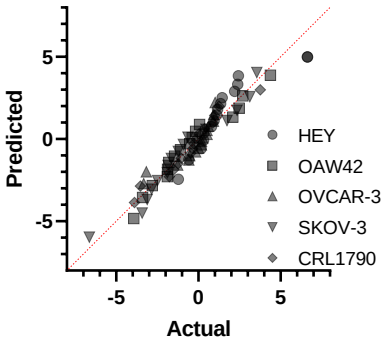



**Fig.A48** Normality test of average promoter methylation for *CGB7* in ovarian group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | ovarian cgb7 |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 6.985 |
| P value | 0.0047 |
| P value summary | ** |
| Statistically significant (P < 0.05)? | Yes |
| Geisser-Greenhouse's epsilon | 0.4468 |
| R squared | 0.3177 |

**Fig.A49** RM one-way ANOVA output of the average promoter methylation for *CGB7* in ovarian group. Produced by Prism8

| RM one-way ANOVA<br>Multiple comparisons | | | | | |
|---|---|---|---|---|---|
| Number of families | 1 | | | | |
| Number of comparisons per family | 10 | | | | |
| Alpha | 0.05 | | | | |
| | | | | | |
| **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| HEY vs. OAW42 | 2.106 | 1.136 to 3.076 | Yes | **** | <0.0001 |
| HEY vs. OVCAR-3 | 1.877 | 0.07284 to 3.681 | Yes | * | 0.0396 |
| HEY vs. SKOV-3 | 2.596 | 1.276 to 3.917 | Yes | *** | 0.0002 |
| HEY vs. CRL1790 | 2.062 | -0.006638 to 4.132 | No | ns | 0.0509 |
| OAW42 vs. OVCAR-3 | -0.2288 | -1.999 to 1.542 | No | ns | 0.9941 |
| OAW42 vs. SKOV-3 | 0.4906 | -0.3234 to 1.305 | No | ns | 0.3777 |
| OAW42 vs. CRL1790 | -0.04327 | -1.821 to 1.734 | No | ns | >0.9999 |
| OVCAR-3 vs. SKOV-3 | 0.7194 | -1.421 to 2.860 | No | ns | 0.8341 |
| OVCAR-3 vs. CRL1790 | 0.1856 | -0.8010 to 1.172 | No | ns | 0.9759 |
| SKOV-3 vs. CRL1790 | -0.5339 | -2.698 to 1.630 | No | ns | 0.9377 |

**Fig.A50** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB7* in ovarian group (Tukey's test). Produced by Prism8

## A6.3.9 Control group (*CGB7*)

| Normality and Lognormality Tests<br>Tabular results | A<br>CRL1790 | B<br>breast-DNA | C<br>cervix-DNA |
|---|---|---|---|
| **Test for normal distribution** | | | |
| **Anderson-Darling test** | | | |
| A2* | 0.3403 | 0.6465 | 0.2515 |
| P value | 0.4504 | 0.0747 | 0.6929 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes |
| P value summary | ns | ns | ns |
| | | | |
| **Number of values** | 16 | 16 | 16 |

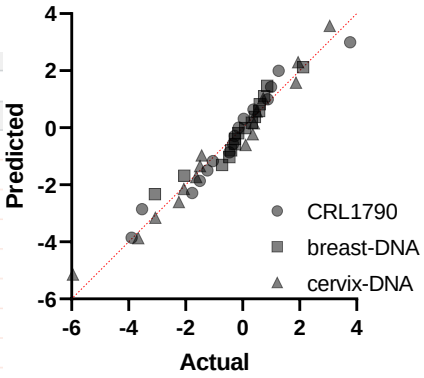



**Fig.A51** Normality test of average promoter methylation for *CGB7* in control group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | control cgb7 |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 1.119 |
| P value | 0.3188 |
| P value summary | ns |
| Statistically significant (P < 0.05)? | No |
| Geisser-Greenhouse's epsilon | 0.6185 |
| R squared | 0.06940 |

**Fig.A52** RM one-way ANOVA output of the average promoter methylation for *CGB7* in control group. Produced by Prism8

## A6.3.10 Cervical group (*CGB7*)

| Normality and Lognormality Tests<br>Tabular results | A<br>C33a | B<br>HT3 | C<br>HeLA | D<br>cervix-DNA |
|---|---|---|---|---|
| **Test for normal distribution** | | | | |
| **Anderson-Darling test** | | | | |
| A2* | 0.2796 | 0.3506 | 0.4217 | 0.2515 |
| P value | 0.5967 | 0.4252 | 0.2835 | 0.6929 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes |
| P value summary | ns | ns | ns | ns |
| | | | | |
| **Number of values** | 16 | 16 | 16 | 16 |

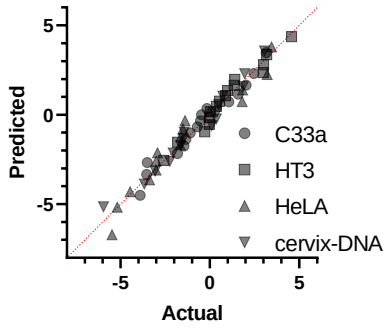



**Fig.A53** Normality test of average promoter methylation for *CGB7* in cervical group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| RM one-way ANOVA<br>ANOVA results | |
|---|---|
| Table Analyzed | cervix cgb7 |
| | |
| **Repeated measures ANOVA summary** | |
| Assume sphericity? | No |
| F | 13.38 |
| P value | <0.0001 |
| P value summary | **** |
| Statistically significant (P < 0.05)? | Yes |
| Geisser-Greenhouse's epsilon | 0.7434 |
| R squared | 0.4715 |

**Fig.A54** RM one-way ANOVA output of the average promoter methylation for *CGB7* in cervical group. Produced by Prism8

| RM one-way ANOVA<br>Multiple comparisons | | | | | |
|---|---|---|---|---|---|
| Number of families | 1 | | | | |
| Number of comparisons per family | 6 | | | | |
| Alpha | 0.05 | | | | |
| | | | | | |
| **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| C33a vs. HT3 | -1.424 | -2.252 to -0.5971 | Yes | *** | 0.0009 |
| C33a vs. HeLA | 0.9456 | 0.1427 to 1.749 | Yes | * | 0.0187 |
| C33a vs. cervix-DNA | 0.2728 | -1.042 to 1.587 | No | ns | 0.9311 |
| HT3 vs. HeLA | 2.370 | 1.445 to 3.295 | Yes | **** | <0.0001 |
| HT3 vs. cervix-DNA | 1.697 | 0.3694 to 3.025 | Yes | * | 0.0106 |
| HeLA vs. cervix-DNA | -0.6728 | -1.991 to 0.6457 | No | ns | 0.4780 |

**Fig.A55** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB7* in cervical group (Tukey's test). Produced by Prism8

## A6.3.11 Breast group (*CGB7*)

| Normality and Lognormality Tests<br>Tabular results | A<br>MCF-7 | B<br>MDA468 | C<br>breast-DNA |
|---|---|---|---|
| **Test for normal distribution** | | | |
| **Anderson-Darling test** | | | |
| A2* | 0.4268 | 0.3483 | 0.6465 |
| P value | 0.2752 | 0.4309 | 0.0747 |
| Passed normality test (alpha=0.05)? | Yes | Yes | Yes |
| P value summary | ns | ns | ns |
| | | | |
| **Number of values** | 16 | 16 | 16 |

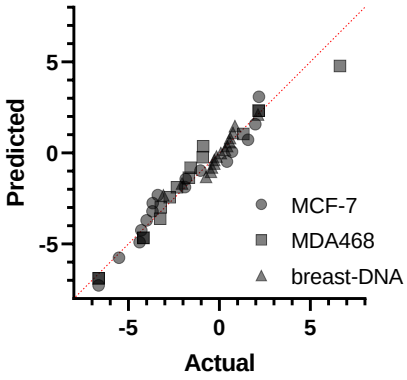



**Fig.A56** Normality test of average promoter methylation for *CGB7* in breast group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

RM one-way ANOVA
ANOVA results

| | | |
|---|---|---|
| 1 | Table Analyzed | breast cgb7 |
| 2 | | |
| 3 | **Repeated measures ANOVA summary** | |
| 4 | Assume sphericity? | No |
| 5 | F | 4.573 |
| 6 | P value | 0.0333 |
| 7 | P value summary | * |
| 8 | Statistically significant (P < 0.05)? | Yes |
| 9 | Geisser-Greenhouse's epsilon | 0.6999 |
| 10 | R squared | 0.2336 |

**Fig.A57** RM one-way ANOVA output of the average promoter methylation for *CGB7* in breast group. Produced by Prism8

RM one-way ANOVA
Multiple comparisons

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | |
| 2 | Number of comparisons per family | 3 | | | | |
| 3 | Alpha | 0.05 | | | | |
| 4 | | | | | | |
| 5 | **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Significant?** | **Summary** | **Adjusted P Value** |
| 6 | MCF-7 vs. MDA468 | -0.4720 | -1.889 to 0.9446 | No | ns | 0.6694 |
| 7 | MCF-7 vs. breast-DNA | -1.986 | -3.494 to -0.4783 | Yes | ** | 0.0100 |
| 8 | MDA468 vs. breast-DNA | -1.514 | -3.806 to 0.7776 | No | ns | 0.2317 |

**Fig.A58** RM one-way ANOVA Multiple comparisons output of the average promoter methylation for *CGB7* in breast group (Tukey's test). Produced by Prism8

## A6.3.12 Colon group (*CGB7*)

Normality and Lognormality Tests
Tabular results

| | | A | B |
|---|---|---|---|
| | | HCT116 | CRL1790 |
| 1 | **Test for normal distribution** | | |
| 2 | **Anderson-Darling test** | | |
| 3 | A2* | 0.4323 | 0.3403 |
| 4 | P value | 0.2665 | 0.4504 |
| 5 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 6 | P value summary | ns | ns |
| 7 | | | |
| 8 | **Number of values** | 16 | 16 |

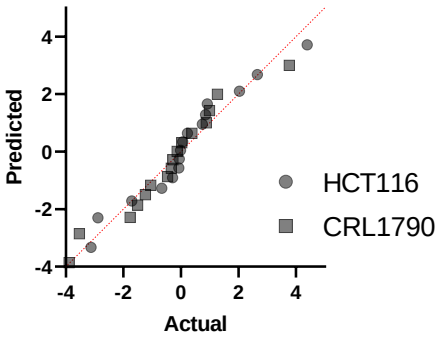



**Fig.A59** Normality test of average promoter methylation for *CGB7* in colon group. Produced by Prism8 (Left: Anderson-Darling test output; right: plot of Normality test)

| | Paired t test | |
|---|---|---|
| 1 | Table Analyzed | colon cgb7 |
| 2 | | |
| 3 | Column B | CRL1790 |
| 4 | vs. | vs. |
| 5 | Column A | HCT116 |
| 6 | | |
| 7 | **Paired t test** | |
| 8 | P value | 0.3159 |
| 9 | P value summary | ns |
| 10 | Significantly different (P < 0.05)? | No |
| 11 | One- or two-tailed P value? | Two-tailed |
| 12 | t, df | t=1.038, df=15 |
| 13 | Number of pairs | 16 |

**Fig.A60** Paired t test output of the average promoter methylation for *CGB7* in colon group. Produced by Prism8

# A7. qRT-PCR data

**Table A6** Cq values obtained from RUN1 of qRT-PCR protocol. Adapted from LightCycler96 software

| RUN1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Analysis Name** | **Position** | **Sample Name** | **Gene Name** | **Cq** | **Cq Mean** | **Cq Error** | **Dye** |
| Rel Quant | A3 | 1 | CGB3-9 | 34.22 | 33.95 | 0.381838 | FAM |
| Rel Quant | A4 | 1 | CGB3-9 | 33.68 | 33.95 | 0.381838 | FAM |
| Rel Quant | E3 | 1 | GAPDH | 15.62 | 15.58 | 0.056569 | FAM |
| Rel Quant | E4 | 1 | GAPDH | 15.54 | 15.58 | 0.056569 | FAM |
| Rel Quant | A5 | 2 | CGB3-9 | 32.71 | 33.04 | 0.46669 | FAM |
| Rel Quant | A6 | 2 | CGB3-9 | 33.37 | 33.04 | 0.46669 | FAM |
| Rel Quant | E5 | 2 | GAPDH | 15.88 | 15.57 | 0.438406 | FAM |
| Rel Quant | E6 | 2 | GAPDH | 15.26 | 15.57 | 0.438406 | FAM |
| Rel Quant | A7 | 3 | CGB3-9 | 32.07 | 32.18 | 0.155563 | FAM |
| Rel Quant | A8 | 3 | CGB3-9 | 32.29 | 32.18 | 0.155563 | FAM |
| Rel Quant | E7 | 3 | GAPDH | 16.11 | 16.12 | 0.014142 | FAM |
| Rel Quant | E8 | 3 | GAPDH | 16.13 | 16.12 | 0.014142 | FAM |
| Rel Quant | A9 | 4 | CGB3-9 | 39.63 | 39.63 | 0 | FAM |
| Rel Quant | A10 | 4 | CGB3-9 | | | | FAM |
| Rel Quant | E9 | 4 | GAPDH | 16.61 | 16.835 | 0.318198 | FAM |
| Rel Quant | E10 | 4 | GAPDH | 17.06 | 16.835 | 0.318198 | FAM |
| Rel Quant | A11 | 5 | CGB3-9 | 30.55 | 30.725 | 0.247487 | FAM |
| Rel Quant | A12 | 5 | CGB3-9 | 30.9 | 30.725 | 0.247487 | FAM |
| Rel Quant | E11 | 5 | GAPDH | 15.26 | 15.51 | 0.353553 | FAM |
| Rel Quant | E12 | 5 | GAPDH | 15.76 | 15.51 | 0.353553 | FAM |
| Rel Quant | B1 | 6 | CGB3-9 | 25.23 | 25.37 | 0.19799 | FAM |
| Rel Quant | B2 | 6 | CGB3-9 | 25.51 | 25.37 | 0.19799 | FAM |
| Rel Quant | F1 | 6 | GAPDH | 15.07 | 14.735 | 0.473762 | FAM |
| Rel Quant | F2 | 6 | GAPDH | 14.4 | 14.735 | 0.473762 | FAM |
| Rel Quant | B3 | 7 | CGB3-9 | 26.12 | 25.79 | 0.46669 | FAM |
| Rel Quant | B4 | 7 | CGB3-9 | 25.46 | 25.79 | 0.46669 | FAM |
| Rel Quant | F3 | 7 | GAPDH | 13.8 | 14.24 | 0.622254 | FAM |
| Rel Quant | F4 | 7 | GAPDH | 14.68 | 14.24 | 0.622254 | FAM |
| Rel Quant | B5 | 8 | CGB3-9 | 29.92 | 29.84 | 0.113137 | FAM |
| Rel Quant | B6 | 8 | CGB3-9 | 29.76 | 29.84 | 0.113137 | FAM |
| Rel Quant | F5 | 8 | GAPDH | 15.79 | 15.9 | 0.155563 | FAM |
| Rel Quant | F6 | 8 | GAPDH | 16.01 | 15.9 | 0.155563 | FAM |
| Rel Quant | B7 | 9 | CGB3-9 | 37.78 | 37.195 | 0.827315 | FAM |
| Rel Quant | B8 | 9 | CGB3-9 | 36.61 | 37.195 | 0.827315 | FAM |
| Rel Quant | F7 | 9 | GAPDH | 15.58 | 15.705 | 0.176777 | FAM |
| Rel Quant | F8 | 9 | GAPDH | 15.83 | 15.705 | 0.176777 | FAM |
| Rel Quant | B9 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | B10 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | F9 | 10 | GAPDH | 15.86 | 15.78 | 0.113137 | FAM |
| Rel Quant | F10 | 10 | GAPDH | 15.7 | 15.78 | 0.113137 | FAM |
| Rel Quant | B11 | 11 | CGB3-9 | 34.84 | 35.24 | 0.565685 | FAM |
| Rel Quant | B12 | 11 | CGB3-9 | 35.64 | 35.24 | 0.565685 | FAM |
| Rel Quant | F11 | 11 | GAPDH | 15.2 | 15.15 | 0.070711 | FAM |
| Rel Quant | F12 | 11 | GAPDH | 15.1 | 15.15 | 0.070711 | FAM |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Rel Quant | C1 | 12 | CGB3-9 | 31.35 | 31.36 | 0.014142 | FAM |
| Rel Quant | C2 | 12 | CGB3-9 | 31.37 | 31.36 | 0.014142 | FAM |
| Rel Quant | G1 | 12 | GAPDH | 16.71 | 16.67 | 0.056569 | FAM |
| Rel Quant | G2 | 12 | GAPDH | 16.63 | 16.67 | 0.056569 | FAM |
| Rel Quant | C3 | 13 | CGB3-9 | 34.55 | 33.095 | 2.057681 | FAM |
| Rel Quant | C4 | 13 | CGB3-9 | 31.64 | 33.095 | 2.057681 | FAM |
| Rel Quant | G3 | 13 | GAPDH | 15.95 | 16.035 | 0.120208 | FAM |
| Rel Quant | G4 | 13 | GAPDH | 16.12 | 16.035 | 0.120208 | FAM |
| Rel Quant | C5 | 14 | CGB3-9 | 30 | 29.985 | 0.021213 | FAM |
| Rel Quant | C6 | 14 | CGB3-9 | 29.97 | 29.985 | 0.021213 | FAM |
| Rel Quant | G5 | 14 | GAPDH | 15.21 | 15.685 | 0.671751 | FAM |
| Rel Quant | G6 | 14 | GAPDH | 16.16 | 15.685 | 0.671751 | FAM |
| Rel Quant | D3 | 19 | CGB3-9 | | | | FAM |
| Rel Quant | D4 | 19 | CGB3-9 | | | | FAM |
| Rel Quant | H3 | 19 | GAPDH | 32.74 | 32.735 | 0.007071 | FAM |
| Rel Quant | H4 | 19 | GAPDH | 32.73 | 32.735 | 0.007071 | FAM |
| Rel Quant | D5 | 20 | CGB3-9 | 34.31 | 34.625 | 0.445477 | FAM |
| Rel Quant | D6 | 20 | CGB3-9 | 34.94 | 34.625 | 0.445477 | FAM |
| Rel Quant | H5 | 20 | GAPDH | 15.98 | 15.885 | 0.13435 | FAM |
| Rel Quant | H6 | 20 | GAPDH | 15.79 | 15.885 | 0.13435 | FAM |
| Rel Quant | A1 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | A2 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | E1 | NTC | GAPDH | | | | FAM |
| Rel Quant | E2 | NTC | GAPDH | | | | FAM |

**Table A7** ΔΔCq values obtained from RUN1.

| **SAMPLE** | | **Cq *CGB*** | | **Cq *GAPDH*** | | **Δ Cq** | **ΔΔCq** | **fold difference** |
|---|---|---|---|---|---|---|---|---|
| ID | Z | Average | SD | Average | SD | | | |
| MCF-7 | 1 | 33.95 | 0.381838 | 15.58 | 0.056569 | 18.37 | -3.12 | 8.69 |
| MDAMB468 | 2 | 33.04 | 0.46669 | 15.57 | 0.438406 | 17.47 | -4.02 | 16.22 |
| C-33a | 3 | 32.18 | 0.155563 | 16.12 | 0.014142 | 16.06 | -5.43 | 43.11 |
| HeLa | 4 | 39.63 | 0 | 16.835 | 0.318198 | 22.795 | 1.305 | 0.40 |
| HT-3 | 5 | 30.725 | 0.247487 | 15.51 | 0.353553 | 15.215 | -6.275 | 77.44 |
| BeWo | 6 | 25.37 | 0.19799 | 14.735 | 0.473762 | 10.635 | -10.855 | 1852.17 |
| JEG-3 | 7 | 25.79 | 0.46669 | 14.24 | 0.622254 | 11.55 | -9.94 | 982.29 |
| HCT116 | 8 | 29.84 | 0.113137 | 15.9 | 0.155563 | 13.94 | -7.55 | 187.40 |
| CRL-1790 | 9 | 37.195 | 0.827315 | 15.705 | 0.176777 | 21.49 | 0 | 1.00 |
| 3T3 | 10 | - | - | 15.78 | 0.113137 | - | - | - |
| HEY | 11 | 35.24 | 0.565685 | 15.15 | 0.070711 | 20.09 | -1.4 | 2.64 |
| OAW42 | 12 | 31.36 | 0.014142 | 16.67 | 0.056569 | 14.69 | -6.8 | 111.43 |
| OVCAR-3 | 13 | 33.095 | 2.057681 | 16.035 | 0.120208 | 17.06 | -4.43 | 21.56 |
| SKOV-3 | 14 | 29.985 | 0.021213 | 15.685 | 0.671751 | 14.3 | -7.19 | 146.02 |
| NRT | 19 | - | - | 32.735 | 0.007071 | - | - | - |
| | 20 | 34.625 | 0.445477 | 15.885 | 0.13435 | 18.74 | -2.75 | 6.73 |

**Table A8** Cq values obtained from RUN2 of qRT-PCR protocol. Adapted from LightCycler96 software

| RUN2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Analysis Name** | **Position** | **Sample Name** | **Gene Name** | **Cq** | **Cq Mean** | **Cq Error** | **Dye** |
| Rel Quant | A3 | 1 | CGB3-9 | 36.65 | 34.94 | 2.418305192 | FAM |
| Rel Quant | A4 | 1 | CGB3-9 | 33.23 | 34.94 | 2.418305192 | FAM |
| Rel Quant | E3 | 1 | GAPDH | 15.14 | 15.16 | 0.028284271 | FAM |
| Rel Quant | E4 | 1 | GAPDH | 15.18 | 15.16 | 0.028284271 | FAM |
| Rel Quant | A5 | 2 | CGB3-9 | 33.66 | 33.76 | 0.141421356 | FAM |
| Rel Quant | A6 | 2 | CGB3-9 | 33.86 | 33.76 | 0.141421356 | FAM |
| Rel Quant | E5 | 2 | GAPDH | 15.29 | 14.96 | 0.466690476 | FAM |
| Rel Quant | E6 | 2 | GAPDH | 14.63 | 14.96 | 0.466690476 | FAM |
| Rel Quant | A7 | 3 | CGB3-9 | 32.88 | 32.325 | 0.784888527 | FAM |
| Rel Quant | A8 | 3 | CGB3-9 | 31.77 | 32.325 | 0.784888527 | FAM |
| Rel Quant | E8 | 3 | GAPDH | 15.29 | 15.49 | 0.282842712 | FAM |
| Rel Quant | H9 | 3 | GAPDH | 15.69 | 15.49 | 0.282842712 | FAM |
| Rel Quant | A9 | 4 | CGB3-9 | 38.22 | 38.22 | 0 | FAM |
| Rel Quant | A10 | 4 | CGB3-9 | | | | FAM |
| Rel Quant | E9 | 4 | GAPDH | 16.21 | 16.2 | 0.014142136 | FAM |
| Rel Quant | E10 | 4 | GAPDH | 16.19 | 16.2 | 0.014142136 | FAM |
| Rel Quant | A11 | 5 | CGB3-9 | 30.22 | 30 | 0.311126984 | FAM |
| Rel Quant | A12 | 5 | CGB3-9 | 29.78 | 30 | 0.311126984 | FAM |
| Rel Quant | E11 | 5 | GAPDH | 14.5 | 14.68 | 0.254558441 | FAM |
| Rel Quant | E12 | 5 | GAPDH | 14.86 | 14.68 | 0.254558441 | FAM |
| Rel Quant | B1 | 6 | CGB3-9 | 24.47 | 24.67 | 0.282842712 | FAM |
| Rel Quant | B2 | 6 | CGB3-9 | 24.87 | 24.67 | 0.282842712 | FAM |
| Rel Quant | F1 | 6 | GAPDH | 14.73 | 14.67 | 0.084852814 | FAM |
| Rel Quant | F2 | 6 | GAPDH | 14.61 | 14.67 | 0.084852814 | FAM |
| Rel Quant | B3 | 7 | CGB3-9 | 25.96 | 25.915 | 0.06363961 | FAM |
| Rel Quant | B4 | 7 | CGB3-9 | 25.87 | 25.915 | 0.06363961 | FAM |
| Rel Quant | F3 | 7 | GAPDH | 14.27 | 14.105 | 0.233345238 | FAM |
| Rel Quant | F4 | 7 | GAPDH | 13.94 | 14.105 | 0.233345238 | FAM |
| Rel Quant | B5 | 8 | CGB3-9 | 29.51 | 28.805 | 0.997020561 | FAM |
| Rel Quant | B6 | 8 | CGB3-9 | 28.1 | 28.805 | 0.997020561 | FAM |
| Rel Quant | F5 | 8 | GAPDH | 15.64 | 15.35 | 0.410121933 | FAM |
| Rel Quant | F6 | 8 | GAPDH | 15.06 | 15.35 | 0.410121933 | FAM |
| Rel Quant | B7 | 9 | CGB3-9 | 34.4 | 35.16 | 1.074802307 | FAM |
| Rel Quant | B8 | 9 | CGB3-9 | 35.92 | 35.16 | 1.074802307 | FAM |
| Rel Quant | F7 | 9 | GAPDH | 15.18 | 15.125 | 0.077781746 | FAM |
| Rel Quant | F8 | 9 | GAPDH | 15.07 | 15.125 | 0.077781746 | FAM |
| Rel Quant | B9 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | B10 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | F9 | 10 | GAPDH | 15.65 | 15.36 | 0.410121933 | FAM |
| Rel Quant | F10 | 10 | GAPDH | 15.07 | 15.36 | 0.410121933 | FAM |
| Rel Quant | B11 | 11 | CGB3-9 | 35.06 | 34.675 | 0.544472222 | FAM |
| Rel Quant | B12 | 11 | CGB3-9 | 34.29 | 34.675 | 0.544472222 | FAM |
| Rel Quant | F11 | 11 | GAPDH | 14.86 | 14.74 | 0.169705627 | FAM |
| Rel Quant | F12 | 11 | GAPDH | 14.62 | 14.74 | 0.169705627 | FAM |
| Rel Quant | C1 | 12 | CGB3-9 | 31.35 | 31.8 | 0.636396103 | FAM |
| Rel Quant | C2 | 12 | CGB3-9 | 32.25 | 31.8 | 0.636396103 | FAM |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Rel Quant | G1 | 12 | GAPDH | 15.7 | 15.925 | 0.318198052 | FAM |
| Rel Quant | G2 | 12 | GAPDH | 16.15 | 15.925 | 0.318198052 | FAM |
| Rel Quant | C3 | 13 | CGB3-9 | 34.56 | 34.735 | 0.247487373 | FAM |
| Rel Quant | C4 | 13 | CGB3-9 | 34.91 | 34.735 | 0.247487373 | FAM |
| Rel Quant | G3 | 13 | GAPDH | 15.24 | 15.28 | 0.056568542 | FAM |
| Rel Quant | G4 | 13 | GAPDH | 15.32 | 15.28 | 0.056568542 | FAM |
| Rel Quant | C5 | 14 | CGB3-9 | 30.06 | 29.515 | 0.770746391 | FAM |
| Rel Quant | C6 | 14 | CGB3-9 | 28.97 | 29.515 | 0.770746391 | FAM |
| Rel Quant | G5 | 14 | GAPDH | 15.02 | 15.195 | 0.247487373 | FAM |
| Rel Quant | G6 | 14 | GAPDH | 15.37 | 15.195 | 0.247487373 | FAM |
| Rel Quant | D3 | 19 | CGB3-9 | | | | FAM |
| Rel Quant | D4 | 19 | CGB3-9 | | | | FAM |
| Rel Quant | H3 | 19 | GAPDH | 32.77 | 33.345 | 0.813172798 | FAM |
| Rel Quant | H4 | 19 | GAPDH | 33.92 | 33.345 | 0.813172798 | FAM |
| Rel Quant | D5 | 20 | CGB3-9 | 26.8 | 28.645 | 2.609224023 | FAM |
| Rel Quant | D6 | 20 | CGB3-9 | 30.49 | 28.645 | 2.609224023 | FAM |
| Rel Quant | H5 | 20 | GAPDH | 15.92 | 15.78 | 0.197989899 | FAM |
| Rel Quant | H6 | 20 | GAPDH | 15.64 | 15.78 | 0.197989899 | FAM |
| Rel Quant | A1 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | A2 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | E1 | NTC | GAPDH | | | | FAM |
| Rel Quant | E2 | NTC | GAPDH | | | | FAM |

**Table A9** ΔΔCq values obtained from RUN2.

| **SAMPLE** | | **Cq CGB** | | **Cq gapdh** | | **Δ Cq** | **ΔΔCq** | **fold difference** |
|---|---|---|---|---|---|---|---|---|
| Z | ID | Average | SD | Average | SD | | | |
| 1 | MCF-7 | 34.94 | 2.418305 | 15.16 | 0.028284 | 19.78 | -0.255 | 1.19 |
| 2 | MDAMB468 | 33.76 | 0.141421 | 14.96 | 0.46669 | 18.8 | -1.235 | 2.35 |
| 3 | C-33a | 32.325 | 0.784889 | 15.49 | 0.282843 | 16.835 | -3.2 | 9.19 |
| 4 | HeLa | 38.22 | | 16.2 | 0.014142 | 22.02 | 1.985 | 0.25 |
| 5 | HT-3 | 30 | 0.311127 | 14.68 | 0.254558 | 15.32 | -4.715 | 26.26 |
| 6 | BeWo | 24.67 | 0.282843 | 14.67 | 0.084853 | 10 | -10.035 | 1049.15 |
| 7 | JEG-3 | 25.915 | 0.06364 | 14.105 | 0.233345 | 11.81 | -8.225 | 299.21 |
| 8 | HCT116 | 28.805 | 0.997021 | 15.35 | 0.410122 | 13.455 | -6.58 | 95.67 |
| 9 | CRL-1790 | 35.16 | 1.074802 | 15.125 | 0.077782 | 20.035 | 0 | 1.00 |
| 10 | 3T3 | | | 15.36 | 0.410122 | | | |
| 11 | HEY | 34.675 | 0.544472 | 14.74 | 0.169706 | 19.935 | -0.1 | 1.07 |
| 12 | OAW42 | 31.8 | 0.636396 | 15.925 | 0.318198 | 15.875 | -4.16 | 17.88 |
| 13 | OVCAR-3 | 34.735 | 0.247487 | 15.28 | 0.056569 | 19.455 | -0.58 | 1.49 |
| 14 | SKOV-3 | 29.515 | 0.770746 | 15.195 | 0.247487 | 14.32 | -5.715 | 52.53 |
| 19 | NRT | | | 33.345 | 0.813173 | | | |
| 20 | | 28.645 | 2.609224 | 15.78 | 0.19799 | 12.865 | -7.17 | 144.01 |

**Table A10** Cq values obtained from RUN3 of qRT-PCR protocol. Adapted from LightCycler96 software

| **Run3** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Analysis Name** | **Position** | **Sample Name** | **Gene Name** | **Cq** | **Cq Mean** | **Cq Error** | **Dye** |
| Rel Quant | A3 | 1 | CGB3-9 | 33.46 | 33.325 | 0.190919 | FAM |
| Rel Quant | A4 | 1 | CGB3-9 | 33.19 | 33.325 | 0.190919 | FAM |
| Rel Quant | E3 | 1 | GAPDH | 15.09 | 15.06 | 0.042426 | FAM |
| Rel Quant | E4 | 1 | GAPDH | 15.03 | 15.06 | 0.042426 | FAM |
| Rel Quant | A5 | 2 | CGB3-9 | 33.79 | 33.82 | 0.042426 | FAM |
| Rel Quant | A6 | 2 | CGB3-9 | 33.85 | 33.82 | 0.042426 | FAM |
| Rel Quant | E5 | 2 | GAPDH | 14.77 | 14.73 | 0.056569 | FAM |
| Rel Quant | E6 | 2 | GAPDH | 14.69 | 14.73 | 0.056569 | FAM |
| Rel Quant | A7 | 3 | CGB3-9 | 32.07 | 32.1 | 0.042426 | FAM |
| Rel Quant | A8 | 3 | CGB3-9 | 32.13 | 32.1 | 0.042426 | FAM |
| Rel Quant | E7 | 3 | GAPDH | 15.74 | 15.725 | 0.021213 | FAM |
| Rel Quant | E8 | 3 | GAPDH | 15.71 | 15.725 | 0.021213 | FAM |
| Rel Quant | A9 | 4 | CGB3-9 | | | | FAM |
| Rel Quant | A10 | 4 | CGB3-9 | 39.51 | 39.51 | 0 | FAM |
| Rel Quant | E9 | 4 | GAPDH | 16.37 | 16.29 | 0.113137 | FAM |
| Rel Quant | E10 | 4 | GAPDH | 16.21 | 16.29 | 0.113137 | FAM |
| Rel Quant | A11 | 5 | CGB3-9 | 31.03 | 31.005 | 0.035355 | FAM |
| Rel Quant | A12 | 5 | CGB3-9 | 30.98 | 31.005 | 0.035355 | FAM |
| Rel Quant | E11 | 5 | GAPDH | 15.27 | 15.43 | 0.226274 | FAM |
| Rel Quant | E12 | 5 | GAPDH | 15.59 | 15.43 | 0.226274 | FAM |
| Rel Quant | B1 | 6 | CGB3-9 | 24.47 | 24.455 | 0.021213 | FAM |
| Rel Quant | B2 | 6 | CGB3-9 | 24.44 | 24.455 | 0.021213 | FAM |
| Rel Quant | F1 | 6 | GAPDH | 14.08 | 14.05 | 0.042426 | FAM |
| Rel Quant | F2 | 6 | GAPDH | 14.02 | 14.05 | 0.042426 | FAM |
| Rel Quant | B3 | 7 | CGB3-9 | 26.54 | 26.65 | 0.155563 | FAM |
| Rel Quant | B4 | 7 | CGB3-9 | 26.76 | 26.65 | 0.155563 | FAM |
| Rel Quant | F3 | 7 | GAPDH | 14.07 | 13.93 | 0.19799 | FAM |
| Rel Quant | F4 | 7 | GAPDH | 13.79 | 13.93 | 0.19799 | FAM |
| Rel Quant | B5 | 8 | CGB3-9 | 29.88 | 29.98 | 0.141421 | FAM |
| Rel Quant | B6 | 8 | CGB3-9 | 30.08 | 29.98 | 0.141421 | FAM |
| Rel Quant | F5 | 8 | GAPDH | 15.1 | 14.965 | 0.190919 | FAM |
| Rel Quant | F6 | 8 | GAPDH | 14.83 | 14.965 | 0.190919 | FAM |
| Rel Quant | B7 | 9 | CGB3-9 | 36.8 | 36.955 | 0.219203 | FAM |
| Rel Quant | B8 | 9 | CGB3-9 | 37.11 | 36.955 | 0.219203 | FAM |
| Rel Quant | F7 | 9 | GAPDH | 15.14 | 15.32 | 0.254558 | FAM |
| Rel Quant | F8 | 9 | GAPDH | 15.5 | 15.32 | 0.254558 | FAM |
| Rel Quant | B9 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | B10 | 10 | CGB3-9 | | | | FAM |
| Rel Quant | F9 | 10 | GAPDH | 15.07 | 15.2 | 0.183848 | FAM |
| Rel Quant | F10 | 10 | GAPDH | 15.33 | 15.2 | 0.183848 | FAM |
| Rel Quant | B11 | 11 | CGB3-9 | 35.31 | 35.32 | 0.014142 | FAM |
| Rel Quant | B12 | 11 | CGB3-9 | 35.33 | 35.32 | 0.014142 | FAM |
| Rel Quant | F11 | 11 | GAPDH | 14.64 | 14.58 | 0.084853 | FAM |
| Rel Quant | F12 | 11 | GAPDH | 14.52 | 14.58 | 0.084853 | FAM |
| Rel Quant | C1 | 12 | CGB3-9 | 31.2 | 31.425 | 0.318198 | FAM |
| Rel Quant | C2 | 12 | CGB3-9 | 31.65 | 31.425 | 0.318198 | FAM |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Rel Quant | G1 | 12 | GAPDH | 15.87 | 15.795 | 0.106066 | FAM |
| Rel Quant | G2 | 12 | GAPDH | 15.72 | 15.795 | 0.106066 | FAM |
| Rel Quant | C3 | 13 | CGB3-9 | 34.97 | 35.445 | 0.671751 | FAM |
| Rel Quant | C4 | 13 | CGB3-9 | 35.92 | 35.445 | 0.671751 | FAM |
| Rel Quant | G3 | 13 | GAPDH | 14.56 | 14.92 | 0.509117 | FAM |
| Rel Quant | G4 | 13 | GAPDH | 15.28 | 14.92 | 0.509117 | FAM |
| Rel Quant | C5 | 14 | CGB3-9 | 29.68 | 29.89 | 0.296985 | FAM |
| Rel Quant | C6 | 14 | CGB3-9 | 30.1 | 29.89 | 0.296985 | FAM |
| Rel Quant | G5 | 14 | GAPDH | 15.3 | 15.625 | 0.459619 | FAM |
| Rel Quant | G6 | 14 | GAPDH | 15.95 | 15.625 | 0.459619 | FAM |
| Rel Quant | D3 | 19 | CGB3-9 | 36.29 | 36.29 | 0 | FAM |
| Rel Quant | D4 | 19 | CGB3-9 | | | | FAM |
| Rel Quant | H3 | 19 | GAPDH | 32.24 | 32.265 | 0.035355 | FAM |
| Rel Quant | H4 | 19 | GAPDH | 32.29 | 32.265 | 0.035355 | FAM |
| Rel Quant | D5 | 20 | CGB3-9 | 34.8 | 34.815 | 0.021213 | FAM |
| Rel Quant | D6 | 20 | CGB3-9 | 34.83 | 34.815 | 0.021213 | FAM |
| Rel Quant | H5 | 20 | GAPDH | 15.18 | 15.17 | 0.014142 | FAM |
| Rel Quant | H6 | 20 | GAPDH | 15.16 | 15.17 | 0.014142 | FAM |
| Rel Quant | A1 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | A2 | NTC | CGB3-9 | | | | FAM |
| Rel Quant | E1 | NTC | GAPDH | | | | FAM |
| Rel Quant | E2 | NTC | GAPDH | | | | FAM |

**Table A11** ΔΔCq values obtained from RUN3

| **SAMPLE** | | **Cq CGB** | | **Cq gapdh** | | **Δ Cq** | **ΔΔCq** | **fold difference** |
|---|---|---|---|---|---|---|---|---|
| Z | ID | Average | SD | Average | SD | | | |
| 1 | MCF-7 | 33.325 | 0.190919 | 15.06 | 0.042426 | 18.265 | -3.37 | 10.34 |
| 2 | MDAMB468 | 33.82 | 0.042426 | 15.725 | 0.056569 | 18.095 | -3.54 | 11.63 |
| 3 | C-33a | 32.1 | 0.042426 | 15.725 | 0.021213 | 16.375 | -5.26 | 38.32 |
| 4 | HeLa | 39.51 | | 15.725 | 0.113137 | 23.785 | 2.15 | 0.23 |
| 5 | HT-3 | 31.005 | 0.035355 | 15.43 | 0.226274 | 15.575 | -6.06 | 66.72 |
| 6 | BeWo | 24.455 | 0.021213 | 14.05 | 0.042426 | 10.405 | -11.23 | 2401.97 |
| 7 | JEG-3 | 26.65 | 0.155563 | 13.93 | 0.19799 | 12.72 | -8.915 | 482.71 |
| 8 | HCT116 | 29.98 | 0.141421 | 14.965 | 0.190919 | 15.015 | -6.62 | 98.36 |
| 9 | CRL-1790 | 36.955 | 0.219203 | 15.32 | 0.254558 | 21.635 | 0 | 1.00 |
| 10 | 3T3 | | | 15.2 | 0.183848 | | | |
| 11 | HEY | 35.32 | 0.014142 | 14.58 | 0.084853 | 20.74 | -0.895 | 1.86 |
| 12 | OAW42 | 31.425 | 0.318198 | 15.795 | 0.106066 | 15.63 | -6.005 | 64.22 |
| 13 | OVCAR-3 | 35.445 | 0.671751 | 14.92 | 0.509117 | 20.525 | -1.11 | 2.16 |
| 14 | SKOV-3 | 29.89 | 0.671751 | 15.625 | 0.459619 | 14.265 | -7.37 | 165.42 |
| 19 | NRT | | | 32.265 | 0.035355 | | | |
| 20 | | 34.815 | 0.021213 | 15.17 | 0.014142 | 19.645 | -1.99 | 3.97 |

# A7.1 Correlation

**Table A12** Association between the transcription data and averaged methylation. Spearman's rho produced in Minitab. The grey rows contain Spearman's coefficient and the no-fill rows contain the p-value for the respective coefficient.

| | **FOLD DIFF.** | | **FOLD DIFF.** | | **FOLD DIFF.** |
|---|---|---|---|---|---|
| **WHOLE** | 0.038 | ***CGB*358_WHOLE** | 0.000 | ***CGB*7_WHOLE** | -0.121 |
| | 0.901 | | 1.000 | | 0.694 |
| **5 REGION** | 0.049 | ***CGB*358_5 REGION** | 0.225 | ***CGB*7_5 REGION** | 0.038 |
| | 0.873 | | 0.459 | | 0.901 |
| **3 REGION** | -0.099 | ***CGB*358_3 REGION** | -0.044 | ***CGB*7_3 REGION** | -0.088 |
| | 0.748 | | 0.887 | | 0.775 |

# A8. ELISA data

**Table A13** Free-beta hCG ELISA OD values used to calculate the concentration of the subunit in media. Values are 450nm reading subtracted from the 650nm reading

| | ID | Run1 | | Run 2 | | |
|---|---|---|---|---|---|---|
| standards | 0 ng/ml | 0.011 | 0.012 | 0.009 | 0.008 | - |
| | 1.25ng/ml | - | - | 0.077 | 0.07 | - |
| | 2.5ng/ml | 0.171 | 0.169 | 0.139 | 0.123 | - |
| | 5ng/ml | 0.298 | 0.302 | 0.261 | 0.198 | - |
| | 10ng/ml | 0.498 | 0.479 | 0.414 | 0.384 | - |
| | 25ng/ml | 0.934 | 0.932 | 0.816 | 0.844 | - |
| | 50ng/ml | 1.421 | 1.462 | 1.219 | 1.225 | - |
| | 100ng/ml | 2.08 | 1.95 | - | - | - |
| | 200ng/ml | 2.377 | 2.298 | - | - | - |
| | low standard | 0.716 | 0.684 | 0.563 | 0.559 | - |
| samples | CRL-1790 | 0.011 | 0.012 | 0.009 | 0.009 | 0.008 |
| | SKOV-3 | 0.194 | 0.197 | 0.192 | 0.177 | 0.174 |
| | C-33a | 0.012 | 0.011 | 0.01 | 0.01 | 0.007 |
| | MDAMB468 | 0.012 | 0.01 | 0.009 | 0.009 | 0.01 |
| | OVCAR-3 | 0.012 | 0.011 | 0.01 | 0.008 | 0.008 |
| | HEY | 0.076 | 0.077 | 0.077 | 0.065 | 0.072 |
| | OAW42 | 0.023 | 0.016 | 0.017 | 0.013 | 0.016 |
| | MCF-7 | 0.011 | 0.01 | 0.012 | 0.007 | 0.008 |
| | BeWo | 0.199 | 0.161 | 0.162 | 0.153 | 0.142 |
| | JEG-3 | 0.091 | 0.093 | 0.085 | 0.072 | 0.078 |
| | HeLa | 0.014 | 0.014 | 0.013 | 0.009 | 0.01 |
| | 3T3 | 0.012 | 0.013 | 0.01 | 0.008 | 0.009 |

**Table A14** Intact hCG ELISA OD values used to calculate the concentration of the hormone in media.

| | ID | OD (460nm-650nm) | |
|---|---|---|---|
| standards | 0 mIU/ml | 0.021 | 0.019 |
| | 5 mIU/ml | 0.021 | 0.019 |
| | 10 mIU/ml | 0.052 | 0.055 |
| | 50 mIU/ml | 0.17 | 0.203 |
| | 200 mIU/ml | 0.899 | 0.932 |
| | 500 mIU/ml | 1.855 | 1.834 |
| Samples | CRL-1790 | 0.018 | 0.017 |
| | SKOV-3 | 0.019 | 0.019 |
| | C-33a | 0.018 | 0.018 |
| | MDAMB468 | 0.02 | 0.019 |
| | OVCAR-3 | 0.019 | 0.018 |
| | HEY | 0.021 | 0.019 |
| | OAW42 | 0.021 | 0.018 |
| | MCF-7 | 0.021 | 0.021 |
| | BeWo | 0.511 | 0.523 |
| | JEG-3 | 0.535 | 0.54 |
| | HeLa | 0.02 | 0.018 |
| | 3T3 | 0.018 | 0.017 |

**Regression Formula:** $y = d + \frac{a-d}{1+(x/c)^b}$

**Inverse Formula:** $x = c(\frac{a-d}{y-d} - 1)^{\frac{1}{b}}$

Where:
y = Response Value eg. OD
x = Dose Value eg. Concentration
and the constants/parameters are as follows:

| a | b | c | d |
|---|---|---|---|
| 2.062 | -1.147 | 36.998 | 0.038 |

**$R^2$ value:** 0.998

**Fig.A61** Regression formula and $R^2$ calculated based on free beta hCG ELISA standards from Run 1 used to calculate concentration of samples. Curve equation calculated by 4 parameter fit model. Calculation done using https://elisaanalysis.com

**Regression Formula:** $y = d + \frac{a-d}{1+(x/c)^b}$

**Inverse Formula:** $x = c(\frac{a-d}{y-d} - 1)^{\frac{1}{b}}$

Where:
y = Response Value eg. OD
x = Dose Value eg. Concentration
and the constants/parameters are as follows:

| a | b | c | d |
|---|---|---|---|
| 2.9654991 | -1.4387708 | 356.9563315 | 0.0242282 |

**$R^2$ value:** 0.9996244

**Fig.A62** Regression formula and $R^2$ calculated based on intact hCG ELISA standards. Curve equation calculated by 4 parameter fit model used to calculate concentration of samples. Calculation done using https://elisaanalysis.com