University of Cyprus

DEPARTMENT OF BIOLOGICAL SCIENCES

&

THE CYPRUS INSTITUTE OF NEUROLOGY & GENETICS

TRANSLATIONAL GENETICS TEAM

IDENTIFICATION AND CHARACTERIZATION OF FETAL SPECIFIC METHYLATED REGIONS FOR NON-INVASIVE PRENATAL DIAGNOSIS

Marios Ioannides

December 2014

University of Cyprus

DEPARTMENT OF BIOLOGICAL SCIENCES

&

THE CYPRUS INSTITUTE OF NEUROLOGY & GENETICS

TRANSLATIONAL GENETICS TEAM

IDENTIFICATION AND CHARACTERIZATION OF FETAL SPECIFIC METHYLATED REGIONS FOR NON-INVASIVE PRENATAL DIAGNOSIS

Marios Ioannides

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December 2014

VALIDATION PAGE

Doctoral Candidate: Marios Ioannides

Doctoral Thesis Title: Identification and Characterization of Fetal Specific Methylated Regions for Non-Invasive Prenatal Diagnosis

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the ***Department of Biological Sciences*** *and was approved on the 5th December 2014 by the members of the* ***Examination Committee****.*

Examination Committee:

Research Supervisor:	Dr. Philippos Patsalis Professor	_____
Committee Member:	Dr. Chrysoula Pitsouli Assistant Professor	_____
Committee Member:	Dr. Leonidas Phylactou Professor	_____
Committee Member:	Dr. Peter Karayiannis Professor	_____
Committee Member:	Dr. Kyriacos Felekkis Assistant Professor	_____

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

……..

……..

ΠΕΡΙΛΗΨΗ

Η προγεννητική διάγνωση έχει ως σκοπό τη διάγνωση των πιο κοινών εμβρυικών ανευπλοειδιών και ανωμαλιών. Σήμερα, ο μη επεμβατικός έλεγχος προσφέρεται σε όλες τις κυήσεις παρέχοντας ένα ρίσκο κινδύνου εμφάνισης των πιο συχνών ανωμαλιών του εμβρύου. Στις κυήσεις υψηλού ρίσκου η διάγνωση πραγματοποιείται με τη συλλογή εμβρυικού ιστού με επεμβατικές μεθόδους. Παρόλο που η διαδικασία αυτή προσφέρει μια υψηλής ακρίβειας διάγνωση, συνδέεται με ένα σημαντικό ποσοστό αποβολής. Επομένως, η ανάγκη για μείωση του ρίσκου αποβολής οδήγησε στην ανακάλυψη της παρουσίας ελεύθερου εμβρυικού DNA στο μητρικό περιφερικό αίμα. Αυτό αποτέλεσε ένα σημαντικό επίτευγμα στη ανάπτυξη μεθόδων μη επεμβατικής προγεννητικής διάγνωσης, οι οποίες είναι ασφαλέστερες και διαθέσιμες σε όλες τις εγκυμοσύνες.

Αξιοποιώντας τις διαφορές μεθυλίωσης μεταξύ εμβρυϊκού DNA από πλακούντα και DNA μητρικού περιφερικού αίματος, αρκετές μελέτες έχουν εντοπίσει ένα αριθμό από εμβρυοειδικά διαφορικές μεθυλιωμένες περιοχές. Παρόλα αυτά η εφαρμογή τους αποτέλεσε πρόκληση λόγω των μικρών ποσοτήτων ελεύθερού εμβρυϊκού DNA εν τη παρουσία αυξημένων ποσοτήτων μητρικού DNA. Η ερευνητική μας ομάδα με τη χρήση της μεθοδολογίας Ανοσοκατακρήμνισης Μεθυλιωμένου DNA (MeDIP) σε συνδυασμό με μικροσυστοιχίες DNA κατάφερε να ταυτοποιήσει χιλιάδες διαφορικά μεθυλιωμένες περιοχές στα χρωμοσώματα 13, 21, 18, X και Y. Αργότερα, χρησιμοποιώντας μια υποομάδα περιοχών κατάφεραν να διαγνώσουν κυήσεις με σύνδρομο Down με υψηλή ευαισθησία και ακρίβεια.

Ο κύριος στόχος αυτής της μελέτης ήταν η ταυτοποίηση και ο χαρακτηρισμός εμβρύοειδικών δεικτών για τα χρωμοσώματα 13, 18, 21, και X για χρήση τους στην ανάπτυξη ενός μη επεμβατικού προγεννητικού τεστ. Στο πρώτο στάδιο, χρησιμοποιήσαμε τα προαναφερθέντα δεδομένα μικροσυστοιχιών και επιλέξαμε ένα αριθμό περιοχών τις οποίες και χαρακτηρίσαμε. Επιβεβαιώσαμε τη διαφορά μεθυλίωσης μεταξύ εμβρυικού και μητρικού DNA σε τρείς, οκτώ και 12 περιοχές στα χρωμοσώματα 13, 18 και 21 αντίστοιχα. Ένα υποσύνολο αποτελούμενο από 15 περιοχές χρησιμοποιήθηκε για περαιτέρω χαρακτηρισμό σε 50 δείγματα χοριακών λαχνών και 50 δείγματα περιφερικού αίματος από μη έγκυες γυναίκες προκειμένου να διερευνηθεί η διακύμανση της μεθυλίωσης μεταξύ διαφορετικών ατόμων. Το συμπέρασμα ήταν ότι, παρά τη μεταβλητότητα της μεθυλίωσης μεταξύ των

δειγμάτων η προσέγγιση μας παρείχε μια σαφή διάκριση μεταξύ του εμβρυικού και μητρικού ιστού.

Το δεύτερο στάδιο αφορούσε τον σχεδιασμό και την εφαρμογή υπερ-υψηλής ανάλυσης μικροσυστοιχιών DNA. Η εφαρμογή αυστηρών κριτηρίων επιλογής επιβεβαίωσαν τη παρουσία διαφορικής μεθυλίωσης μεταξύ εμβρυικού και μητρικού ιστού σε 31, 22, 46 και δύο περιοχών στα χρωμοσώματα 13, 18, 21 και X αντίστοιχα. Η πλειονότητα των περιοχών αυτών βρέθηκε να βρίσκεται σε γονίδια πολλά από τα οποία συνδέονται με ασθένειες.

Το τελευταίο στάδιο αφορούσε την επιβεβαίωση των αποτελεσμάτων χρησιμοποιώντας την μεθοδολογία αλληλούχισης επόμενης γενιάς σε συνδυασμό με τη μεθοδολογία MeDIP. Οι δύο προσεγγίσεις βρέθηκαν να είναι όμοιες αφού παρείχαν ακριβή αποτελέσματα σε σχέση με την ταυτοποίηση διαφορικά μεθυλιωμένων περιοχών.

Συμπερασματικά, η εργασία μας παρέχει μια επέκταση στο πάνελ εμβρυοειδικών δεικτών που διατίθενται για NIPD του συνδρόμου Down και ειναι δυνατόν να αποτελέσει το σημείο εκκίνησης για την ανάπτυξη μεθόδων ανίχνευσης άλλων συχνών ανευπλοειδιών. Επιπλέον, τα στοιχεία μας δείχνουν ότι παρά το γεγονός ότι η διακύμανση της μεθυλίωσης είναι εμφανής, η διαφορά μεθυλίωσης στους δυο ιστούς είναι αρκετά μεγάλη ώστε να επιτρέπει την ιστο-ειδική ταυτοποίηση μεθυλίωσης.

ABSTRACT

Prenatal diagnosis aims to detect the most common aneuploidies and fetal abnormalities. It is routinely offered today as a non-invasive prenatal screening provided to all pregnancies or as an invasive procedure offered only to high risk pregnancies. While the former provides only a risk factor, the latter provides a more definitive diagnosis albeit associated with a significant rate of spontaneous miscarriages. Thus, the need of minimizing potential pregnancy risks has led to the discovery of cell free fetal DNA (cffDNA) in the maternal circulation, an important achievement towards the development of non-invasive prenatal diagnosis (NIPD) methodologies that are safer, more effective and available for all pregnancies.

Taking advantage of the methylation differences between placenta derived cffDNA and the maternal blood, several studies have successfully identified a number of fetal specific differentially methylated regions (DMRs). Even though the implementation of DMRs in the detection of aneuploidies has proven to be a challenge due to the limited amount of cffDNA in the high maternal background, our group has previously combined methylated DNA immunoprecipitation (MeDIP) with high resolution array Comparative Genomic Hybridization (aCGH) and successfully identified thousands of candidate DMRs on chromosomes 13, 21 18, X and Y. Using a subset of these DMRs they were able to correctly classify Down syndrome pregnancies with high sensitivity and specificity.

The main objective of this study was the identification and characterization of fetal specific biomarkers for chromosomes 13, 18, 21, and X for their potential utilization towards the development of an NIPD test. In the first stage we screened previously obtained aCGH data for the characterization of new fetal specific DMRs. We confirmed three, eight and 12 DMRs on chromosomes 13, 18 and 21 respectively. A subset was further validated in a set of 50 chorionic villus samplings (CVS) and 50 non-pregnant peripheral blood samples (WBF) in order to investigate the inter-individual methylation variability. It was concluded that despite the variability in the methylation between samples our approach provided a clear distinction between the fetal and maternal tissue.

The second stage involved the design and implementation of ultra-high resolution aCGH. Implementing stringent selection criteria we confirmed the differential

methylation pattern between CVS and WBF in 31, 22, 46 and two DMRs on chromosomes 13, 18, 21 and X respectively. The majority of DMRs were found to be located on genes many of which are associated with diseases.

The last stage involved the confirmation of the MeDIP-Chip results using the MeDIP-seq approach. The two approaches were found to be highly consistent as they provided accurate and robust results in regards to the discovery of differentially methylated regions.

In conclusion, our work provides an expansion in the biomarker panel available for NIPD for Down syndrome and can eventually provide the starting point towards the development for assays towards the detection of all common chromosomal aneuploidies. Furthermore, our data indicate that inter-experimental and inter-individual variation in methylation is apparent, yet the difference in methylation status across tissues is large enough to allow for robust tissue specific methylation identification.

ACKNOWLEDGEMENTS

I would like to express my gratitude and my deepest respect to my advisor Prof. Philippos C. Patsalis for his guidance and the faith he has shown in me all these years. He has been a source of inspiration to me, a role model in my professional and personal life. His valuable, constructive advice, positive and enthusiastic spirit has given me the opportunity to develop my skills, excel as a scientist and reach the maximum level of my education.

I would like to thank my colleagues at the Translational Genetics Team at the Cyprus Institute of Neurology and Genetics and the NIPD Genetics Ltd for their help and positive interaction during all the stages of my PhD project. I would like to express my warm thanks to Dr. George Koumbaris and Dr. Michael Hadjidaniel for their dynamic scientific input towards the completion of my project.

Furthermore, I would like to thank my family for their help and encouragement. Your love, sacrifices and moral support have shaped my character and provided me with the necessary equipment to develop to the person I am today. My gratitude is endless.

Lastly, but always first in my life, I would like to express my gratefulness to my other half, my wife Juliana Saavedra. Without you nothing could be accomplished. You are the source of my strength and confidence. Thank you for being in my life, your love and support and the endless amount of patience and encouragement you have given me in order for me to pursuit my dream. I am lucky to have you in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFP: alpha fetoprotein
aCGH: array comparative genomic hybridization
BAC: bacterial artificial chromosome
CNC: copy number changes
cffDNA: cell-free fetal DNA
CNV: copy number variant
COBRA: combined bisulfite restriction assay
CVS: chorionic villus sampling
DMRs: differentially methylated regions
DNMTs: DNA methyltransferases
FISH: fluorescence in situ hybridization
IN: input
HCG: human chorionic gonadotropin
IP: immunoprecipitated
LM PCR: ligation-mediated PCR
MBD: methyl-binding protein
MeDIP: methylation DNA immunoprecipitation
MeDIP-Chip: MeDIP in combination with aCGH
MeDIP-seq: MeDIP in combination with NGS
MHC: major histocompatibility complex
MLPA: multiplex ligation-dependent probe amplification
MPS: massively parallel sequencing
MS: mass spectrometry
NIPD: non-invasive prenatal diagnosis
NGS: next-generation sequencing
PAPP-A: pregnancy-associated plasma protein A
PCR: polymerase chain reaction
PGD: preimplantation Genetic Diagnosis
qPCR: quantitative PCR
RhD: rhesus D
SNP: single nucleotide polymorphism
SRY: sex determining region Y
WBF: whole blood female (non-pregnant peripheral blood)

1 INTRODUCTION

1.1 Chromosomal Abnormalities and Associated Syndromes

Normal chromosomal distribution is largely dependent on the inheritance of 23 homologous pairs of chromosomes –one from each parent- as well as on the precise cell division of daughter cells after fertilization. Any disruption on the dosage of the genetic material would result in chromosomal abnormalities with serious effects on the normal cell growth pattern and/or normal expression of genes. In general, there are two types of chromosomal abnormalities: the structural abnormalities involve the gain or loss of chromosomal material due to deletions, duplications, insertions, inversions, or translocations [1, 2]; numerical abnormalities are characterized by the presence or absence of one or more chromosomes. If one or more chromosome sets are present, then the phenomenon is called polyploidy whereas the presence or absence of one or more chromosomes is called aneuploidy. The most well-known autosomal aneuploidies include Down syndrome (trisomy 21), Edwards syndrome (trisomy 18) and Patau syndrome (trisomy 13). Sex chromosomal aneuploidies include Turner Syndrome (monosomy X), Klinefelter syndrome (47, XXY), and trisomy X (47, XXX).

Chromosomal abnormalities result from non-disjunction during meiosis or mitosis. Usually, failure of segregation of a homologous pair because of meiotic non-disjunction either during Anaphase I or II would result in aneuploidy. On the other hand, complete failure of disjunction during meiosis usually results in polyploidy. Furthermore, mosaicism, the presence of two or more cell lines, is a result of mitotic errors during somatic cell division [1].

1.1.1 Down Syndrome

Down syndrome or trisomy 21 is characterized by the presence of a third chromosome 21. It is one of the most common birth defects but its frequency has decreased dramatically due to routine prenatal testing, with prevalence of one in

700 births [1, 3, 4] with the risk increasing as maternal age increases. Studies have shown that trisomy 21 is also one of the most frequent chromosomal abnormalities causing early pregnancy loss [1].

Meiotic non-disjunction during meiosis I (80%) or meiosis II (20%) is the cause of about 95% of trisomy 21 cases, resulting in the inheritance of an entire extra chromosome from the parents due to the presence of two copies of chromosome 21 in one of the parental gametes [1, 5]. Less frequently, the fetus inherits a translocated chromosome from a parent with balanced translocation that involves chromosome 21 with another chromosome. Furthermore, in rare cases, affected fetuses appear to have two different cell lines present, one consisting of the normal 46 chromosomes and the other cell line consisting of an extra chromosome 21. This is possible due to mitotic errors in the zygote. In these cases, the clinical manifestation of the disease varies depending at which developmental stage the error had occurred.

Patients with Down syndrome are presented with flattened face, slanted almond shaped eyes and absence of nasal bone. Other common physical features of Down syndrome include short neck, small hands with bilateral single palmar crease, short stature and clinodactyly. One of the major complications of trisomy 21 is mental retardation that varies from mild to severe, while other health issues are strongly associated with trisomy 21 patients such as high risk for psychiatric and social problems, respiratory, hearing and heart complications, Alzheimer's disease, anemia and leukemia in early childhood [1, 5, 6].

1.1.2 Edwards Syndrome

Edwards syndrome or trisomy 18 is a chromosomal abnormality that is associated with the presence of an extra chromosome 18. Its prevalence is one in 6000 live births with a higher survival rate in females. 90% of the infants die within the first year of life [7]. During conception the frequency is much higher but in 95% of the cases the embryo is spontaneously aborted. The cause of trisomy 18 is usually due to errors in meiosis I in the mother and less frequently during meiotic non-disjunction in the father [8]. In rare cases the cause is the inheritance of a derivative chromosome from a parent that is a carrier of a balanced translocation. Mosaic cases have also been reported with milder symptoms.

Edwards syndrome is characterized by multiple dysmorphic features such us dysplastic ears, clenched hands, prominent head and sternum [9]. Affected newborns also have kidney and urinary tract malformations and frequently heart dysplasia [9].

1.1.3 Patau Syndrome

Patau syndrome or trisomy 13 is caused by the presence of an extra chromosome 13 with a prevalence of one in 8000 live births. Its frequency appears to be higher but in 95% of the cases the embryo is spontaneously aborted. Maternal age plays a role in the risk of the disease. Usually affected newborns die within the first year of age, while there are reports of cases of prolonged survival [1].

As mentioned previously, most of the cases of Patau syndrome are associated with the presence of an extra chromosome 13. Furthermore the syndrome may be associated with Robertsonian translocation- that is, the attachment of chromosome 13 to another acrocentric chromosome- and less frequently with the reciprocal translocation between chromosome 13 and a non-acrocentric chromosome [1, 2]. Like Down and Edwards syndromes, Patau syndrome can be detected prenatally by ultrasound and can be confirmed by cytogenetic analysis. Common features include holoprocencephaly, cleft lip, cleft palate, urogenital and heart malformations [1].

1.1.4 Klinefelter Syndrome

Klineftelter syndrome is a chromosomal disorder that affects males. It is characterized by the presence of an extra chromosome X and is one of the most common disorders in the male population with a prevalence of one in 500 to 1000 newborn males [10, 11].

The additional X chromosome is the result of meiotic non-disjunction in one of the parents during gametogenesis. In poly-X karyotypes the extra chromosome X is known to be inactive because of the X inactivation process. Studies though have shown some X-linked genes escape the X-inactivation resulting in various degrees of mental retardation [12].

Affected males appear to have tall stature with sparse facial and body hair. Common characteristics also include gynecomastia, infertility, developmental delay and mild mental retardation [1, 7].

1.1.5 Turner Syndrome

Turner syndrome is a chromosomal disorder associated with the absence of chromosome X (monosomy X). The prevalence is one in 5000 live births [8]. Even though its frequency is more common, 99% of the affected embryos do not survive to term.

Affected females appear to have heterogeneous phenotypes with short stature being the most common characteristic. Other features of the disease include premature ovarian failure and ovarian tissue degeneration even though in rare cases normal ovarian function is retained [13] .

Meiotic non-disjunction during spermatogenesis is the cause of Turner syndrome in most cases. Furthermore, during fetal development, mitotic errors may result in mosaic cases while in rare occasions patients have partial deletions or ring chromosome X [1].

1.2 Prenatal Diagnosis

Prenatal diagnosis is the screening of an embryo that aims to detect of the most common aneuploidies, i.e Down syndrome, Edwards syndrome, Patau syndrome, as well as other fetal abnormalities. In the early 80s the screening was solely based on maternal age [14], while later specific proteins in the maternal blood were utilized for the detection of fetal abnormalities [15]. Today, prenatal screening can be performed invasively or non-invasively usually during the first and second trimesters of pregnancy.

1.2.1 Non Invasive Prenatal Screening

Non-invasive prenatal screening is based on the detection of fetal abnormality using ultrasound findings combined with specific markers in maternal serum. It does not give a definitive diagnosis but it provides a risk factor for potential abnormal pregnancies [16, 17]. The basic marker used during the first trimester of pregnancy (11-13 weeks of gestation) for the detection of Down syndrome is the measurement of fetal nuchal translucency which is “the maximum thickness of the subcutaneous translucency between the skin and the soft tissue overlying the cervical spine of the fetus” [18, 19] . This marker in combination with the maternal age and the levels of free beta-human chorionic gonadotropin (β-hCG) and Pregnancy associated Plasma Protein (PAPP-A) in the maternal serum has been associated with a Down syndrome detection rate of 90% [20].

During the second trimester, chromosomal abnormalities were found to be associated with several sonographic markers such as absent or hypoplastic nasal bone, increased nuchal fold thickness, intracardiac hyperechogenic focus, and aberrant right subclavian artery (ARSA) [21]. Additionally, algorithms have been developed that allowed the estimation of the risk for aneuploidies by combining it with fetal markers such us beta-hCG ,PAPP-A and AFP, and inhibitin A [22]. The ability to correctly detect abnormalities such as Down and Edwards syndromes is around 80% and 64% respectively [15, 23], while integration of the two trimester screening results increases the detection rates to 94% [15].

1.2.2 Invasive Prenatal Diagnosis

While non-invasive screening provides only a risk factor for fetal abnormalities detection, a conclusive diagnosis can only be made from fetal tissue. Thus, high risk pregnancies, or couples with family history of inherited diseases go through the invasive procedures in order to collect fetal tissue for further genetic testing.

Invasive procedures include chorionic villus sampling, amniocentesis and chordocentesis. All of them are medical procedures that involve the insertion of the needle trans-abdominally under ultrasound guidance into the uterus to aspirate amniotic fluid or withdraw placental tissue or fetal blood (Figure 1.1). Chorionic villus sampling (CVS) can also be performed trans-cervically, but it is more technically demanding with more failures in the attempt to obtain adequate sample amount [24] . CVS is usually performed between the 11th and 13th week of gestation while amniocentesis is performed between the 13th and 16th [25]. Both procedures are associated with a spontaneous abortion risk of 0.5-1% [24, 26].

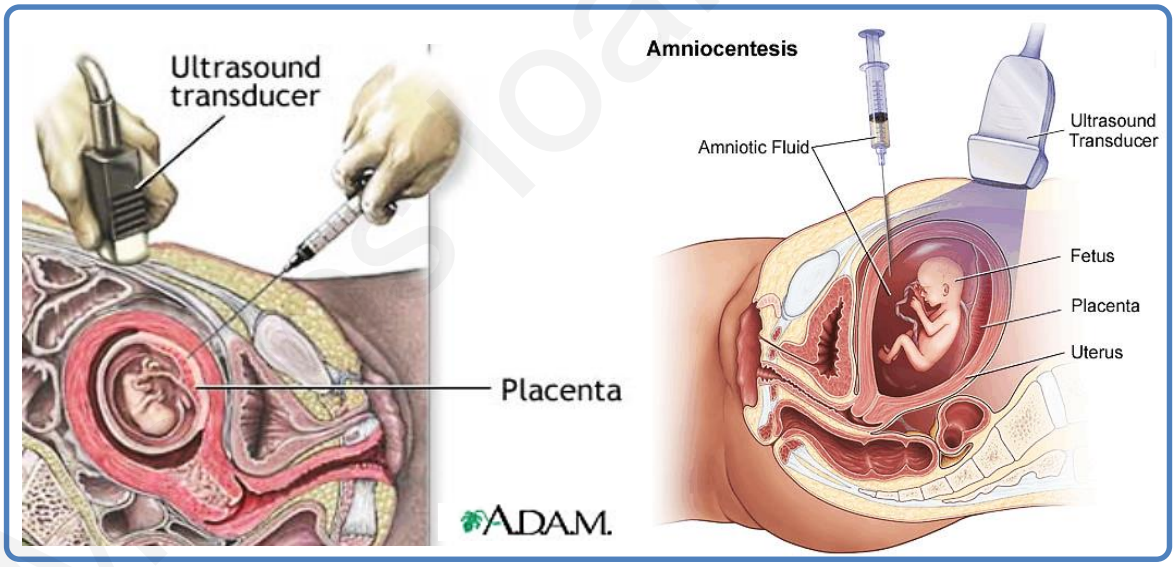


Figure 1.1 Invasive procedures for fetal tissue collection. A long needle is inserted under ultrasound guidance trans-abdominally into the uterus in order to aspirate CVS (Left) or amniotic fluid (Right) depending on the gestational age.

1.2.2.1 Classical Cytogenetic Analysis-Karyotyping

Cytogenetics is the field of genetics that studies the number and structures of chromosomes. Classical cytogenetic analysis usually refers to karyotyping. Karyotype is the classification of homologous chromosomes based on their size, centromere location and chromosomal banding. Chromosomes from cultured cells are arrested during metaphase stage and fixed on a glass slide. Subsequent treatment with trypsin and specific dyes allows the under the microscope visualization of banding patterns specific for each chromosome. The most commonly used dye is the Giemsa in which the A,T rich regions absorb the dye resulting in dark banding as compared to the gene G,C rich regions that appear lighter (G-banding) [27] (Figure 1.2).

Karyotyping allows the simultaneous analysis of all chromosomes and the detection of chromosomal aneuploidies such us polyploidies, trisomies, monosomies and mosaicism; that is why until today it is the preferred method for prenatal diagnosis. Structural chromosomal rearrangements such as deletions, duplications, inversions or translocations can be detected if the breakpoints are larger than 5-10 Mbases. Disadvantages of this method include failure to culture, culture contamination, or suboptimal chromosome preparations [28].

Figure 1.2 Normal male karyotype. Chromosomes are arranged in pairs of homologues according to banding pattern and size. (Photo courtesy of Department of Cytogenetics and genomics)

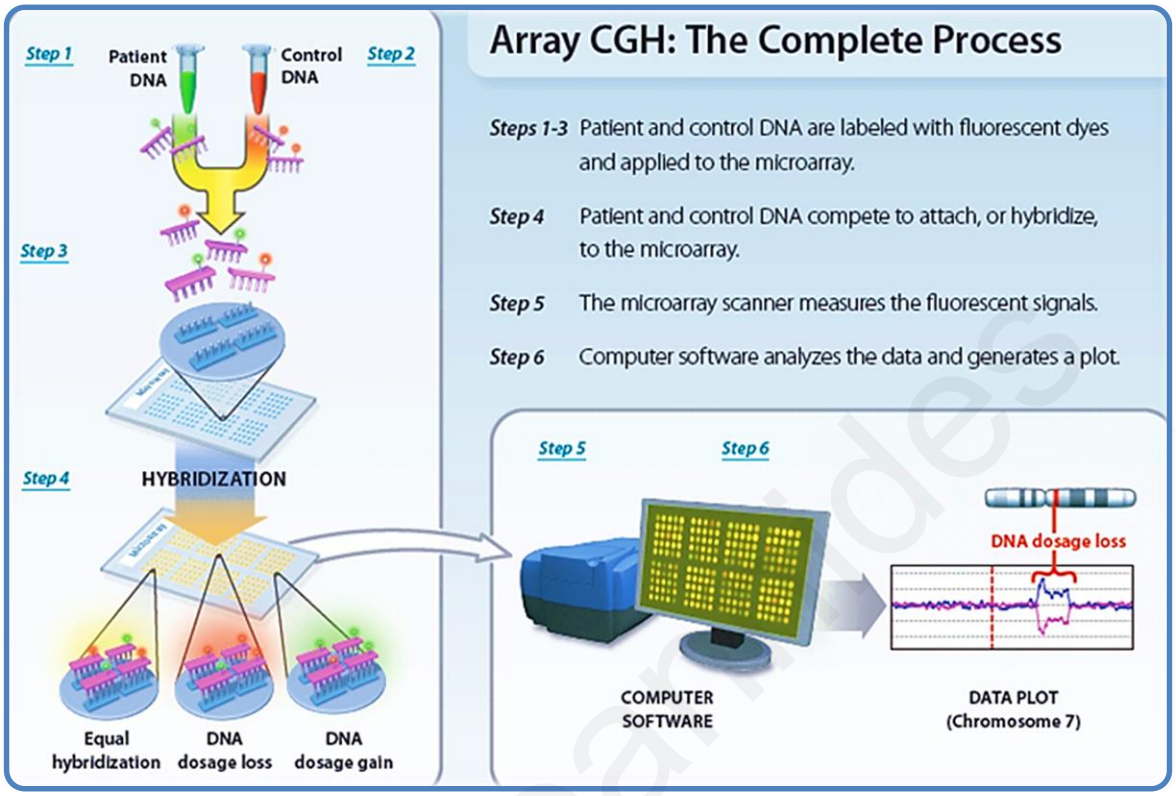
1.2.2.2 Array Comparative Genomic Hybridization-(aCGH)

Array Comparative Genomic Hybridization is a method based on the same principle as Comparative Genomic Hybridization (CGH) that aims to detect copy number changes between test and reference DNA. In CGH, the two samples are differentially labelled with fluorescent dyes and are hybridized simultaneously on a glass slide containing chromosome metaphase spreads of a normal individual. Due to the low resolution of the technique though (5-10Mb), the last two decades CGH has been replaced by aCGH. Initially, bacterial artificial chromosome clones (BAC) mapped onto the human genome were spotted on a glass slide (array slide) robotically. The number of clones and their distance between two consecutive clones is what determined the resolution of the array chip [29]. Due to prime demand of aCGH many improvements have been introduced in the technology of array. The resolution has been increased dramatically not only by increasing the number of features (hybridization targets) on the array slide but also by decreasing their size substantially from a few kilobases to oligonucleotides ranging from 25-80bp [29]. Today, aCGH is extensively utilized for the detection of copy number changes in all fields of molecular genetics such us the detection and diagnosis of cancer, prenatal diagnosis of subtle chromosomal aneuploidies and syndromes, postnatal diagnosis, uniparental disomies (UPD) and preimplantation genetic screening and diagnosis [30-35]. Thus, depending on the application there are different array platforms that can be used ranging from whole genome arrays to targeted arrays that include only specific regions of the genome.

As mentioned before, the procedure involves the differential labelling of test and reference using equal amounts of DNA. The two labelled DNA samples are then mixed and hybridized together on the array chip. Following hybridization, slides are scanned and imported in the appropriate software for assessment of the data and extraction of the fluorescent signal intensities for analysis. Comparative analysis of the fluorescent signal intensity is statistically processed in order to identify copy number changes (CNCs) (Figure 1.3).

Despite the many advantages of this method, a major drawback is that it only detects unbalanced chromosomal abnormalities. Balanced chromosomal abnormalities such as inversions or translocations that have no influence on the

copy number state will not be detected by aCGH that is why other techniques are used to complement aCGH when necessary.



http://www.nature.com/scitable/topicpage/microarray-based-comparative-genomic-hybridization-acgh-45432#

Figure 1.3 Outline of the array CGH procedure.

1.2.2.3 Quantitative Fluorescent PCR (QF-PCR)

QF-PCR is a fast and accurate method to detect chromosomal abnormalities. It has been introduced in the prenatal setting as a rapid and high confidence test for the detection of trisomy 13, 18, 21 as well as sex chromosomal abnormalities. It can be performed on DNA derived from amniotic fluid, CVS or peripheral blood. Usually, it is offered in conjunction to karyotype analysis since QF-PCR only targets specific chromosomes, even though studies have shown that QF-PCR directed to common aneuploidies could replace karyotyping [36]. The first clinical report was in 2001 and since then many countries in Europe offer it routinely as a diagnostic test [37, 38].

QF-PCR is based on the assumption that the amount of DNA product within the exponential phase of the PCR reaction is proportional to the quantity of the initial target. Thus, PCR reaction cycles must be limited to avoid reaching the amplification plateau. In addition, the regions selected to be amplified must be highly polymorphic,

such as short tandem repeats (STRs) [38]. STRs are DNA sequences of 2-5 nucleotides long that are repeated from few to many times in tandem order giving a high rate of heterozygosity among different allelic forms and individuals [39]. Fluorescently labelled primers flanking these regions are designed and are PCR amplified. Multiplexing of primer pairs often takes place in order to cover as many loci on the chromosomes under investigation as possible. The fragments are then separated by capillary electrophoresis using DNA sequencer and the data produced are analyzed in order to detect dosage ratios. In normal heterozygotes the dosage ratio should be within 0.8-1.4, while trisomies would be detected either as three peaks with 1:1:1 ratio or as a pattern of two peaks with a ratio of about 2:1 (Figure1.4).

Conclusively, QF-PCR is an inexpensive, fast, accurate and relatively high throughput methodology. On the other hand, some of its major caveats are its inability to detect low levels of mosaicism and the fact that the clinical significance of results is sometimes debatable, especially when findings involve chromosome X, while its targeted nature prevents it from becoming a stand-alone test.

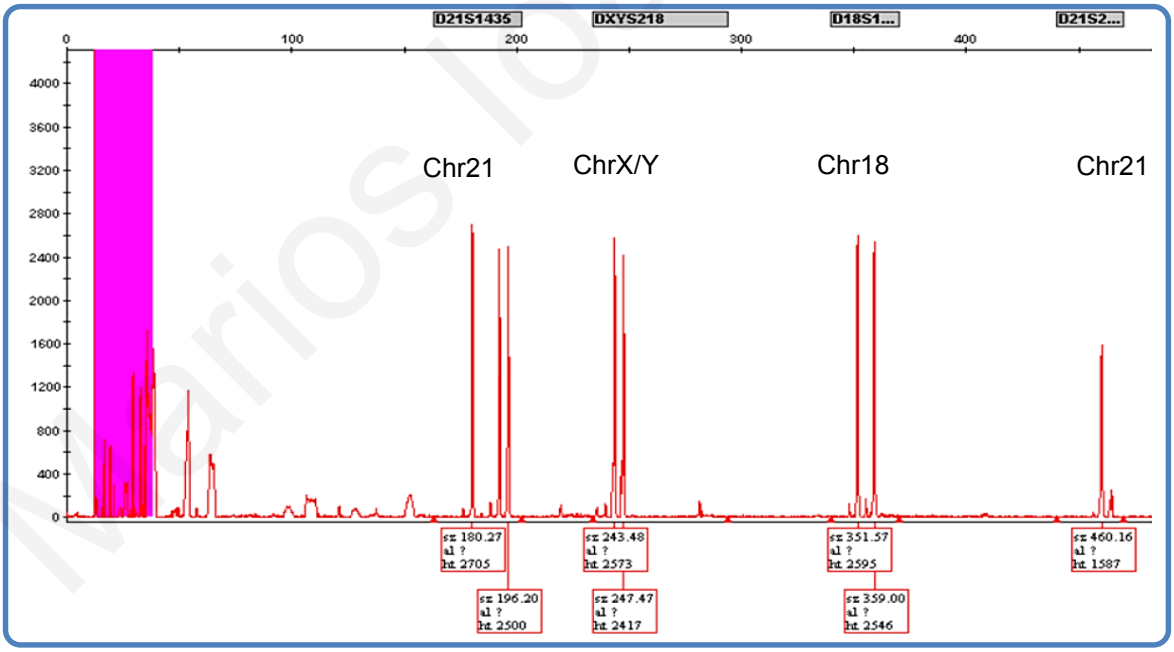


Figure 1.4 QF-PCR analysis performed on a trisomy 21 case.
Highly polymorphic markers on chromosomes 18, 21 X and Y are PCR amplified and analyzed by capillary electrophoresis. Biallllelic ratios (1:1) indicate normal copy number for the specific chromosome while triallelic ratio (1:1:1 or 2:1) indicate presence of an extra copy (trisomy 21) (Photo courtesy of Department of Cytogenetics and genomics).

1.2.2.4 Multiplex-Ligation Probe Amplification (MLPA)

MLPA is a relatively new molecular method used for diagnosis. Similarly to QF-PCR, it is a fast and accurate method for the prenatal diagnosis of the most common chromosomal aneuploidies. Its principle lies in the simultaneous amplification of more than 60 probes specific for different regions in the genome, using only one set of primers. More specifically, for each region to be amplified there are two sequence specific probes hybridized adjacent to each other. Each one, in addition to the target specific oligonucleotide, it consists of an overhang that is used as a recognition sequence for the universal primers. Furthermore, in order for each amplicon to have a unique size, a stuffer sequence is present on the right probe, flanked by the target specific and the primer recognition sequences (Figure 1.5A). After denaturation of the test DNA, the probe mix is hybridized and the two probes are ligated. Then, probes are simultaneously amplified using the same primer pair, one primer of which is fluorescently labelled [40]. The amplicons are then separated by fragment electrophoresis and visualized in an electropherogram (Figure 1.5B). The height of each peak is compared to a control region or a reference normal sample. Deletion is apparent as a decrease in the peak height, while duplication is reflected by an increase in the fragment's size.

The MLPA approach is routinely used in the prenatal diagnosis as well as in the diagnosis of a large panel of microdeletion/microduplication syndromes such as Prader Willi, Angelman, di George, Williams as well as other common diseases such as Acute Lymphoblastic Leukemia, Polycystic kidney disease etc.

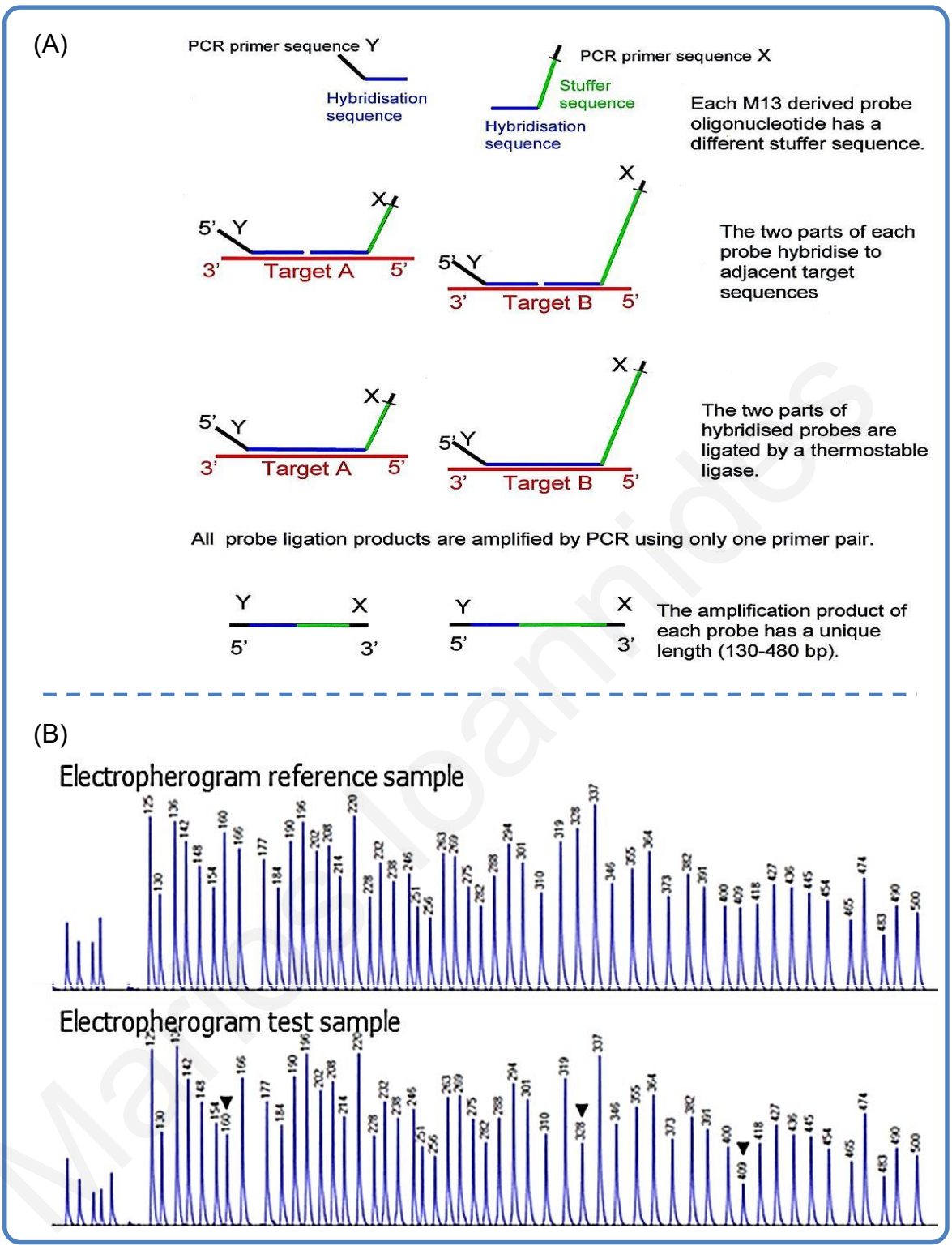


Figure 1.5 A: Outline of the MLPA procedure.**B:** Electropherogram of amplified products separated by capillary electrophoresis. Fragments indicated by an arrow appear to be deleted as compared to the reference control [41].

1.2.2.5 Fluorescent in-situ Hybridization (FISH)

FISH is a diagnostic method that combines cytogenetic and molecular techniques. Fluorescently labelled probes, specific for the regions/chromosomes under investigation, are hybridized on metaphase chromosomes (metaphase FISH) or nuclei (interphase FISH) fixed on a glass slide. Subsequent washes are then performed to remove unincorporated probes and the chromosomes are counterstained with a fluorescent dye that binds non-specifically to double stranded DNA. The fluorescent signal is then visualized under a fluorescent microscope that is capable of exciting the probe's fluorescent dye. A control probe is always used in order to ensure successful probe hybridization and avoid any false positive results (Figure 1.6). Absence of probe signal would suggest deletion, while amplifications or presence of a double signal would suggest duplication of the specific region. Due to its relatively low resolution, FISH is now used as confirmatory test of karyotyping or rarely as a diagnostic test of suspected abnormalities that are beyond the resolution of conventional karyotyping. Commercially available probes allow the detection and diagnosis of chromosomal aneuploidies on chromosomes 13, 18, 21, X and Y in addition to common microdeletion/microduplication syndromes and subtelomeric regions.

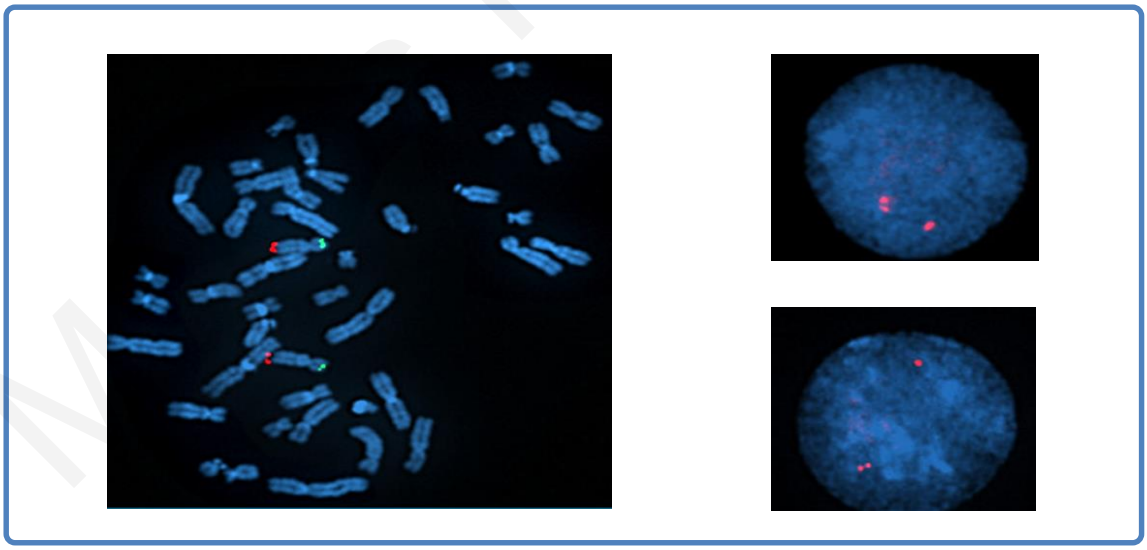


Figure 1.6 FISH performed on metaphase chromosomes (metaphase FISH) and nuclei (interphase FISH). Probes specific for the subtelomeric regions of chromosome 10 are labeled with two different fluorescent dyes (p arm with FITC and q arm with TRITC). In this case, the subtelomeric region of the q-arm (long arm) appears to be duplicated. Due to chromosomes being in a highly condensed form during metaphase, duplication is clearer on the interphase cells. (Photo courtesy of the Department of Cytogenetics and Genomics).

1.3 Non-Invasive Prenatal Diagnosis

As mentioned previously, ultrasonography combined with biochemical screening of maternal serum from the first or second trimester give an early relative risk detection of an abnormal pregnancy, taking into consideration the maternal age, reducing in this way the parental anxiety of high risk pregnancies. Even though there is no side effect for either the mother or the baby, the main disadvantage of the screening methods is that there is a 5% of false positive results, while not all affected fetuses can be detected [16, 17]. At present, the gold standard in prenatal diagnosis involves invasive procedures with CVS being the method of choice for the first trimester and amniocentesis for the second. Invasive procedures are offered only at high risk pregnancies and are associated with a significant rate of spontaneous miscarriages despite relatively high detection rates [24]. Thus, the need for minimizing any potential pregnancy risks has led to the discovery of fetal cells and free fetal DNA in the maternal circulation, an important achievement towards the development of non-invasive prenatal diagnosis (NIPD) methodologies that are safer, more effective and available for all pregnancies.

1.3.1 Fetal Cells in Maternal Circulation

Several studies focused their interest in the detection of fetal cells in the maternal circulation during pregnancy. Initially scientists have turned their investigations in the fetal unique cell types, such as trophoblasts, but their low amount in the maternal blood and the absence of efficient antibody specificity for the surface antigens caused their further use difficult [42, 43]. Later, nucleated erythrocyte cells became the fetal cells of choice due to their abundance during the first trimester and their clearance shortly after pregnancy [44-46]. Despite that, it was shown that their numbers increase during certain conditions such as preeclampsia, in addition to the fact that these nucleated cells are not purely of fetal origin [47, 48]. Later, the presence of Y chromosome specific sequences after delivery was detected in women with previous male pregnancies [49], an alarming finding for the implementation in NIPD.

In an attempt to measure the amount of fetal cells in the maternal blood using molecular techniques such as FISH and PCR, scientists found that the total

nucleated cells varied from as low as two to six cells per milliliter of maternal blood in normal pregnancies while this number was significantly higher in abnormal pregnancies [50, 51].

In conclusion, the presence of fetal cells in the maternal circulation, even though promising, the small number of fetal cells, the persistence of these cells for long periods of time in the maternal circulation, as well as the lack of reliable methods to isolate pure fetal cells, rendered this approach inefficacious for further investigation and consequently inappropriate for its clinical application.

1.3.2 Cell Free Fetal DNA (cffDNA)

In 1997, Lo et al. discovered the presence of free fetal DNA in maternal plasma [52]. This breakthrough came as alternative to the discovery of small amount of free fetal cells in the maternal circulation and played a dramatic role in the development of NIPD.

In the maternal circulation the majority of the total free DNA is of maternal apoptotic hematopoetic cells [53]. Different sources of fetal tissue have been speculated to be the origin of total cffDNA including apoptotic trophoblastic cells or fetal hematopoetic cells [54]. Evidence coming from studies that used anebryonic pregnancies have shown that the source of fetal DNA is of placental origin most probably derived from the bidirectional trafficking of apoptotic placenta cells into the maternal circulation [54-57].

CffDNA can be detected as early as 5-7 weeks of gestation with high specificity and sensitivity showing variable amounts between pregnancies [58, 59]. It was demonstrated that the cffDNA fraction increases as the pregnancy progresses ranging from 3.4% during early to 6.2% during late pregnancy [60]. These values though could be underestimated since more recent studies reported a median cffDNA fraction of 10% and 11.6% in first trimester pregnancies using droplet digital PCR and massive parallel sequencing (MPS) respectively [61, 62].

In addition to its reliability and ease of detection, cffDNA has been shown to persist only for a few minutes after delivery [63-65]. This property added to its usefulness

in NIPD since future pregnancies would not have a risk of false positive results due to possible carryover DNA from previous pregnancies.

1.3.3 DNA Methylation

DNA methylation in vertebrates is a conserved epigenetic mark that involves the addition of a methyl group on carbon 5 of cytosines present in CpG dinucleotides. The methylation process is mediated by three DNA methyltransferases. DNMT3A and DNMT3B are responsible for *de novo* methylation of DNA, essential for embryonic development. They have also been associated with methylations of retrotransposon sequences and satellite repeats in pericentric regions. DNMT1 is a maintenance methyltranferase responsible for copying existing methylation during DNA replication in somatic cells [66].

Mammalian genomes appear to be globally methylated -that is, all categories of DNA sequences are potential targets of CpG methylation - with the exception of regions with high density of CpG dinucleotides, namely CpG islands (CGIs) [67]. Frequent methylation targets include regulatory regions of genes such as the promoter regions and first exons. In general, promoters and gene bodies of actively transcribed genes have been shown to be unmethylated and methylated respectively, whereas gene silencing has been correlated with methylation of the regulatory regions [68, 69]. Supporting evidence comes from studies on the human X chromosomes where it was shown that even though X_i is methylated on promoter CGIs, a situation that maintains silencing, X_a was shown to be more methylated, mainly in regions such as gene bodies, hinting that gene body methylation is indeed involved in the transcriptional activity of genes [70, 71]. It is suggested that transcriptional repression is caused by DNA methylation in combination with histone deacetylation and complex nucleosomal remodeling factors that render the chromatin inaccessible to transcription factors [72, 73] (Figure 1.7).

As a consequence, DNA methylation plays a critical role in transcriptional regulation, genomic imprinting, X inactivation, chromosomal rearrangements and aging [72]. Furthermore, early embryonic lethality of mice with inactive methyltranferases suggest implication of DNA methylation in embryonic development [74]. Disruption of normal DNA methylation is a critical step in the development of diseases, hence, a variety of complex disease states ranging from neurological to psychiatric

diseases have been associated with aberrations in DNA methylation [75, 76]. In cancer, global hypomethylation of cancerous tissues ultimately cause oncogene activation, loss of imprinting [77-79], while studies on breast and colon cancer have shown that tissue specific hypermethylation of CGI and CGI shores, or transcriptional activation of retrotransposons, result in gene silencing of tumor suppressor genes and chromosomal instability [66].

Due to its regulatory and gene expression implications, DNA methylation has also been associated with tissue specificity. Tissue specific methylation heterogeneity was initially suggested by Rakyan et al. who analyzed the methylation profile of seven tissues from 32 individuals, focusing their interest in 90 genes within the major histocompatibility complex (MHC) [80]. Later, different studies confirmed that methylation patterns across different tissues vary significantly [81-84]. Determination of tissue specific differentially methylated regions (tDMRs) has been utilized for the discovery of biomarkers for disease progression and prognosis, especially in the field of cancer research, disease detection and response to treatment [76, 85-88]. Furthermore, several studies have shown that despite tissue specific patterns, methylation status varies between individuals. It has been suggested that several factors such as diet, stress, age and other stochastic events influence the methylation patterns of individuals [89-91]. This variation has been shown to be related to gene expression levels and as such has been associated with phenotypic variation, promoting population fitness [92, 93].

A variety of techniques have been developed towards the detection and analysis of methylation patterns across the genome. Currently there are three main approaches for the determination of the DNA methylation status. First, with the utilization of methylation sensitive restriction digestion, scientists enrich for methylated regions that contain the enzyme's restriction site. Even though readily implemented, it limits the number of regions suitable for testing to the regions containing the restriction sites and is highly susceptible to false positive results due to incomplete digestion of the DNA [94]. A second approach involves bisulfite treatment of DNA in order to convert unmethylated cytosines to uracils, leaving the methylcytosines unaffected. Bisulfite conversion is usually followed by methylation specific PCR or sequencing. It has become the gold standard for methylation analysis because of its base pair resolution of DNA methylation and quantitative nature [95] , but its implementation faces several drawbacks. Bisulfite treated DNA has reduced sequence complexity

making the sequencing alignment cumbersome due to the conversion of unmethylated cytosines to thymines after PCR amplification. Furthermore, the chemical treatment of DNA causes a substantial degradation and incomplete conversion of methylated cytosines to uracils causing false positive methylation calls [96]. The third approach is an affinity based enrichment of methylated DNA using either antibodies specific to methylcytosines -methylated DNA Immunoprecipitation (MeDIP)- or use of methyl-binding proteins (MBD). The MeDIP approach has been used extensively as it is an inexpensive, reproducible method for detection of whole genome methylated regions [97, 98]. All of the aforementioned approaches have been widely utilized for methylome analysis and biomarker discovery in combination with other techniques in order to assess enrichment such as microarrays (MeDIP-Chip [99]), qPCR (MeDIP-qPCR [100]), sequencing (MeDIP-seq [101], Bis-seq [102]) etc.

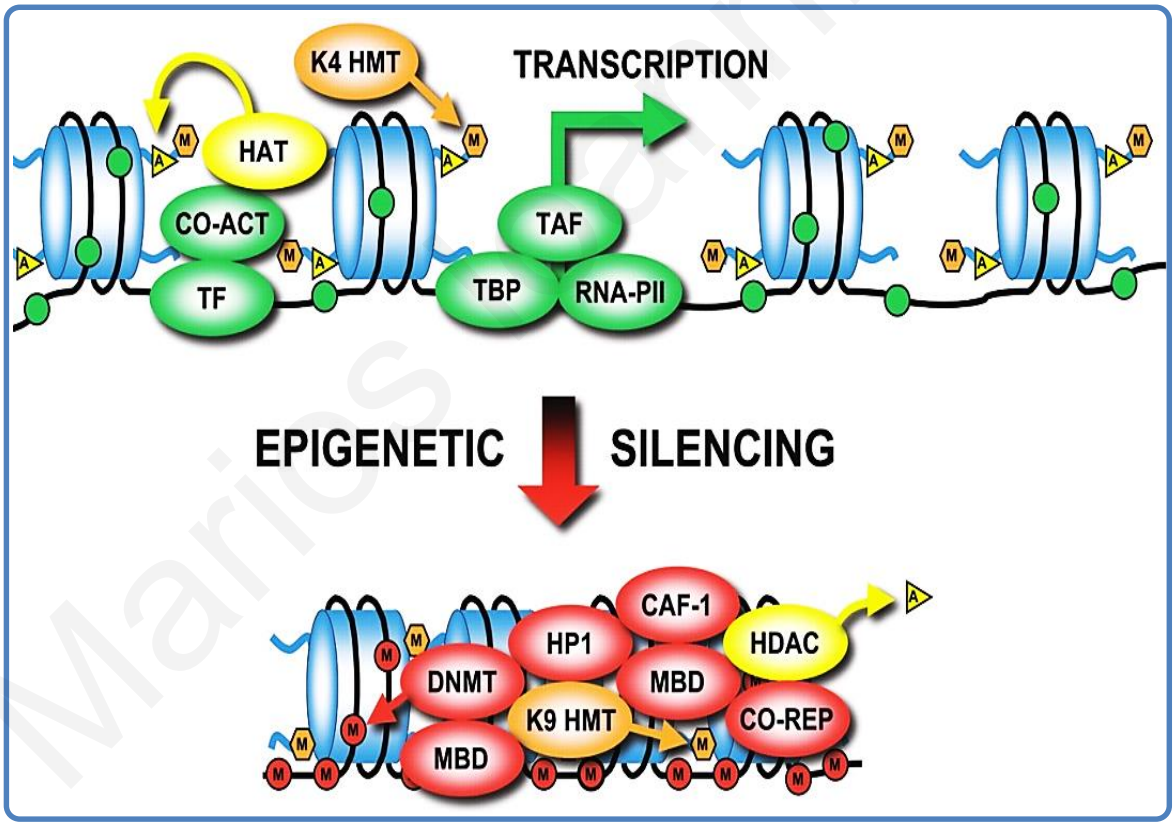


Figure 1.7 Suggested model for transcriptional silencing due to epigenetic changes. Top: Transcriptionally active gene promoter indicated by loosely spaced nucleosomes (Blue cylinders). Bottom: Transcriptional inactivation indicated by chromatin remodeling with nucleosomes more tightly packed. DNA methyltransferases (DNMT-red oval) and histone deacetylases (yellow oval) along with nucleosomal remodeling render the transcription factors inaccessible [103].

1.3.4 Fetal Specific Biomarkers-a step towards NIPD

In the prenatal setting, biomarker investigation and DMR identification was launched after the discovery of cffDNA in maternal circulation during pregnancy. Despite its low concentration in the high maternal background, scientists focused their investigation in the prenatal diagnosis of several single gene inherited diseases and disorders using unique DNA sequences that could be readily distinguishable in the maternal circulation. The first large-scale application of non-invasive prenatal test was the fetal RhD genotyping from maternal RhD negative plasma [104]. Fast and reliable results, especially during the second trimester of pregnancy, led to the incorporation of the test in the UK and later in other centers in Europe [105-107]. NIPD has also been used as a tool for the exclusion of sex linked disorders. Different research groups have applied real time qPCR targeting SRY or DYS-14 sequences, in order to detect male fetuses. This approach was found to be highly specific and sensitive and its use in clinical practice was imperative as possibly affected male fetuses from a carrier mother could easily be detected [108-110]. In order to overcome the limitation of detecting only 50% of the affected pregnancies and the presumption that absence of the sequences constitute a female fetus, hence decreasing any false negative results, panels of polymorphic sequences that are present in the fetus but not the mother are used to confirm the presence of the predicted female fetal DNA in the plasma [111, 112]. Furthermore, similar approaches are implemented for detection paternally inherited alleles for NIPD of several monogenic autosomal dominant diseases such as achondroplasia, myotonic dystrophy and Huntington's disease. On the other hand autosomal recessive diseases such as β-thalassemia and cystic fibrosis, cannot be detected as easily especially in couples that share an identical mutation. In such cases, determination of the heterozygote state of the fetus is determined by ascertaining the presence of the paternal wild type allele [113, 114].

The need to discover fetal biomarkers that can be detected in all pregnancies irrespective of fetal gender and polymorphisms still remains a challenge. Towards this goal several groups, taking advantage of the methylation differences between placenta derived cffDNA and cell free maternal DNA have focused their investigation on the detection of highly specific fetal biomarkers that can be potentially be used in the detection of fetal aneuploidies. In an initial effort Chim et al. [64] analyzed the methylation status of the promoter region of the *SERPINB5*

gene on chromosome 18, known to be hypomethylated in the placenta and hypermethylated in the whole blood. Using bisulfite sequencing they confirmed the methylation differences between the two tissues. Furthermore, using methylation specific qPCR (qMSP) on the bisulfite converted plasma derived-DNA they successfully demonstrated that the hypomethylated fetal derived *SERPINB5* sequences could be detected in the maternal plasma in all three trimesters of pregnancy. The unmethylated fetal specific sequences were cleared from maternal plasma 24 hours after delivery while the (methylated) maternal background was present before and after delivery [64].

Later, Chiu et al. [115] used a similar approach in order to investigate the *RASSF1A* gene located on chromosome 3. Placenta and maternal blood samples were obtained from all three trimesters of pregnancy and it was shown that unlike the *SERPINB5*, *RASSF1A* is hypermethylated in placenta when compared to the maternal blood cells**.** In another study [116], the same research team combined bisulfite sequencing with single nucleotide primer extension and mass spectrometry with which they identified 22 candidate CGIs that were differentially methylated in placental tissue and maternal blood. Further selection criteria allowed them to distinguish two regions (U-*PDE9A* and U-CGI-137) as potential biomarkers. They demonstrated that both regions can be used as fetal biomarkers as they were shown to be unmethylated in the placenta and methylated in the maternal plasma [116]. In search for markers on chromosome 21, Old et al. adopted a different method. Their initial screening was based on three different approaches, with a common criterion being the presence of one or more methylation sensitive restriction sites. Confirmation of the selected regions using bisulfite sequencing and qMSP allowed them to identify one biomarker upstream of the *AIRE* gene showing hypermethylation in placenta samples and hypomethylation in maternal peripheral blood cells [117].

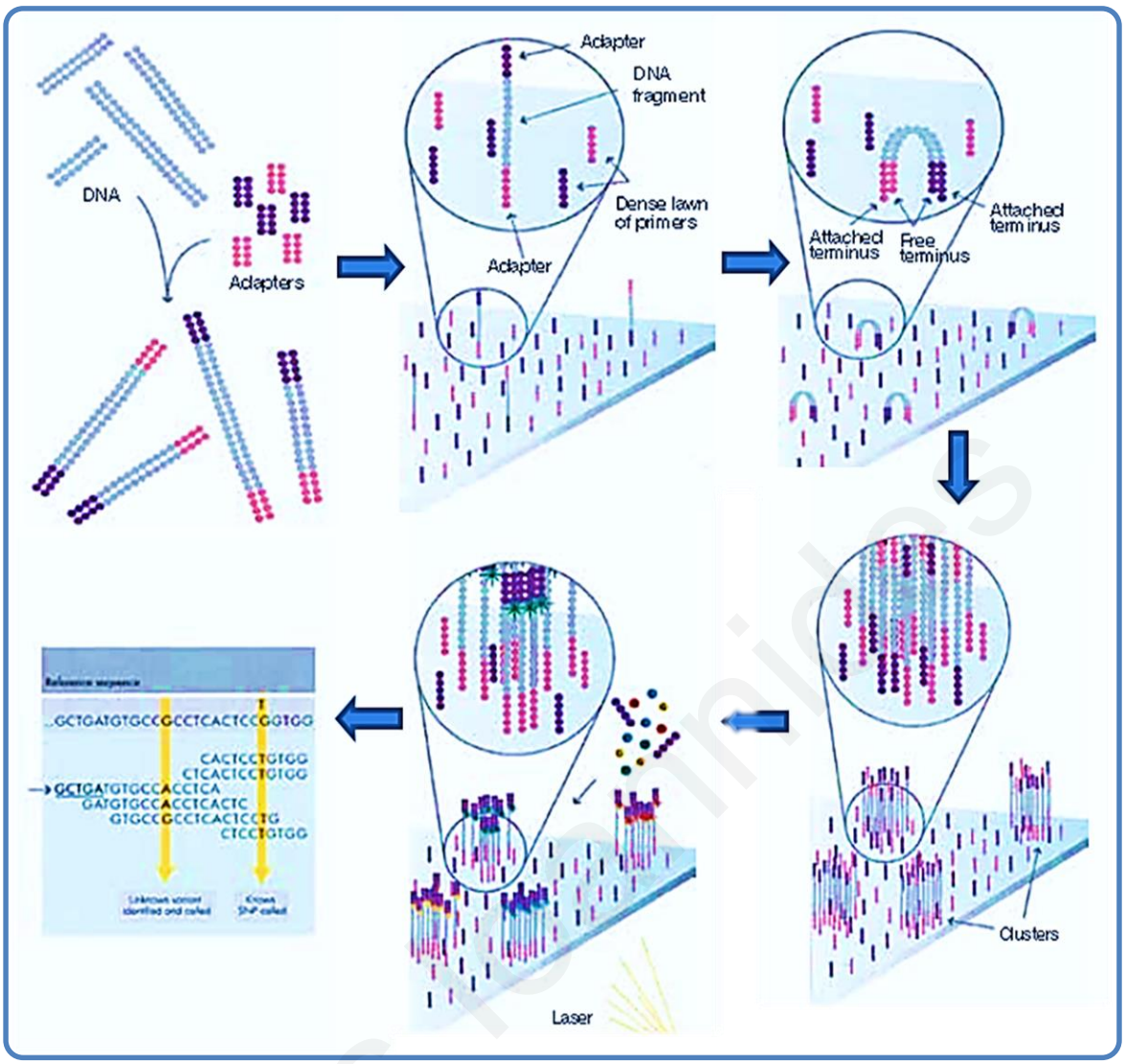
In order to overcome the limitations of bisulfite sequencing and restriction digestion Papageorgiou et al. utilized the methylation DNA immunoprecipitaion approach coupled with high resolution tiling oligonucleotide analysis (MeDIP-Chip) for the enrichment of hypermethylated fetal DNA [100]. In this study, chromosome wide detection of differentially methylated regions was performed on chromosomes 13, 18, 21, X and Y. For this, DNA from normal female peripheral blood and placental sample from first and third trimester pregnancies were subjected to

immunoprecipitation and then each sample (genomic and precipitated) was co-hybridized on an array chip specific for each chromosome tested. The results clearly showed differential methylation between female peripheral blood and fetal DNA. Interestingly, approximately equal number of regions were hypo and hypermethylated and relatively constant between first and third trimester. This approach allowed the identification of more than 10,000 DMRs on chromosomes 13, 18, 21, X, and Y, independent from methylation sensitive restriction sites, CGIs, or promoter regions. Nine of these DMRs on chromosomes 18 and 21 were screened with qPCR and were found to have higher relative fold enrichment in the 1st and 3rd trimester placenta than female peripheral blood.

Recently, in another attempt to identify DMRs between placental and maternal blood, affinity base enrichment using MBD was used to enrich methylated regions. Subsequent amplification and hybridization on a CGI microarray revealed a great number of genome wide DMRs. The top 7 DMRs were located within disease causing genes as confirmed by bisulfite sequencing and combined bisulfite restriction analysis (COBRA). Based on their findings the authors concluded that the DMRs identified can potentially be developed into molecular biomarkers for the detection of monogenic diseases, but first a larger sample set needs to be screened in order to confirm the results reported [118].

Recently, the advent of next generation sequencing (NGS) changed the landscape of the field. This high throughput technology does not require enrichment of DNA. Instead, millions of fragments of DNA are clonally amplified in clusters and then both strands are sequenced in parallel. The resulting reads (sequenced fragments) are computationally aligned to the reference genome and depending on the project's objectives, downstream bioinformatics and biostatistical analyses are performed [119]. The most widely used sequencing technique is sequencing by synthesis. Briefly, for each sequencing cycle all four deoxyribonucleotides (dNTP) are present, each one labeled with a different dye. The dye also serves as a transient terminator of polymerization. Once each dNTP is incorporated to the nucleic acid chain, the fluorescence is imaged and then the terminator is enzymatically cleaved in order to allow the next dNTP to be incorporated during the next cycle. This allows the parallel sequencing of thousands of fragments with high precision (Figure 1.8).

Two different research teams combined bisulfite conversion [102] and MeDIP [120] with NGS in order to shed light on the methylation profiles of maternal and fetal tissues. First, Lun et al.[102] were able to recover the fetal methylome non-invasively from pregnant women using as a reference the methylome of non-pregnants and 1st and 3rd trimester placentas. Also, a great number of genome wide DMRs were identified, most of them being hypomethylated in the placenta, confirming the hypomethylation status of the placenta. As the authors conclude the non-invasive deduction of the fetal methylome can be used in the future as a platform not only for pathophysiological studies of pregnancies but also as a direct method for NIPT [102]. The second study [120] using the MeDIP-seq approach interrogated maternal blood and paired 1st trimester placentas in order to identify whole genome DMRs. DMRs were observed in all chromosomes most of which were hypermethylated in the placenta. In addition, the DNA methylation profiles for the two tissues were assessed using beadchip arrays resulting in identification of thousands of differentially methylated CpG sites in the placenta. In an effort to identify DNA markers associated with chromosomal aneuploidies, they identified two DMRs on chromosome 13, three DMRs on chromosome 18 and three DMRs on chromosome 21 that were confirmed with both platforms. Furthermore, they report DMRs associated with chromosomal diseases on various chromosomes, a step towards the improvement of NIPT in microdeletion syndromes [120].



(http://support.illumina.com/content/dam/illumina-arketing/documents/products/techspotlights/techspotlight_sequencing.pdf)

Figure 1.8 Outline of Next Generation Sequencing procedure.
Adapters are initially ligated on randomly fragmented DNA. Following denaturation, single stranded DNA binds randomly to the flow cell consisting of a lawn of primers complementary to the adapters. DNA fragments are then clonally amplified (solid phase bridge amplification) to generate clusters consisting of more than 1000 identical copies of each DNA fragment. Each subsequent sequencing cycle begins by addition of all nucleotides bound to a reversible terminator. Once the nucleotides are incorporated they emit a specific fluorescent signal which is captured resulting in the identification of each base. Bioinformatics alignment on the reference genome is then performed for further investigation according to the project's objectives.

Table1.1 Fetal specific DMRs for aneuploidy detection and methods of their identification

DMR	Method of identification / confirmation	Chromosomes Involved	Characteristics
SERPINB5	Bisulfite Conversion-Methylation specific qPCR	18	Hypomethylated in fetus
RASSF1A	Bisulfite Conversion-Methylation specific qPCR	3	Hypermethylated in fetus
UPDEA	Bisulfite sequencing-single primer extension	21	Hypomethylated in fetus
U-CGI-137	Bisulfite sequencing-single primer extension	21	Hypomethylated in fetus
AIRE	Restriction Digestion-Bisulfite conversion-Methylation specific qPCR	21	Hypermethylated in fetus
Thousands of DMRs	MeDIP-aCGH/qPCR	13,18,21,X,Y	Hypermethylated in fetus
7 DMRs	MBD-aCGH/COBRA		Hypermethylated in fetus
Hundreds of DMRs	Bisulfite conversion-NGS	whole genome	Hypermethylated in fetus
8 DMRs	MeDIP-NGS/Bead chip array	13, 18, 21	Hypermethylated in fetus

1.3.5 Non-Invasive Prenatal Diagnosis of Fetal Aneuploidies

Although identification of fetal specific biomarkers has been possible, the limited amount of cffDNA in the excess of maternal background has proven to be a challenge in the detection of aneuploidies [121]. Taking advantage of the physical and molecular characteristics of the cffDNA, research groups have managed to select fetal specific loci in order to assess for the chromosome dosage of the fetus as compared to the normal maternal DNA. In an initial effort Lo et al. used placenta specific *PLAC4* mRNA informative heterozygous SNPs (RNA-SNP approach) in order to detect chromosome 21 dosage differences. They utilized single primer extension in combination with mass spectrometry and they successfully detected the trisomy 21 cases with 90% sensitivity and 96.5% [122]. To overcome the limitation of applying this approach only to pregnant women with a fetus heterozygous for the SNP, they suggested that combination of the RNA-SNP approach with the measurement of total *PLAC4* mRNA can be used potentially in all pregnancies for the detection of trisomy 21, as the mRNA concentration would increase in the presence of an extra chromosome 21 [123]. The technological advance of digital PCR and its implementation in the prenatal setting has allowed discriminating fine quantitative differences for the assessment of chromosome dosage. A pilot study utilized digital PCR by comparing relative chromosome dosages (RCD) between chromosome 21 and a reference chromosome. Classification of trisomy 21 cases was successful for all abnormal placenta tissues. Despite this, in order to demonstrate the application of this method in samples containing low fetal DNA concentrations, it was shown that although trisomy 21 was correctly classifiable, a large number of digital PCR reactions were required [124]. In a proof of concept study, Tong et al. [65] applied the epigenetic-genetic approach (EGG) for chromosome dosage analysis. Methylation restriction enzyme digestion was used in combination with digital PCR to interrogate the fetal-specific hypermethylated promoter region of the *HLCS* gene and the *ZFY* locus on chromosome Y. EGG analysis of eight maternal plasma samples from each of the three trimesters of pregnancy supported this approach as a potential method to detect trisomy 21, since all but one sample were correctly identified. As the authors discuss, EGG has an advantage over the previous approaches in which implementation was restricted due to the rarity of having informative SNPs present.

Later, the same group used paternally derived fetal SNPs on chromosome 14 to explore the possibility of applying this method to both male and female fetuses [125].

In 2011, a MeDIP-qPCR approach was proposed for the NIPT of trisomy 21 [126]. Initially, they selected 12 previously identified DMRs that showed hypermethylation in the placenta and hypomethylation in the female peripheral blood [100], and applied the MeDIP-qPCR methods on 20 normal and 20 trisomy 21 pregnancies. Based on the DNA methylation ratio of each DMR a diagnostic formula was then developed using the eight most statistically significant DMRs. This formula provided correct classification of 40 additional pregnancies, 14 of which were trisomy 21, with 100% sensitivity and 100% specificity opening the way for the development of a new NIPD for trisomy 21 and at the same time emphasizing the need of multiple DMRs for accurate diagnosis. Recently, a new version of the diagnostic formula was developed using 75 samples and the accuracy and reproducibility of the approach was confirmed, by correctly classifying 100 cases with 99.2% sensitivity and 100% specificity [127].

Several groups focused their investigation on the development of NIPT using NGS. Initially, two independent groups, used shotgun sequencing of maternal plasma to accurately diagnose all fetal aneuploidies in their sample sets by normalizing the number of sequence tags on the chromosome of interest to the number of tags on all autosomes using z-score analysis [128, 129]. In a blind validation study using 753 samples, Chiu et al., in order to compare the sensitivity of the method, they showed that low throughput strategies would detect trisomy 21 pregnancies with higher statistical confidence resulting in substantially higher sensitivity and specificity [130]. In accordance with the above studies, Palomaki et al. and Ehrich et al., using a minor protocol modification from the previous studies and sample multiplexing, carried out NGS on 449 and 1696 high risk pregnancies respectively with a successful classification of trisomy 21 with less than a 1% false positive rate [131, 132]. More challenging than trisomy 21 appeared to be the detection of trisomy 13 and 18, due to the relative lower average of GC content of these two chromosomes [129, 133-135]. Despite that, the sensitivity and specificity of the method was dramatically improved by utilizing different statistical algorithms than the z-score calculation [135], or incorporating a GC correction in the z-score calculations [133, 134]. In order to reduce sequencing costs, complexities and avoid a large amount of redundant data generated by whole genome NGS, researchers

developed a novel approach for applying NGS in selected regions [136-139]. In a proof of concept study, Sparks et al. utilized the digital analysis of selected regions for the detection (DANSR) of trisomy 18 and 21. DANSR is a ligation based method that enables the simultaneous amplification and NGS of several regions of the genome. The scientists ensured hybridization specificity via the use of three locus specific oligos for each region and simultaneous amplification of the selected region by incorporation of universal primer recognition sites onto the flanking oligos [139]. Using DANSR, they interrogated 384 loci on chromosomes 18 and 21 and successfully classified all aneuploidy samples in a total of 298 high risk pregnancies, including 39 trisomy 21 and seven trisomy 18 samples. Subsequent validation studies of the same group showed high sensitivity of the method with less than 1% false positive rate. Furthermore, inclusion of polymorphic loci in the assay enabled the estimation of the fetal fraction, which in combination with a novel algorithm improved the assay performance and increased the confidence calls of abnormal samples [137, 138].

Table1.2 Different approaches used for prenatal diagnosis

Type of Prenatal Diagnosis	Method	Advantages	Disadvantages
Non-Invasive Screening	Estimation of levels of specific markers in maternal serum in combination with U/S findings	No risk for the fetus Offered to all pregnancies	High false positive rate
Invasive Diagnosis	Cytogenetic and other molecular tests	Definitive diagnosis	0.5-1% risk of fetal loss Offered to high risk pregnancies
Non-Invasive Diagnosis	NGS	No risk for the fetus Highly validated method Clinically implemented	Expensive Complex Bio statistical and Bioinformatics analyses
	MeDIP-qPCR	Validated method Cheap	Not clinically implemented
	RNA-SNP approach	No risk for the fetus	Low sensitivity and specificity Polymorphism Dependent Not validated
	Relative Chromosome Dosage approach(RCD)	No risk for the fetus Polymorphism independent	Not validated Low sensitivity and specificity Multiple replicates needed for statistically significant results
	Epigenetic Genetic approach(EGG)	No risk for the fetus Polymorphism independent	Not validated

1.4 Aim and Hypothesis

The main objective of our study is the identification and characterization of fetal specific biomarkers for chromosomes 13, 18, 21 and X that can be used in NIPD of most common aneuploidies associated with high risk pregnancies.

The study is separated into three main stages: The first stage of the study involves the characterization of new DMRs and the investigation of tissue specificity and methylation variability using existing array data sets. The second stage involves the design of ultra-high resolution aCGH chip for chromosomes 13, 18, 21 and X and the characterization of new DMRs using MeDIP in combination with aCGH (MeDIP-Chip). Finally, the third stage involves the utilization of MeDIP in combination with NGS (MeDIP-seq) in order to confirm the DMRs obtained from the MeDIP–Chip.

Based on the aforementioned project's objectives we hypothesized that, for the first stage of the study, several DMRs would be confirmed with the potential to be incorporated in the future in a diagnostic formula for the NIPD of all trisomies. In addition, it was expected that the methylation status between the fetal and maternal tissues would be distinct despite the degree of methylation variability between the two tissues. By utilizing the ultra-high resolution aCGH in the second stage we predicted that the DMRs identified in the first stage would be confirmed and in addition, a more robust collection of DMRs would be obtained from all chromosomes under investigation and therefore our panel of DMRs towards their implementation in NIPD would be expanded. Finally, for the third stage of the study we hypothesized that MeDIP-seq and MeDIP-Chip approaches would provide consistent results and hence, most of the DMRs characterized using the MeDIP-Chip approach would be confirmed. Furthermore, since both approaches are affinity based, it was expected that they would detect methylated sites independent from CGIs or promoters regions in a wide range of CpG sites, reaffirming in this way the robust and reproducible nature of the MeDIP methodology.

2 MATERIALS AND METHODS

2.1 Human Samples and DNA preparation

Female non-pregnant peripheral blood samples (WBF) were obtained anonymously from normal non-pregnant females 20-40 years of age. 1st trimester (normal, trisomy 13, trisomy 18, trisomy 21 and monosomy X) chorionic villi samplings (CVS) were obtained from the Department of Cytogenetics and Genomics at the Cyprus Institute of Neurology and Genetics (Nicosia, Cyprus). Third trimester placentas were collected anonymously immediately after delivery. Protocols used for collecting samples for our study were approved by the appropriate Bioethics Committees, and informed consent was obtained from all participants. WBF, CVS and placenta samples were used to extract DNA using the QIAamp DNA blood midi kit or the QIAmp DNA mini kit according to the manufacturer's instructions (QIAGEN, Hilden, Germany). All CVS and placenta samples underwent Quantitative-Fluorescent PCR (QF-PCR) analysis in order to confirm their status.

For the DMR screening and characterization of inter-individual methylation variability of the DMRs obtained from previous array results [100], DNA from 50 WBF and 50 CVS were used. For the purposes of aCGH assays five WBF, five normal CVS, five of each abnormal group CVS (trisomy 13, 18, 21, monosomy X) and five normal 3rd trimester placentas were used. New DMRs were screened on DNA from six to eight WBF, normal CVS and abnormal CVS (only for chromosome 21).

2.2 MeDIP assays

2.2.1 Ligation Mediated PCR (LM PCR)

LM PCR and MeDIP assays for the first part of the project were conducted as described previously [100] (Figure 2.1A). Briefly, 2.5μg of genomic DNA were sonicated using the Bioruptor Twin sonication system (UCD-400, Diagenode, Liege, Belgium) into fragments, 300-1000bp in size. Fragment size was verified using

agarose gel electrophoresis (Figure 2.2). The fragments were blunt-ended using HPLC water, 1X NEB buffer 2 (New England BioLabs, Ipswich, UK), 1.2 µl of 10X bovine serum albumin (New England BioLabs) 1.2µl of 100 mmol/l dNTP mix (GE Healthcare, Little Chalfont, UK) and 1µl of T4 DNA polymerase (3U/µl; New England BioLabs) in a total volume of 121µl. 40µl of linkers (JW102 and JW 103) were then ligated onto the blunt ends by overnight incubation at 16˚C with 5µl T4 DNA ligase (New England) and 1X T4 DNA ligase buffer (New England) in a total volume of 71µl. Overhangs were subsequently filled in by incubating at 72˚C for 10 minutes with 1µl of 100 mmol/L dNTP mix (GE Healthcare), 1X PCR gold buffer (Roche, Mannheim, Germany), 1.5mmol/l MgCl2 (Roche), 11µl HPLC water and 0.5µl AmpliTaq DNA polymerase (Applied Biosystems, Branchburg, New Jersey, USA) in 20.5µl reaction mixture. 50ng of ligated DNA was removed and kept as input DNA (IN). The remaining ligated DNA (800-1200ng) was subjected to MeDIP using 3µg mouse anti-5'methylCytosine (a-5mC) antibody (Eurogentec , Saraing, Belgium). Hypermethylated DNA bound to a-5mC antibodies was magnetically captured using Dynabeads® M-280 Sheep Anti-Mouse IgG magnetic beads (Life technologies, Carlsbad, California, USA) and subsequently released using 1µl of proteinase K (Roche). LM PCR was performed on 12ng of each input and MeDIP DNA with 1X Advantage GC-melt buffer (Clontech, Mountain View, CA, USA), dNTP mix, Advnatage GC Genomic LA polymerase (Clonetech) and 1.2µM JW-102 primer as described in Papageorgiou et al [100].

Alternatively, for the purposes of MeDIP-Chip and subsequent screening of DMRs the MagMeDIP kit (Diagenode) was used according to the manufacturer's instructions with the following modifications: the DNA was fragmented and underwent the same procedure of linker ligation as previously mentioned. The DNA was subsequently subjected to immunoprecipitation as follows: 50µl of DNA (24ng/µl) was denatured using the appropriate buffers in a total volume of 90µl at 95^{o}C for five minutes and immediately placed on ice for another five minutes. 12µl were removed for input DNA and the rest (75µl) was incubated with 5µl of antibody mix and 20µl of washed magnetic beads. The IP-mix was placed on a rotating wheel at 4^{o}C overnight at 20rpm. The DNA-Ab-bead complex was subsequently washed four times with iced cold buffer as indicated (Diagenode). The tubes were kept on ice at all times. Recovery of the immunoprecipated fraction was completed using the Ipure kit (Diagenode). IP DNA was eluted from the magnetics beads twice by adding 50µl of elution buffer and rotating at 40rpm at room temperature for 15

minutes each time, while IN DNA underwent the elution process once by adding 12μl of elution buffer. In order to improve binding of DNA on the beads and reduce sample loss, 2μl of Carrier DNA was added along with 100 μl of 100% isopropanol and 15μl of beads in each IN and IP sample. Tubes were incubated for one hour at room temperature by rotating at 40rpm. Samples were then washed with the appropriate buffers (Diagenode) and finally eluted from the magnetic beads in 50μl of supplied buffer. 18.5μl of purified DNA was used for LM PCR in a 25μl reaction mixture at the following final concentrations: 1μM JW102 primer, 2.5U Accuprime Taq Polymerase (Life technologies) and 1X Accuprime PCR buffer II. The PCR was performed under the following conditions: initial denaturation at 95^{o}C for three minutes, 35 cycles consisting of denaturation step at 95^{o}C for one minute, annealing at 95^{o}C for one minute and extension for one minute at 72^{o}C and a final extension step at 72^{o}C for 10 minutes.

2.2.2 No Ligation Mediated PCR (Non-LM PCR)

For validation/characterization of DMRs without LM PCR (Non-LM PCR) linkers were not added to the sonicated DNA, thus samples underwent only the MeDIP procedure as described in the previous section (Figure 2.1B).

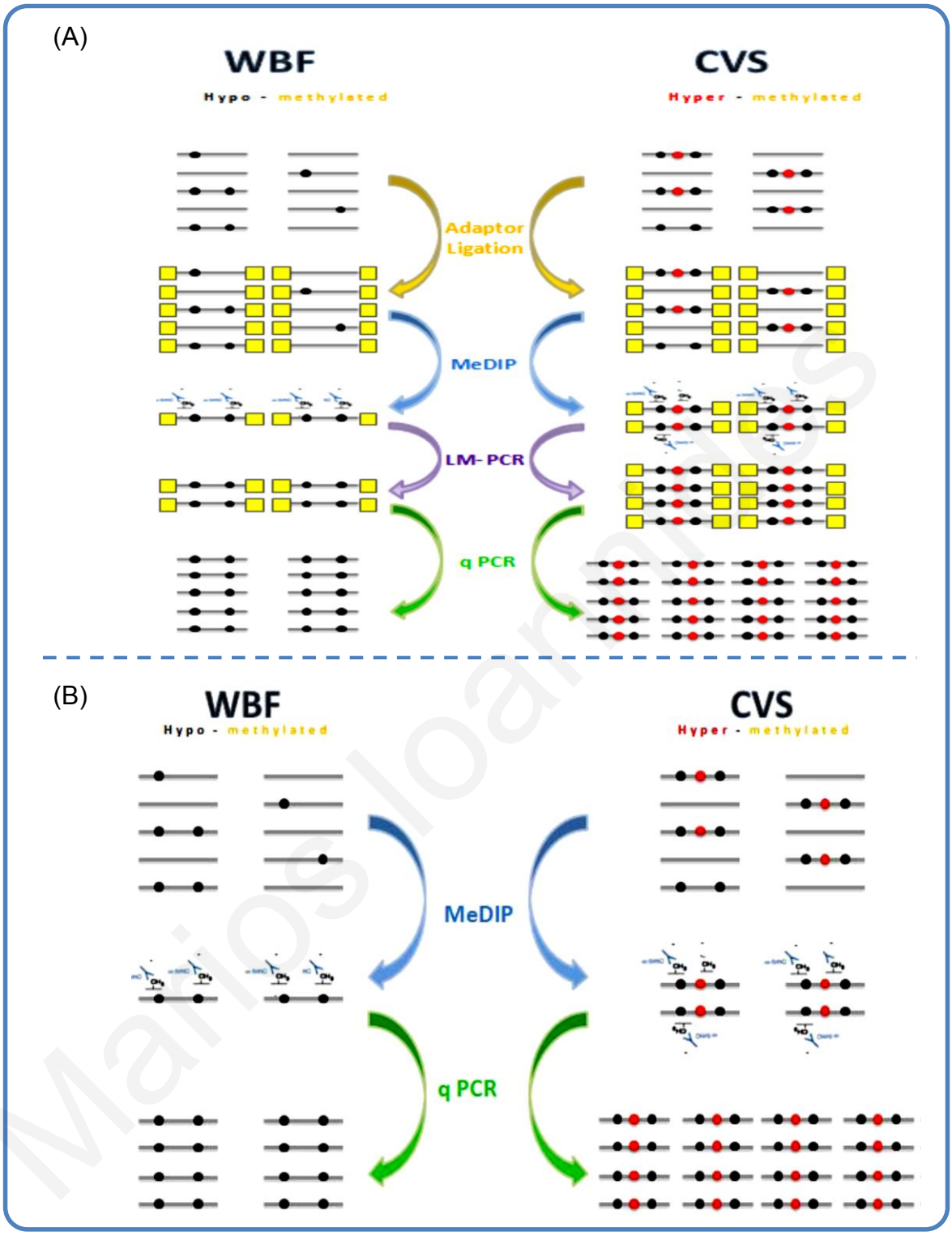


Figure 2.1 Methylated DNA Immunoprecipitation using the LM PCR and Non-LM PCR protocols. **A:** For the LM PCR protocol, linkers were added to the blunt-ended sonicated DNA from WBF and CVS following incubation with antibody specific to the methylated cytosines. The antibody-methylated DNA complex was then collected using magnetic beads and subjected to LM PCR. Enrichment values for each one of the selected DMRs was calculated following qPCR amplification. **B:** For the Non-LM PCR protocol the immunoprecipitated DNA was directly quantified by qPCR. (Red Dots: hypermethylation, Black dots: hypomethylation)

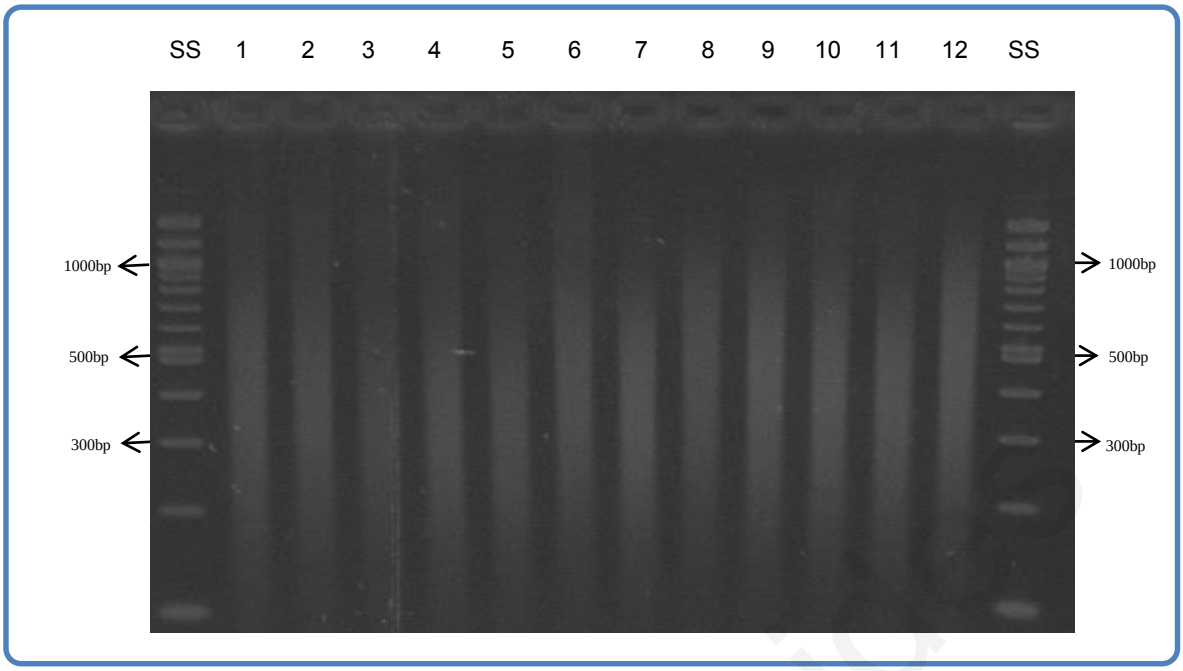


Figure 2.2 DNA fragment size verification prior to the MeDIP procedure. Sonicated DNA samples were run in a 2% gel agarose gel in order to ensure DNA fragmentation. SS: 100bp size standard, 1-12: Sonicated DNA samples

2.3 MeDIP-Array Comparative Genomic Hybridization (MeDIP-Chip)

In order to expand the prenatal panel of DMRs for the detection of all aneuploidies associated with high risk pregnancies, an ultra-high resolution aCGH chip was designed specifically for chromosomes 13, 18, 21, and X. The principle underlying aCGH is the detection of copy number changes or fold changes in the genome of a subject by the differential labelling of equal quantities of the subject's DNA (IP DNA) with a normal DNA sample which in our project is the subject's untreated DNA (IN DNA) following co-hybridization onto a microarray platform containing genomic sequences. Relative signal ratios are then assessed to determine whether there is a gain (hypermethytion) or loss (hypomethylation) for each DNA sequence represented on the array (Figure 2.3).

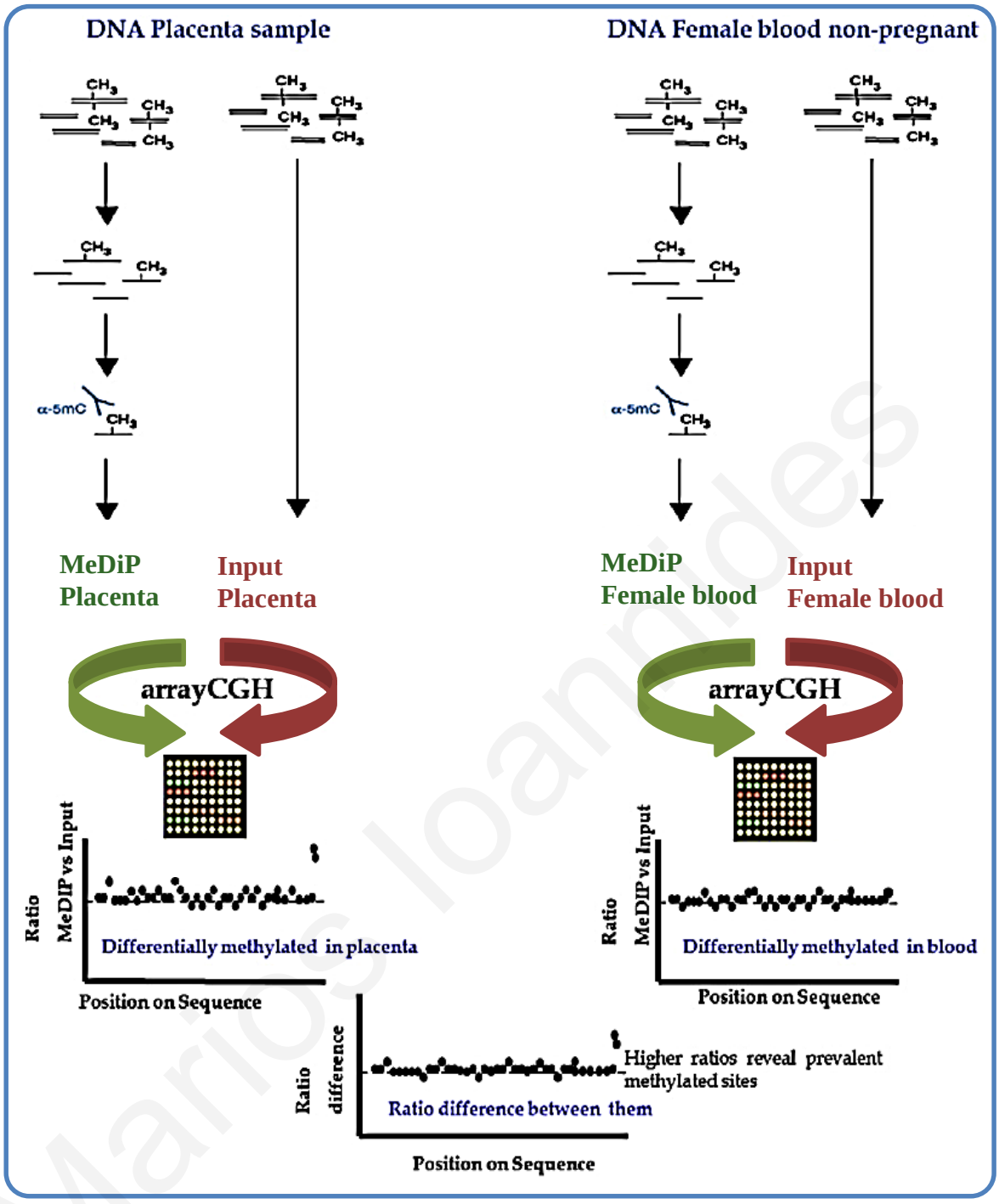


Figure 2.3 Summary of MeDIP-Chip procedure. Equal quantities of the MeDIP (IP) and unprocessed DNA (IN) were differentially labeled and co-hybridized on ultra-high resolution oligo arrays specific for the chromosomes 13, 18, 21 and X. Positive and negative signal intensity ratios indicate hypermethylated and hypomethylated probes respectively. Figure adopted from Patsalis P.C., 2012 [140].

2.3.1 Array Design

Custom ultra-high resolution tiling oligonucleotide array was designed for each one of the chromosomes 13, 18, 21 and X, in collaboration with Oxford Genome Technology (Oxford, UK). The custom design array platforms included one million probes (1M), with a mean probe spacing of 48bp (960K), 38bp (954K), 18bp (910K) and 56bp (971K) for chromosome 13, 18, 21 and X respectively. All probes were high quality, unique probes and assigned the highest quality score mainly based on GC content. The array design covered indiscriminately all regions of the chromosome except repeat genomic regions as indicated by the UCSC genome browser database [141] (Figure 2.4). Each array also included 10 000 backbone probes covering the rest of the autosomes for normalization purposes.

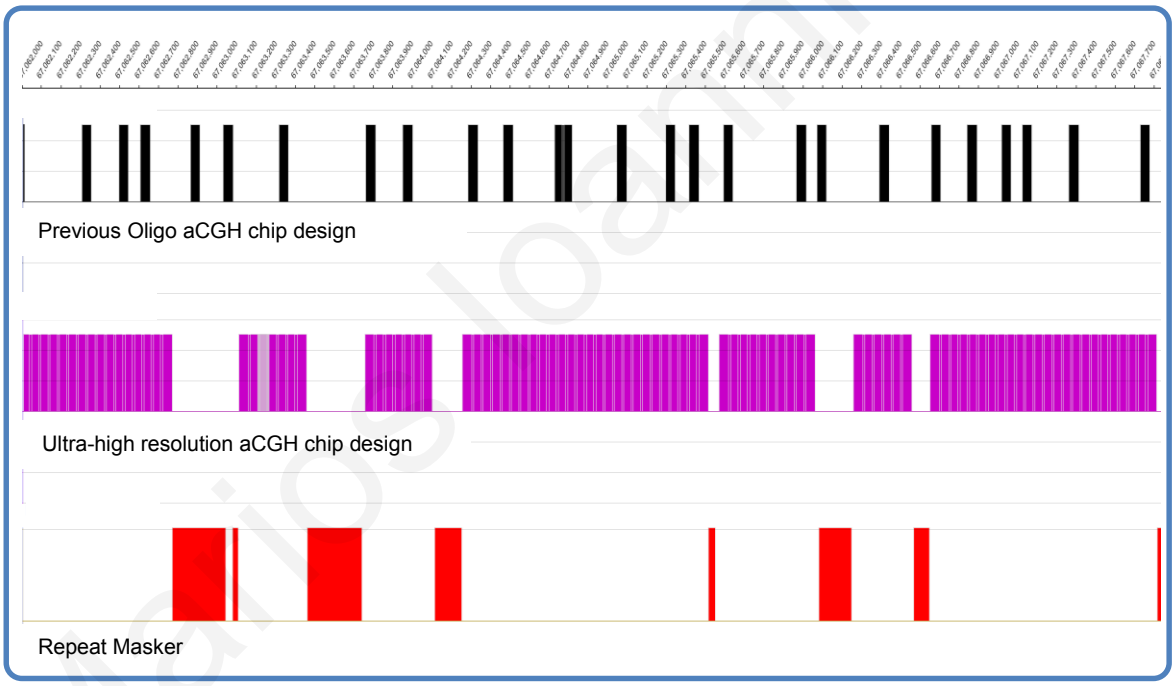


Figure 2.4 Ultra-high resolution aCGH chip design. One array chip for each chromosome tested was printed with 1 million unique probes (purple boxes) for the selection of new DMRs (middle track). The new design is of higher resolution as compared to previous high resolution array chips [100] (black bars- top track) and covered all regions of the chromosome excluding known repeat genomic regions (red boxes-bottom track).

2.3.2 Array Comparative Genomic Hybridization of CVS and WBF

During the MeDIP and LM PCR procedure, each sample group was pooled as shown in Figure 2.5 prior to aCGH, in order to surpress any inter-individual and inter-experimental variability. For the aCGH procedure 1μg of each of MeDIP (pooled) and INPUT DNA (pooled) were denatured at 100^{o}C for 15 minutes in a solution containing 1X Random Primer solution (Life Technologies) and HPLC water in a total volume of 131μl. While on ice the samples were differentially labeled by adding 19μl of labeling reaction consisting of 1X dNTP mix, 0.25nmol Cy3 or Cy5-dUTP Fluorescent dye (GE Health Care, New Jersey, USA) and 90U Klenow fragment (Life Technologies). The mix was incubated overnight at 37^{o}C. The reaction was stopped with 15μl of stop buffer. Unlabeled nucleotides were removed using the Amicon Ultra Centrifugal Filter kit (Millipore, Hertfordshire, UK) and the labeled DNA was eluted in 158μl of 1X TE buffer pH 8 (Promega, Madison Wisconsin, USA). The DNA was then mixed with 0.01mg/ml Cot-1 DNA (Life Tecnologies), 1X Blocking Agent (Agilent) and 1X Agilent Hybridization buffer (Agilent) in a total volume of 520μl. The reaction mix was incubated at 95^{o}C for five minutes and immediately after at 37^{o}C for 30 minutes. After blocking, 490μl of the reaction mix was placed on the array slide and incubated at 65^{o}C in a rotating hybridization oven for 24 hours. Unhybridized DNA was subsequently washed away using wash buffers 1 and 2 as directed by the manufacturer (Agilent). Next the slides were scanned using the Agilent Microarray Scanner 2565GA at 3μm resolution. Data were then extracted using the Agilent Feature Extraction software. The array design was first imported and the program then assigned a default grid template and protocol for each extraction set. This way the program gives each feature its chromosomal position. Following extraction, the files were loaded into the Agilent Workbench software (V7.0.4.0) where the $\log_2$ ratios of each probe were calculated. Final analysis and biomarker discovery was performed using the SignalMap Viewer Software.

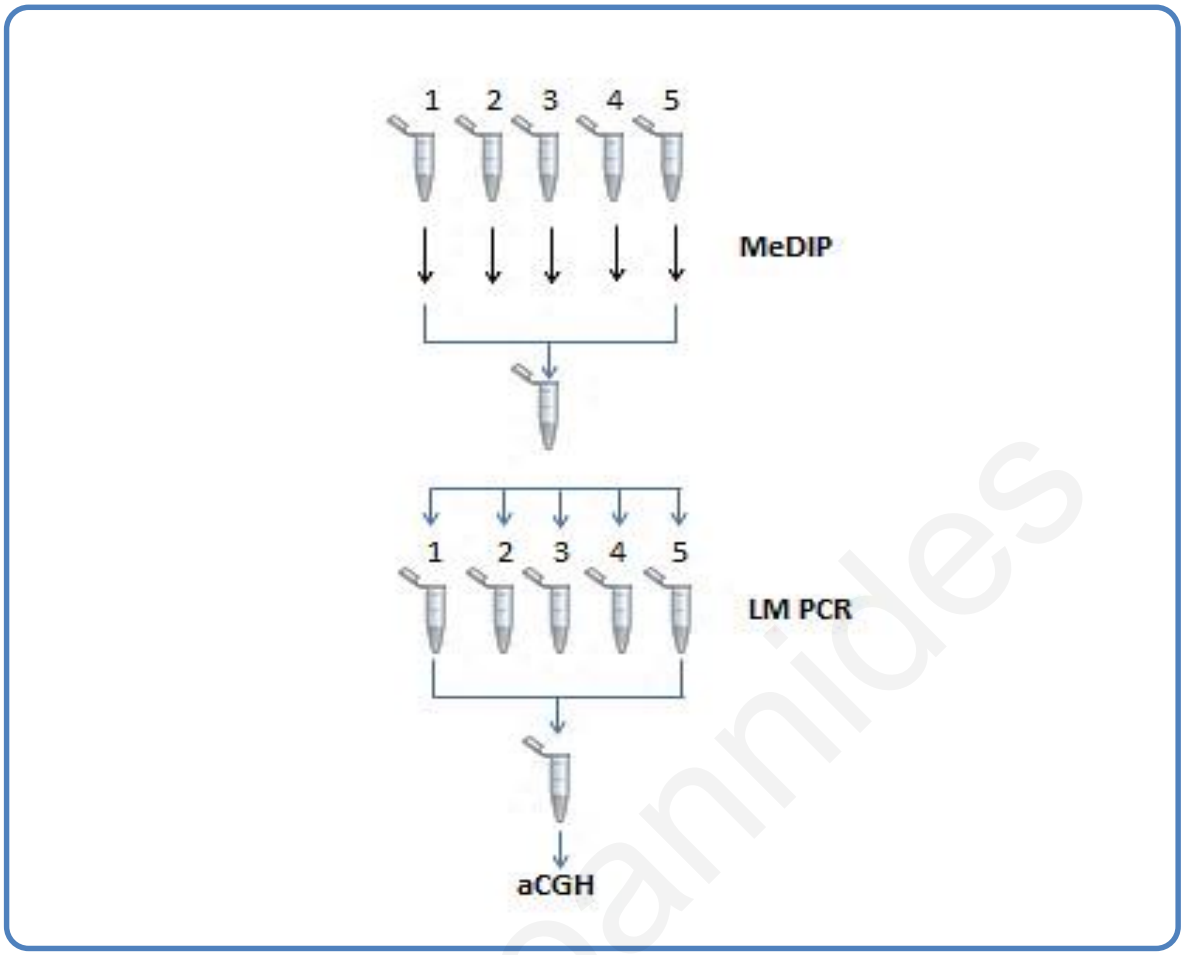


Figure 2.5 Pooling of MeDIP/IN samples prior to aCGH

2.4 Biomarker Discovery (MeDIP-Chip)

Data tracks from the array results were loaded and viewed in parallel with tracks downloaded from the UCSC genome browser. Data tracks included ReapeatMasker, CpG content, CGI tracks, Refseq genes and promoter region tracks–defined as the region up to 2kb upstream of the 5' end of each gene- and they were utilized in order to characterize and correlate the DMRs based on their location.

2.4.1 Selection Criteria for DMRs

The criteria used for screening the previous aCGH data set [100] were: (a) Inclusion of at least three consecutive probes as indicated by the array results in order to increase the possibility of having a true biological event, (b) consistent DNA hypermethylation in both first and third trimester placentas (CVS) and

hypomethylation in the female normal peripheral blood (WBF), and (c) exclusion of repeat genomic regions, segmental duplications and copy number variable regions based on the Database for Genomic Variants (DGV) (http://projects.tcag.ca/variation).

The initial criteria for the 1M whole chromosome custom aCGH were the same as the aforementioned except for (b) where the hypermethylation should be consistent in first, abnormal and third trimester placentas, and hypomethylation in the female normal peripheral blood.

DMRs were retrieved automatically using the MS-AND algorithm [142]. The software takes data (log_2 ratios) from three array sets at a time (i.e. 1st trimester CVS, 3rd trimester CVS and WBF) in order to identify the DMRs. While these criteria were used for chromosome 21, for chromosomes 13, 18 and X more stringent criteria were applied for DMR identification (see Results section 3.3.2.2)

"*Abnormality unique DMRs*" were identified using the following criteria: (a) Inclusion of at least three consecutive probes as indicated by the array results (b) DNA hypermethylation in abnormal placentas and consistent hypomethylation in 1st trimester CVS and WBF (c) exclusion of repeat genomic regions, segmental duplications and copy number variable regions.

2.5 Real-Time PCR (qPCR)

Selected candidate DMRs obtained from the aCGH data, were initially screened on IP DNA obtained from six to eight WBF, normal CVS and abnormal CVS (only in the case of trisomy 21). Furthermore, for the characterization of the inter-individual variability on a subset of confirmed markers on chromosomes 18 and 21, qPCR was performed on 50 WBF and 50 CVS normal samples**.**

2.5.1 Primer Design

Specific primers were designed for each of one of the selected regions (DMRs). Primer3 [143] or OligoAnalyzer 3.1 [144] software was used for the primer design

using the default settings. Each PCR product was designed to be 80-150bp long, with melting temperature of about 60ºC. The difference in the Tm between forward and reverse primers was chosen not to exceed 4ºC with GC content of 50-60%. The specificity of each primer was confirmed using the Primer-BLAST tool [145] as well as the *in-silico* PCR tool available at the UCSC Genome browser [141].

2.5.2 Standard curve analysis and qPCR Efficiency

Each primer set was tested initially on 8ng of control DNA, to ensure single product amplification as depicted by the melting curve analysis. Then, standard curves were constructed for each one of the primer sets to determine the efficiency of the PCR, its linear dynamic range and reproducibility. Briefly, 200ng of control DNA was used for six subsequent serial dilution dilutions (1:2) in order to reach the smallest amount of 3.125ng. 2μl from each dilution was used as a template for the standard curve construction. Each reaction was performed in quadruplicates in order to ensure higher precision in the calculations. The PCR mix consisted of 5μl of SYBR Green PCR mastermix (Eurogentec), 2μl of each primer mix (each one at a final concentration of 0.2nM) and 2μl of DNA. The samples were subjected to qPCR using the BIORAD CFX 384 Real time system (BIORAD, Hercules, California) under the following conditions:

Table 2.1 qPCR conditions

Temperature	Time	Cycles	Step
95 ˚C	10min	1 cycle	Denaturation
95 ˚C	15sec	40 cycles	Denaturation
60 ˚C	15sec		Extension
95 ˚C	15sec	1 cycle	
60 ˚C	15sec	1cycle	Melting curve analysis
95 ˚C	0.5 ˚C/sec		

Standard curve analysis was performed using the Bio-Rad CFX manager software (BIORAD). The efficiency of the assay should be 90-110%, which translates to a standard curve slope of -3.32± 0.25. Primers that failed to reach the specific range were not used for downstream experiments

2.5.3 DMR Confirmation

The methylation status of the selected DMRs was confirmed on MeDIP or IN DNA from CVS and WBF samples using qPCR. Each reaction was performed in triplicates on 8ng template DNA in a final volume of 10μl as described in the previous section.

2.6 Statistical calculations

MeDIP enrichment values of the 50 CVS and 50 WBF samples utilized in the characterization of the inter-individual methylation variability (first stage of the study) were calculated for each region using the following equation:

$$\textbf{Enrichment} = e^{\Delta Ct} \quad (1)$$

where e corresponds to the efficiency obtained in each real-time PCR reaction $e = 10^{(-1/\text{slope of STD curve})}$ and ΔCt indicates the cycle difference between input DNA and MeDIP DNA $[Ct_{(IN)} - Ct_{(IP)}]$. The mean enrichment values of each DMR were compared between WBF and CVS samples using the Mann-Whitney U tests and the corresponding p-values were used to decide whether there was significant evidence to claim that the mean enrichments of the two groups were different [146]. Hierarchical clustering of the DMRs was conducted using an iterative algorithm that joins similar clusters based on the set of dissimilarities of the 100 individuals (calculating the Euclidean distanced between clusters) and re-computing their distances at each stage by the Lance-Williams dissimilarity update formula [147].

2.7 MeDIP-Next Generation Sequencing (MeDIP-seq)

Pooled WBF and pooled normal CVS MeDIP samples that previously underwent the MeDIP-Chip procedure were also used for next generation sequencing (NGS). The rationale was to use NGS in order to confirm the DMRs identified using the MeDIP-aCGH methodology. For the purpose of this stage of the study, whole genome methylome recovered after MeDIP underwent adapter ligation and clonal

amplification before sequencing. The resulting sequenced reads were then aligned to the reference genome, followed by overlapping analysis between the confirmed DMRs using the MEDIP-chip and the MEDIP-seq result.

2.7.1 MeDIP-Seq Library Preparation and Sequencing

MeDIP WBF and CVS were blunt ended using 1X NEB buffer 2 (New England BioLabs), 0.12 mM dNTP mix, 1mM ATP (New England BioLabs), 20U T4 polynucleotide kinase (New England BioLabs), 6U T4 polymerase (New England BioLabs), 0.8mg/ml bovine serum albumin and HPLC water in a final volume of 50ul. The mix was incubated at 25^{o}C for 15 minutes followed by 15 minute incubation at 12^{o}C. Adapters were ligated on the blunt ended DNA using 1X Quick Ligase Buffer (New England BioLAbs), 2.5μM Illumina adapter mix (Illumina, San Diego, California, USA), 5U Quick Ligase (New England Biolabs), and 18μl of template DNA in a total volume of 40μl. Samples were incubated at room temperature for 20 minutes. Adapter overhangs were filled in using 33μl of DNA with 1X Thermopol buffer (New England BioLabs), 0.25mM dNTP mix and 16U Bst Polymerase (New England BioLabs) in a total volume of 40μl and the mix was incubated for 20 minutes at 65^{o}C with subsequent enzyme inactivation at 80^{o}C for 20 minutes. 6μl of DNA was then subjected to 50μl Indexing PCR raction using the following reagents at final concentrations: 1X Herculase II reaction buffer (Agilent Technologies, Santa Clara, California, USA), 0.25mM dNTP mix, 0.2μM of each of illumina_P5 and _P7 Index Primers (Illumina) and HPLC water. The PCR was performed under the following cycling conditions: initial denaturation at 95^{o}C for 5 minutes, 12 cycles consisting of denaturation step at 95^{o}C for 30 seconds, annealing at 60^{o}C for 30 seconds and extension for 30 seconds at 72^{o}C and a final extension step for three minutes at 72^{o}C. DNA products were purified using Mini elute columns (Qiagen) after each library preparation step [119]. PCR products were then assessed using an Agilent 2200 tape station D1000. Subsequently, DNA concentrations of the sample was measured (Qubit, Invitrogen) in order to ensure that equimolar amount of DNA was used for cluster generation according to the following formula:

X ng/uL * 1X10^6 UL/L * bp mol/ 600g * 1/350 = Y nM, where X is the measured DNA concentration in ng/μl and Y the converted concentration in nmol. Following cluster generation the samples were set for sequencing on an Illumina Hiseq 2500 according to the manufacturer's instructions.

2.7.2 Sequence alignment

Resulting sequence reads were aligned to the human reference genome (hg19) using the Burrows-Wheeler aligner. After subsequent removal of duplicate reads using the Picard MarkDuplicate tool, the alignment files were used as input to MEDIPS software for differential methylation analysis. Regions with at least three consecutive windows that showed statistically significant differential methylation were considered to be DMRs.

3 RESULTS

3.1 Fetal Specific Biomarker Discovery Using Existing Array Data

Whole chromosome oligo aCGH was previously performed on chromosomes 13, 18 and 21 with median probe spacing of 225bp, 170bp and 70bp respectively [100]. Using the following criteria: (a) Inclusion of at least three consecutive probes, (b) consistent DNA hypermethylation in both first and third trimester placentas (CVS) and hypomethylation in the female normal peripheral blood (WBF), and (c) exclusion of repeat genomic regions, segmental duplications and copy number variable regions based on the Database for Genomic Variants (DGV) we selected 11 regions on chromosome 13, 22 regions on chromosome 18 and 34 on chromosome 21 in order to investigate the differential methylation status between CVS and WBF. All primers were designed and checked for specificity using the BLAST and the *in-silico* PCR tool available at the UCSC browser. A unique PCR product, as indicated by a single peak during melting curve analysis ensured the high specificity of the primer set tested (Figure 3.1A). Furthermore, in order to estimate the PCR amplification efficiency, that is, the ability of the each primer to double the amount of PCR product after each PCR cycle, primers were subjected to standard curve analysis. A slope within the acceptable values of -3.32± 0.25 suggests near 100% efficiency during the PCR reaction (Figure 3.1B). Sequences of all primer sets are shown in Tables 3.1-3.3. Screening of the selected regions was performed initially in a cohort of six 1st trimester CVS and six WBF. Relative fold enrichment (equation 1) was calculated for all samples tested. Relative fold enrichment graphs (Appendix I) constructed for each DMR amplicon showed that three, eight and 12 regions had differential methylation between the two tissues for chromosomes 13, 18, 21 respectively. In addition, we correlated the loci of the confirmed DMRs with position of genes, promoter regions and CpG islands as shown in tables 3.4-3.6.

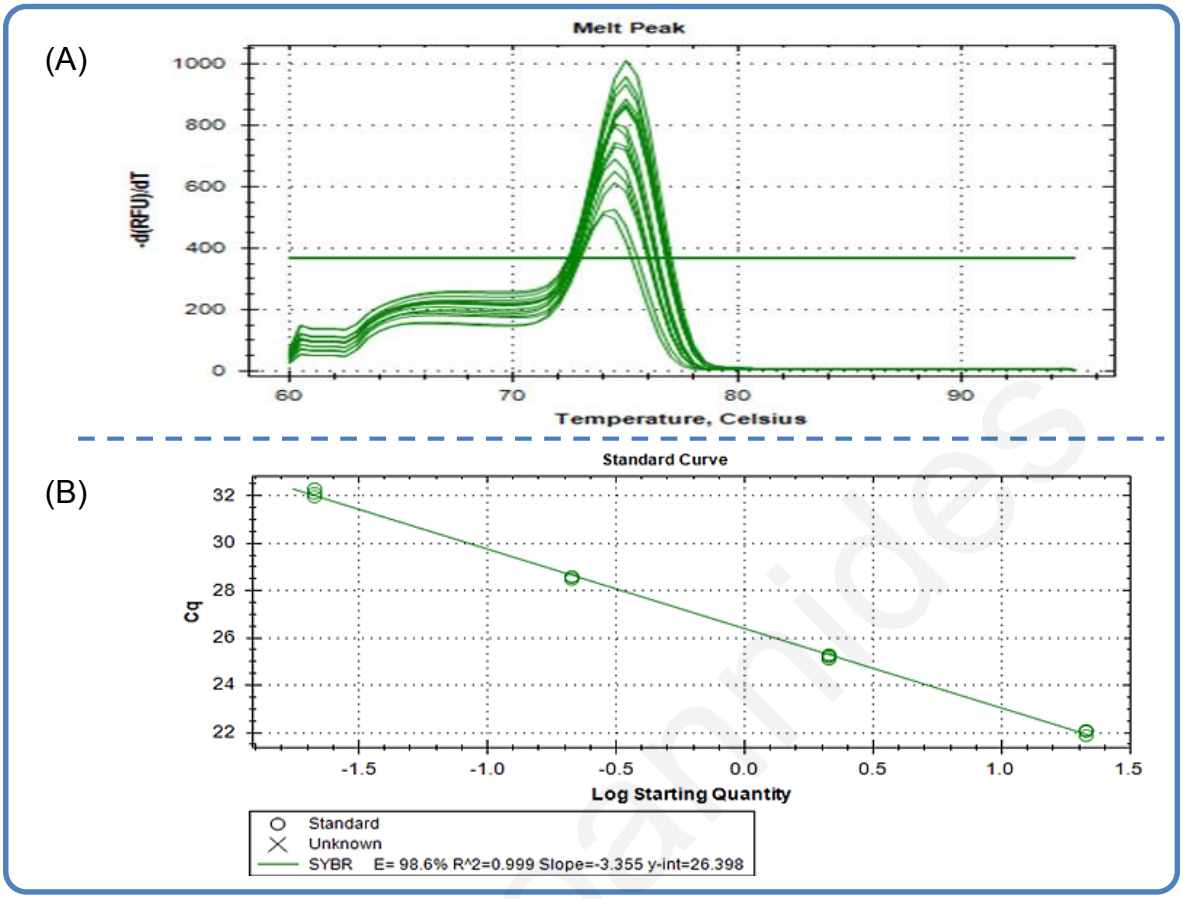


Figure 3.1 A: Example of melting curve analysis confirming a single PCR product as indicated by the presence of a single peak. **B**: Standard curves were constructed for each one of the primer sets in order to ensure efficient PCR amplification on control DNA samples prior to DMR screening on IP DNA**.**

Table 3.1 Primer sequences tested on selected DMRs for chromosome 13.

DMR	START	STOP	FORWARD	REVERSE
R2	32382658	32382798	TCCACTGCTAGTTGACCCTTG	TCAATTATGACGGATGGAGA
R4	39191374	39191499	TTGAACTCCCACCACCTGAT	CAAAATCAGCTTGACGAGCA
R7	45826843	45826949	CCACGGCACTCTTGACTCTT	CCTCAGGGAGCTTGCTTATC
R8	46211344	46211432	GGGGAACCTGGTCCAGTAGT	AAAATCACAATCGCCAGAGG
R9	47870040	47870153	GGTGGGTCATTAACCTTTGG	TCACTTGCCTTCCTGAGTTGT
R11	57236495	57236585	TGAGAAAGCAGTAATTGCAAGA	CACGCTGCCTGTAGAAACTG
R12	58216479	58216617	GCTGCCAAAGGGAGAAATTA	CCAGCATGGAGTCACTGAGA
R13	58347384	58347521	GTCCACTTAGAGGGCACAGC	GCTGCCTCCTGAAGACAGAA
AV2a	34943966	34944088	CGTATGCACGCGTATGTATGT	AGAAGGGGGCAAGGATCTAA
AV2b	34944219	34944325	CATTTGTCCGTTCTGCACTTT	ACCCAGGAGGTCAGGATACC
AV3a	42046960	42047069	GGCCTCTCACCTTTTCTGGT	GTGAAGGCGCTCATCTTTTT
AV3b	42047398	42047507	CCAACCTTCGGAAAGATGAA	GCATAAACCACCAGAAAGCA
AV5a	99345329	99345412	TTTCGCTGGGCAGATAGAAC	GGGTGCAAAAGTGTGAATGA
AV5b	99345914	99345995	GGGGTGGTGTTCGTTTGTTT	ACGTCGTGGGCAATTCTTAC
AV5c	99345628	99345743	GCTGCCCAATTTTCTAGCAC	TTTAAAACCTCTTGCCCTTCC

Table 3.2 Primer sequences tested on selected DMRs for chromosome 18.

DMR	START	STOP	FORWARD	REVERSE
VA1	509190	509291	GGCTCTCCCTCCTGTTCTCT	CGTGGTCTGCTTCACATTTC
VA2a	662501	662625	GATGGGTTAGGGATTGTGGA	TGGTGCCTCATCAGGTGTAA
VA2b	662656	662735	ACCCCTCACTCTCGATCTGT	TGTCGATACCTTGCCGTAGG
VA3a	1161302	1161411	CCTCATGGGTCAGTTAGACCA	CAGCCCTTAGAAGGCCAAGT
VA3b	1161302	1161444	CCTCATGGGTCAGTTAGACCA	ACCACACAGGCCTAAGGAAA
VA4a	2940946	2941089	AACAAGCCCAAACCTCACAC	TGAGCTCCGATGATGACAAG
VA4b	2941253	2941333	TGTTTCATCTCCACCCGATT	GCCACCTCACCAATTCCTAC
VA6a	7917547	7917683	CTTCTGTGTACGCGGTTGTG	TCACACGGAGTCTCCAATCA
VA6b	7917378	7917547	TGCCTGTCCCTTCTAACCAG	AAAAATGAGCAGCCTCCAGA
VA8a	23200797	23200890	CTGCACCATCACTCAAGAGC	GAGACACCGTAATGGGTGGT
VA8c	23200722	23200858	CCCATTACGGTGTCTCAAGG	TGGCACCAAGTGCAAACTTA
VA9a	32298091	32298192	ATCTGTGGCCAGGTTGATAA	GCATGCAGTTCATTTCTTCG
VA9b	32297898	32298098	CTTTTGGGAAACATGCCACT	CCTTTGCTAATCCCTTCACC
VA10	35041837	35041938	TCCTTCCTTCTTTCGGTCAA	AGATGCAGAGCTTGGAGGAG
VA11a	35509214	35509293	CGCAGCAGGAGATACAGGAC	CTCAAAGGCGAGCCATTTCT
VA11b	35509276	35509395	AAATGGCTCGCCTTTGAGTA	TTTCGCCAGTGATGAAAAGC
VA12a	40374064	40374199	TGAGCCACATCCGTAGATTG	CTTGGGTTCCCTCTTTACCC
VA12b	40374286	40374372	AAACACGTGCAGTCCCACTC	CAATCTACGGATGTGGCTCA
VA12c	40374353	40374459	GTGCAGCAGCCTGACAATAA	ATTGGGAGCAATTGCAGAAG
VA14a	50564454	50564561	CATCAGATGGCAAGAATATCCA	CGTCCAGAAACCAAAGTCAA
VA14b	50564256	50564448	CGGTCGGATGAGACAAAAAT	GGGCTCAGGTCTTACCACAA
VA15	58815097	58815212	GCACGTTCTCCAGTGGTTCT	GGGGTCAGGACACACTTCAC
VA16a	58823995	58824075	CATTTCAGTGTCTCGGAGGAG	AATCAAGCCTGAAGGGGAAT
VA16b	58824180	58824360	AGCTGATGGCAAGTGTGATG	AGGAAAATGGGATGGGATTC
VA17b	59109462	59109553	CGCATATGATTGGCTTCCAT	TTCAGTGTCCCCAGTCAAAA
VA17c	59109553	59109730	GGGAACGTGACCTAGTTGGA	TGTGGACCTGTTGACACTTGA
VA18b	62176850	62176953	TGGCTCTTCTAACGGTTCTTC	CCCACAGGGTATTCGATTGT
VA18c	62176713	62176866	GCTCAGGCTGGCTCTTCTAA	TCTGATCTTCAGCCCAACCT
VA19a	62530349	62530429	TGTCAACTTACGCACGCTTC	AGCACGAGAATTGTCATCAGA
VA19b	62530408	62530509	TTCTGATGACAATTCTCGTGCT	TGCAGGTTGTACTTGCCACT
VA20a	68905606	68905689	TCGTCCAGAAGGTGACTGAA	AGGACACGCTTGGTGAGATT
VA21a	70759966	70760094	TCCCCACTCATGAGGAACAC	GTCCAAGCGAGGCAGAATTA
VA21d	70760038	70760188	GGGAGGGAAGTAGCAGGAAC	TTCCTCATGAGTGGGGATCT
VA22a	72971413	72971497	AGGTGACAAACCGCTCTCAA	GCCTCTGTGCCTTTTCTGTC
VA22b	72971691	72971770	GGACCACCACTGACCAAACT	CGTCGCTTGTTTTACACCTG
All2	55090427	55090524	TGTGCCTCTCCCTTGAGACT	AAATTGCAGCCAATGCTTCT
B3	44166131	44166263	TGTGGTTTCAAACATGCACA	CTGAAAAGGCCACTCTGAGG
C1	58956266	58956391	GTGAGAGAGAACGCCAGGAG	TGAGCCAACTCTGGTGTCAG

Table 3.3 Primer sequences tested on selected DMRs for chromosome 21.

DMR	START	STOP	FORWARD	REVERSE
M1	15253407	15253501	CACATCCAACCTGGTGCTAA	AGGCCTAAGGGTTGCCTCT
M2	15282925	15283027	ATGGAACAACACTGGGAACC	CATGAAGTCCACTCCAAAATGA
M3	15688345	15688471	TGGTGCAATCTGAACTCTCG	TTTTTGCGGCTATCGTGTAA
M4	16061365	16061493	CCACTGTTTTCCTCCTGTCA	AGGCCAGAGTGCATTCAAAC
M5B	16615785	16615897	CTGGAGGGTGTCTCTGTTCC	AAGGCTCATTGCTCCTGCTA
M6	17490929	17491023	TGGCATTTTAACTGGAAAGCA	GCTACTTCATGTAAACATTCTCC
M7	17516633	17516760	CCATACTGGTGACTTATTTGTTCC	TTCCAAATTTTCACAACTCTTCAA
M8	30742372	30742451	TTGGTAGAAACAGCTGAAGAGG	CTTTACCCTTCGTCCCAACA
M9	30776236	30776376	TTATCTCTCCTGCACCCACA	GGGAAATAAACAGGCACACAA
M10	31443934	31444061	GCTTGTTGGTGGTTATGCTCT	TGTTTATCTAACTACCAGGCCACA
M11	14886844	14886949	GCATTTCCGTTTCCACATTT	TCAAGGAAAAAGTGAGAATGACC
M12	16092149	16092228	GCCATTATAGGCATGCAAGG	CTCCTCTGAATACCGGAAACC
M13	28729693	28729827	TCAGCAATTACAAAAGCTTCCA	TTGATGTCCTCCTCCTGACC
M14A	29833003	29833108	TTGAAATGTTCGTTTTGTTCC	GGCCTGGTTTTGAAATCAAGTA
M14B	29832972	29833108	TTCTTGTGTATCAGCATGCAA	GGCCTGGTTTTGAAATCAAGTA
M15	30178776	30178924	GGGTGCCAATGTTTCTAACC	TGTCTCGTCTTGAAGGTGGA
M16A	22009911	22010058	AGCAGGCAGAATTGATGACC	AGGTGGGATGAACCACAAAA
M16B	22009889	22010019	CCCCAAGGCCTAGAGAAGAC	GAAATTCCTGGGAGCGAAAT
M18	15331818	15331945	GATGGATGGCCTTTTGGTAA	TATTTGGTTTGCCCCTTCCT
M19	15000495	15000603	AGGGTGGGGTAAACATGGAG	TCCATTTAGAGGCAGATAAACCA
M20	15178413	5178497	CATTAGCGGGTCAGCTAGGA	TGGCAATTACATCTGCCATTA
M22	27690397	27690481	AGGAAGAACTGAGGTGGGATA	CCATCCGAACTCTTAAATCAGG
M23	28576046	28576130	GACGACACTGTGTGCAAGAGA	AAAGAGGTCACAATCTCTTCGTCT
M24	29751035	29751114	CACTGAAGTGACTGGTTTCAGC	GGCCTAATGATGCCATTTCT
M25	37692864	37692974	TTGTCTGCCCGTATGGAAGT	ATGGTTGTAGGGCTCATTCA
M26	39209277	39209382	TCCGGAGCTGAATCTCTTTC	GCCAAGCTTTCAGAAACCAA
M27	42178812	42178948	ATACGTGTCCTGCCTTCCAC	GCTTTGAGCAGAGAGGGAAA
M28	45171107	45171192	CCCAGAAATTCCATTTGCAG	GAAAGGCTCAACCAACCAAC
M29	14990100	14990201	GCATGCAGTAGATTGGCACT	TGTACCTGATCAGCCATCCA
M1E	44953674	44953772	TCGCACTGAGGCTTCCTACT	AAGTTGTGGGCTGGGATTTT
Nn2	31427008	31427139	ACCATTGTGGATCACAGCAG	GCTCCGAGGATTAGGGAAAG
On2	34492982	34493090	CTCCTGACCCACTCCCAATA	GGAAACTCAGGGTCAAACGA
Fd1	42006045	42006153	ATGTTGCCTGGGATATGCTT	AACTGGCTGCGTGAGGATA
El-3	42355352	42355484	GCCTTGGGACAAAAATGACA	TGGGCACAGCCCTAACTAAC
El-4	42355802	42355908	GGCCAGGTTGTTTCAGATTG	TTCCGGCAGAGTTTATTTGG
Id2	42753720	42753866	ACCGTATCATTTCCCCAGGT	TGACCACATTTCCACCACAG
A5	**39279856**	**39280004**	**GCTGGACCAGAAAGTGTTGAG**	**GTGTGCTGCTTTGCAATGTG**
C5	**33320735**	**33320829**	**CTGTTGCATGAGAGCAGAGG**	**CGTCCCCCTCGCTACTATCT**
D2	**42189557**	**42189683**	**TGCAGGATATTTGGCAAGGT**	**CTGTGCCGGTAGAAATGGTT**
El-2	**42355712**	**42355815**	**TGAATCAGTTCACCGACAGC**	**GAAACAACCTGGCCATTCTC**
Eli-1	**42357215**	**42357341**	**CCGTTATATGGATGCCTTGG**	**AAACTGTTGGGCTGAACTGC**
H2	**32268843**	**32268943**	**CCACATCCTGGCCATCTACT**	**TTCCACAGACAGCAGAGACG**
J2	**37841284**	**37841411**	**ATTCTCCACAGGGCAATGAG**	**TTATGTGGCCTTTCCTCCTG**

Regions in bolt indicate previously validated regions used as comparison standards [126, 127].

Table 3.4 Details on confirmed regions on chromosome 13 that showed differential enrichment between CVS and WBF.

Chromosomal Region	Position (hg18)	Location Type	Gene
AV2	34943985-34944319	CGI-promoter	NBEA
AV5	99345306-99346014	CGI-promoter	CLYBL
R2	32382488-32382831	Intergenic	

Table 3.5 Details on confirmed regions on chromosome 18 that showed differential enrichment between CVS and WBF.

Chromosomal Region	Position (hg18)	Location Type	Gene
VA1	508280-509314	Intergenic	
VA4	2940888-2941347	Intragenic	LPIN2
VA15	58814900-58815220	Intergenic	
VA17	5910936-59109737	Intragenic	BCL2
VA22	72971389-72971902	Intragenic	MBP
AII	55090284-55090605	Intragenic	RAX
B	44165984-44166275	Intergenic	
C	58955844-58956604	Intragenic	BCL2

Table 3.6 Details on confirmed regions on chromosome 21 that showed differential enrichment between CVS and WBF.

Chromosomal Region	Position (hg18)	Location Type	Gene
Nn	31426757-31427146	Intragenic	TIAM1
On	34492714-34493203	Intergenic	
Fd	42005961-42006216	Intragenic	C21orf129
EI	42355366-42355908	Intergenic	
EII	42357141-42357401	Intergenic	
Id	42753677-42754026	Intergenic	
M1E	44953640-44953854	Intragenic	TSPEAR
M28	45171015-45171225	Intragenic	ITGB2
M18	15331818-15331945	Intragenic	NR1P1
M20	15178413-15178497	Intergenic	
M25	37692864-37692974	Intergenic	DYRK1A
M27	42178808-42179008	Intragenic	C2CD2

3.2 Validation and Characterization of DMRs Using LM PCR vs Non-LM PCR approaches

Based on this initial screening we selected 15 regions for further validation / characterization, using seven previously validated DMRs by Papageorgiou et al. [126] and Tsaliki et al. [127], as a comparison standard (Table 3.7). This DMR validation study was conducted on a set of 50 CVS and 50 WBF samples using MeDIP with subsequent amplification of IP and IN products (LM PCR), followed by qPCR. The efficiency of the methodology was monitored using one hypermethylated (HYPER) and one hypomethylated (HYPO) control regions. The HYPER is a region that showed hypermethylation for both CVS and WBF, while the HYPO is a region that showed hypomethylation for the two tissues [100]. As expected, enrichment values for HYPO were 0.46 and 0.50 in WBF and CVS samples while the HYPER control region showed enrichment for CVS and WBF with mean enrichment values of 3.12 and 3.22 respectively, indicating that the MeDIP procedure was highly specific for the methylated regions. Moreover, the previously validated DMRs performed as previously described [100], exhibiting distinctively different enrichment between CVS and WBF. All tested DMRs showed a significant enrichment ($p<0.01$) in CVS samples compared to those of WBF (Table 3.7). We compared the performance of the 15 newly selected DMRs with the previously validated set and we were able to determine that 11 of 15 DMRs showed enrichment values higher than the lowest of the previously validated DMRs, ranging from 1.9 to 6.4. Additional comparison of the two DMR sets also illustrated that for 11 of these 15 regions the difference of means (mean enrichment CVS – mean enrichment WBF) was again higher than the respective values of the validated DMRs (ranging from 1.6 to 6.4) (Table 3.7). Furthermore, all DMRs tested resulted in inter-individual methylation variability as reflected by the range of enrichment values obtained from each DMR (Figure 3.2) as well as the coefficient of variation values (Table 3.7). CVS samples appeared to be less variable than the WBF with mean coefficient of variation value of 0.36 and 0.88 respectively. Only four of the 15 new DMRs, despite the significant difference in enrichment ($p<0.01$), showed an overlap in their enrichment values as shown by the box blots (Figure 3.2)

To better investigate tissue specificity (CVS-WBF) in the 15 newly selected DMRs in relation to the previously validated DMRs, we also constructed a heat map and

hierarchical clustering of the 50 CVS and 50 WBF samples based on the obtained enrichment values (Figure 3.3). This analysis showed a clear differentiation between the two tissue types based on the obtained enrichment values. Furthermore, DMR clustering analysis showed that there was no distinct clustering separation between the newly selected and the previously validated DMRs.

The same subset of DMRs that was validated with LM PCR was subsequently subjected to MeDIP-qPCR without amplification following the MeDIP procedure (Non-LM PCR) (Figure 2.2B) in order to observe if the amplification bias added by the LM PCR affects the enrichment and specificity between the two tissues. Our control regions showed as expected similar enrichment values for both tissues with the enrichment of the HYPER control region being an order of magnitude higher than the HYPO control marker, an indication that the MeDIP procedure was successful (Table 3.8). Despite the low relative fold enrichment values obtained for each DMR, there is fetal specific enrichment in all CVS samples as compared to the female peripheral blood samples. All DMRs tested showed a statistically significant enrichment difference ($p<0.01$) (Table 3.8). Despite this, the enrichment between CVS and WBF samples showed overlapping values as indicated by the box blots (Figure 3.4). Furthermore, the two tissues exhibited tissue specificity in regards to their methylation status and similarly to the LM PCR results; there was no clustering separation between the newly selected and the previously validated DMRs (Figure 3.5). A correlation study between Non-LM PCR and LM PCR for the mean enrichment values for each marker was also performed. Results suggest that there is a high degree of correlation between the two approaches for the WBF samples (Pearson correlation $r=0.967$) whereas no correlation was indicated for the CVS samples ($r=-0.069$) (Table 3.9).

Table 3.7 Ranking of DMRs tested according to the difference between mean enrichment values for each DMR using the LM PCR protocol.

Marker	Mean WBF	Mean CVS	Mean Difference	SD WBF	SD CVS	p value	Coefficient of Variation WBF	Coefficient of Variation CVS
EI-4	0.017	6.384	6.367	0.022	2.143	1.53E-17	1.294	0.336
EII-1	**0.065**	**5.319**	**5.254**	**0.071**	**1.937**	**7.06E-18**	**1.092**	**0.364**
H2	**0.135**	**4.068**	**3.933**	**0.093**	**1.252**	**7.06E-18**	**0.689**	**0.308**
EI-2	**0.064**	**3.894**	**3.83**	**0.2**	**1.338**	**7.50E-18**	**3.125**	**0.344**
EI-3	0.116	3.905	3.789	0.312	1.556	1.96E-17	2.690	0.398
B3	0.126	3.86	3.734	0.1	1.268	2.29E-17	0.794	0.328
M27	0.532	4.113	3.581	0.2	1.386	7.06E-18	0.376	0.337
D2	**0.317**	**3.364**	**3.047**	**0.179**	**1.301**	**7.06E-18**	**0.565**	**0.387**
M28	0.189	2.777	2.588	0.125	0.919	7.06E-18	0.661	0.331
M1E	0.149	2.636	2.487	0.097	0.742	3.44E-17	0.651	0.281
Id1	0.398	2.682	2.284	0.155	0.932	1.04E-17	0.389	0.348
A5	**0.337**	**2.505**	**2.168**	**0.159**	**0.986**	**7.06E-18**	**0.472**	**0.394**
C5	**0.18**	**2.321**	**2.141**	**0.106**	**0.84**	**7.06E-18**	**0.589**	**0.362**
C1	0.106	2.229	2.123	0.083	0.635	7.06E-18	0.783	0.285
AII-2	0.065	2.003	1.938	0.084	0.933	1.23E-16	1.292	0.466
On2	0.281	1.993	1.712	0.138	0.552	7.06E-18	0.491	0.277
Nn2	0.245	1.924	1.679	0.107	0.78	7.07E-18	0.437	0.405
J2	**0.116**	**1.707**	**1.591**	**0.079**	**0.519**	**7.06E-18**	**0.681**	**0.304**
Fd1	0.135	1.676	1.541	0.101	0.513	7.06E-18	0.748	0.306
M25	1.038	1.822	0.784	0.452	0.655	1.88E-09	0.435	0.359
M20	0.42	0.796	0.376	0.186	0.303	2.37E-10	0.443	0.381
M18	0.097	0.22	0.123	0.07	0.151	6.36E-08	0.722	0.686
HYPER	3.124	3.226	0.102	0.68	0.982	0.951	0.218	0.304
HYPO	0.469	0.508	0.039	0.945	0.948	0.119	2.015	1.866

Mean Average Enrichment				Mean Coefficient of Variation	
0.233	2.83			0.882	0.363

Regions in bolt indicate previously validated regions used as comparison standards [126, 127]

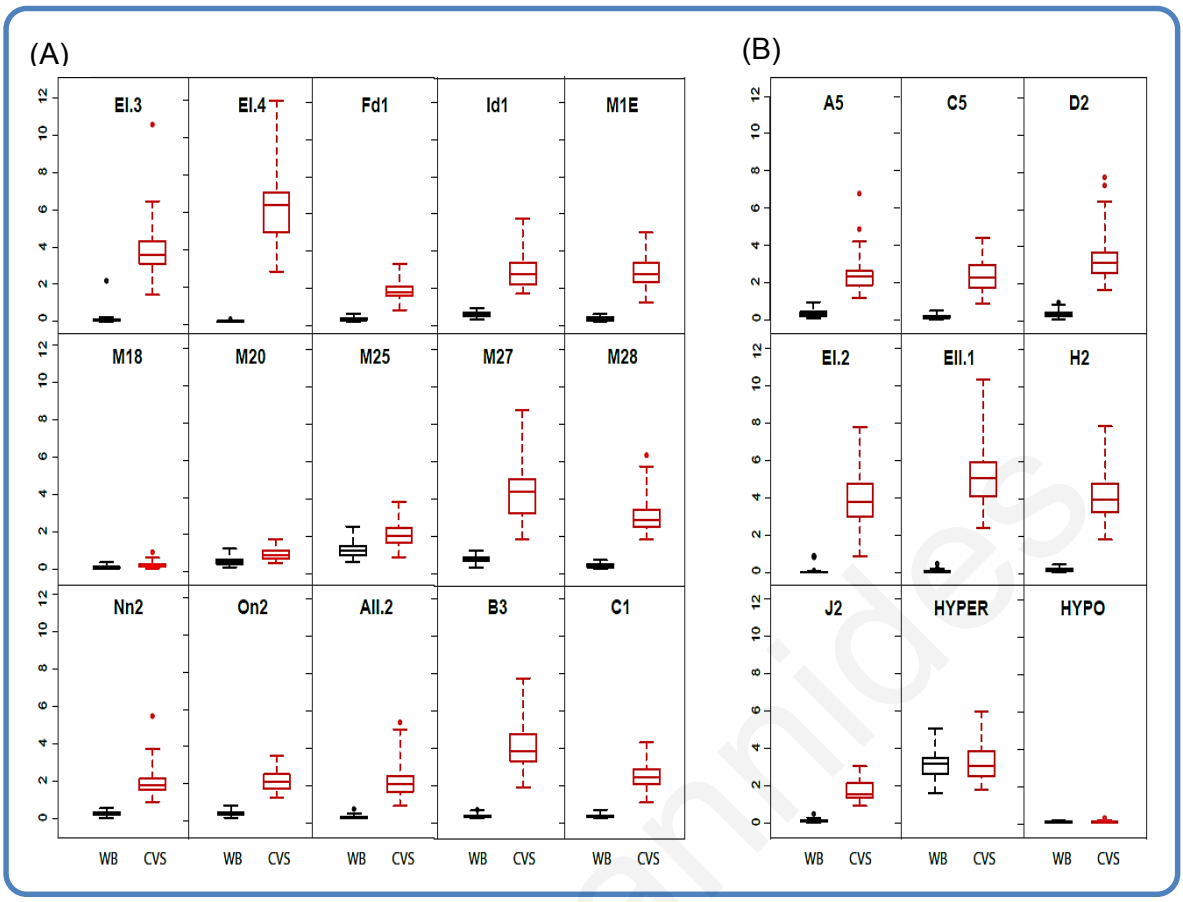


Figure 3.2 Enrichment profile for selected DMRs on 50 WBF and 50 first trimester CVS using the LM PCR protocol. Box plots show the distribution of the relative fold enrichment values for WBF (black) and CVS (red) for each DMR. The median value is represented by a horizontal line. The bottom of the box indicates the 25th percentile (lower quartile) and the top the 75th percentile (upper quartile). Whisker lines indicate the outlier boundaries [top: median+ 3(75%-25%); bottom: median - 3(75%-25%)]. As depicted by the plots, there is a clear separation of the methylation enrichment values between WBF and CVS despite the methylation variability. Fd1, M18, M20, M25 show overlap between the interquartile values, thus they were not characterized as DMRs. **A:** New DMRs, **B**: Previously validated DMRs. HYPER: hypermethylated marker for both tissues, HYPO: Hypomethylated marker for both tissues.

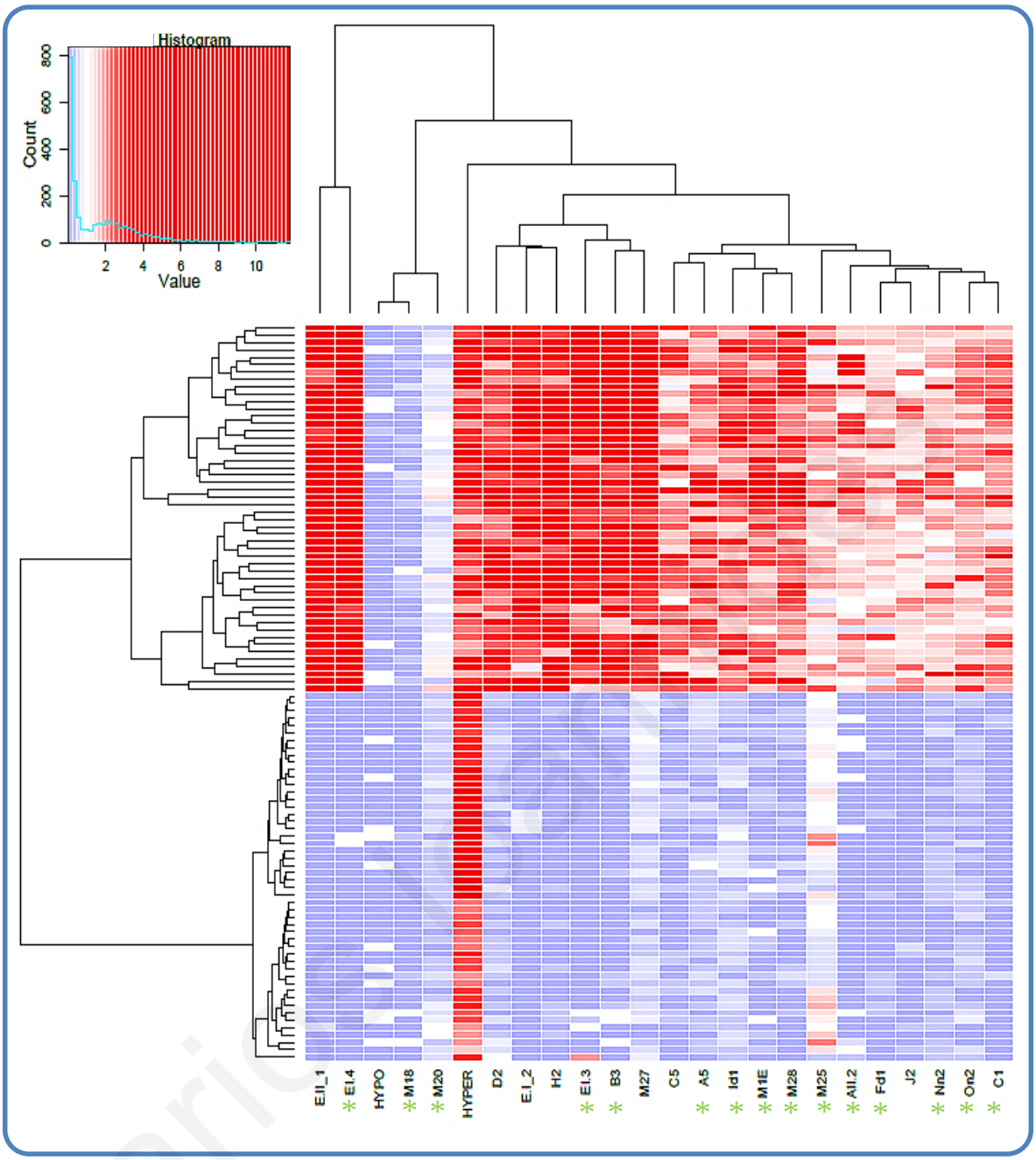


Figure 3.3 DMRs exhibiting tissue specificity between CVS and WBF using the LM PCR protocol. A heat map constructed based on the relative fold enrichment values obtained for the two tissues tested (CVS and WBF) with the MeDIP-qPCR approach shows clustering of the hypermethylated CVS samples (high enrichment-red) and the hypomethylated WBF samples (low enrichment–blue). Furthermore, DMR clustering analysis showed no cluster distinction between the 15 newly selected DMRs and the previously validated DMRs that were used in this study as a comparison standard.

Table 3.8 Enrichment values obtained from the MeDIP Non-LM PCR protocol.

Marker	Mean WBF	Mean CVS	Mean Difference	SD WBF	SD CVS	pvalue	Coefficient of Variation WBF	Coefficient of Variation CVS
B3	4.30E-09	4.69E-08	4.26E-08	4.81E-09	1.91E-08	4.70E-17	1.119	0.407
M1E	4.50E-09	3.93E-08	3.48E-08	3.43E-09	1.65E-08	6.42E-17	0.762	0.420
Nn2	6.90E-09	3.79E-08	3.10E-08	6.59E-09	1.31E-08	2.06E-23	0.955	0.346
C1	2.88E-09	2.08E-08	1.79E-08	2.90E-09	5.98E-09	6.82E-17	1.007	0.288
Id1	7.52E-09	2.43E-08	1.68E-08	5.59E-09	6.90E-09	1.04E-14	0.743	0.284
H2	1.17E-09	1.60E-08	1.48E-08	9.35E-10	5.50E-09	3.23E-17	0.799	0.344
J2	1.88E-09	1.52E-08	1.33E-08	1.52E-09	6.94E-09	6.42E-17	0.809	0.457
AII_2	1.76E-09	1.50E-08	1.32E-08	2.42E-09	6.65E-09	3.37E-16	1.375	0.443
Fd1	3.28E-09	1.45E-08	1.12E-08	3.45E-09	4.67E-09	1.60E-15	1.052	0.322
On2	3.23E-09	1.41E-08	1.09E-08	2.08E-09	4.19E-09	1.39E-16	0.644	0.297
E.I_3	4.88E-10	8.36E-09	7.87E-09	9.44E-10	2.25E-09	6.51E-17	1.934	0.269
D2	1.98E-09	9.49E-09	7.51E-09	1.09E-09	3.59E-09	1.19E-15	0.551	0.378
M27	1.52E-09	7.42E-09	5.90E-09	8.33E-10	2.18E-09	9.31E-17	0.548	0.294
M25	1.20E-08	1.52E-08	3.20E-09	3.51E-09	3.99E-09	2.64E-06	0.293	0.263
M28	6.18E-10	3.47E-09	2.85E-09	4.26E-10	1.06E-09	4.70E-17	0.689	0.305
M20	5.66E-09	8.26E-09	2.60E-09	1.68E-09	2.51E-09	3.33E-04	0.297	0.304
A5	9.38E-10	3.22E-09	2.28E-09	4.67E-10	1.03E-09	9.25E-15	0.498	0.320
E1 II	5.50E-11	1.96E-09	1.91E-09	1.16E-10	7.84E-10	3.44E-17	2.109	0.400
C5	2.45E-10	2.02E-09	1.78E-09	2.15E-10	8.11E-10	9.31E-17	0.878	0.401
M18	2.11E-09	3.60E-09	1.49E-09	2.17E-09	1.68E-09	2.22E-06	1.028	0.467
EII 1	2.18E-11	8.85E-10	8.63E-10	4.73E-11	3.53E-10	3.90E-17	2.170	0.399
EI 4	9.45E-12	3.64E-10	3.55E-10	2.22E-11	1.30E-10	3.44E-17	2.349	0.357
HYPO	4.63E-09	4.08E-09	-5.50E-10	6.64E-09	3.79E-09	1.28E-01	1.434	0.929
HYPER	3.90E-08	3.29E-08	-6.10E-09	1.20E-08	9.03E-09	1.13E-02	0.308	0.274

DMRs are ranked with respect to the difference between the mean enrichment values obtained from the WBF and CVS for each DMRs tested. All DMRs subjected to the Non-LM PCR protocol showed statistically higher enrichment ($p<0.01$) in the CVS samples than in WBF, confirming the results obtained from the LM PCR procedure. HYPER: Hypermethylated marker for both tissues, HYPO: Hypomethylated marker for both tissues.

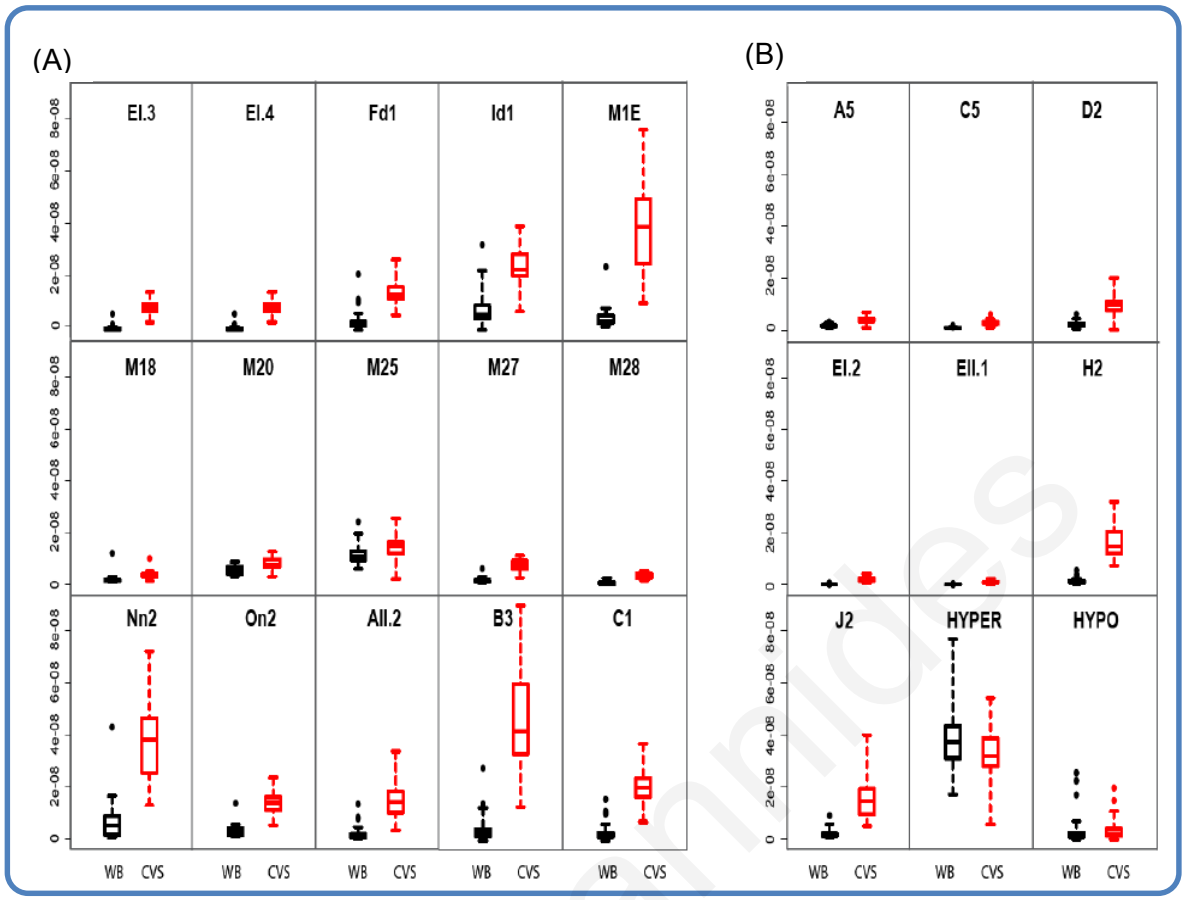


Figure 3.4 Enrichment profile for selected DMRs on 50 WBF and 50 first trimester CVS samples using the Non-LM PCR protocol. The plots show the distribution of the relative fold enrichment values for WBF (black) and CVS (red) for each DMR. Horizontal bars represent the median value. It is evident that even though the difference in the enrichment values between the two tissues at a given DMR is suppressed, the enrichment in the CVS is higher than in the WBF samples. **A**: New DMRs, **B**: Previously validated DMRs. HYPER: hypermethylated marker for both tissues, HYPO: Hypomethylated marker for both tissues.

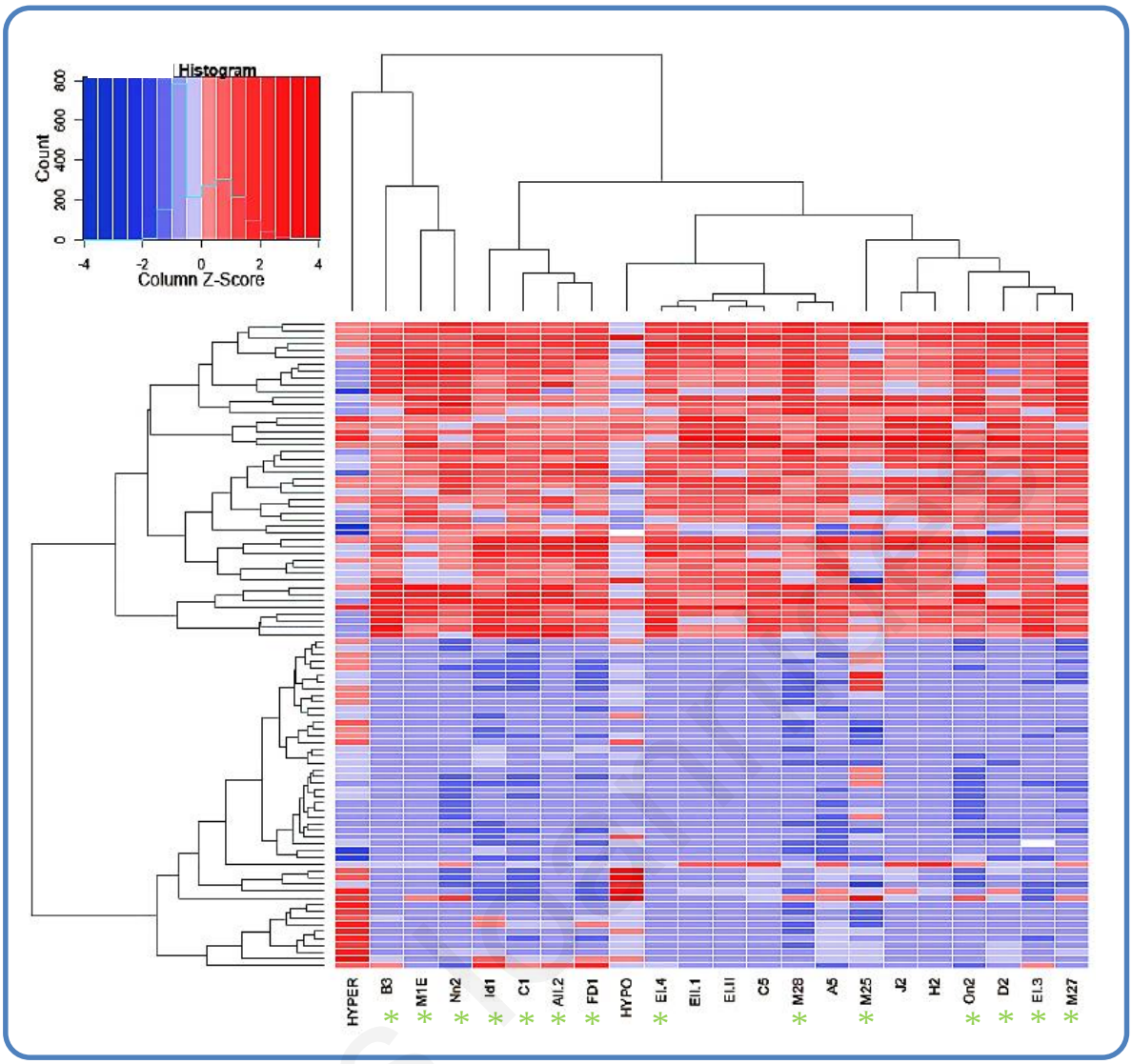


Figure 3.5 DMRs exhibiting tissue specificity for CVS and WBF with the Non LM-PCR protocol. Similarly to the LM PCR approach, CVS samples show higher enrichment (hypermethylation – red) for the DMRs tested as compared to the WBF samples (hypomethylation-blue). * Newly selected DMRs

Table 3.9 Correlation between LM PCR and Non-LM PCR protocols.

		r (Pearson correlation)	P-value	Result: Evidence for an association	Strength of the correlation
NON- LM PCR vs. LM PCR	**WBF**	0.967	1.71×10^{-14}	YES	High
	CVS	-0.069	0.749	NO	-

Pearson correlation between the two protocols indicated that there is no correlation between the enrichment values obtained for the CVS samples on the DMRs tested whereas WBF appear to be strongly associated.

3.3 Biomarker Discovery Using MeDIP-Chip Approach on Chromosome wide 1M Ultra High Resolution Custom aCGH

3.3.1 Identification of New DMRs

For each custom ultra-high resolution aCGH experiment five samples of each set tested were pooled after the MeDIP and LM PCR procedures. Next, 1μg of the pooled PCR products was used for the aCGH protocol (Figure 2.3). After MeDIP, each one of the samples underwent quality control using qPCR with a control marker in order to ensure that the MeDIP procedure was successful. Following aCGH, the features (probes) of each chip were extracted and assigned their chromosomal location using the Feature Extraction Software. Next, the signal intensities of the normalized $\log_2$ (IP/IN) were calculated automatically by the Agilent Genomic Workbench software. Probes that did not perform well in any one of the array chips of the same chromosome were excluded from all array sets.

For final analysis and selection of DMRs, data files (gff files) were imported in the Signal Map Viewer. Parallel screening of the aCGH results allowed the simultaneous correlation of the identified DMRs with CGIs, CpG content, promoter regions, genes and CNV regions. We have identified 371, 682, 2825 and 537 DMRs on chromosomes 13, 18, 21 and X respectively, that showed hypermethylation in 1st trimester CVS, 1st trimester abnormal CVS (i.e trisomy 13, 18, 21, or monosomy X) and hypomethylation in WBF (Table 3.10). From these, 47-71% were located on intragenic regions (gene bodies) and only a small number was located on CGIs or promoter regions. We have then correlated the total number of DMRs found with the GC content of each chromosome tested. As shown in Figure 3.6, with increasing GC content more DMRs were identified using the MeDIP-Chip approach.

Table 3.10 DMRs identified using 1M custom aCGH for chromosomes 13,18 and 21.

Chromosome	DMRs	Gene Bodies	Gene Bodies (%)	CGIs	CGIs (%)	Promoter	Promoter (%)
chr13	371	263	70.89	3	0.81	6	1.62
chr18	682	484	70.97	4	0.59	12	1.76
chr21	2825	1320	46.73	51	1.81	45	1.59
chrX	537	355	66.11	0	0.00	7	1.30

All DMRs identified were excluded from copy number variable (CNV) and repeat regions. The majority of them were located on gene bodies and only a few of them on CGIs and promoter regions.

A subset of the identified DMRs was then selected according to the established criteria and compared to 3rd trimester CVS, in order to confirm similar methylation status between the two tissues. (In the case of chromosome X aCGH was not performed on 3rd trimester CVS due to the fact that the pooled samples used from comprised of male and female samples).

To assess the performance of the procedure, aCGH was initially implemented on the chromosome 21 chip and previously validated DMRs [126, 127] were checked for the methylation patterns between the two tissues. All seven DMRs were confirmed as they showed hypermethylation in CVS samples (1st and 3rd trimester and trisomy 21) and hypomethylation in WBF (Figure 3.7 and Appendix II). We also compared the methylation patterns of the new DMRs confirmed using the old array set (Tables 3.4-3.6). The methylation status of all DMRs was consistent between the two array platforms. (Appendix II).

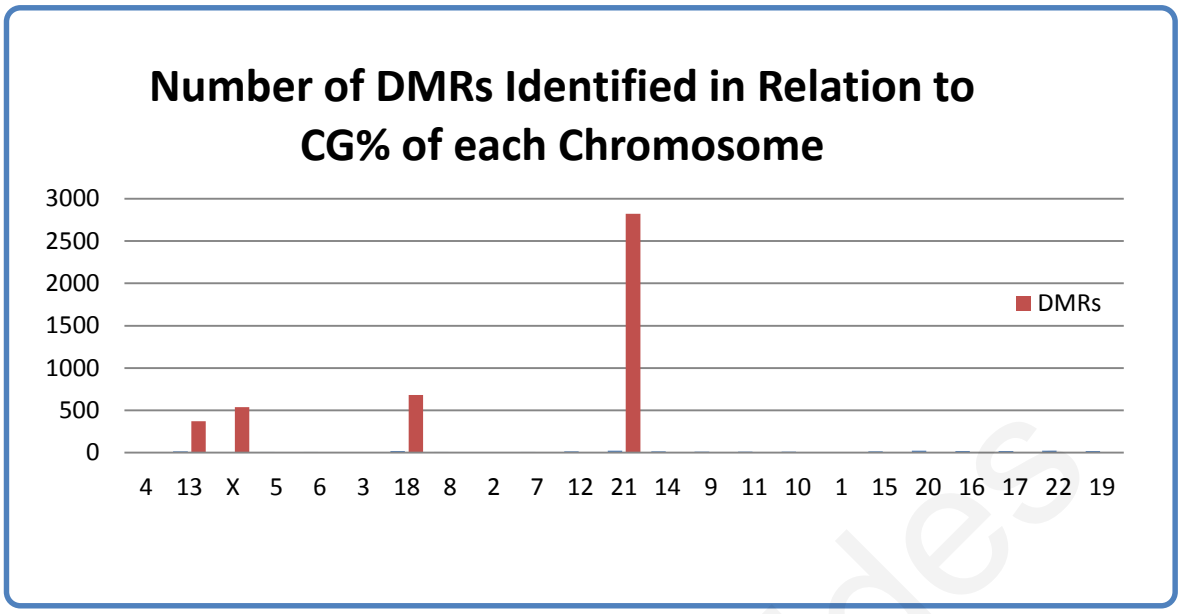


Figure 3.6 Association of the number of DMRs identified with the GC content of each one of the chromosomes under investigation. It is apparent that the number of DMRs identified increased with increasing CG content.

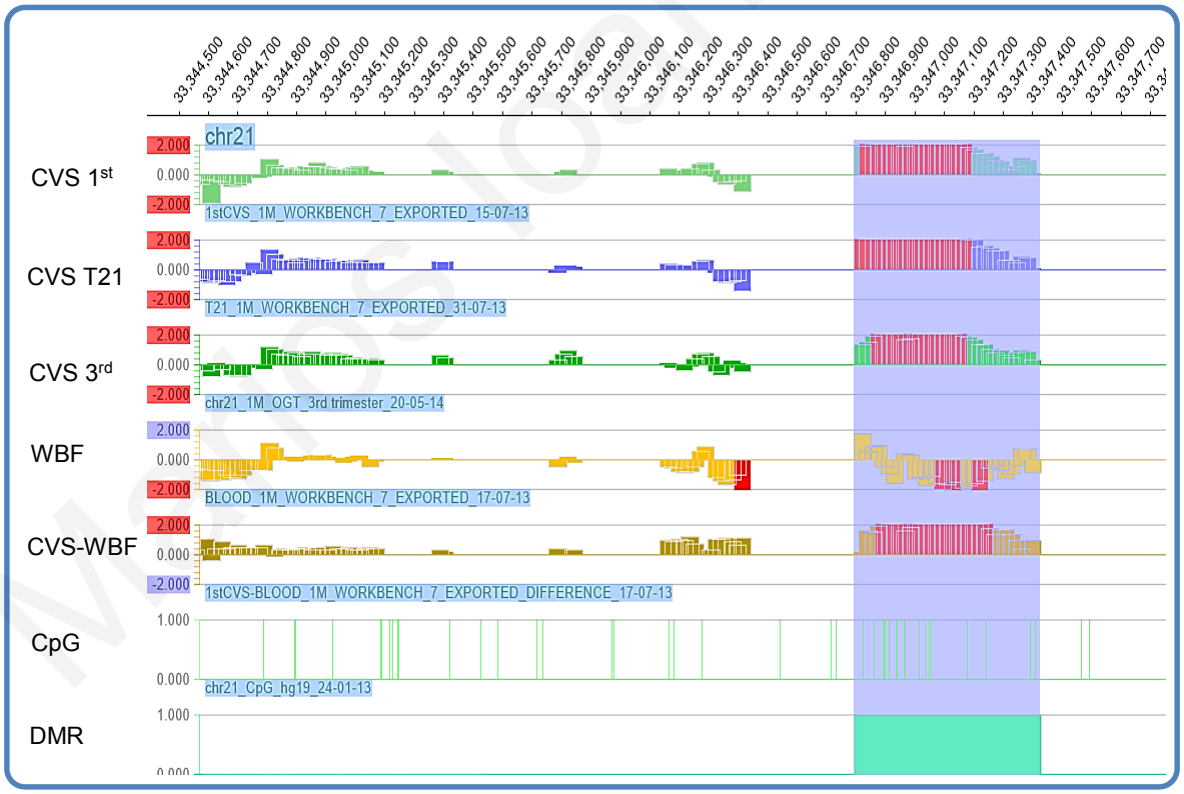


Figure 3.7 A representative (blue shaded) region (Region H) which was identified and confirmed using the previous array set. CVS hypermethylation as indicated by positive $\log_2$ ratio is apparent in 1st, 3rd trimester and trisomy 21 CVS (T21) while WBF is shown to be hypomethylated. The positive difference between CVS and WBF (5th track labeled as CVS-WBF) confirmed the hypermethylation status of CVS.

3.3.2 Screening and Confirmation of Selected DMRs using MeDIP-qPCR

3.3.2.1 Chromosome 21

Next, based on the aforementioned criteria, we have selected 118 new DMRs for screening using MeDIP-qPCR on multiple 1st trimester CVS and WBF samples. Trisomy 21 CVS samples were also utilized in order to confirm the consistency of the differential methylation and potential trisomy 21 specificity. Based on the qPCR results obtained, we have classified the new validated DMRs in four categories:

1) ***Not DMRs***: qPCR results showed no distinction in the methylation levels between CVS and WBF (Table 3.11 and Appendix III)
2) ***Bad DMRs***: qPCR results showed methylation variability among the samples tested and/or average Ct differences in values between CVS and WBF samples less than 2.6 (Δ_1Ct<2.6) (Table 3.12 and Appendix III)
3) ***Good DMRs***: qPCR results showed clear distinction of the methylation status between the two tissues (Δ_1Ct>2.6), but no difference between normal CVS and trisomy T21 CVS (Δ_2Ct<1.7) (Table 3.13 and Appendix III)
4) ***Good DMRs- T21 Specific***: qPCR results showed clear distinction of the methylation status between the WBF and CVS samples(Δ_1Ct>2.6) and clear distinction between normal CVS and T21 (Δ_2Ct>1.8) (Table 3.14 and Appendix III)

In total, we confirmed the differential methylation status between CVS and WBF in 46 new DMRs (Table 3.13) including four that appeared to be more methylated in the trisomy 21 samples (*T21 specific* DMRs) (Table 3.14). The ten best performing DMR primer sets (i.e primer sets that showed the ten highest Δ_1Ct values as shown in Table 3.13) in addition to all the *Good DMRs- T21 Specific* (Table 3.14) were correlated with respect to CGI location, promoter region and gene location (Table 3.15).

3.3.2.2 Chromosomes 13, 18 and X

Based on the findings on chromosome 21, more stringent criteria for DMR selection for chromosomes 13 and 18 were implemented. In addition to the established ones, the presence of at least two CpGs in the region in order to be selected as candidate DMR was necessary. Further criteria included the $\log_2$ (IP/IN) value of WBF to be at

least -0.5 and the difference value between $\log_2$ (CVS-WBF) to be at least 1.5. Applying these criteria we selected and designed 79 primer sets corresponding to 70 DMRs on chromosome 13 and 62 primer sets corresponding to 55 DMRs on chromosome 18. Interestingly, all DMRs selected showed differential methylation in CVS as shown by the Δ_1Ct values. Despite this, according to our classification requirements only 31 DMR amplicons on chromosomes 13 and 22 DMR amplicons on chromosome 18 were classified as "*Good DMRs*". Due to lack of trisomy 13 and 18 samples, we were not able to select for trisomy (13 and 18) specific DMRs as in the case of trisomy 21. The primer sets considered and Δ_1Cts for all the regions are summarized in Tables 3.16-3.19. The location of the ten best performing DMR primer sets (i.e primer sets that showed the ten highest Δ_1Ct values as shown in Tables 3.17 and 3.19) of each one of the chromosomes was further correlated with the CGI locations, promoter and genic regions (Tables 3.20 and 3.21).

Comparison of the confirmed DMRs from the old array set (shown in tables 3.4 Table 3.5) as well as control region *SERPINB5* located on chromosome 18 [64] showed differential methylation status for all DMRs for both chromosomes (Appendix II). It was apparent though that certain markers on chromosomes 18 (i.e VA1, VA4, VA17, VA22) showed increased methylation status for WBF as indicated by the positive $\log_2$ ratio. aCGH is a hybridization based method that is largely influenced by hybridization dynamics. Given that the probes were designed so that they would have optimal hybridization characteristics, the high GC content of the DMRs, as well as the high density of the probes would cause tighter binding of the probes and/or low signal to noise ratio respectively, resulting in increased background. To further investigate this, we took advantage of the fact that chromosome 21 DMRs were selected with more relaxed criteria, resulting in many of the DMRs to fail to be confirmed. We correlated the location of the DMRs that were classified as "*Good DMRs*" to the locations of the DMRs that failed to be confirmed (classified as *"Bad DMRs"* and "*Not DMRs*"). It was apparent that 28 of 46 of the '*Good DMRs*" were located on the part of the chromosome with intermediate GC content whereas 46 of 72 of the DMRs that failed to be confirmed were located in a region of the chromosome with high GC content (Figure 3.8).

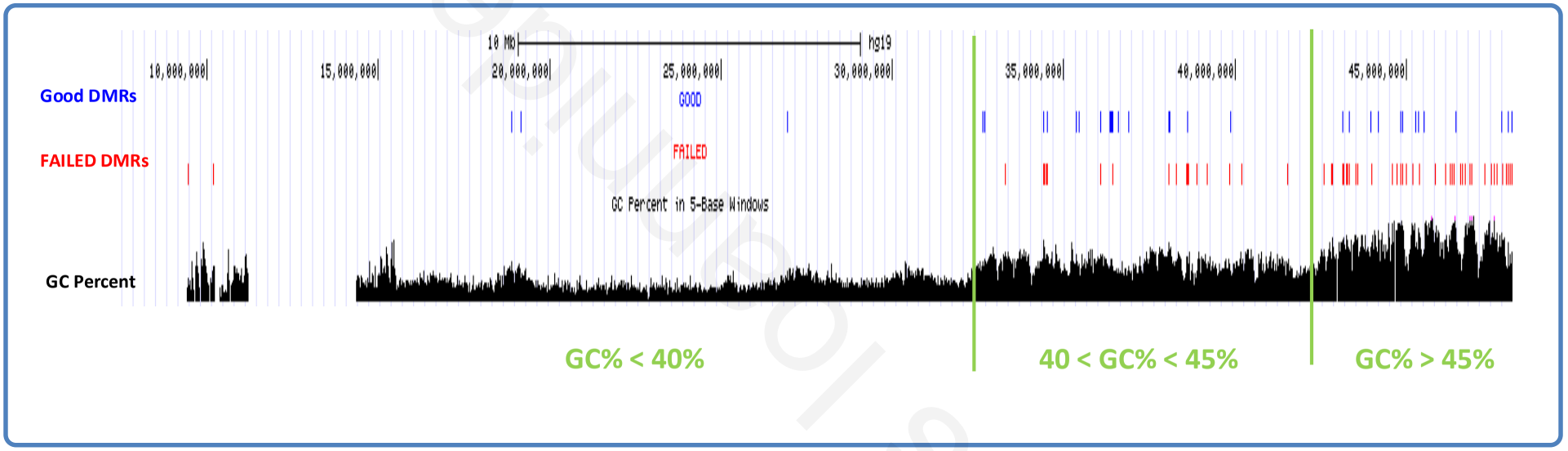


Figure 3.8 Graphical view of the DMRs of chromosome 21 in relation to the GC content. Most of the DMRs that failed to be confirm (Red Track) are located on the part of the chromosome that is high in GC% (GC% > 45%). Most of the *"Good DMRs"* (Blue track) were located in the range of GC% of 40-45%. Green vertical lines highlight the three different GC% levels of the chromosome.

Table 3.11 Primer sequences of DMRs on chromosome 21 classified as "*Not DMRs*".

DMR	START	STOP	FORWARD	REVERSE
1T21	9439496	9439579	CATCTTCATTGGCAGCCTTC	TCCAGGGTCGGTATCTTCT
7T21	10164107	10164251	GGCAGTGAAACGGGTTCTA	TGGGACAGATGGGCAAAC
9T21	34444182	34444306	GATTCACCTTTCGTTTCCCTTTC	GAGACGCCTGGTTACCTATTT
11T21	34511428	34511524	TTCCTAATGAGAACAAGCGTAGAA	TGCGGGTGTGCATCTTC
12T21	38267125	38267257	CCCAAGTTTCCTAAGGCAGAG	GCTGTCCAGCCTGTGTTT
13T21	38599868	38599983	TTCTGAGCTCTGTGACCCA	TACACTGGAGACCACCAACT
14T21	40194893	40194973	CCAGGAAAGGCAGGATTGAA	TCAAGAGGCGTGGTTTGG
16T21	41516081	41516189	AGCAAATAGCATTTCTGCATTCTC	GACTCTTCCTCATTCATTTCAGGTA
20T21	42830604	42830701	CAGCGACAAGCGGAAGT	GCTGGACAGAGTGTGGTT
26T21	43255583	43255690	GCGGAAGGTCTTCTCACAA	CGGCATCTGTGGACTCTG
27T21	43274773	43274880	TCTCTGGAGGTGGTGAAGTA	GGATGACTGCAACTGGATGA
28T21	43278710	43278820	GACAGAGGCGGACATGAG	GCTGAAACCCTTGGTCCT
29T21	43530937	43531055	TCTTACTCGTGGGTGGGAT	CAAGACCTGGGACCGTTAC
30T21	43580614	43580720	GGGGCAGTAACTGAGGTAGC	AGGCTCCAGGACTCAGAGAG
32T21	43984808	43984893	TCCGTTGCAGGTGGGAT	GCGAGCAGCAGCATCAG
33T21	44576708	44576837	CAGGTGAAGCACTCAGGAC	TCGCTGTGTTCAGAAGAATGA
35T21	44882049	44882166	ATCGCCGCTGTTGGTTT	GGCTCTACACTTGCTTCCAC
36T21	45825827	45825918	CTCTCTGGACTTCATCCTGTTC	TCCGCTTCACAATGATGATCT
38T21	46273936	46274029	AGCCACCCAGCGTTTAC	CGTCAGGTGCAGGTGATT
39T21	46323816	46323915	CTCATCAGCAGCCACATAGA	GGGCTGGGATGGAGAAAG
42T21	46567282	46567362	CCTGTGGGTTGCTTTTGACG	CCTGCTGGGAGTTCGGGA
43T21	46640922	46640997	ACGGCAAGGTACTGAGG	TAGGTGTGAGGACAGACAAGC
44T21	46697875	46697991	CGCAGTTGCCCAAATCCA	CTCAGTTAGTGTATGTTGCCCA
45T21	46834353	46834411	ATGACTGAGAATGTTGTGGGAGG	GCTGAATTGGGACAGGTGGTC
46T21	46860364	46860419	CTCCGTGCCATGTTCCC	GAACGTGGCTCAAACCC
47T21	47287232	47287328	AAAGCAGCAACCACCTAGAA	GTCCTGTATGCTGGTGACTAC
51T21	47470109	47470247	CCATCGAGAAACCATCCTCTG	ACCAGCCGAGTTTGGAAAG
58T21	47795998	47796105	CCTCTACCCTCTCCGTCTG	CCAGGCGTCCCAAACTAC
55T21	47626491	47626594	CCTCGGGCAAGGGGGA	AGGTCTTCGGTGAGTGGTC
59T21	47808920	47809042	CTGCTAGTTTCCGCACTTACA	ACAAGGAAGCCACAGAACTC
60T21	48018636	48018720	ACTTGAATCGCATGGGTCAC	TGGAAAAAGCAACTCCATCA
61T21	48069718	48069807	GCCAACCTCAGGATCTCAAG	GCCAATGTCCACACAAGAAA
114T21	44985468	44985578	AAGGCCGTCACCAAGTACAC	TTTTGAATGCTGCCTTTTCA
115T21	45182037	45182136	TTCGTTTGCCTGGAGTGG	GGACCAGCACGAAACAGAT
116T21	45362406	45362512	GCCACTGGTGTTCAGTATGT	GGTCTCTGAAGATGAACGAGAAG
120T21	46897304	46897406	CACCTCTGCTTCTCTTCCCA	CTTACCTCTTTCACGCGGC
121T21	47546035	47546121	GTGAAAACCTCTACTCCATCGC	TGAACTTCTCAGCGACCAG
Un11	42587945	42588086	TGAGGAACCTGGAAGGAGCC	GGACCATCACCCACTTTCAC
Un14	43321279	43321397	CCACGGGGGACACTACAATG	CAGTCACGGGGAACTTTCA
Un20	44710288	44710383	GACCTTCATCTCCTGCTTCTG	CTTGAATGGACAGGGATGC
ZH7	38592859	38592938	CAGGGTCGCTATAACGAGGT	CGCAGCCATGTGAACATTAG
ZH8	39169894	39169989	CGCCTGTGGTTTGCATTT	CCCGAGTTGTGAGGATGTT
ZH11	46572645	46572743	CTTCAGTGTGAAGCATTTGTC	AAATAGGAAGTAACTGGCATCA
HZ8	38886643	38886761	TTGGGGTGTGACACTTTTGA	AAACAGGTTATCGGCAGAGG

Table 3.12 Primer sequences of DMRs on chromosome 21 classified as "*Bad DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ1Ct
112T21	42798520	42798617	CTCTAGTAAGCGGGGAAGGG	GAAGTCGACGGAAAGCAGTG	5.1*
113T21	43145690	43145798	CGCTCGCAGCTACACAC	GGAATAGTGCTCAGGCTTGG	3.7*
104T21	46128769	46128936	CCTTTCCTGGCGTGGATG	GCCTCAGCCCCTGAGCTAC	3.3*
75T21	34398403	34398486	GACCCAGAATGCTCCGATAC	GGGAGGACAGAAGAGCAAAG	3.2*
92T21	38631075	38631180	TGAGACCTGTCCAGTATCGAA	GTTCTCCACTTGGTGTTTGTTG	2.1*
Un7	36422650	36422749	CCACTGTGGTCACACGTATT	AGCCTTTCTTCATCTTGGAACT	2.8
Un12	43167085	43167191	GCTTCCTGTAACCCTCAC	AACACACTGACTCCTTCTGGTAG	2.6
Un5	36080201	36080318	CTCACGGTGGCCCATAAA	CCGTCTTGACTTCACCATCA	2.5
84T21	38066228	38066335	CGTGCCTCTTCTTACGGATT	GACACAAACTCGTTTGGTTTGA	2.5
86T21	38069366	38069444	GTTCTTGTTACTTGTCGGGTTTC	CCTGGGCTTGAAAGGGATTA	2.4
76T21	34396450	34396559	CCAGAAGGCACGAGACTG	ACAAAGAGAAACCCACGTTATTG	2.3
Un18	47971308	47971402	TCCTCTGAGATGCGTCGTG	GCTACAAGGTTTATGCTAAGTGACAA	2.3
72T21	34396934	34397040	TTGCTGCACACGGTCTC	GATGAGGAATTAAGCCCGAACT	2.2
74T21	34401595	34401713	GCACTAGAGACACACATGCATTA	GAAAGGCAGTTCCGGGTTAG	2.2
90T21	38592853	38592978	TCCTCACAGGGTCGCTATAA	AGTTTCAGTAGATACCGGCTATTG	2.2
52T21	47477660	47477747	CTCCTGCCTAGAAAGATCCAAG	GCTGTGAGCAGGGATCAG	2.1
Un3	33282057	33282156	ATAAAGCCAGCAAAGGCTATTT	CCAAAGAAACGATGTGGGTAAA	2
78T21	34482254	34482337	GAGGAAGTGCCAGGTCATAAA	GATTAGCGACTGGGAGGTTTC	1.8
108T21	44835128	44835234	CCCTCAAGTAACTGGGTCAATC	CTTCAGCTACTTCGGCTTCTT	1.8
105T21	46391167	46391268	GTTAGGCTCCAATTCCTCCTC	GAGTGCTGGGCGGTTAC	1.7
107T21	47918892	47918984	ACACAACCCAGATAACCCATAC	ATTAGCCCATTTCTTTCAAGTGC	1.5
122T21	46129779	46129885	CGTTCCAAATGCAGCTCAATA	CCTGAAGCTACGTTGCAAAT	1.5
123T21	46572510	46572608	ATTGGCGTGTCTGCATGG	GCTTCTTGCGCCTCCTTT	1.5
10T21	34482514	34482600	CAAACGGTCCTTGTGTGTATAAG	CGTACCGCTGCAATGTAAAG	1.5
94T21	39844980	39845075	CGCTGCCTTCTGAGACGTAG	TCCTCCATTTCATTTTGAAGTATG	1.4
100T21	44838749	44838847	CTGTGTGGTCTCAGGGATTAAA	GTGAGGGTGGGCTAATTTAGAA	1.1
111T21	38267121	38267268	CAGACCCAAGTTTCCTAAGGC	CTTCACCACACGCTGTCCAG	0.3
117T21	45550922	45551071	GGTATTCACATGAACTTTGTATGAGAC	CAGCAGGATACAACATAACCCT	-0.3

*Indicates highly variable Ct values among the samples tested.

Table 3.13 Primer sequences of DMRs on chromosome 21 classified as "*Good DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ_1Ct	Δ_2Ct
102T21	45336762	45336881	GGTGGTGTTGAGCGGATAG	CCTAGAGAAGCAACAGACCTTAC	5.9	0.6
Un25	45336784	45336895	GGGAAGTTGGAGTCTTGTTTGT	CCACCGTCACTGTTCCTAGA	5.7	0.8
UN19	48087545	48087629	TTCTCCTGGAGGACTGTGTT	GGCGTTATGCTGACCACTT	5.5	1.4
Un8	36577515	36577613	TGATCACAGGAAGTTGAGAAGAG	GCTGCTAAACGGCTGATAGA	4.9	0.9
Un16	44167086	44167191	GGACTCTTGCTTCCAAAACGG	GCGGCTGGGGATGAGACC	4.9	1.1
Un17	47788835	47788938	GCTTCCTCCTTCCTGACTGTG	GGTTCCACTTTTCTCCTGTTCTTAC	4.7	0.8
Un4	34524282	34524379	CAAAGACAACCATCGAGGAGAG	TTGTAAGAGGTGCTGTGTTCC	4.6	1
Un9	38598152	38598226	TTTGTCCTCTCATCTTGGAACA	TTTAGGAAAGCAAAGGGACAGG	4.6	0.9
Un10	39869802	39869989	CAGCAGCCACTTGCCGAGTATC	TATCCCTAACGGTGCCTGGA	4.4	1
Un23	45509249	45509355	ACATATTCAGCCTCACATGGATTA	GACAAAGTGCTTTCAGTCATGTATT	4.3	0.3
Un22	45254848	45254983	GTCTCCGGAGGAAAGCATAC	TTGTGAGGGTCCATGAAATACT	4.2	0.7
101T21	45254816	45254963	AGGTCTGGAAATGGTCTGAATAG	CTTGAGTTACCAGGATGTTCATTT	4	0.5
65T21	26935493	26935632	AGGTCAAGTGGACATTATGTGTAT	AAGTCTGGCTGCCCTAATTC	4	0.9
ZH1	32699157	32699267	AGAGGACCAGATGTGAAAGTTATT	CATATGACCCTGTTAGCCTGTT	3.8	0.7
ZH2	35349029	35349143	GATGACACTCCCTCGTTCTAATAC	AACTGCATTATTCACGGTAACATC	3.8	0.3
109T21	38083165	38083258	TCCTCCACTTCCAGCTC	CCCAGCACTTAGTTGTCC	3.8	0.9
99T21	44834725	44834834	GCTTCAAAGAGGAAACACACAA	GTAGCACAACTGATTCCAACAC	3.7	1.4
Un15	43947866	43947967	CAGAGCCTCCTAAACGCAC	CTTGAAAGGTAAAGAAACCCACA	3.7	0.4
Un13	43319784	43319858	AGATGCGAGAGAGAGAGCC	TAAACCCGAAACACCCACGG	3.7	1.1
79T21	35349029	35349143	GATGACACTCCCTCGTTCTAATAC	AACTGCATTATTCACGGTAACATC	3.6	0.8
67T21	32699157	32699267	AGAGGACCAGATGTGAAAGTTATT	CATATGACCCTGTTAGCCTGTT	3.6	1
71T21	34393240	34393327	TGGGAGGGGTAGAGGAAAAG	GTCGCCTGTCAGGATCAAGT	3.5	0.5
81T21	36356697	36356838	GTCCCATATGAGGGCAGAAA	CCTCTGAAATAGTGAATCTCAAACC	3.5	0.3
ZH5	36901433	36901544	ATCAGCGTTCCCTCAAATACA	TCAAGCAGCCTGGTTTAGTT	3.5	1
82T21	36421827	36421952	CTCAATGGTCTTTGCTGATTTAGT	TTGATGCCAGCGTTGAATTAC	3.5	0.3
Un21	44876857	44876957	TTTCACCTCCATTTCCTTGTAGA	CGTCTGCTGTTTGGCACC	3.4	1.4
23T21	43146159	43146257	GCCAGTCTGGGTCTGTTT	GGAACTGACAAAGGCAACTTTA	3.4	0.7
ZH6	38063551	38063668	TGTAACTTCACTTGGAGCCTTT	GCCACAGCGATCCCTTAAA	3.3	1.7
64T21	19162216	19162298	TAGTTACTCCCACCAACGTAAA	TTGTTCAGATGTAAGGCAAAGA	3.3	1.6
ZH4	36341270	36341387	AGAGTTCTGCCTCCCCATTT	GGGCGCAAGAGATAAATGAA	3.2	0.1
Un1	32637891	32638002	TCAGAGGAGTTGCCGACTAA	CCTTTCTAACTTGGTCCTTCCC	3.2	1.2
62T21	18886734	18886874	GGGCTGAAATGTCCAAAGTAAG	CACACTGTGTTCTCAACTTCATC	3	1.1
69T21	34392812	34392897	GGGAAATTGTAGCACTGAAGC	TGATAAGACTGTTGCCCAAAGA	2.9	1
106T21	46451073	46451163	AAACCGCCTTCTCCTTCAG	AGTGTTCGTGGCTTCCTTT	2.9	1.2
110T21	34400342	34400429	GTTCTCTCCGGGACCTGATC	CACACAGCGGTACCTTTTCA	2.8	1.3
8T21	34400764	34400864	GGTCAATCCACACCCTCTTAG	TGAAGAAGGAACATCCACAGATT	2.7	1.7
70T21	34395009	34395094	CGCCATTTGTTTGCAGAGTT	CGCTCCTGGTCGTGAATG	2.7	1.4
Un6	36398129	36398227	GGAAACTCATATGGTTCTCTGC	TCATTCCTTAATCTTCCTGTGTTC	2.6	0.8
ZH3	35449681	35449783	AATTACGTCTCGTGCCCTTT	CCTGAACGCCAATGAGAAGA	2.6	0.8
Un5	36080201	36080318	CTCACGGTGGCCCATAAA	CCGTCTTGACTTCACCATCA	2.6	1.1
84T21	38066228	38066335	CGTGCCTCTTCTTACGGATT	GACACAAACTCGTTTGGTTTGA	2.6	1.6
Un18	47971308	47971402	TCCTCTGAGATGCGTCGTG	GCTACAAGGTTTATGCTAAGTGACAA	2.6	1.7

Table 3.14 Primer sequences of DMRs on chromosome 21 classified as "*Good DMRs-T21 specific".*

DMR	START	STOP	FORWARD	REVERSE	Δ_1Ct	Δ_2Ct
TS1	42354430	42355570	TTTCCTTTCTGGGCACCTT	CGTCGTAGAGTTCGTCTAACAG	4.3	1.9
TS2	38074232	38074350	TCGAAGCAGCAATCCAAAGA	ATCCCTCAAACAGCGAACAG	3.8	2.1
TS3	38077238	38077344	AGGATTGAAGCGTGCAGAG	TAGTTGCTCTAGGACAGCCTAT	3.8	1.8
TS4	38079788	38079936	GTGGAAATTAGCCACCTCCTC	CCTTCCCTATTCGGCAACTAAA	2.7	2.3

Table 3.15 Details on the best selected DMRs regions on chromosome 21

DMRS	START	LOCATION	GENE	OMIIM
102T21	45336712-45336889	Intragenic	AGPAT3	614894
Un4	34524265-34524468	Intergenic		
UN19	48087516-48087719	CGI		
Un8	36577416-36577619	Intragenic	RUNX1	151385
Un16	44166999-44167310	Intragenic	PDE9A	602973
Un17	47788699-47788992	Intragenic	PCNT	605925
Un9	38598143-38598274	Intragenic	DSCR9	
Un10	39869594-39870031	Intragenic	ERG	165080
Un25	45336694-45336915	Intragenic	AGPAT3	614794
Un23	45509229-45509396	Intragenic	TRAPPC10	602103
TS1	42354413-42355423	Intragenic-CGI	DSCAM	602523
TS2	38074043-38074372	Intragenic	SIM2	600892
TS3	38077214-38077399	Intragenic-CGI	SIM2	600892
TS4	44834618-44834929	Intragenic	SIM2	600892

Table 3.16 Primer sequences of DMRs on chromosome 13 classified as "*Bad DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ_1Ct
V1.42_2	88326732	88326834	TCTAGGCATCGGGAGGTATT	CTACCAAACCGCATCTTTAAGC	5.1*
V1.42_1	88326914	88327001	AGGACGCTGATGTGCTTTAATA	CGGAGCTTTGAGTAAGGAGAAA	4.7*
V2.34_5	112859697	112859834	AGTTGAGAAGTGACCTGTTGTT	CAGATGGTTCATTCATGGCT	4.4*
V2.25_4	99777306	99777388	ACAATTATGATTACTCCATGACCAG	TGAGACACAGAAATATGAATGCAG	4*
V1.6_1	28503642	28503760	TCTCACCTTGAAGAGCTAAATCC	CTGGAGCGAAGAGCAAGTTAT	3.8*
V2.43_3	112859088	112859231	GCAGGTGGTTGGAATACTATGA	GTAATTCATGAGCGCACCTTTC	3.8*
V1.13	33591560	33591658	GGCGTGGAATTAGGAGAGAAA	CTCACTGCGGAAACCTATCA	3.7*
V1.5	28498982	28499114	AAACCCGCTCTCTCAGG	AACAACGCCAATCCAGTATC	3.5*
V1.24_1	51288372	51288480	GAAGCGTTGTTTCATCCACATTA	GGAGTTGAGATCCCAGAAAGAG	3.3*
V2.2_1	22630507	22630651	GTCTTCCGTATAGTGTTTAGTTGTT	AAAGTCACTTCTTCCTCTTCTTTAG	3.2* **
V2.8_4B	33002948	33003057	CTTGGAGAACTCTTTCAGATCCC	ATGGGTACTCCCTAGTTCCTG	3.2*
V2.32_2B	111854266	111854412	TGGGCTTCTTTTATCCCTGA	TGATCCGATACTCCCAAAATG	3.2*
V1.51	108865425	108865527	CAGCTTCTACTGAGTGGGTTAG	CAGATTCAGAGAGAGGAAACAA	3*
V1.39_1	88325982	88326124	GATTCGCAGGCACACTGATA	GCAACTCCAAACTAAAGCCAAG	2.9**
V1.19	44808241	44808384	ACACACCAAGGAATGACATACA	CCCTCACTGAGAAGGAATTG	2.6
V2.11_2	44808241	44808383	ACACACCAAGGAATGACATACA	CCTCACTGAGAAGGAATTGG	2.6
V2.26_1	99777523	99777641	CAGTCCCAGACTCTCAAATCTTTC	CTCTACAATGTATCATTTCATCAAACAC	2.6
V1.12	33484751	33484844	CTGAATGAAACATAGGCAGCTTAC	TTGCCATAGAAGCCTGAAGTTA	2.5
V2.18_5	52735098	52735205	TGTCAACCAGAGTCAGAGAGA	CAGGACACAATTCAAATGCTTC	2.5
V1.57	111819548	111819627	GGTCATATAGCGTGAAAGTTGG	GACGAGTCCTCAGTTTCTTCTAA	2.5
V2.35_5	113000906	113001002	TGAGATAACCTCGGCAAATTAG	AGGAGAGTTTATGGAGTTACAATCTG	2.5
V2.18_5	52735098	52735205	TGTCAACCAGAGTCAGAGAGA	CAGGACACAATTCAAATGCTTC	2.5
V2.19_1	78406432	78406533	TGACTGAAGCCATTGTACCC	GGATAAGGCCTAAGAAGTAAGCA	2.3
V2.19_1	78406432	78406533	TGACTGAAGCCATTGTACCC	GGATAAGGCCTAAGAAGTAAGCA	2.3
V1.1	21619455	21619553	GCCAGTCTCTTTCCACGATAG	TCTGATCTGTTCTTGTGGAGATG	2.2
V2.1_1	22624823	22624929	GAGGTTCCTCGTGTAGGTTATC	GACAAAGAGGTGAAAGCAGTTAG	2.2
V2.44_2	113236611	113236739	ACGTCAGGCCCTAACAATATG	GCTGTCATAGTGCCCTCATT	2.2
V1.39_2	88325803	88325936	GATCCAGCGGAGAACTGAAA	AGTGCATAAACCTCGCAAGTA	2.1
V1.27	52395997	52396126	GCACAGCCCTGGGTAAAT	CACCTGTTATTTGCCTGCTATTC	2
V1.28	52396273	52396405	CCTACATAGGGCAAACAGAAGA	GAGCAGATCCCACAAGCATA	2
V2.24_1	99715192	99715337	CTCTGCGGGTGTTACCATTT	GCTTTCCTGGACCTCTTGTATT	2
V1.45	99856203	99856301	TCTGCAATTATGACCGGGTA	CAACGTAGAGCTGAGTGACAG	2
V2.41_2	99934117	99934215	GAATTTCAAACTGCGGTCTC	TAGACTTCTGGCAACAGCTTAC	2
V1.21	45082492	45082581	ACACATGGTTCCGATGTTCA	ACCATCTGCAGCACAACTG	1.8
V1.22	45116554	45116680	CAAATCTTTATTTATCACACACCACAG	TTTCTCCAGTGAGCTCTCATATTT	1.8
V2.7_1	31712502	31712622	CGTAAGTGAACCCAGCAGTAAA	TGTCATGAATGCTCAGGCTAAA	1.7
V2.9_1	33922317	33922454	TTCAAACCTCATACCGCTATTT	TCCTGGGAAATAAACTGTAGATAGA	1.7
V2.12_5	46929017	46929117	AGTTGGTGGAGCGATAGTC	GGGAACATGTCTCTTTGGTT	1.7
V2.14_5	48986368	48986470	CAAGTCTGACATTGCCAAGTTAAT	TGGGTTAATATCCAATTGTGTTGC	1.7
V2.37_2	52316603	52316689	CCAGTTATATTTGCAAGCATCAGTTAC	CAAGGTAGGCATGGGACAATAAG	1.7
V2.22_1	98825105	98825188	ACGTTTCTGAAAGCTAGGATGAG	TTGGTCTGCCTTTGCTACTG	1.6
V2.21_3	97906928	97907072	TGAGACCCTCTAAACTTCCAAA	GTTTCAGTCGTGTTTGCTCATA	1.4
V2.38_2	73430102	73430239	CACTTGATAGCTACTGTGTGGTT	GTTGCAAATGTACATACAAGTAGGA	1.1
V1.36_1	77097199	77097314	TGCCTTATATCGACACATTTCCT	GCTCTTTCAACCTGCCATTC	1.1
V2.29_5	110867356	110867450	GTAAATAGAGTAACAAATAGCGAGGTG	GCACTGAGGCTGTTGCTAA	1.1
V2.42_2	107352658	107352750	TTTACTTCCGCGGTATCCAG	TCATATCATTGTTGAGGCTCATTTC	1
V1.36_2	77097302	77097405	AGGTTGAAAGAGCCTTGAGTC	GACTCGTTTCTATCGAACTGCT	0.9
V2.22_4B	98825268	98825391	GCATTCTGGTTGCTTCATAAAT	AAGAGCTCTATGCCAAACTGAT	0.7

*Indicates highly variable Ct values among the samples tested.

Table 3.17 Primer sequences of DMRs on chromosome 13 classified as "*Good DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ_tCt
V2.15_5	49079919	49080022	ACAGGATACAGGCAGCAATC	TTTAGTCACGGAAGGAAGCC	7.2
V2.27_1	102572101	102572206	CCTCATGATGGGAAGGAAAT	AAGAAAGAGTATCTCTCGGTTATGG	6.6
V2.15_4	49079743	49079855	GGCAAGCCTCCTAAGTTACA	TGGCCCTAGAACTACACTCTG	6.2
V1.29_1	60026134	60026245	AGGGATACAGGAGAGTCAAGAT	TTCCCTCACTCTTTAGGAAACTG	5.9
V1.48	107143538	107143655	CAGAAATAAACGCCGGAACATC	TCCTCCGAAGGCATCTCA	5.8
V1.49	107146286	107146380	AAATGTTCAGCTCGTAACTCGT	GCTGGCAGTCTGTTCGG	5.4
V2.30_1	110993222	110993323	GGACTCTGTTCCTTTCTGACTC	CTGCCAGGTCACTGAAATCT	5.4
V1.55	110994208	110994350	GGCCAGATGACATCACATAGAA	TGAGTCCTGTCTAAGCAACAAG	5.4
V2.39_2	84453616	84453711	CCATTGTGCCAGTAGGAAGAG	GAACCGAAAGCGGTCCAA	5
V2.5_2	28492817	28492907	ACTCTCAAAGGTCTGGATGTTG	GAGAGAGCTTGAACCACAAAGA	4.8
V1.41	84455402	84455509	GTTAGCGAACTCATTAGGGAAA	AGGGCTTCACAAGTCTGC	4.6
V2.28_2	102572377	102572478	TTTAGGTTAATCCCTGCGTTT	ACAGCAGATGGCTCTCTTG	4.6
V2.3_5	28001035	28001166	TGAGAAGTTGTGATGTTTATGTTG	AAAGCTTAAACTGCCTAGTGTT	4.5
V2.20_3	95356205	95356351	CCGCATAGTGATGACTTCCAA	CTTTCTCCGTTTGCCCTGA	4.5
V1.52	109550453	109550548	CTGTGAGAGGAAACAACAGCTA	CCTGTGTGTGACGTGGTC	4.4
V1.6_2	28503429	28503543	TGGTGGAGTATCTGGGAGAG	CACTATGCGATCCGACACTAC	4.1
V1.40_1	78397059	78397176	CCCTGTGTTAACTATCTCCAACC	CCTGCTAAACACTGAAGCCTAT	3.9
V2.33_5	112331116	112331208	AAGCGAAAGTGCTCCGT	CCTTATTACTGAGACAACCCTGT	3.8
V2.6_2	28502804	28502914	GTTGGGCTTCGGGAAAGA	GAAGAACAGAGGAGAGCTGACAG	3.6
V1.23	45149767	45149875	ATCTTCCGTGTGAGATTCATC	GATAACTAGCGTTACTCCTGCTC	3.6
V2.41_4	99933956	99934081	TGCCTGAGGACTGATTAGGA	TAGCCAGCAATCCATAAGGTAAG	3.6
V2.4_1	28006523	28006606	AGTGATTCAGAATGGCTTAGGG	ACTCTGCTTTCATGGCTACAA	3.5
V1.3	28000256	28000340	TGGGCTCTGTATTACCAGTAGA	GGTAGCGCCAGTATCCAAA	3.4
V2.36_5	50697752	50697844	CCTGAATTATTGCTCGTGGTTC	TTTCTGGAGTAGATGGAGGATTT	3.4
V2.16_4	50706273	50706396	CCTGATTTCGTTTACGAGTTTC	TCTGAAAGAGACAAGGTCTAGAAAG	3.4
V1.7	28529299	28529424	TCCACCCAAGAGAACCAATCA	TTCCTGGGCGTGATCACAA	3.3
V2.17_3	52352533	52352639	CCTGCCTTGTTCGTATTTGTG	CGGCGGTGAATACAAAGAAAG	3.2
V2.23_1	99706203	99706348	CTGACTACCACATCTCAAAGAAGA	TGCAGTGTCTTACAGTGGAA	3.1
V2.10_1	41237992	41238085	CGCATTAACCACAGGGATGA	TTGGACTCCACGGCATAATAAA	3
V2.13_3	47253644	47253771	GATCTCTGGACCTCCTAATTC	AGTCACACACAAGTTTGAGAAA	3
V1.50	107160183	107160303	ACGGACTATGCCACAAGATG	AGCAGAGCTGTTAGACAAAGAA	3

Table 3.18 Primer sequences of DMRs on chromosome 18 classified as "*Bad DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ1Ct
V2.72	56939311	56939425	GCACAGGGAACATCGAAGA	AATTGCAGCCAATGCTTCTC	4.7*
V3.1_2	813481	813616	TCAGAATCATCTTCGACGGATAG	CGCCTACCTTTCGTTCATTTAG	4.5*
V3.32_4	72783004	72783124	GCACGTCCCTCTACCTACTTC	CCTGACCACACATGACAAGAC	3.6*
V3.3_B	3060562	3060702	TGCAGCCCAGATAAGAAACA	GCAGGAAATGTGTATTGGCTTC	2.7
V3.5_3	3500025	3500117	GGAGGCCATTGATATTTCCTACA	CTGGGCCACCACAGTTATC	2.6
V3.45_B	60606844	60606946	TGGTCTGGATAGCCGTTCTC	AAAAGCAGCTTCCCCAAAAT	2.6
V2.148	51777710	51777816	GAGAAACATATCATGTATCACGTTGG	GCTGTAATTCTGTAGGCACAAAC	2.5
V2.148	51777710	51777816	GAGAAACATATCATGTATCACGTTGG	GCTGTAATTCTGTAGGCACAAAC	2.5
V3.29_5	29330214	29330305	CTCACATCACGTATCTTCTCATT	TACGCAGATTACTCTCCCAAG	2.4
V2.70	56931042	56931177	AGCAATTAAGGCGTGGTAAAG	GTGCTCTCTGTGACCTTGTC	2.3
V3.15_5	13382194	13382323	AGGCTGTTTCACTTCTCCATT	CAAACTGAACTGTGTCCGTATTATATC	2.3
V3.31_1	60617941	60618058	TGAAACCAGGAGCAGTCATTAT	TATCGAGCCTTGTCCCAAAC	2.1
V3.35_B	44181782	44181884	GCTGCCAGATGAATTCTGTT	CAAAAGAGCAAGCGAAATGA	2.1
V3.24_3	60369871	60369987	CAAGTCCTACCAAAGACCTCAC	GAGACAGAATGGAGCTGAATC	1.9
V3.40_2	9240335	9240411	AAGTCCATGTTACCTTCCTCTAT	ACCGTAGCAAGAAGCAATC	1.9
V2.52_1	43802215	43802364	CACTGTGAAACTGGTAGTCTGG	AGAGGGCACTTAATGCTTGTTA	1.8
V2.168_2	74163629	74163745	TGTTTGAGTATACTGTGTGTGTTTG	AGCTTGTCCAGGATGTTAATGA	1.8
V3.5_B4	3500159	3500306	AGTTGACACAGGATGAAAGAGG	GTGGTCTAGACGCCATTCTTT	1.8
V2.52_2	43802419	43802512	CTGTACGCTACAAGGTCTAGATG	CTTGGTGCTTCAACATGAGAAA	1.7
V2.168_1	74163817	74163935	TTCCTCTCCCTACAAACCATTC	GCCGTACTTACTCTCCCAATAG	1.7
V3.21_1	42745711	42745807	AACCAACTGCGTAAAGGAAAG	CACCTAGACTCAGAGAGGAAGAA	1.7
V3.33_1	5516005	5516080	CGCAAGTAAGTGCCCATAGA	TATCTTTCTACCCTAAGGACTCCA	1.6
V3.37_3	23557480	23557583	AATACACAACTGGGCTCAGAAA	AATCCAGGACTCTAGGCTCTC	1.6
V3.4_3	3061634	3061733	AGAAGCCTGTGAGCAAAGAG	TATGTGAATGCTCCCTCCAT	1.5
V3.22_1	43563438	43563521	ATGTGGCAAGGGATCAAATAAAG	AGGTTCTTGGTAATTTGGGATTTG	1.5
V3.6_1	5455636	5455732	GGTTCCTGCCTTCATTTGTTG	TCACTGGAAAGCCATCATCTAC	1.4
V3.9_1	9861823	9861920	TCCATTTCACTGGCTGAGTTT	GCACAACTCCACTGTCATTCT	1.4
V3.12_4	12840121	12840213	GGCAAGAAGATGACCAATTC	ACGTACTGATTACCCTTAGTTGAG	1.4
V3.9_B3	9861909	9862029	GTGGAGTTGTGCAAGCATTT	TTCGGGCCTTCAACCTTT	1.3
V3.13_B	13087627	13087719	CAAGACAGCAGTATCGCAGGT	CAAATCCCAAGCCACAGAAG	1.3
V3.27_3	60820095	60820169	GTCCTTCAGCGGAGACTACA	CAAGTTGGACGCCGTCAG	1.2
V3.36_3	60820095	60820169	GTCCTTCAGCGGAGACTACA	CAAGTTGGACGCCGTCAG	1.2
V3.26_B	60512813	60512954	GATTGAGGTTTCCAACATATCG	CATGCTAACAAAAGGCAGCA	1.2
V3.25_B	60497979	60498125	TGTGCATTTTACCTACTGGACA	TCTAAATTCACGAAATTAAGGATGAG	1.1
V3.18_2	22720284	22720426	CCAAAGAAAGAGATAAAGCAGGAAC	CTGAAACCTCTGCCATGTCTTA	1
V3.19_2	23978350	23978450	CTTCTCTTCCACTTCAGGCAA	CTCTAAACACCAGGACGACAC	0.8
V3.28_2	72604810	72604910	AATGCTCATGCCTAAGTCTGT	ATCACAACCTTTCCAGTCATACTA	0.8
V3.8_2	9574764	9574861	CTCTGTACTTCCTGGACAACTC	TTGTACTTGCTGCTCTCTGG	0.6
V3.23_B	46158064	46158144	TAGCTGTGGTTGAGGCCTTT	GGAAACGTGGGAAGTTTGTT	0.5
V3.41_2	13756146	13756254	CCTCTAGAAGATGGCTTTAAATGAG	CAGGCATTTCAATTGGGATTT	0.4

*indicates highly variable Ct values among the samples tested

Table 3.19 Primer sequences of DMRs on chromosome 18 classified as "*Good DMRs*".

DMR	START	STOP	FORWARD	REVERSE	Δ1Ct
V2.57	45912025	45912158	GGACATATGCAGAGAAGTAACTGA	CAGCCTTGTGCATGTTTGAA	5.8
V2.82	60765938	60766068	TCCTGCCTGTTTCCACTTTC	GTGCTAACCACTGGGTGAATTA	5.6
V3.1_4	813609	813699	GGTAGGCGTAGGCGTTAAA	CCTGAGACAGTCAATGACAAATG	4.8
V2.81	60765702	60765812	GAAAGGAGGAATACTACTGTTGATTTG	CCGGCAAACGAGCTACTTA	4.4
V2.4	2960420	2960563	TTCTGCAGTCGAGCTTTGT	CCTCTATGAATAAATCGCTATTGTGAG	4.3
V3.43_2	55862966	55863055	GTAAGTGCCACCGAAGGTT	AGCTAATGGATTCACGGTAAGAG	4.1
V2.149	54157257	54157350	GAAGACAGAACTAACGTCAGCA	GCAAGTGTTCAGACCCATCT	4
V2.132	29971329	29971467	CAGCAGGATGCAGTCTCTAAA	GCTGTATTCATCTATGAGGAAATTACG	3.9
V2.47_2	42339249	42339355	AATCAGCAGACCTTTATGGCT	CAGGGAGGGATTATGCTTTACA	3.8
V3.38_3	46455499	46455580	CATGCAGTCGAATCATTTCTC	GGGTAACAGTGGGAACCTG	3.8
V3.11_5	11863670	11863768	CTGCCAAATGTACCAGGGAT	GCACATTTCTGTTTCCTCCAAG	3.7
V2.47_1	42339344	42339425	AATCCCTCCCTGAAACCTTAAAT	AACAATCCACCTCACTGAGAAA	3.6
V3.17_1	20772020	20772121	GTCCTACTTTCCTCTAGCACATC	TCTCAGTCCAAGTCCTCTCA	3.4
V2.12	9244559	9244690	GAGAACTTTGTGAAGAGTGTTAC	GTCTCTGTTCTTGTTCATCAG	3.2
V3.2_1	2971219	2971326	CGTAAGAGGAGCAAGGATGTT	CTTTGCCTTTGTGCTCTGG	3.2
V3.39_4	56933569	56933671	GCTCAGGAAAGCACAGCTA	ACATTGCTGATAAGTGCTCCT	3.1
V2.126	22805352	22805462	TTCCACAGTTGGGTGTCTT	AGAAGTACAACTGCAAGTTCTGT	3
V3.6_3	5455753	5455866	GCAGAAACATTGGGCTTGATT	CGTTGGCTTAGTGACCTGAA	3
V2.51	43785239	43785360	TGTTCACAGGTAATACGGAAAGG	ACCAAAGACCAAGCACTATCG	2.9
V3.42_5	32440347	32440494	CAGTTTGGGTGCCAACTATG	AGAACTTGCTTCATTTCGTCTT	2.8
V3.20_1	38970191	38970287	TGAAACAGCCAGAAGAGAGTAAG	GTTTCGGATGCGGAAATAAAG	2.7
V3.3_B	3060562	3060702	TGCAGCCCAGATAAGAAACA	GCAGGAAATGTGTATTGGCTTC	2.7

Table 3.20 Details on the best selected DMRs regions on chromosome 13.

DMR	START	LOCATION	GENE	OMIM
V2.5	28492779-28492985	Intragenic	PDX1-AS1	600733
V2.15	49079547-49080145	Intragenic	RCBTB2	603524
V1.29	60026124-60026428	Intrergenic		
V2.39	84453560-84453717	Intragenic	SLITRK1	609678
V2.27	102572058-102572215	Intragenic	FGF14	601515
V1.48	107143491-107143795	Intragenic	EFNB2	600527
V1.49	107146281-107146536	Intragenic	EFNB2	600527
V2.30	110993171-110993328	Intragenic	COL4A2	120090
V1.55	110994151-110994406	Intragenic	COL4A2	120090

Table 3.21 Details on the best selected DMRs regions on chromosome 18.

DMR	START	LOCATION	GENE	OMIM
V2.4	2960419-2960592	Intragenic	LPIN2	605519
V2.132	29971249-29971574	Intragenic	GAREM	
V2.47	42339235-42339510	Intragenic	SETBP1	611060
V2.57	45911996-45912169	Intragenic	ZBTB7C	
V2.149	54157193-54157480	Intergenic		
V2.81	60765646-60765849	Intergenic		
V2.82	60765866-60766077	Intergenic		
V3.1	813609-813787	Promoter	YES1	164880
V3.43	55862966-55863101	Intragenic	NEDD4L	606384
V3.38	46455499-46455617	Intragenic	SMAD7	602932

Chromosome X DMR identification was performed on pools of five WBF, five female CVS samples and five monosomy X CVS samples. Since the presence of a single chromosome X in both monosomy X and normal males would suggest active X chromosome (Xa), the monosomy X CVS samples would represent the Xa for both male and Turner syndrome cases. All Turner syndrome cases were karyotypically analyzed in order to confirm the presence of an intact chromosome X. As shown in Table 3.10, 537 regions were identified as potential DMRs, 355 of which were located on gene bodies and seven on promoter regions. We have designed 84 primers that corresponded to 79 different regions some of which are located on genes that escape X inactivation. After qPCR screening, the same classification criteria were applied, as previously mentioned. As shown in Tables 3.22 and 3.23 all but two regions -chrX-318 and chrX-397- failed to be confirmed with qPCR (Appendix III). Region ChrX-318 is located in the gene body of the *SMPX* gene and chrX-397 is located on the gene body of *ARX* gene (Table 3.24).

Table 3.22 Primer sequences of DMRs on chromosome X classified as "*Not DMRs*".

DMRs	START	STOP	FORWARD	REVERSE	Δ1Ct
chrX_525A	44794193	44794339	AAGCAAGGATTTTGTAAGGGTG	CCCAGAAAAGGCCAAGTAAG	3*
chrX_364A	24751282	24751380	AAACATGCAGAGAAGGGAGTATG	CCAGGCTATTCTCATCACCATT	2.1
chrX_448	31272044	31272189	GCATCATTCAGGTGAGCATC	GCCCTCTTAACCACAAGTGTT	2
chrX_185	13350643	13350762	AGGCCACATTTCTGGTATCTG	GCACATGATAAACAGTTCACGT	2
chrX_1307	154313787	1.54E+08	GAAAGACCCAACCAGAGGAA	AGTTGGAGGTCGCTTATCTGTT	1.9
chrX_441	31233809	31233900	TGATAGAAGGCTTACGTGCTG	GCTTTCCACACGGAGTAATG	1.9
chrX_136	10472045	10472132	GCCATTCACGATGTTGTTGA	GCCCTTGCTCCCTCTAAATTA	1.8
chrX_139	10474915	10475003	GCACTCATTTGCAACTGACT	CGCTAGATGTCACGTGTTTT	1.8
chrX_118	9526917	9527032	GATTTACATAGCTGGACTTTGGTG	TGGACTACATACAACGCTATCCT	1.7
chrX_135	10470338	10470436	GCCATCCATAGACTTTCAGAA	TGTTTCCGTCTGTGTACGAA	1.7
chrX_507	41548310	41548438	CTCCTTTTACTTCAGGCGAAC	GCCTTATCACCATGAATCACG	1.7
chrX_976	113819900	1.14E+08	ATTGCGAAAAAGTCTCCAGC	TTATGTGACATCAGGAACGGc	1.7
chrX_250A	17519995	17520097	CCTTGTGAGAAAATGCAGCT	AAACGTCAGAAGTAACCCCA	1.6
chrX_322	21869068	21869215	GGATCCCTATTAGACCAGCTTt	CCTAGACCTTCCGCTGAATT	1.6
chrX_394	25018348	25018436	TAACCGGTAATCACTCCCAT	GTCTTGTTATCCCTCCCTGG	1.6
chrX_91	7185232	7185317	TGGAAGAAGATTAGGGCATTGT	GAGATGCATTCTCCTTCCTTGA	1.5
chrX_92	7185232	7185318	AGATTAGGGCATTGTATGAGTGtc	AGAGATGCATTCTCCTTCCTTG	1.5
chrX_181	13339268	13339407	TTGCAAGCTCCTATCAAGAGA	CGTCACTTACCTCGATGTCTGA	1.5
chrX_298	19902536	19902684	TGTCAGTAACGCCAAAAGAAct	AATGGAAGAAAAGGCTCCCA	1.5
chrX_525B	44794025	44794156	TGGTGGTATGATATGAAGCAAG	TTAGCTCCATGCATTGGTTT	1.5
chrX_172	12938668	12938755	TCTTCCTGACATTGCCTGTT	GACATGAGGAATGGCTGAAA	1.5
chrX_306A	19984711	19984846	CCGTGTGAATGTCTTTGCTT	GCAACTGTTTTTGGTAGGTCA	1.4
chrX_307	19985852	19985945	CTCTTAGAAGCTGCCATGGAGT	TCTGCCTCCTTTCTTGTAGGG	1.4
chrX_364B	24751125	24751230	GAGCTTGGCCAGTTACCTAA	TAATCTCTGCTTCTGGGCTG	1.4
chrX_648	55755829	55755978	TGTCAGTAATCGTGGCATCTAG	ACAGAGGGCATTTTCACAATt	1.4
chrX_1183	134494436	1.34E+08	GTTTTGCTCGGAAGTCACAA	CTCCCCTGTATGAAGTCGTTG	1.4
chrX_245	17490933	17491021	TGCAGAACAGTTTTCATGGGT	TGCTGTTTGGAACTTGAATTC	1.4
chrX_219	**15471178**	**15471309**	**AGACCCAGTCATTCCTTGTTAAT**	**TCAAGAAAGGACACCAGTAAGAG**	1.4
chrX_66	6609667	6609771	TTCCCCAACACATAATTGCA	CGTAGCTTGTTTTTTGAAGACCC	1.3
chrX_166	11646774	11646912	ATAGCCATTTCCTCCAGAGC	TGCCAAACGTTTCTTATATCCA	1.3
chrX_250B	17520167	17520257	GTTCGATATGACAGGCCAAC	GACCCGAGGAGACAAATAGG	1.3
chrX_257	17687479	17687596	GCAGAAATCAGTTGCAGGAA	AGTCCCCAGAAACATCAGACc	1.3
chrX_303	19983270	19983399	TGGTTACTGGAAGAATTGCTCT	agAAAGACAGCATTGCCTGTAC	1.3
chrX_907	106331663	1.06E+08	TACCTTGTTTGGTTCCCCTG	CGCAGCCAATAGAATTTGTC	1.3
chrX_1185	134710965	1.35E+08	AATCTGGCATGGGATTTAAAAA,	TCATTAAAGACCAAAATTCCACTAT	1.3
chrX_1195	135307880	1.35E+08	CATGATCAGTAAGAGGACAGCC	ATGCTAGTTCCTGGTGGTTTTC	1.3
chrX_60	6515547	6515654	AGCTCAATTCTCCCTCATGG	TGTGTTAGACAGATACGGCAGAT	1.2
chrX_157	10727085	10727178	CCTTTATCAGGAGGTCCCAAAT	AGCTATACCTTCAGTGCGTAAT	1.2
chrX_689	67886228	67886335	ACAGTTAGGATTGGGGTTGC	GGTTCTGGTTGTCTCTTCCA	1.2
chrX_829	96470824	96470913	CGTGTACTGTTGAGAGTGAGGA	ATCAATTTCGCACTTGCATC	1.2
chrX_920b	108625578	1.09E+08	TTTCACAAGAGGAGGAGACG	CACCTTTCTATTGCATGGCC	1.2
chrX_65	6609608	6609692	ATGCAGCAGTAGCGTTTCTTAA	TTGGTATGCAATTATGTGTTGG	1.2
chrX_1284	149936611	1.5E+08	GATGCTCGTTCTTGACATCG	ATCATTTCTGCACCGGAGAT	1.1

DMRs	START	STOP	FORWARD	REVERSE	Δ1Ct
chrX_1291	153293400	1.53E+08	GAGACTCTACAAACAGAACGGG	GCACTTCCTTGCCAGTCTTT	1.1
chrX_171	12937103	12937191	CAAAACACCTGGCAAATACTG	AGCATTGACGACTGAAGGAA	1
chrX_218	15431594	15431693	AGAATTGGTGACATAAGTTGGG	AAGATGGGGGTTAAATGCCT	1
chrX_269	18807252	18807395	TGATACCACCAACGGAAACA	CTACCTGTTCCCATTCATGC	1
chrX_306B	19984861	19984966	CTCGGCTTCTTCCACTCAAA	GAGTATGCTTGTTGGTTGGG	1
chrX_836	96522751	96522891	CCTGGGGAATAAATTAGGCG	TTCATGAGGGTCAAAGTCAGTC	1
chrX_1053	122821577	1.23E+08	CACTTAGAAAACCAAATCCTCCA	TGAGAAGCGGAAGAAAATGTG	1
chrX_1290	150577738	1.51E+08	TAACCGTTTTCTGGATGGATTT	TCAGGACACAGCAAAGTAGCAC	1
chrX_80	6978244	6978359	CTCAGTTCTCCTTCCTGCAA	CCTGGTACTGCAACACATCTt	0.9
chrX_170	12842377	12842496	GACGGAGTGATGTGTTGCAGTA	CACCTTCAAAATGGAAAACAGC	0.9
chrX_323	21887178	21887297	GGAAAATTACCTTTGTGAGCAA	GCCTTTCCTATATCCTTAACGC	0.9
chrX_258	17739869	17739960	CATCAAAATACAAACCCAGAGG	TTACATAACAAGTGGCCCGC	0.9
chrX_109	8957678	8957777	CCAAAGCTTCAGGTGAAACA	CTGTGATCAATGCTCTTCCG	0.8
chrX_289	19735876	19735961	TGCCTCTAACTGCTTTCCTC	TTGTTCATATATAGCCCTGCTC	0.8
chrX_546	46784182	46784278	CGTTTGTCTTGCTTTCTTCC	TGCTCTAATTTTCAGTGACGC,	0.8
chrX_920a	108625350	1.09E+08	GAGCTCTGTTCTTCCTCGAA	TCTCTCCTTCTTTTGCAGCT	0.8
chrX_400	27294016	27294132	TGCCTTCATTGTCTTCAAGC	CTCACAGGTCTGCAGTCATG,	0.8
chrX_498	**40534109**	**40534223**	**CCTACTGAACATCGCTAGACATT**	**GAAATACCATCCGCAGGATCT**	0.8
chrX_698	69562402	69562523	TTTGATAGCACGACTGTTGC	AAGCAGCTTATCTTTCCTTGAC	0.7
chrX_787	85393680	85393809	ACTGGAAATCGTCGATATGTGA	CCTAAAACTGTAAGTTTCTCGCAA	0.7
chrX_1206	135739046	1.36E+08	GATCAATAGGTCCCAAGACGT	TGGTGGCAGAACGTATATGG	0.7
chrX_58	6459845	6459932	ACCAGTCCACTCAAGGTTCTAAG	CCCGAGAAGGTATTTCATGAC	0.7
chrX_84	**7019434**	**7019534**	**GGCTGGTTGTTCACCTTCTA**	**ATCCCATCAACGTAGTGTCTTC**	0.7
chrX_69	6620617	6620705	AACGAGATAAATGTAATGCCAGA	GTAGAGGCACAGAAGCAATGGT	0.6
chrX_203	13790960	13791096	AGAGCTGTAAATACGTCGGCc	CAAAATGCAGGCTTACCAGG	0.6
chrX_343	**24186872**	**24186963**	**CATGATTCTCGTAACTGGGAAG**	**GTCTTACTATCAAACGCCCTCAA**	0.6
chrX_93	**7201345**	**7201439**	**ACTGGGAAGACAGACTCAGAG**	**TCTCTGAAGTACTCGTTGGTATTG**	0.6
chrX_509	41828327	41828421	CATTTTGGGCTTCGTACCTT	GGCAGCTACCGATGTAAGAA	0.5
chrX_1041	120301165	1.2E+08	GGTTGTTTTGGGGCATTATAAT	ACCGAGAAATTCTGTCGTACCT	0.5
chrX_344	**24204080**	**24204208**	**CCCCTAAAACACCTATATTGATTTC**	**GTTGACTACCGATAATGGTATCTTC**	0.5
chrX_1015	**118610036**	**1.19E+08**	**TCCGTATGCAGGTTGAGATTTAT**	**GATAGGAAAGCTCCTCCAAACA**	0.5
chrX_1073	**129288394**	**1.29E+08**	**GGGTGGAACTGATGGGAAA**	**GATTGCTTTCTGTAATGCCTCTC**	0.5
chrX_1072	**129270439**	**1.29E+08**	**CCTACTCTGGCTCTCATCATTC**	**CTCCTTCATTTGACACCAGTAGA**	0.4
chrX_71	6685579	6685683	CAGTTCGCCCTTCATAGCAG	CCAGACCTGTATTGTGGAAGC	0.1
chrX_107	8854443	8854587	AATCTGTGCCCATTAGACCTAG	GTTAATGTCAAGTCTGGCCG	0
chrX_744	73326692	73326829	GAATGGCTTGTGTTGGTCATAT	GTCCACCAACATGGGGTC	0
chrX_1070	128857665	1.29E+08	GTCAAAACCCGGAACAAACA	TCAGGTGGTAATGGATGATTCA	0
chrX_53	6198381	6198477	ATTCACCTTGAGCGGCTTTA	CTCCCAGAAACATGGACAGT	-0.8
chrX_51	4734882	4734985	CGGGCCGTAGAAAAAAAA	GTAAGTGGCTTGTTCTCTGAAAA	-1

*Indicates highly variable Ct values among the samples tested.

Regions in bold indicate regions that coincide with genes that escape X inactivation.

Table 3.23 Primer sequences of DMRs on chromosome X classified as "*Good DMRs*".

DMRS	START	STOP	FORWARD	REVERSE	Δ_1Ct	Δ_2Ct
chrX_318	21766050-	21766134	AATTCTACAGGGTTGCATTTTG	ATCTTGCTCCGTTCAGAGAA	2.7	2.5
chrX_397	25028044	25028141	AGGAGAAGTAGATCGGTGGATT	AGGATCAGAACAAGGGCTTTa	3.4	2.3

Table3.24 Details on the best selected DMRs regions on chromosome X

DMR	START	LOCATION	GENE	OMIM
chrX-318	21766050-21766227	Intragenic	SMPX	300695
chrX-397	25027911-25028147	Intragenic	ARX	300382

3.3.3 Identification of Abnormality Unique DMRs

In order to identify methylation differences associated with chromosomal abnormalities, thus further expanding the DMR panel to specifically detect abnormal fetuses, different criteria were established for the DMRs selection. Towards this goal, hypermethylation of the abnormal CVS (Trisomy 13, 18, 21, monosomy X) was compared to hypomethylated WBF and 1st trimester CVS. As Table 3.25 shows, several "*Abnormality Unique DMRs'* were identified for all chromosomes under investigation. Selection of a subset of these DMRs and screening on individual samples is necessary in order to confirm their differential methylation status.

Table 3.25 DMR identification of *"Abnormality Unique*" DMRs

Chromosome	Total DMRs	Promoters	Gene Bodies
chr13	68	1	37
chr18	246	5	124
chr21	8474	48	3450
chrX	427	8	202

3.4 MeDIP-seq of CVS and WBF

CVS and WBF pooled samples used for MeDIP-Chip were also subjected to whole genome MeDIP-seq for confirmation of the validated DMRs. Samples were sequenced on a separate lane on the Illumina Hiseq 2500 sequencer (high output run). The raw sequencing reads were assessed for quality using the open source software FASTQC [148]. The reads passing the quality filters were aligned using the Burrows-Wheeler aligner [149]. A total of 235724800 paired-end reads were aligned to the hg19 reference genome (build 37) for the CVS sample and 240039592 reads for the WBF sample. The resulting alignment file (.sam) was compressed to its binary form (.bam file) using samtools [150] for subsequent analysis. Duplicate reads were removed using the Picard MarkDuplicate tool. Indel realignment and quality score recalibration was performed using the GATK toolkit [151]. The resulting alignment files, (.bam files) were subsequently used as input to the R package MEDIPS [152]. MEDIPS calculates differential coverage between groups of samples derived from MeDIP experiments. The first step of MEDIPS is a saturation analysis of the CVS and WBF alignment files to verify that they were adequate and reproducible for comparison and the effectiveness of the MeDIP enrichment is assessed by calculating the overall CpG enrichment. This CpG density is used as a factor for normalization of the alignment counts and to calculate the relative methylation scores of the two files. Subsequently the regions of differential coverage are detected by estimating the variability of the relative methylation score counts in the data using negative binomial distribution. A total of 13267 hypermethylated DMRs and 5733 hypomethylated DMRs were identified in CVS as compared to WBF. Regions on chromosomes X and Y were not included in this analysis as the CVS samples used were consisting of both male and female samples. Furthermore, no correlation was observed between the number of DMRs (hypermethylated on CVS) identified and the GC content, chromosome size, number of genes and number of CGIs on each autosome (Figures 3.9-3.12). We initially compared our MeDIP-seq results with DMRs identified in previous publications as well as the DMRs confirmed using the previous array data. As shown Appendix IV, 30 of 33 DMRs showed consistent methylation levels. DMRs VA17 located on chromosome 18 was not confirmed with the NGS approach. In addition, there were no reads aligned for two DMRs on chromosome 21 (M18 and M20).

Table 3.26 Total number of DMRs identified using MeDIP-seq

Chromosome	Total		Gene Bodies		CGIs		Promoters	
	Hyper	Hypo	Hyper	Hypo	Hyper	Hypo	Hyper	Hypo
chr1	1344	497	940	266	142	18	98	43
chr2	1013	414	633	174	164	12	50	23
chr3	628	247	489	120	75	17	43	30
chr4	524	181	322	80	141	10	18	25
chr5	821	280	524	140	143	14	74	24
chr6	891	309	574	145	125	19	73	12
chr7	664	227	491	120	62	18	26	19
chr8	658	219	468	75	111	2	20	5
chr9	328	108	259	70	11	8	12	4
chr10	653	459	459	213	54	22	34	34
chr11	719	376	553	202	51	14	26	35
chr12	804	266	545	154	137	4	52	32
chr13	352	192	231	82	64	3	18	4
chr14	661	209	511	97	48	10	26	7
chr15	450	229	318	125	38	6	24	13
chr16	539	269	443	135	35	1	10	15
chr17	578	330	455	147	95	2	49	31
chr18	230	146	168	89	31	2	19	5
chr19	396	209	351	137	56	9	15	23
chr20	612	261	431	117	55	7	43	10
chr21	198	138	122	68	16	5	13	0
chr22	204	167	173	90	3	2	8	7
GENOMEWIDE	13267	5733	9460	2846	1657	205	751	401

Unlike MeDIP-Chip, MeDIP-seq provides information for methylated sites across genome. Hyper: Region that showed hypermethylation in the CVS. Hypo: Region that showed hypomethylation in CVS.

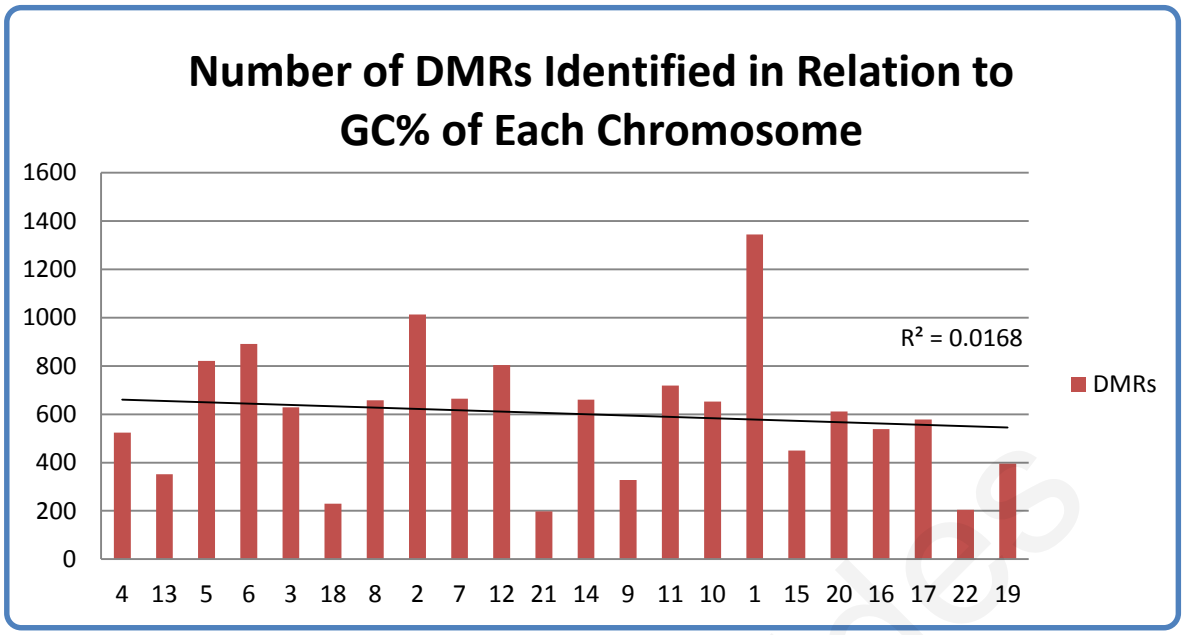


Figure 3.9 Correlation between the number of DMRs found in each chromosome and the GC content of each chromosome shows no association between the two values.

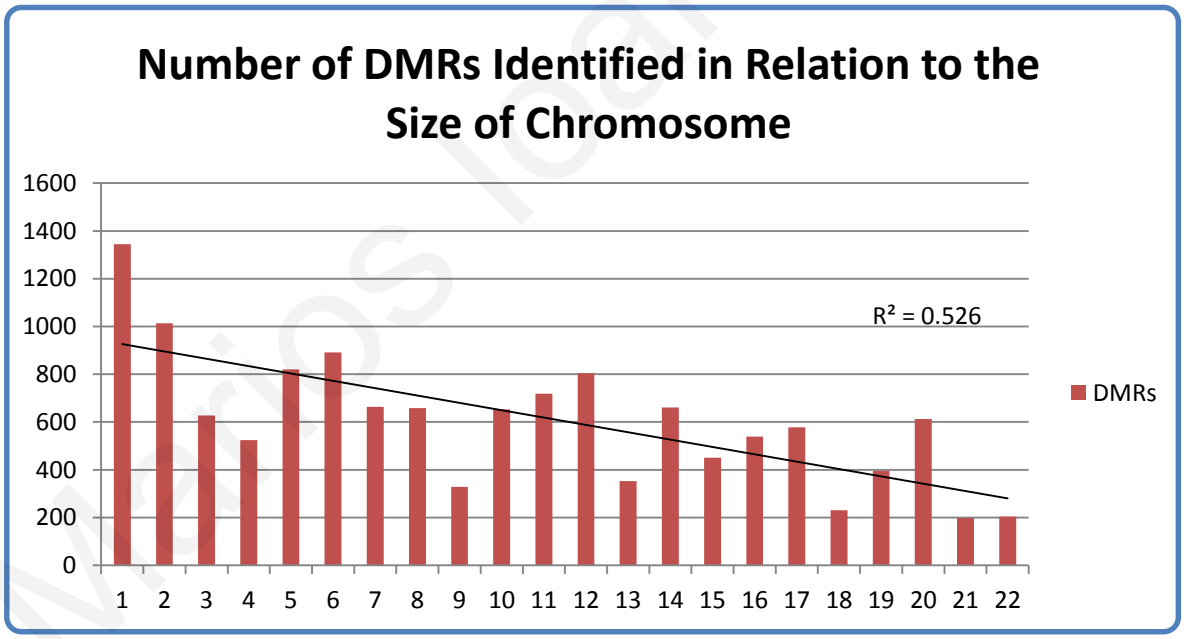


Figure 3.10 Correlation between the number of DMRs found in each chromosome and the size of each chromosome shows no association between the two values.

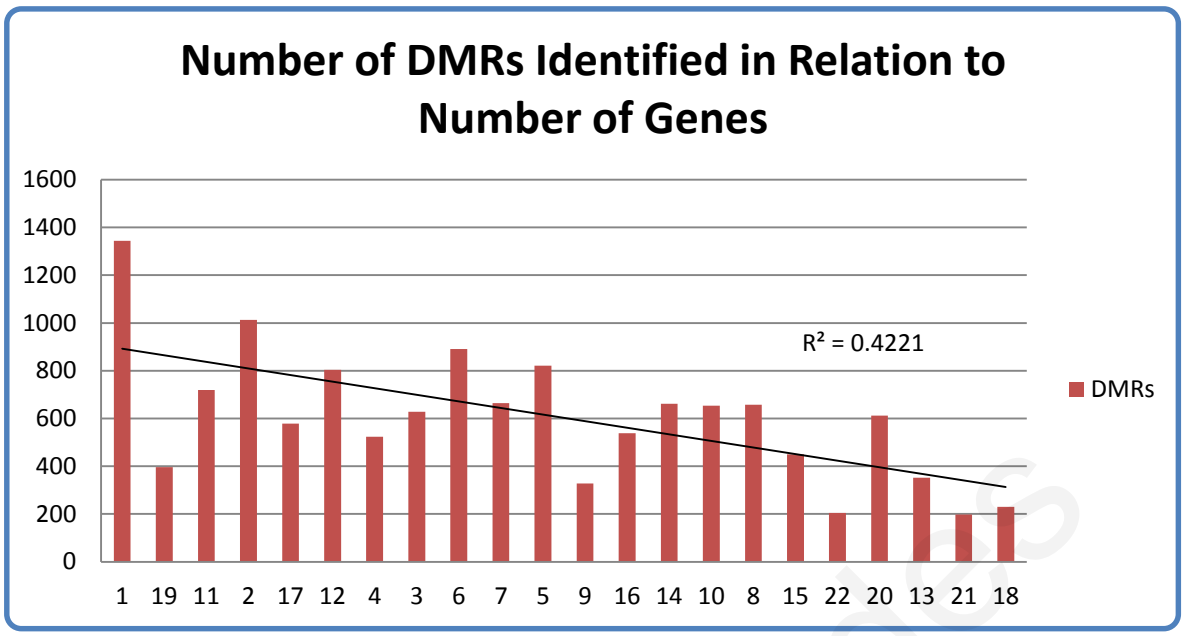


Figure 3.11 Correlation between the number of DMRs found in each chromosome and the number of genes in each chromosome shows no association between the two values.

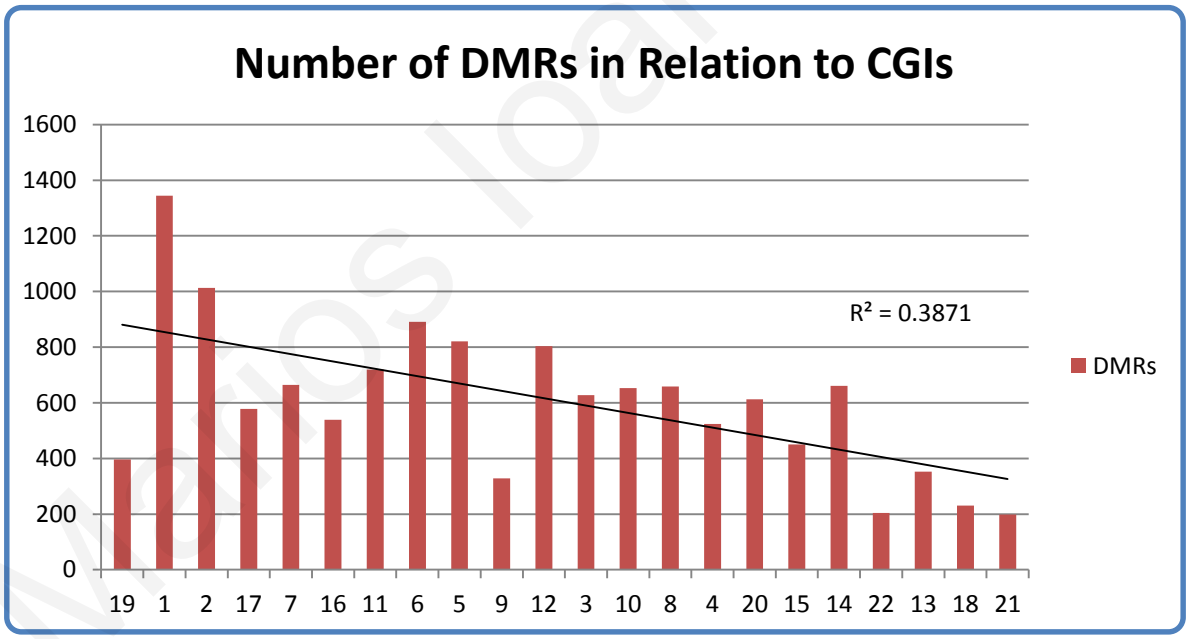


Figure 3.12 Correlation between the number of DMRs found in each chromosome and number of CGIs shows no association between the two values.

We subsequently performed overlapping analysis of the confirmed regions classified as "*Good DMRs*" and *"Good DMRs-T21 specific*" confirmed with ultra-high resolution MeDIP-Chip approach and the MeDIP-seq for the chromosomes 13, 18, 21. Most regions had consistent results between the two approaches. Regions that were not confirmed was due to absence of reads for the specific DMRs (Appendix V). Additionally, only a limited number of DMRs showed overlapping between regions confirmed by MeDIP-Chip and regions detected by the MEDIPS software. Specifically, six, seven and seven DMRs on chromosomes 13, 18 and 21 respectively were identified by both approaches. (Figure 3.13 –Appendix VI). It is worth noting that many of the best performing DMRs were also detected by MEDIPS software.

Figure 3.13 Overlapping analysis between DMRs identified by MeDIP-Chip and MeDIP-seq. Regions identified by MEDIPS (Green track) software were overlapped with regions confirmed by MeDIP-Chip and qPCR (Blue track). In this example CVS (top panel) shows higher read depth-hypermethylation- for the highlighted region (DMR TS1-chromosome 21) as compared to the WBF (lower panel).

4 DISCUSSION

The main objective of this study was the identification and characterization of differentially methylated regions between fetal DNA represented by CVS samples, and maternal DNA represented by DNA isolated from female non-pregnant peripheral blood**.** Several studies have focused their research on the discovery of fetal specific biomarkers and fetal DNA enrichment methodologies that could potentially distinguish fetal DNA from the high maternal background and ultimately quantify the small dosage differences in the trisomic pregnancies. Even though many have succeeded, several limitations prevented their implementation in NIPD. With the utilization of the MeDIP strategy we were able to achieve direct enrichment of whole methylome, unrestricted from form CGI and restriction sites.

4.1 DMR Identification and Inter- Individual Methylation Variability Using Existing Array Data

Previously using the MeDIP assay followed by high resolution array CGH, a large number of DMRs that showed hypermethylation in placenta and hypomethylation in the maternal peripheral blood [100] were identified. A subset of these DMRs was further validated and implemented in the quantification and consequently correct classification of Down syndrome pregnancies with high sensitivity and specificity [126, 127]. Thus, this study was based on an already validated and well established methodology. Using the same MeDIP-Chip data as previously described, we have confirmed the differential methylation status in three, eight and 12 DMRs on chromosomes 13, 18 and 21 respectively utilizing the MeDIP-qPCR with the incorporation of the amplification of MeDIP products (LM PCR).

We then selected 15 DMRs for further validation in 50 WBF and 50 CVS samples. Based on our results, 11 of the 15 DMRs were strongly and consistently hypermethylated in CVS samples. The ability of these DMRs to distinguish between CVS and WBF was equivalent to that of the seven previously validated DMRs used as performance standards. In fact, the tissue discriminating performance of the DMRs tested, shows close similarities with the previously validated DMRs as

illustrated by our heat map distribution and the unsupervised clustering patterns obtained (Figure 3.3).

DMR enrichment values showed variability among the different samples. This is likely caused by a combination of both inter-experimental technical variability and inter-individual methylation variability. The presence of variability in MeDIP based assays has previously been described by Butcher et al. [153], while several other studies have emphasized the issue of inter-individual methylation variability. Rakyan et al. examined the DNA methylation profile of the human major histocompatibility complex on six different tissues and demonstrated that methylation was heterogeneous for several loci, while inter-individual methylation variability was evident in all regions tested, with 118 loci having more than 50% median methylation difference in at least one tissue [80]. A quantitative study by Bock et al. of the methylation pattern utilizing a large dataset from the Human Epigenome Project (HEP) documented that regions with low CpG density, as the DMRs under investigation, showed higher inter-individual variability as compared to regions with high CpG density, such as CGIs [154] probably due to the fact that the latter maybe influenced by the methylation status of nearby CpG [93]. In addition, the high inter-individual variability has been attributed to a variety of factors including environmental conditions, diet, age and psychosocial factors [90, 91, 155]. Another possible explanation of the variability of methylation between the WBF and CVS samples can be attributed to the fact that the former is a reflection of the methylation profile of total peripheral leucocytes and not of pure cell population, whereas the latter is a rich source of differentiating cells that their methylation landscape may be more flexible to changes [156].

Others have also shown that methylation variability can coincide with tissue specific DMRs without obscuring the tissue discriminating properties of those DMRs [92]. It is therefore of no surprise that despite the DNA methylation variability in our study, the newly validated set of 11 DMRs clearly distinguish between CVS and WBF tissues.

The MeDIP-qPCR approach was also performed on the same subset of DMRs without amplification following the MeDIP procedure (Non-LM PCR). It is apparent that even though the differential methylation status between WBF and CVS is statistically significant ($p<0.01$), the methylation enrichments show overlap between

the two tissues (Figure 3.4). Thus, it can be suggested that LM PCR adds an amplification bias to the methylation enrichment which enables us to better discriminate between WBF and CVS tissues. The positive bias can also be concluded by the studies of Papageorgiou et al. and Tsaliki et al. [126, 127] where the MeDIP LM PCR approach in combination with qPCR allowed them to successfully detect the trisomy 21 cases with high sensitivity and specificity. Correlation analysis between LM PCR and Non-LM PCR approaches for the enrichment values of the DMRs tested revealed strong correlation between the enrichment values obtained from WBF, whereas no correlation was observed in the enrichment values obtained CVS. It has been previously reported that regions with low CpG content- as the ones described herein- demonstrate technical variability after MeDIP. In the case of CVS samples, this effect may be amplified during the LM PCR and is essentially reflected by the qPCR amplification. On the other hand, due to hypomethylation of the WBF, the amount of DNA recovered after MeDIP is small enough so that PCR amplification has no effect. Additionally, the qPCR in the LM PCR approach is performed on pre-amplified double stranded DNA, whereas in the case of the Non-LM PCR approach the DNA template is single stranded, the stability of which is unknown.

Overall, we were able to show that the selected regions had distinct methylation patterns between fetal and maternal tissues, despite inter-individual and inter-experimental variability. Furthermore, amplification of the MeDIP products prior to qPCR, further increases the methylation enrichment differences between the tissues making the distinction even more robust.

4.2 DMR Identification Using Custom 1M Ultra-High Resolution aCGH

4.2.1 Chromosomes 13, 18, 21

For this stage of the study we utilized a custom ultra-high resolution aCGH design specific for chromosomes 13, 18, 21 and X with a mean probe spacing of 48bp, 38bp, 18bp and 56bp respectively. The design included one million probes (1M) covering indiscriminately all regions of the specific chromosomes, excluding the

repetitive portion of the chromosomes. Prior to hybridization we performed pooling of five samples per group at different stages of the experimental procedure (Figure 2.3). This way we aimed to suppress any individual specific methylation variability between WBF and CVS samples and any potential technical variability in order to identify a set of more robust DMRs that are common to all samples. As a result, we identified a total of 371, 682, 2825 and 537 DMRs on chromosomes 13, 18, 21 and X respectively, that showed hypermethylation in 1st trimester CVS and 1st trimester abnormal CVS (i.e trisomy 13, 18, 21, or monosomy X) and hypomethylation in WBF samples. Initially, we performed aCGH on chromosome 21 in order to assess the efficiency of DMR discovery, by using the previously identified DMRs as control standards. Then, the newly selected regions were tested on multiple samples to confirm that the methylation pattern between the two tissues is a true biological event rather than hybridization artifact. Using the initial set of criteria, out of 118 regions tested, 72 failed to be confirmed as DMRs. Further stringent criteria were then applied in order to minimize the false positive DMR discovery for the rest of the chromosomes (chromosomes 13, 18 and X) based on the performance of *"Good DMRs"* on MeDIP-Chip for chromosome 21. We confirmed tissue specific differential methylation in all regions tested on chromosomes 13 and 18. According to the classification requirement though, only 53 of 141 selected regions were classified as *"Good DMRs"*. None of the selected DMRs was tested against individual trisomy 13 and 18 samples as these trisomies are very rare and difficult to be obtained. Further testing on abnormal samples hence, is necessary to confirm trisomy specificity.

GC content association with the number of DMRs identified on chromosomes 13, 18 and 21 revealed that chromosome 21 has much higher number of DMRs identified than the other chromosomes. Even though having information from all chromosome sets would have been more conclusive, we used only chromosome 21 to further investigate potential GC bias. Since less stringent/biased criteria were applied for DMR selection, we overlaid the DMRs that failed to be confirmed ("*No DMRS", Bad DMRs*") and "*Good DMRs"* with the GC% for the specific chromosome. As shown in Figure 3.8, most of the DMRs which failed to be confirmed by qPCR are located in the chromosome region with high GC%, whereas most confirmed DMRs are located in the chromosome region having a GC% close to the average human genome GC% (~44%). MeDIP in conjunction with the increased GC content

of the regions resulted in a potentially higher rate of non-specific hybridization to the array. Consequently this caused tighter binding of the non-specific GC-rich sequences to the probes hence increasing the array background. It is also suggested that high resolution arrays reduce the signal to noise ratio resulting in variable $\log_2$ ratios [29]. Furthermore, it has been reported that if the average probe spacing is smaller than the hybridized DNA fragments, neighboring probes may be affected, resulting in a potential background noise [91]. Any combination of the above reasons can possibly explain the fact that many of the chromosome 21 DMRs failed to be confirmed. This may as well explain the presence of high background for some of the confirmed regions especially on chromosome 18 (Appendix II). One way to circumvent the hybridization variability was to include multiple probes per location; this would improve the specificity of the assay in one hand, but in the other it would have lowered the resolution of the array. Using more stringent criteria in the DMR selection of chromosomes 13 and 18 we managed to overcome the imposed limitations.

MeDIP has extensively been used for comparison of DNA methylation profiles and tissue specific differentially methylated regions [81, 120, 157, 158]. In agreement with our observations, these studies demonstrate the robust nature of MeDIP to recover whole the methylome in a sequence- independent manner, irrespective from restriction sites and CpG density. Our study also reaffirms the reproducibility of the MeDIP-chip approach previously used to interrogate chromosome specific differentially methylated regions [126]. As shown in Appendix II all seven previously validated regions on chromosome 21 and the newly identified regions from the previous aCGH set were identified using the ultra-high resolution aCGH. Interestingly, regions M18, M20 and M25 that failed the validation due to overlapping enrichment values between the two tissues, showed weak differences in the 1M array data that would have prevented their selection for further investigation. Region Fd1, with a high difference in methylation status between WBF and CVS shown in the aCGH, was one of the four regions that showed the smallest difference in enrichment levels between the two tissues. Fd1 is a small region flanked by repeat sequences. qPCR may have introduced some noise since, in addition to the small size of the region, one of the primer sequences is partially located in the repeats. Moreover, confirmed regions on chromosomes 13 and 18 using the previous array data showed the expected difference between WBF and CVS samples.

A large number of identified DMRs (28-38%) was located in intergenic regions. This is in agreement with Rakyan et al. who found a range of tissue specific DMRs in different types of genomic regions including intergenic regions [81]. Methylation of these regions was discussed in a study by Eckhardt et al. where it was suggested that regions as such could be associated with enhancer or silencer activity [68]. Moreover, it is suggested that promoter methylation and gene body methylation is positively and negatively correlated with gene transcription respectively [70, 135]. In his review Jones discusses this paradox and proposes that the transcription initiation is sensitive to methylation while elongation is not [159]. Thus methylation within genes may have a role in splicing regulation [159] by promoting suppression of unwanted splicing within actively transcribed genes [81]. Our results are in agreement with the above findings as the majority of the DMRs were located on gene bodies while only 1.1-2.7% of DMRs in promoter region showed hypermethylation. This also supports the notion that at a global level placental tissues are associated with high transcriptional activity as it is demonstrated by the correlation of promoters and gene bodies with gene expression [160] .

Correlation of the best DMRs from each of the chromosomes 13, 18 and 21 identified by the 1M aCGH within gene bodies, promoter regions and CGIs revealed that most of these DMRs are located within gene bodies and only a few in promoter regions and CGIs. Most of the genes are assigned an OMIM database entry and some of which are shown to be associated with fetal development and diseases. As mentioned previously, placental DNA is expected to be hypermethylated in gene bodies and hypomethylated in promoter regions in order to be transcriptionally active. We speculate that this may be the reason why the specific DMRs are hypermethylated at their genic regions suggesting transcriptional activity. Nevertheless, this study focuses on methylation differences at the DNA level thus we cannot provide definitive explanation of this phenomenon. Functional and gene expression analyses of candidate genes and associated diseases should be performed in order to shed light on the role of these genes in fetal aneuploidies. Interestingly, four regions that have been classified as *"Good DMRs-T21 specific*" have shown to be located on genes that are associated with Down syndrome. Specifically, DMR TS1 was found to be located in *DSCAM* (Down Syndrome Cell Adhesion Molecule), a neural cell adhesion molecule that has been demonstrated to be overexpressed in Down syndrome patients [161]. This study suggests that

the congenital gut diseases and mental retardation of Down syndrome may be partly explained by the altered adhesive properties of neural cells and synapse growth and migration due to overexpression of adhesion molecules such as DSCAM. Supporting the implication of *DSCAM* as a candidate gene for Down syndrome, other groups suggest that the cause of congenital heart disease found in patients with Down syndrome may be due to the presence of three copies of DSCAM [162]. Furthermore, DMRs TS2-TS4 were found to be located on *SIM2* gene. Chen et. al suggested that *SIM2* can contribute to the pathophysiology of Down syndrome [163] since it is the only gene that was identified in the Down syndrome's critical region (DSCR) involved in brain and nervous system development.

4.2.2 Chromosome X

It has been shown that methylation is directly implicated in the random X inactivation. This takes place in females in order to compensate for the dosage of gene imbalance due to the presence of two X chromosomes as compared to a single X chromosome present in males [72]. X inactivation starts early in embryonic development [164] with the transcription of the *XIST* gene that acts in cis-fashion to transiently inactivate one of two X chromosomes [165]. X inactivation though is complete after promoter DNA methylation [72]. In our study, due to the limitation of the procedure to distinguish between the two X chromosomes, genes that are altered by X inactivation could not be investigated. This presented a challenge in the DMR identification between the WBF and CVS samples since it is suggested that local promoter regions of X-linked genes show hypermethylation [72]. Nevertheless, about 15% of chromosome X genes are thought to escape X inactivation [166] and as such they are being expressed by both the active X (Xa) and inactive X (Xi), in a manner similar to autosomal genes [167]. Thus, for DMR selection, in addition to previously reported gene subsets that escape X inactivation [12] we also included other regions that showed differential methylation between WBF and CVS (normal and abnormal) assuming that they would behave in the same way as in the case of autosomes. None of the escapee genes nor the additional regions investigated confirmed the differential methylation pattern after the qPCR screening. Only DMRs chrX-318 and chrX-397, were confirmed as DMRs. Chr-318 is located in the gene body of the *SMPX* gene. It is proposed that it functions in the development of sensory hair cells [168] as gene mutations have been shown to be

associated with X-linked deafness [169]. The next DMR, chrX-397, is located on the gene body of *ARX* gene, shown to be predominantly expressed in fetal and adult brain. It has been associated with mental retardation as well as epileptic seizures in both males and females [170- 172].

In conclusion, chromosome X had the lowest rate of DMR identification, mainly because of the increase in the methylation levels of WBF due to random X inactivation. Furthermore, the technical difficulty of the aCGH to detect low levels of methylation due to heterogeneity in X inactivation [12] have rendered the DMRs identification challenging and not as successful as expected. Nevertheless we have identified two regions that appear to be very promising and can further be utilized in the non-invasive detection of X-linked aneuploidies and sex determination.

4.2.3 Abnormality Unique DMRs

Due to the integral relationship between DNA methylation and gene expression, biomarker discovery has been associated with identification of unique targets for treatment and diagnosis, especially in the field of cancer. Differences in the DNA methylation patterns between healthy and Acute Myeloid Leukemia (AML) [173] cell lines revealed thousands of AML specific DMRs [174]. In prostate cancer several methylation markers are being used for prognosis [175] while research groups identified new candidate markers by comparing the differential methylation between prostate cancer and non-malignant specimens [176]. We implemented a similar approach in order to identify abnormality unique DMRs, that is, regions that showed hypomethylation in both WBF and normal CVS and hypermethylation only in the abnormal CVS. As the main goal of this project was to characterize DMRs identified between hypomethylated WBF and hypermethylated CVS, abnormality unique DMR characterization was not pursuit further. Nevertheless, we hereby identified more than 9000 DMRs present in chromosomes 13, 18, 21 and X that show hypermethylation only in the abnormal tissue. These DMRs can potentially disrupt the methylation levels in trisomy cases and as such can serve as diagnostic markers that can directly identify a potential trisomy without the need of complex statistical analyses.

4.3 MeDIP-seq

The last stage of our study was the utilization of the MeDIP-seq approach for confirmation of DMRs validated in the previous stages. Due to the nature of this method we were able to identify DMRs on all autosomal chromosomes and correlate them based on their location. Our study agrees with previous reports which identified high number of hypermethylated DMRs in placenta as compared to hypomethylated ones using the MeDIP-seq analysis [120]. Furthermore, the presence of more DMRs in the gene bodies, as in the case of the aCGH, suggests the implication of the placenta in high transcriptional activity [160]. NGS data shows no preferential increase of DMRs with GC content. In support to the unbiased nature of the NGS method, no correlation was found between the number of DMRs identified on each one of the autosomes and the size, number of genes or CGIs. Furthermore, the relative objectivity of the MeDIP methodology reaffirms the MeDIP-Chip results as well as previous studies that demonstrated that the MeDIP procedure allowed enrichment of regions with a wide range of CpG density [81, 100, 153].

All control regions (Appendix IV) have been confirmed using the MeDIP-seq approach. Inspection of the DMRs which were confirmed in the first stage of the study showed highly consistent results except from one unconfirmed region (VA 17). Due to high maternal methylation in this region, higher read depth would be necessary in order to determine the methylation differences between the two tissues. Indication of the high accuracy of this method -similarly to the 1M aCGH results- is the fact that DMRs that failed to be confirmed in the first stage of the study because of low enrichment difference values between WBF and CVS also failed to be confirmed utilizing the MeDIP-seq.

Comparison of the classified as *"Good DMRs*" and *"Good DMRs-T21 specific*" using MeDIP-seq and MeDIP-Chip, showed both methods to be highly consistent. Specifically, after inspection of the regions we confirmed the differential methylation between the two tissues in 74 of 99 regions (75%), whereas absence of reads was observed for only 25 DMRs. Under representation of reads has been previously reported and attributed to the low CpG density of the regions [177].

Although substantial numbers of DMRs were identified using the MEDIPS software, comparison analysis between these DMRs and the confirmed MeDIP-Chip DMRs indicated a low percentage of overlap. Specifically, of the 99 regions confirmed in MeDIP-Chip only 20 showed overlap between the two platforms. The two methodologies have the same basic principle (MeDIP) and beyond doubt they showed highly consistent results, thus potential heterogeneity between the two approaches is not considered a possibility. On the other hand the efficiency of the MEDIPS software to correctly classify a region as DMR has been deemed low despite the software's ability to combine the MeDIP results and CpG coverage. It has been reported that MEDIPS uses algorithms that average out the reads aligned to the methylated areas resulting in an "*in-silico*" small dynamic range and as such it underestimates the methylation differences between two samples [178].

Overall, MeDIP-seq and MeDIP-Chip provide accurate, robust, and consistent results with regards to the discovery of differentially methylated regions. Since they are both affinity based procedures they can comparably detect methylated sites independent from CGIs or promoters regions on a wide range of CpG sites. We have shown that MeDIP-Chip, even though affected by the GC content due to its hybridization dependent nature, can increase its specificity with the use of more stringent selection criteria. MeDIP-seq analysis provides a more bias-free, comprehensive identification of DMRs but its high cost and bioinformatics implication may still be prohibitive and not readily available for most laboratories. Furthermore, due to the ultra-high resolution of the array chip, we managed to reach the same resolution of the MeDIP-seq approach. Fewer probes could have been incorporated on the array designs in order to cover the whole genome but since the scope of this study was the identification of DMRs on the chromosomes involved in the most common aneuploidies we did not compromise the sensitivity of the assay. On the other hand the high resolution may have caused a low signal to noise ratio, but correct establishment of criteria minimized this issue. The major advantage of MeDIP-seq is that it provides whole methylome information, whereas a great number of array chips would be needed to achieve this resolution for all the chromosomes. Despite this, both approaches are low throughput since not many samples can be run/analyzed simultaneously. Further validation of the DMRs with a higher throughput and inexpensive technology, was however, necessary, in order to confirm and characterize the results. Finally, none of the two approaches give a

base pair resolution of the methylated cytosine residue. Still, for our purposes further information beyond the achieved resolution would not provide any extra benefit.

4.4 Conclusions

NIPD has gained a lot of interest the last few years. Several groups have utilized different approaches for the identification of fetal specific biomarkers for the implementation in the NIPD. Using the MeDIP approach we have successfully identified and characterized fetal specific methylated regions on chromosomes 13, 18, 21 and X.

Initially, the dataset obtained from previous arrays was used for characterization of DMRs using MeDIP with subsequent amplification (LM PCR) in combination with qPCR. We confirmed the differential methylation status in three, eight and 12 DMRs on chromosomes 13, 18 and 21. Furthermore, we showed statistically significant differential enrichment DMRs for both LM and NON-LM PCR approaches in a subset of 15 DMRs. It was concluded that the LM PCR adds a positive bias to the data, resulting in a more robust distinction of the methylation enrichment between the two tissues. Moreover, we reaffirmed the presence of variability in the enrichment values among the different samples. This is likely caused by a combination of both inter-experimental technical variability and inter-individual methylation variability. Despite the DNA methylation variability though, the newly validated set of DMRs was clearly distinguished between CVS and WBF samples in a similar manner as the DMRs used as performance standards.

Next, we have implemented ultra-high resolution aCGH on pooled samples for expansion of the DMRs set on chromosomes 13, 18, 21 and X. We have identified hundreds of DMRs on each one of the chromosomes; most of them located on gene bodies. It is suggested that gene body methylation is positively correlated with gene transcription; hence our results support previous findings that showed that placenta tissue is globally associated with high translational activity. In addition, we demonstrated the robust nature and reproducibility of MeDIP to recover methylated regions, independent from restriction sites and CpG density since most of the confirmed DMRs were not located on high CpG density regions, such as CGIs. Due to the hybridization nature of the aCGH we also demonstrated that that the specificity of the aCGH may be influenced either by the high probe density of the array or the high GC content. Despite this, for the purposes of DMR discovery, implementation of stringent criteria enabled us to overcome these limitations. In total

we have classified as "*Good DMRs*" 99 regions on chromosomes 13, 18 and 21 and two on chromosome X. The focus of this study was the differential methylation differences between CVS and WBF. Correlation of the best DMRs with gene and promoter regions indicated that most of the regions are associated with disease genes. Interestingly, four of our selected regions on chromosome 21 have been found to be correlated with genes that may play a role in the pathophysiology of the disease. Even though promising, further functional and gene analysis study should be performed in order to investigate the association of these genes with the specific diseases.

Finally, we have utilized the MeDIP-seq approach in order to confirm our MeDIP-Chip results. Due to the holistic nature of this approach we were able to identify DMRs on all autosomes. Most of them were found to be hypermethylated in the placenta and located on gene bodies, similarly to the MeDIP-Chip. Furthermore, no correlation was found between the number of DMRs identified on each chromosome with number of genes, chromosome size, number of CGIs and GC content supporting the unbiased character of the methodology. In addition, all of the control regions and most of the newly identified DMRs confirmed the differential methylation status between fetal and maternal tissue. Nevertheless, higher read depth may have been necessary for all the DMRs to be confirmed. Even though a great number of DMRs was identified with the MeDIP-seq using the MEDIPS software only 20% showed overlap between the two platforms. Since both methodologies used the same samples and same principle for DMR identification with highly consistent results, we, as well as previous studies, have attributed this low DMR overlapping to the inefficiency of MEDIPS to correctly classify a region as DMR. Overall, MeDIP-seq and MeDIP-Chip provide consistent, robust and accurate results for DMR identification. Due to the common affinity based methylation enrichment (MeDIP), they both detect methylated regions in a sequence independent manner on a wide range of CpG sites.

In conclusion, we have performed methylation analysis of chromosomes 13, 18, 21 and X on fetal and maternal tissue. We have identified a great number of DMRs that can potentially be implemented in the non-invasive prenatal diagnosis of the most common aneuploidies associated with pregnancies. Furthermore, we provide a great number of DMRs on all autosomes that can be used in the investigation of fetal abnormalities associated with microdeletion and microduplication syndromes

(Section 5-Future plans). Finally, using two different platforms we confirmed the distinct methylation patterns between the two tissues and reaffirmed the robust nature of MeDIP in the methylation enrichment.

5 FUTURE PLANS

We have successfully identified a great number of DMRs on chromosomes 13, 18 21 and X using robust and highly validated methodologies. This way we expanded the panel of existing DMRs on chromosome 21 and characterized a new set of markers that can potentially be used in the development of assays towards the detection of Patau and Edwards syndromes and chromosome X aneuploidies. Since the scope of this study was DMR discovery on the chromosomes involved in the most common aneuploidies, DMR identification on other chromosomes was overlooked. Thus, an immediate future plan is to revisit the MeDIP–seq data with emphasis on DMRs that overlap microdeletion or microduplication syndromes' critical regions. Primer design and validation of these regions would potentially allow us to further increase our NIPD panel in order to provide a more comphenisve non-invasive diagnostic test available for all pregnancies.

Many studies have failed to quantify the minute differences between normal and trisomy cases due the presence of high maternal background. It is estimated that for 10% fetal concentration, in order to distinguish normal from trisomy case the analytical approach used needs to be able to detect 5% differences between the two samples. Quantification using qPCR after enrichment has showed highly consistent, sensitive and specific results. Despite that, its resolving power requires 2-fold change for copy number dosage, while absolute quantification requires the construction of standard curve for comparison, or normalization to control loci to quantify the unknowns. In order to target all ranges of fetal concentrations in the maternal circulation with higher confidence and less statistical analysis, we will attempt to combine the MeDIP approach with digital PCR. Digital PCR is an end point PCR that allows the distribution of small amounts of DNA in thousands of droplets. Each droplet represents a PCR reaction providing high depth coverage of the region. Absolute copy counts will enable us to distinguish the minute differences between normal and trisomy cases. Furthermore, multiplexing of DMRs will also be considered since low or variable copies may be recovered after MeDIP for specific DMRs. Combination of DMRs in the same reaction will allow us to even out the experimental/individual variability and also avoid inefficient quantification due to the low copy number recovery.

6 REFERENCES

1. Connor, M. and M. Ferguson-Smith, *Medical genetics*. 1997, Blackwell Science Ltd.
2. Gardner, R.M., G.R. Sutherland, and L.G. Shaffer, *Chromosome abnormalities and genetic counseling*. 2004: Cambridge Univ Press.
3. Goddijn, M. and N. Leschot, *Genetic aspects of miscarriage.* Best Practice & Research Clinical Obstetrics & Gynaecology, 2000. **14**(5): p. 855-865.
4. Parker, S.E., et al., *Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006.* Birth Defects Research Part A: Clinical and Molecular Teratology, 2010. **88**(12): p. 1008-1016.
5. Touraine, D.F.B.D.F.a.R., *Down syndrome*. 2014.
6. *CDC -Birth Defects-Facts about Down Syndrome*. 2014.
7. Verloes, A., *Edwards syndrome.* 2014.
8. Skupski, M.P., et al., *The Genome Sequence DataBase: towards an integrated functional genomics resource.* Nucleic Acids Res, 1999. **27**(1): p. 35-8.
9. Wu, J., A. Springett, and J.K. Morris, *Survival of trisomy 18 (Edwards syndrome) and trisomy 13 (Patau Syndrome) in England and Wales: 2004-2011.* Am J Med Genet A, 2013. **161A**(10): p. 2512-8.
10. Jeannine, G.J.a.V., *Klineftler syndrome.* 2006.
11. *Klinefelter syndrome.* 2014.
12. Zhang, Y., et al., *Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving.* Mol Biol Evol, 2013. **30**(12): p. 2588-601.
13. *Turner syndrome.* 2014.
14. Adams, M.M., et al., *Down's syndrome. Recent trends in the United States.* JAMA, 1981. **246**(7): p. 758-60.
15. *ACOG Practice Bulletin No. 77: screening for fetal chromosomal abnormalities.* Obstet Gynecol, 2007. **109**(1): p. 217-27.
16. Malone, F.D., et al., *First-trimester or second-trimester screening, or both, for Down's syndrome.* N Engl J Med, 2005. **353**(19): p. 2001-11.
17. Norton, M.E., et al., *First-trimester combined screening: experience with an instant results approach.* Am J Obstet Gynecol, 2007. **196**(6): p. 606 e1-5; discussion 606 e5.
18. Benacerraf, B.R., *The history of the second-trimester sonographic markers for detecting fetal Down syndrome, and their current role in obstetric practice.* Prenat Diagn, 2010. **30**(7): p. 644-52.
19. Nicolaides, K.H., et al., *Fetal nuchal translucency: ultrasound screening for chromosomal defects in first trimester of pregnancy.* BMJ, 1992. **304**(6831): p. 867-9.
20. Cicero, S., et al., *Maternal serum biochemistry at 11-13(+6) weeks in relation to the presence or absence of the fetal nasal bone on ultrasonography in chromosomally abnormal fetuses: an updated analysis of integrated ultrasound and biochemical screening.* Prenat Diagn, 2005. **25**(11): p. 977-83.
21. Agathokleous, M., et al., *Meta-analysis of second-trimester markers for trisomy 21.* Ultrasound Obstet Gynecol, 2013. **41**(3): p. 247-61.
22. Bahado-Singh, R.O., et al., *New Down syndrome screening algorithm: ultrasonographic biometry and multiple serum markers combined with maternal age.* Am J Obstet Gynecol, 1998. **179**(6 Pt 1): p. 1627-31.
23. Spencer, K., et al., *Temporal changes in maternal serum biochemical markers of trisomy 21 across the first and second trimester of pregnancy.* Ann Clin Biochem, 2002. **39**(Pt 6): p. 567-76.
24. Alfirevic, Z., K. Sundberg, and S. Brigham, *Amniocentesis and chorionic villus sampling for prenatal diagnosis.* Cochrane Database Syst Rev, 2003(3): p. CD003252.

25. Grimshaw, G., et al., *Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.* Health Technology Assessment, 2003. **7**(10).
26. Wilson, R.D., *Amniocentesis and chorionic villus sampling.* Curr Opin Obstet Gynecol, 2000. **12**(2): p. 81-6.
27. Czepulkowski, B. and D. Rooney, *Human Cytogenetics: A Practical Approach*. 1987: IRL Press.
28. Robberecht, C., et al., *Diagnosis of miscarriages by molecular karyotyping: benefits and pitfalls.* Genet Med, 2009. **11**(9): p. 646-54.
29. Carter, N.P., *Methods and strategies for analyzing copy number variation using DNA microarrays.* Nat Genet, 2007. **39**(7 Suppl): p. S16-21.
30. Munne, S., *Preimplantation genetic diagnosis for aneuploidy and translocations using array comparative genomic hybridization.* Curr Genomics, 2012. **13**(6): p. 463-70.
31. Evangelidou, P., et al., *Implementation of high resolution whole genome array CGH in the prenatal clinical setting: advantages, challenges, and review of the literature.* Biomed Res Int, 2013. **2013**: p. 346762.
32. Grigori, P., et al., *21 Mb deletion in chromosome band 13q22.2q32.1 associated with mild/moderate psychomotor retardation, growth hormone insufficiency, short neck, micrognathia, hypotonia, dysplastic ears and other dysmorphic features.* Eur J Med Genet, 2011. **54**(3): p. 365-8.
33. Rubio, C., et al., *Use of array comparative genomic hybridization (array-CGH) for embryo assessment: clinical results.* Fertil Steril, 2013. **99**(4): p. 1044-8.
34. Sismani, C., et al., *Cryptic genomic imbalances in patients with de novo or familial apparently balanced translocations and abnormal phenotype.* Mol Cytogenet, 2008. **1**: p. 15.
35. Verbeke, S.L., et al., *Array CGH analysis identifies two distinct subgroups of primary angiosarcoma of bone.* Genes Chromosomes Cancer, 2014.
36. Comas, C., et al., *Rapid aneuploidy testing versus traditional karyotyping in amniocentesis for certain referral indications.* J Matern Fetal Neonatal Med, 2010. **23**(9): p. 949-55.
37. Cirigliano, V., et al., *Clinical application of multiplex quantitative fluorescent polymerase chain reaction (QF-PCR) for the rapid prenatal detection of common chromosome aneuploidies.* Mol Hum Reprod, 2001. **7**(10): p. 1001-6.
38. Mann, K. and C.M. Ogilvie, *QF-PCR: application, overview and review of the literature.* Prenat Diagn, 2012. **32**(4): p. 309-14.
39. Vogel, F., Motulsky A., *Human Genetics. Problems and Approaches*. 4rth ed. 2010. 1006.
40. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res, 2002. **30**(12): p. e57.
41. MRC-Holland, *MLPA DNA protocol version MDTP-v001.* 2013.
42. Bertero, M.T., et al., *Circulating 'trophoblast' cells in pregnancy have maternal genetic markers.* Prenat Diagn, 1988. **8**(8): p. 585-90.
43. Covone, A.E., et al., *Analysis of peripheral maternal blood samples for the presence of placenta-derived cells using Y-specific probes and McAb H315.* Prenat Diagn, 1988. **8**(8): p. 591-607.
44. Bianchi, D.W., et al., *Isolation of fetal DNA from nucleated erythrocytes in maternal blood.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3279-83.
45. Bianchi, D.W., *Current knowledge about fetal blood cells in the maternal circulation.* J Perinat Med, 1998. **26**(3): p. 175-85.
46. Bianchi, D.W., et al., *Erythroid-specific antibodies enhance detection of fetal nucleated erythrocytes in maternal blood.* Prenat Diagn, 1993. **13**(4): p. 293-300.
47. Holzgreve, W., et al., *Disturbed feto-maternal cell traffic in preeclampsia.* Obstet Gynecol, 1998. **91**(5 Pt 1): p. 669-72.
48. Slunga-Tallberg, A., et al., *Maternal origin of nucleated erythrocytes in peripheral venous blood of pregnant women.* Hum Genet, 1995. **96**(1): p. 53-7.
49. Bianchi, D.W., et al., *Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum.* Proc Natl Acad Sci U S A, 1996. **93**(2): p. 705-8.

50. Bianchi, D.W., et al., *PCR quantitation of fetal cells in maternal blood in normal and aneuploid pregnancies.* Am J Hum Genet, 1997. **61**(4): p. 822-9.
51. Krabchi, K., et al., *Quantification of all fetal nucleated cells in maternal blood between the 18th and 22nd weeks of pregnancy using molecular cytogenetic techniques.* Clin Genet, 2001. **60**(2): p. 145-50.
52. Lo, Y.M., et al., *Presence of fetal DNA in maternal plasma and serum.* Lancet, 1997. **350**(9076): p. 485-7.
53. Lui, Y.Y., et al., *Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation.* Clin Chem, 2002. **48**(3): p. 421-7.
54. Alberry, M., et al., *Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast.* Prenat Diagn, 2007. **27**(5): p. 415-8.
55. Wataganara, T., et al., *Placental volume, as measured by 3-dimensional sonography and levels of maternal plasma cell-free fetal DNA.* Am J Obstet Gynecol, 2005. **193**(2): p. 496-500.
56. Sekizawa, A., et al., *Evaluation of bidirectional transfer of plasma DNA through placenta.* Hum Genet, 2003. **113**(4): p. 307-10.
57. Lo, Y.M., et al., *Quantitative analysis of the bidirectional fetomaternal transfer of nucleated cells and plasma DNA.* Clin Chem, 2000. **46**(9): p. 1301-9.
58. Devaney, S.A., et al., *Noninvasive fetal sex determination using cell-free fetal DNA: a systematic review and meta-analysis.* JAMA, 2011. **306**(6): p. 627-36.
59. Birch, L., et al., *Accurate and robust quantification of circulating fetal and total DNA in maternal plasma from 5 to 41 weeks of gestation.* Clin Chem, 2005. **51**(2): p. 312-20.
60. Lo, Y.M., et al., *Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis.* Am J Hum Genet, 1998. **62**(4): p. 768-75.
61. Ashoor, G., et al., *Trisomy 13 detection in the first trimester of pregnancy using a chromosome-selective cell-free DNA analysis method.* Ultrasound Obstet Gynecol, 2013. **41**(1): p. 21-5.
62. Lun, F.M., et al., *Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma.* Clinical chemistry, 2008. **54**(10): p. 1664-72.
63. Lo, Y.M., et al., *Rapid clearance of fetal DNA from maternal plasma.* Am J Hum Genet, 1999. **64**(1): p. 218-24.
64. Chim, S.S., et al., *Detection of the placental epigenetic signature of the maspin gene in maternal plasma.* Proc Natl Acad Sci U S A, 2005. **102**(41): p. 14753-8.
65. Tong, Y.K., et al., *Noninvasive prenatal detection of trisomy 21 by an epigenetic-genetic chromosome-dosage approach.* Clinical chemistry, 2010. **56**(1): p. 90-8.
66. Gopalakrishnan, S., B.O. Van Emburgh, and K.D. Robertson, *DNA methylation in development and human disease.* Mutat Res, 2008. **647**(1-2): p. 30-8.
67. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.
68. Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nat Genet, 2006. **38**(12): p. 1378-85.
69. Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis.* Nat Rev Genet, 2010. **11**(3): p. 191-203.
70. Mohandas, T., R.S. Sparkes, and L.J. Shapiro, *Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation.* Science, 1981. **211**(4480): p. 393-6.
71. Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-3.
72. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
73. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.
74. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

75. Iwamoto K, et al., *Neurons show distinctive DNA methylation profile and higher interindividual variations compared with non-neurons.* Genome Research, 2011. **5**(21): p. 688-96.
76. Szyf, M., *Epigenetics, a key for unlocking complex CNS disorders? Therapeutic implications.* Eur Neuropsychopharmacol, 2014.
77. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* Nature, 1983. **301**(5895): p. 89-92.
78. Callinan, P.A. and A.P. Feinberg, *The emerging science of epigenomics.* Human molecular genetics, 2006. **15 Spec No 1**: p. R95-101.
79. Feinberg, A.P., *Phenotypic plasticity and the epigenetics of human disease.* Nature, 2007. **447**(7143): p. 433-40.
80. Rakyan, V.K., et al., *DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project.* PLoS Biol, 2004. **2**(12): p. e405.
81. Rakyan, V.K., et al., *An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs).* Genome Res, 2008. **18**(9): p. 1518-29.
82. Kitamura, E., et al., *Analysis of tissue-specific differentially methylated regions (TDMs) in humans.* Genomics, 2007. **89**(3): p. 326-37.
83. Lokk, K., et al., *DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns.* Genome Biol, 2014. **15**(4): p. r54.
84. Fan, S. and X. Zhang, *CpG island methylation pattern in different human tissues and its correlation with gene expression.* Biochem Biophys Res Commun, 2009. **383**(4): p. 421-5.
85. Lee, J., et al., *Epigenetic mechanisms of neurodegeneration in Huntington's disease.* Neurotherapeutics, 2013. **10**(4): p. 664-76.
86. Doi, A., et al., *Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.* Nature genetics, 2009. **41**(12): p. 1350-3.
87. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia.* Cancer cell, 2010. **17**(1): p. 13-27.
88. Decock, A., et al., *Genome-wide promoter methylation analysis in neuroblastoma identifies prognostic methylation biomarkers.* Genome biology, 2012. **13**(10): p. R95.
89. Li, C.C., et al., *A sustained dietary change increases epigenetic variation in isogenic mice.* PLoS Genet, 2011. **7**(4): p. e1001380.
90. Wong, C.C., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-26.
91. Lam, L.L., et al., *Factors underlying variable DNA methylation in a human community cohort.* Proc Natl Acad Sci U S A, 2012. **109 Suppl 2**: p. 17253-60.
92. Feinberg, A.P. and R.A. Irizarry, *Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1757-64.
93. Bell, J.T., et al., *DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines.* Genome Biol, 2011. **12**(1): p. R10.
94. Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis.* Nature reviews. Genetics, 2010. **11**(3): p. 191-203.
95. Fouse, S.D., R.O. Nagarajan, and J.F. Costello, *Genome-scale DNA methylation analysis.* Epigenomics, 2010. **2**(1): p. 105-17.
96. Darst, R.P., et al., *Bisulfite sequencing of DNA.* Curr Protoc Mol Biol, 2010. **Chapter 7**: p. Unit 7 9 1-17.
97. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): p. 457-66.
98. Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nat Genet, 2005. **37**(8): p. 853-62.

99. Yalcin, A., et al., *MeDIP coupled with a promoter tiling array as a platform to investigate global DNA methylation patterns in AML cells.* Leuk Res, 2013. **37**(1): p. 102-11.
100. Papageorgiou, E.A., et al., *Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies.* The American journal of pathology, 2009. **174**(5): p. 1609-18.
101. Down, T.A., et al., *A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis.* Nat Biotechnol, 2008. **26**(7): p. 779-85.
102. Lun, F.M., et al., *Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA.* Clin Chem, 2013. **59**(11): p. 1583-94.
103. Laird, P.W., *Cancer epigenetics.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R65-76.
104. Mersy, E., et al., *Noninvasive detection of fetal trisomy 21: systematic review and report of quality and outcomes of diagnostic accuracy studies performed between 1997 and 2012.* Hum Reprod Update, 2013. **19**(4): p. 318-29.
105. Bianchi, D.W., et al., *Noninvasive prenatal diagnosis of fetal Rhesus D: ready for Prime(r) Time.* Obstet Gynecol, 2005. **106**(4): p. 841-4.
106. Finning, K.M., et al., *Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RHD genotyping service.* Transfusion, 2002. **42**(8): p. 1079-85.
107. Lo, Y.M., et al., *Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma.* N Engl J Med, 1998. **339**(24): p. 1734-8.
108. Costa, J.M., et al., *[First trimester fetal sex determination in maternal serum using real-time PCR].* Gynecol Obstet Fertil, 2002. **30**(12): p. 953-7.
109. Scheffer, P.G., et al., *Reliability of fetal sex determination using maternal plasma.* Obstet Gynecol, 2010. **115**(1): p. 117-26.
110. Sekizawa, A., et al., *Accuracy of fetal gender determination by analysis of DNA in maternal plasma.* Clin Chem, 2001. **47**(10): p. 1856-8.
111. Page-Christiaens, G.C., et al., *Use of bi-allelic insertion/deletion polymorphisms as a positive control for fetal genotyping in maternal blood: first clinical experience.* Ann N Y Acad Sci, 2006. **1075**: p. 123-9.
112. Finning, K., P. Martin, and G. Daniels, *A clinical service in the UK to predict fetal Rh (Rhesus) D blood group using free fetal DNA in maternal plasma.* Ann N Y Acad Sci, 2004. **1022**: p. 119-23.
113. Chiu, R.W., C.R. Cantor, and Y.M. Lo, *Non-invasive prenatal diagnosis by single molecule counting technologies.* Trends in genetics : TIG, 2009. **25**(7): p. 324-31.
114. Wright, C.F. and H. Burton, *The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis.* Hum Reprod Update, 2009. **15**(1): p. 139-51.
115. Chiu, R.W., et al., *Hypermethylation of RASSF1A in human and rhesus placentas.* The American journal of pathology, 2007. **170**(3): p. 941-50.
116. Chim, S.S., et al., *Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test for fetal trisomy 21.* Clinical chemistry, 2008. **54**(3): p. 500-11.
117. Old, R.W., et al., *Candidate epigenetic biomarkers for non-invasive prenatal diagnosis of Down syndrome.* Reproductive biomedicine online, 2007. **15**(2): p. 227-35.
118. Yin, A., et al., *Screening significantly hypermethylated genes in fetal tissues compared with maternal blood using a methylated-CpG island recovery assay-based microarray.* BMC Med Genomics, 2012. **5**: p. 26.
119. Meyer, M. and M. Kircher, *Illumina sequencing library preparation for highly multiplexed target capture and sequencing.* Cold Spring Harb Protoc, 2010. **2010**(6): p. pdb.prot5448.
120. Xiang, Y., et al., *DNA methylome profiling of maternal peripheral blood and placentas reveal potential fetal DNA markers for non-invasive prenatal testing.* Mol Hum Reprod, 2014. **20**(9): p. 875-84.
121. Lo, Y.M., et al., *Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis.* American journal of human genetics, 1998. **62**(4): p. 768-75.

122. Lo, Y.M., et al., *Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection.* Nature medicine, 2007. **13**(2): p. 218-23.
123. Tsui, N.B., et al., *Synergy of total PLAC4 RNA concentration and measurement of the RNA single-nucleotide polymorphism allelic ratio for the noninvasive prenatal detection of trisomy 21.* Clinical chemistry, 2010. **56**(1): p. 73-81.
124. Lo, Y.M., et al., *Digital PCR for the molecular detection of fetal chromosomal aneuploidy.* Proc Natl Acad Sci U S A, 2007. **104**(32): p. 13116-21.
125. Tong, Y.K., et al., *Epigenetic-genetic chromosome dosage approach for fetal trisomy 21 detection using an autosomal genetic reference marker.* PLoS One, 2010. **5**(12): p. e15244.
126. Papageorgiou, E.A., et al., *Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21.* Nature medicine, 2011. **17**(4): p. 510-3.
127. Tsaliki, E., et al., *MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21.* Prenatal diagnosis, 2012. **32**(10): p. 996-1001.
128. Chiu, R.W., et al., *Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20458-63.
129. Fan, H.C., et al., *Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood.* Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16266-71.
130. Chiu, R.W., et al., *Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study.* BMJ, 2011. **342**: p. c7401.
131. Ehrich, M., et al., *Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting.* Am J Obstet Gynecol, 2011. **204**(3): p. 205 e1-11.
132. Palomaki, G.E., et al., *DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study.* Genet Med, 2011. **13**(11): p. 913-20.
133. Chen, E.Z., et al., *Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing.* PLoS One, 2011. **6**(7): p. e21791.
134. Palomaki, G.E., et al., *DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study.* Genet Med, 2012. **14**(3): p. 296-305.
135. Sehnert, A.J., et al., *Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood.* Clin Chem, 2011. **57**(7): p. 1042-9.
136. Ashoor, G., et al., *Trisomy 13 detection in the first trimester of pregnancy using a chromosome-selective cell-free DNA analysis method.* Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 2013. **41**(1): p. 21-5.
137. Norton, M.E., et al., *Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18.* American journal of obstetrics and gynecology, 2012. **207**(2): p. 137 e1-8.
138. Sparks, A.B., et al., *Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18.* American journal of obstetrics and gynecology, 2012. **206**(4): p. 319 e1-9.
139. Sparks, A.B., et al., *Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy.* Prenatal diagnosis, 2012. **32**(1): p. 3-9.
140. Patsalis, P.C., *A new method for non-invasive prenatal diagnosis of Down syndrome using MeDIP real time qPCR.* Applied & Translational Genomics, 2012. **1**(0): p. 3-8.
141. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
142. Pavlos Antoniou, S.M., Marios Ioannides, Elisavet A. Papageorgiou and Philippos C. Patsalis, *Identifying Differentially Methylated Regions by efficient bit-masking of DNA microarrays to use as markers for Non Invasive Prenatal Diagnosis.* Engineering Intelligent Systems, 2014.

143. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.* Methods Mol Biol, 2000. **132**: p. 365-86.
144. Owczarzy, R., et al., *IDT SciTools: a suite for analysis and design of nucleic acid oligomers.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W163-9.
145. Ye, J., et al., *Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.* BMC bioinformatics, 2012. **13**(1): p. 134.
146. Mann, H.B. and D.R. Whitney, *On a test of whether one of two random variables is stochastically larger than the other.* The annals of mathematical statistics, 1947: p. 50-60.
147. Lance, G.N. and W.T. Williams, *A general theory of classificatory sorting strategies II. Clustering systems.* The computer journal, 1967. **10**(3): p. 271-277.
148. Andrews, S., *FastQC: A quality control tool for high throughput sequence data.* Reference Source, 2010.
149. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-1760.
150. Li, H., et al., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
151. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Research, 2010. **20**(9): p. 1297-1303.
152. Lienhard, M., et al., *MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments.* Bioinformatics, 2014. **30**(2): p. 284-286.
153. Butcher, L.M. and S. Beck, *AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay.* Methods, 2010. **52**(3): p. 223-31.
154. Bock, C., et al., *Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping.* Nucleic Acids Res, 2008. **36**(10): p. e55.
155. Schneider, E., et al., *Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns.* Nucleic Acids Res, 2010. **38**(12): p. 3880-90.
156. Pujadas, E. and A.P. Feinberg, *Regulated noise in the epigenetic landscape of development and disease.* Cell, 2012. **148**(6): p. 1123-31.
157. Zhang, X., et al., *Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis.* Cell, 2006. **126**(6): p. 1189-201.
158. Palmke, N., D. Santacruz, and J. Walter, *Comprehensive analysis of DNA-methylation in mammalian tissues using MeDIP-chip.* Methods, 2011. **53**(2): p. 175-84.
159. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
160. Chu, T., et al., *Structural and regulatory characterization of the placental epigenome at its maternal interface.* PLoS One, 2011. **6**(2): p. e14723.
161. Yamakawa, K., et al., *DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system.* Human Molecular Genetics, 1998. **7**(2): p. 227-237.
162. Barlow, G.M., et al., *Down syndrome congenital heart disease: a narrowed region and a candidate gene.* Genetics in Medicine, 2001. **3**(2): p. 91-101.
163. Chen, H., et al., *Single-minded and Down syndrome?* : Nat Genet. 1995 May;10(1):9-10.
164. Kay, G.F., *Xist and X chromosome inactivation.* Molecular and cellular endocrinology, 1998. **140**(1-2): p. 71-6.
165. Wutz, A. and J. Gribnau, *X inactivation Xplained.* Current opinion in genetics & development, 2007. **17**(5): p. 387-393.
166. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. **434**(7031): p. 400-4.
167. Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nature genetics, 2007. **39**(4): p. 457-66.
168. Schraders, M., et al., *Next-Generation Sequencing Identifies Mutations of< i> SMPX,</i> which Encodes the Small Muscle Protein, X-Linked, as a Cause of Progressive Hearing Impairment.* The American Journal of Human Genetics, 2011. **88**(5): p. 628-634.

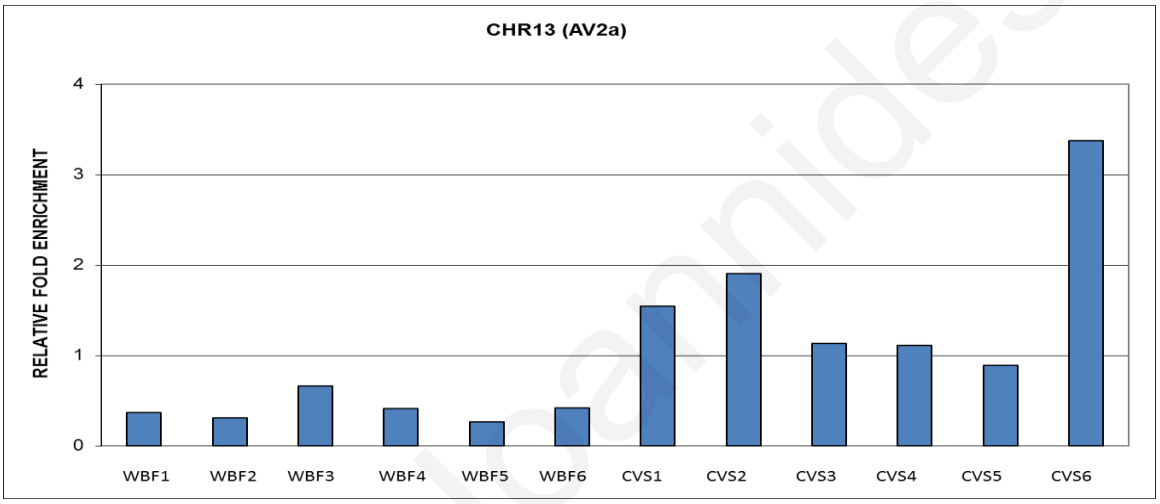
169. Abdelfatah, N., et al., *A Novel Deletion in SMPX Causes a Rare form of X-Linked Progressive Hearing Loss in Two Families Due to a Founder Effect.* Human mutation, 2013. **34**(1): p. 66-69.
170. Scheffer, I.E., et al., *X-linked myoclonic epilepsy with spasticity and intellectual disability: mutation in the homeobox gene ARX.* Neurology, 2002. **59**(3): p. 348-56.
171. Kato, M., et al., *Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation.* Hum Mutat, 2004. **23**(2): p. 147-59.
172. Proud, V.K., C. Levine, and N.J. Carpenter, *New X-linked syndrome with seizures, acquired micrencephaly, and agenesis of the corpus callosum.* Am J Med Genet, 1992. **43**(1-2): p. 458-66.
173. Ghebremichael, L.T., T.L. Veith, and J.M. Hamlett, *Integrated watershed- and farm-scale modeling framework for targeting critical source areas while maintaining farm economic viability.* J Environ Manage, 2013. **114**: p. 381-94.
174. Yalcin, A., et al., *MeDIP coupled with a promoter tiling array as a platform to investigate global DNA methylation patterns in AML cells.* Leukemia research, 2013. **37**(1): p. 102-11.
175. Strand, S.H., T.F. Orntoft, and K.D. Sorensen, *Prognostic DNA Methylation Markers for Prostate Cancer.* Int J Mol Sci, 2014. **15**(9): p. 16544-16576.
176. Kim, J.W., et al., *Identification of new differentially methylated genes that have potential functional consequences in prostate cancer.* PLoS One, 2012. **7**(10): p. e48455.
177. Taiwo, O., et al., *Methylome analysis using MeDIP-seq with low DNA concentrations.* Nat Protoc, 2012. **7**(4): p. 617-36.
178. Clark, C., et al., *A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling.* PLoS One, 2012. **7**(11): p. e50233.

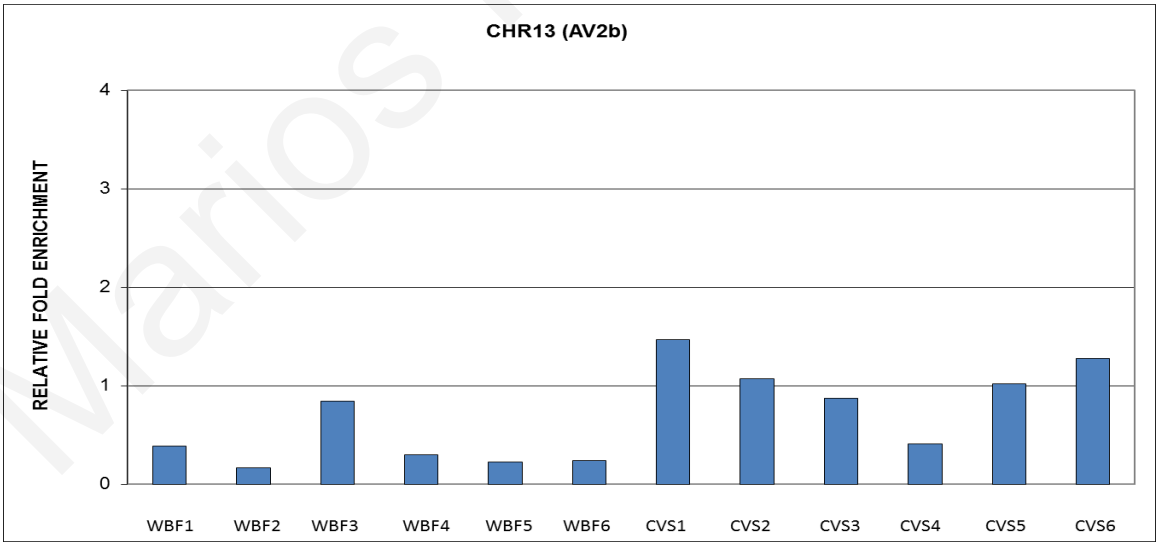
7 APPENDICES

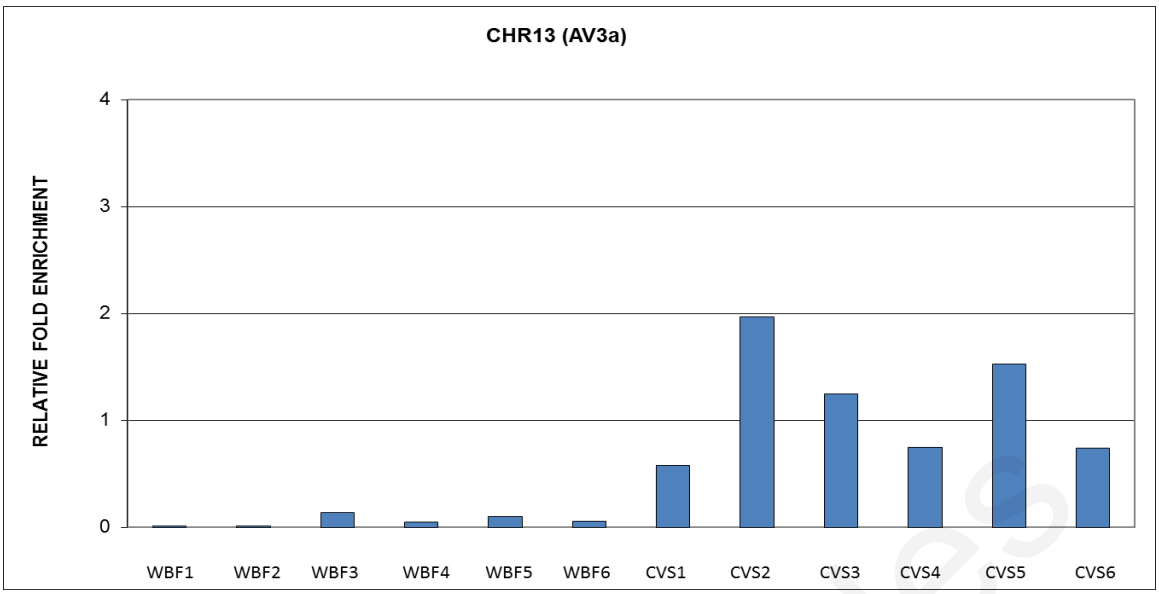
Appendix I

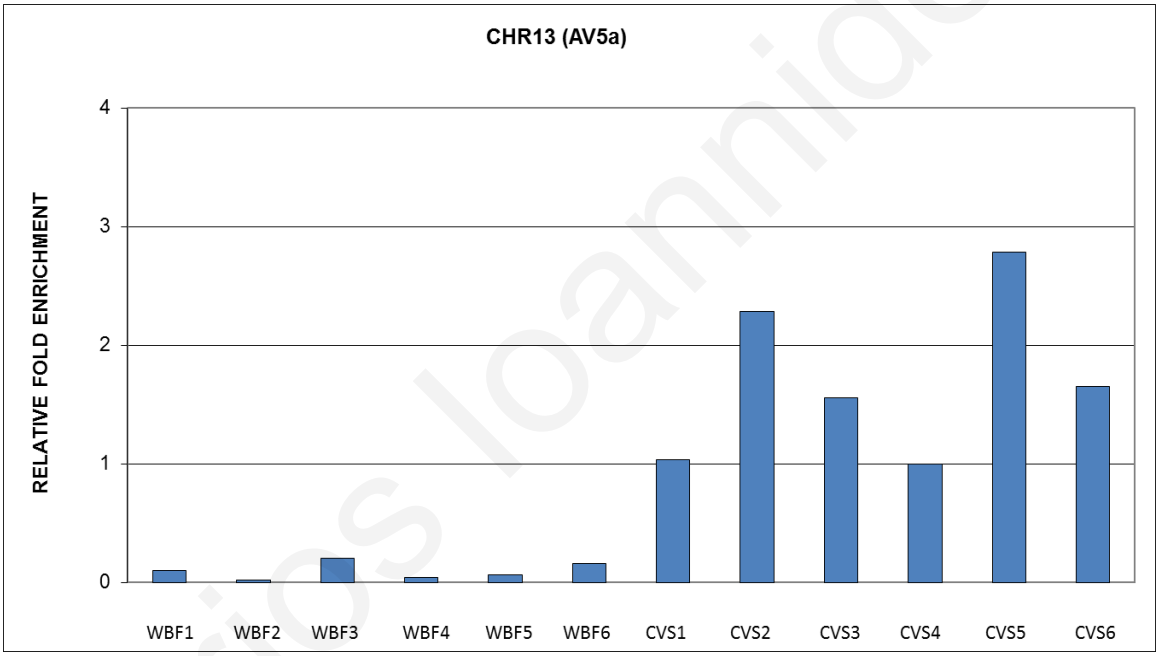
Methylation enrichment on six CVS and six WBF for chromosomes 13, 18 and 21 using the existing aCGH data.

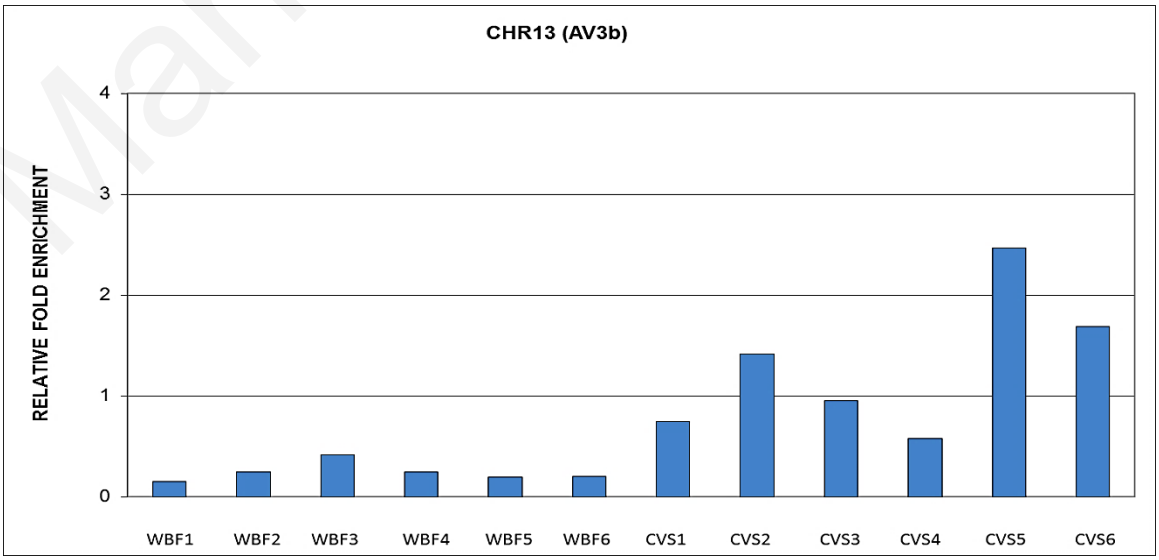
A. Chromosome 13

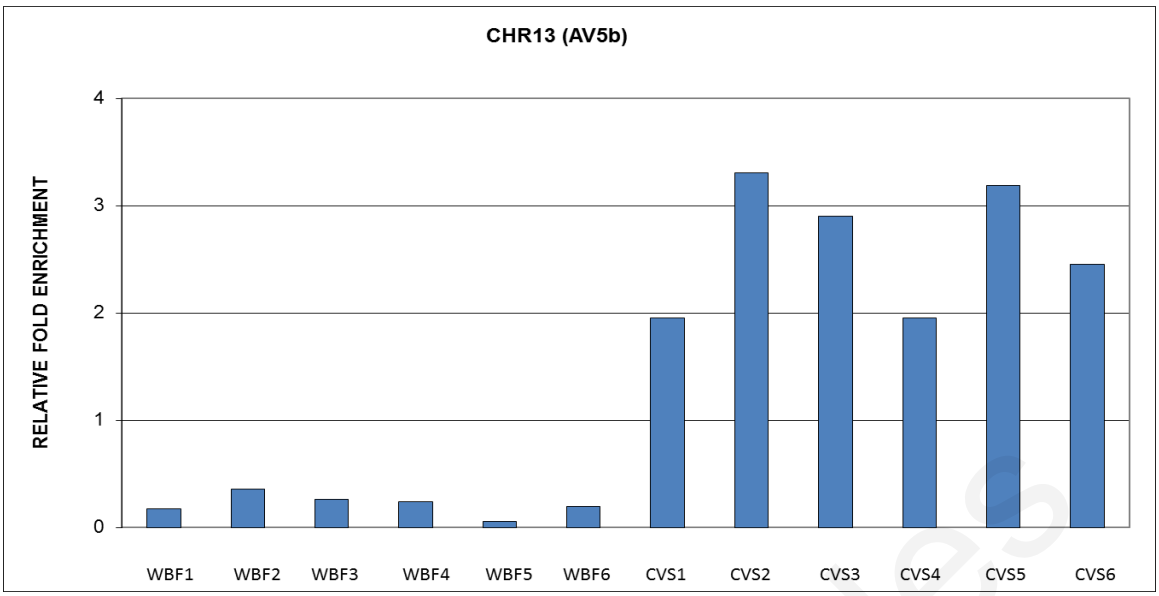


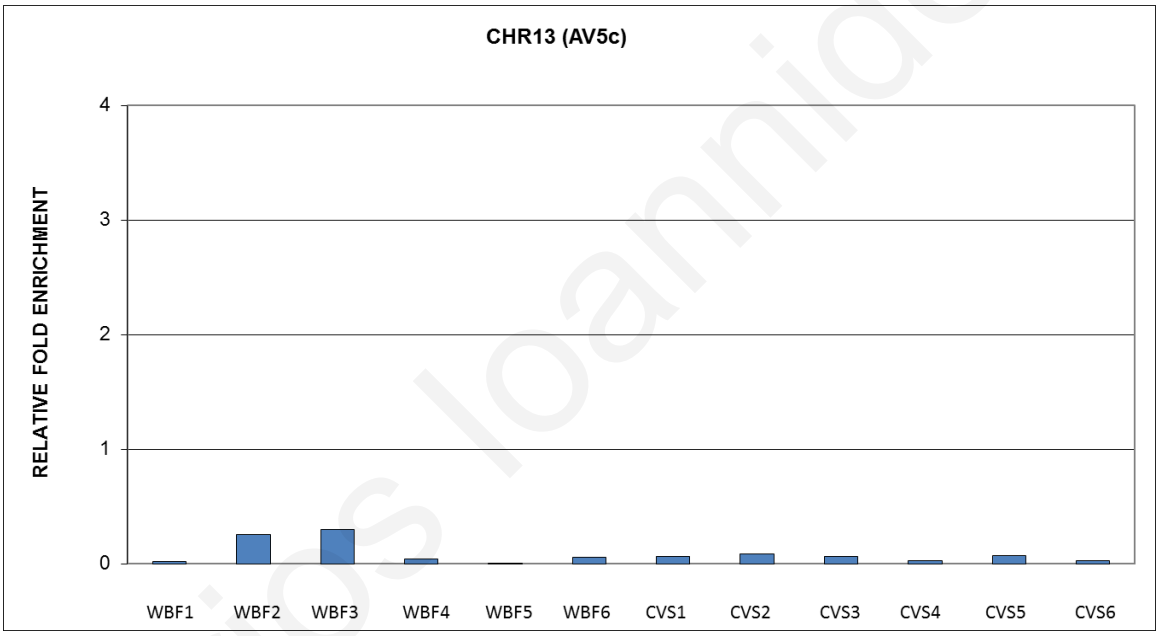


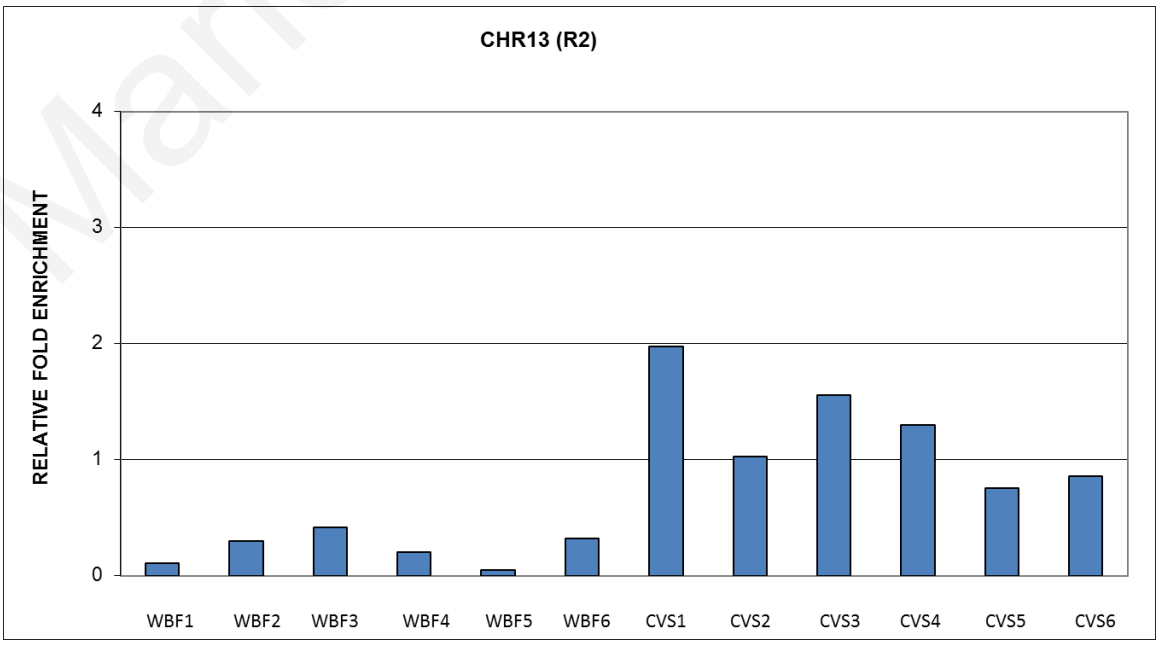


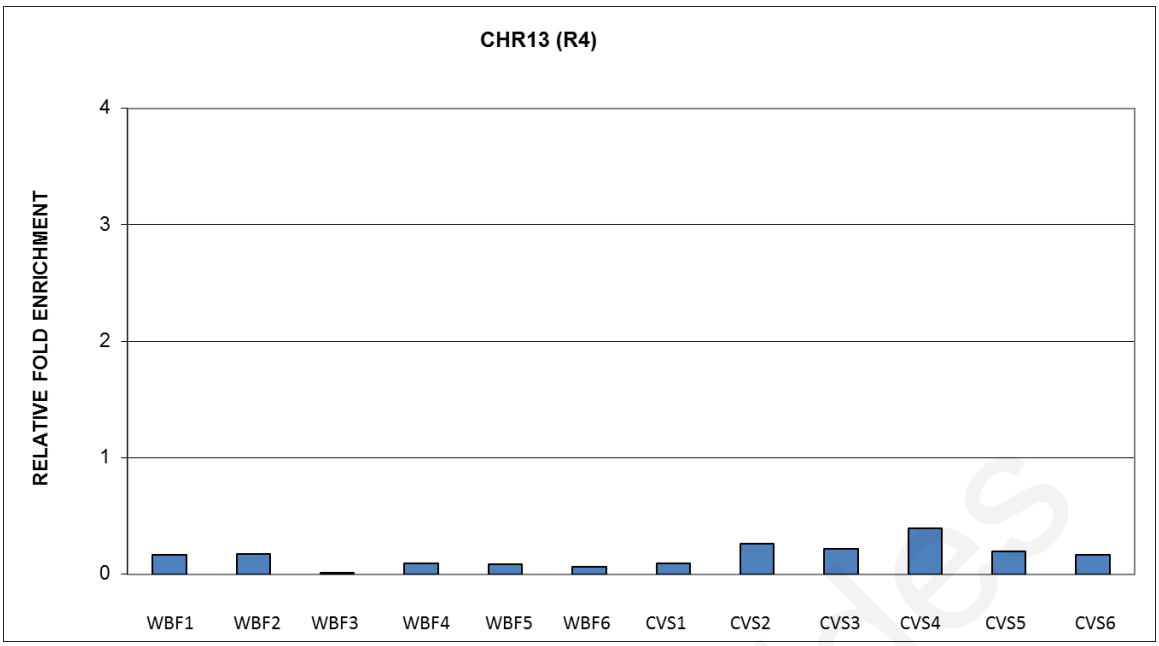


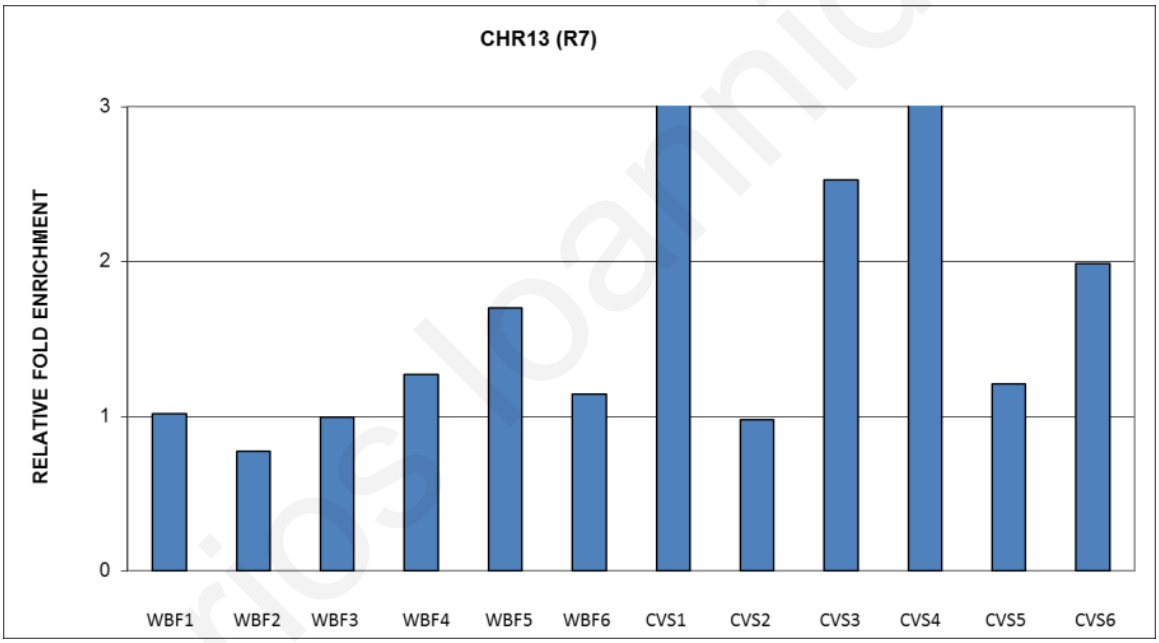


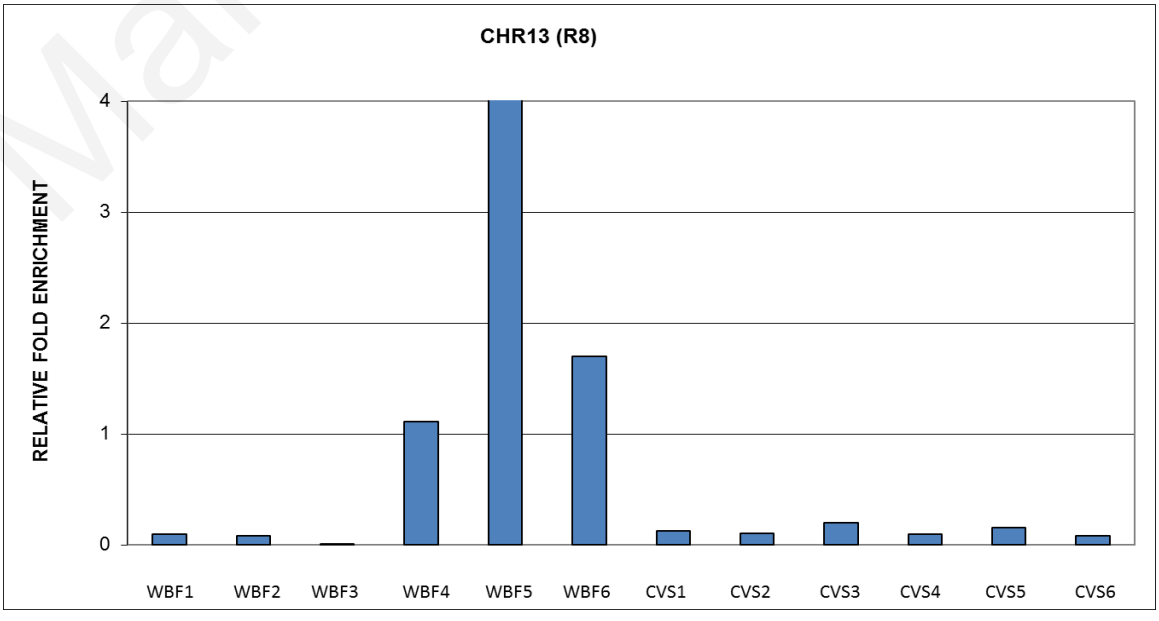


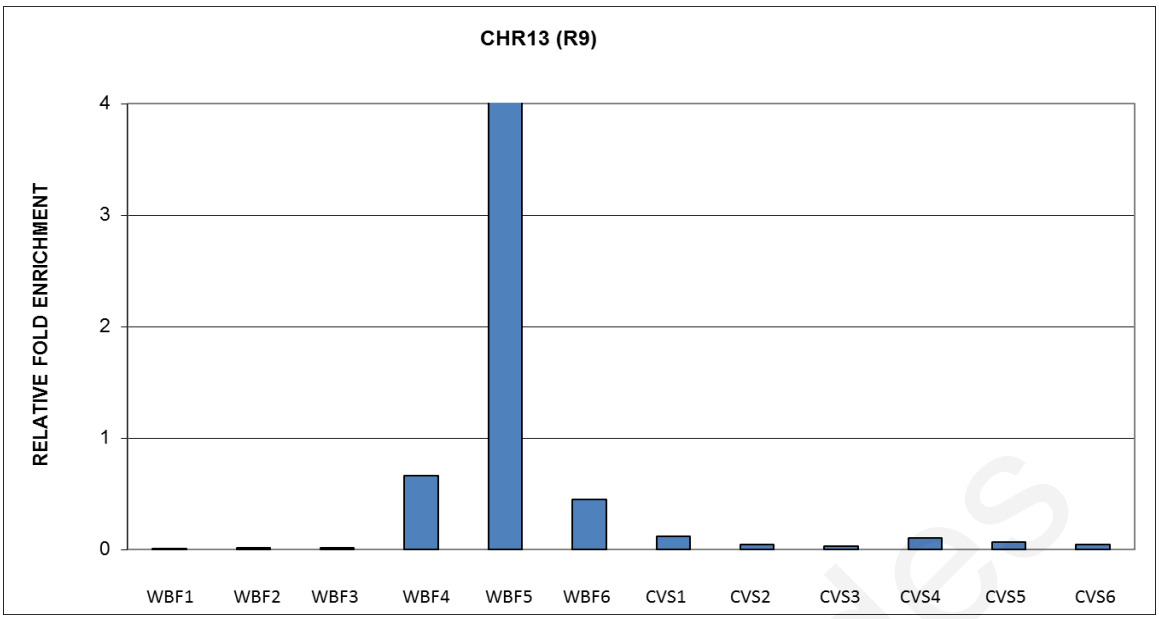


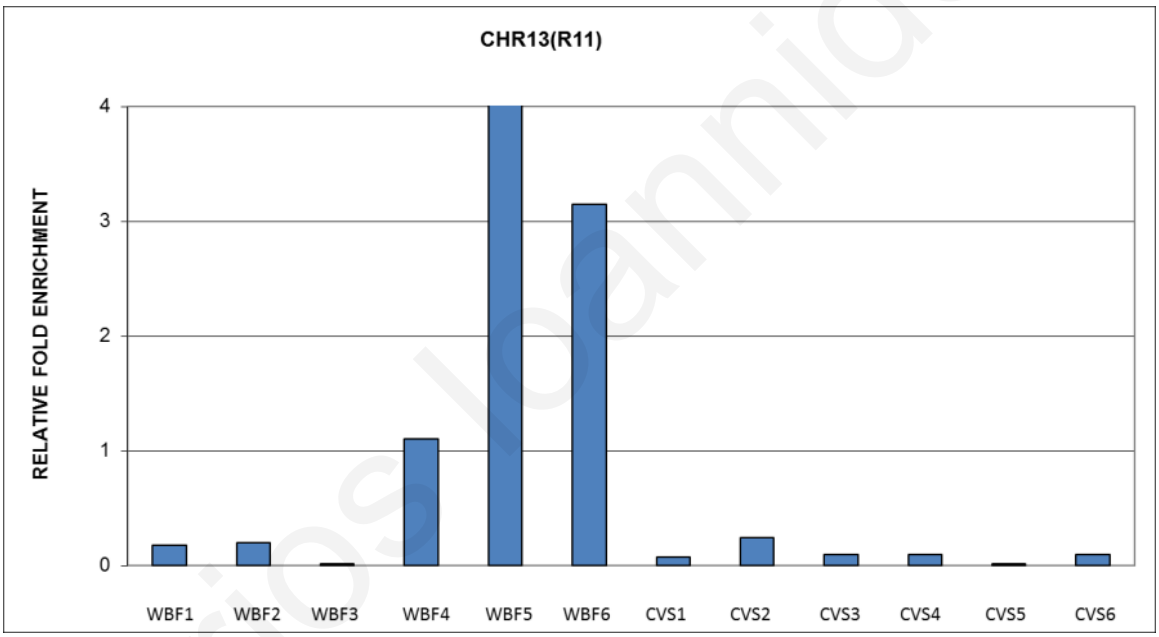


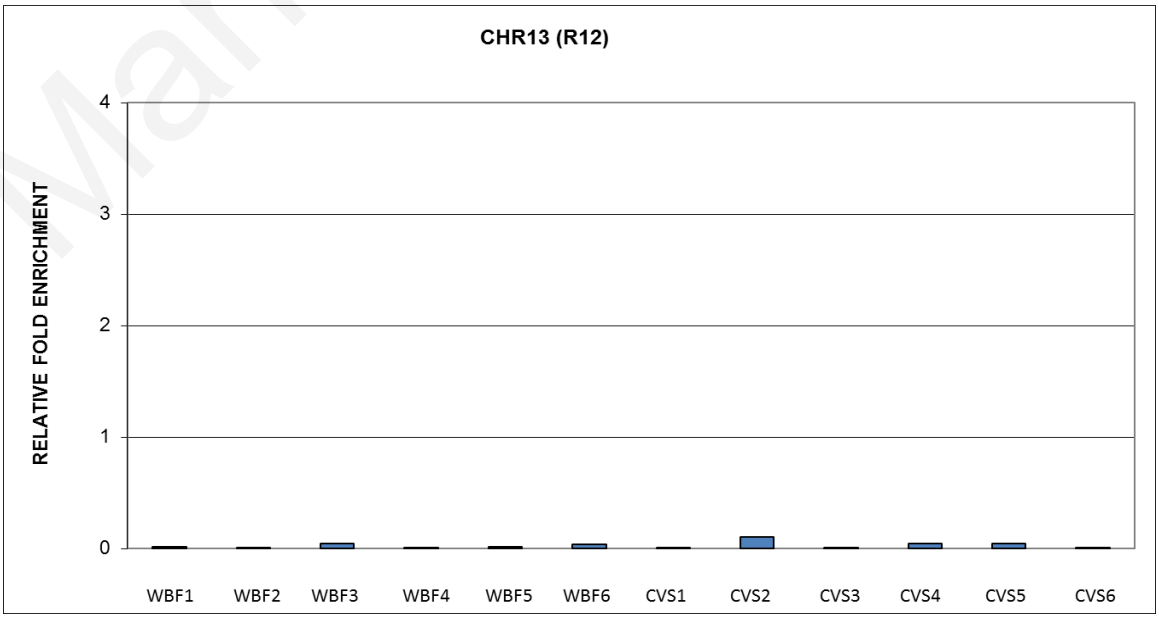


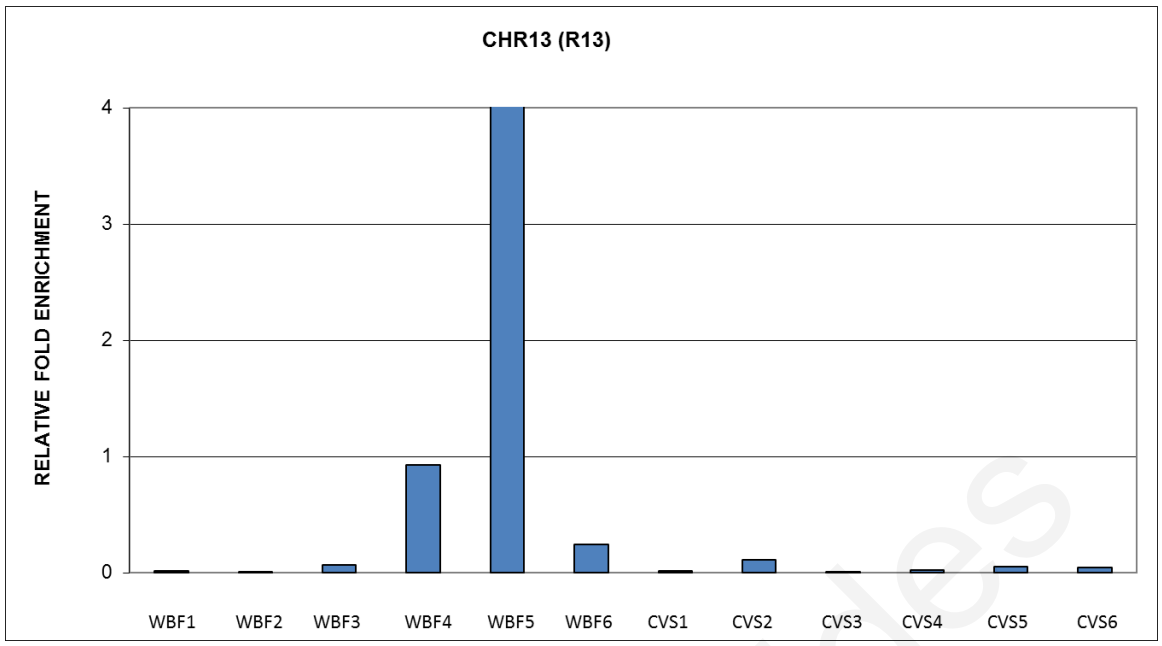




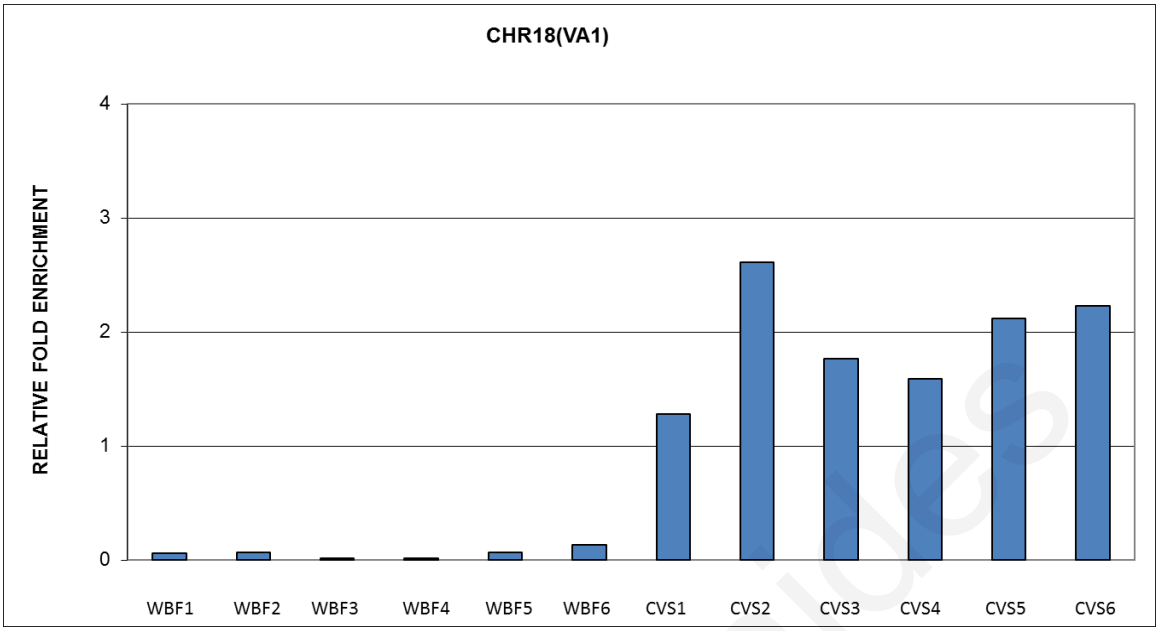


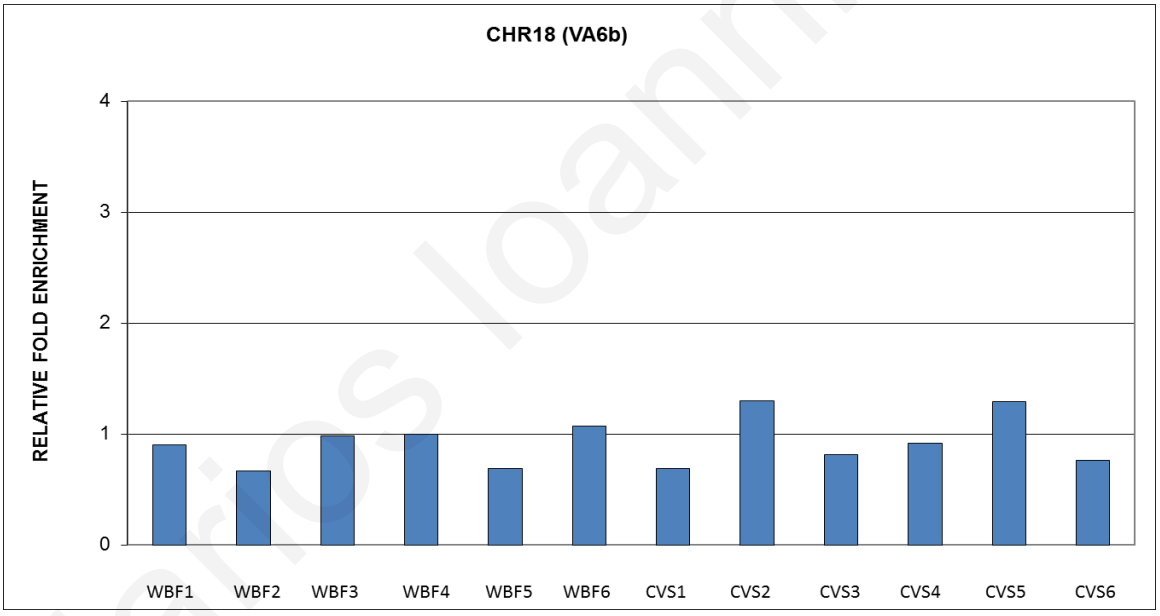


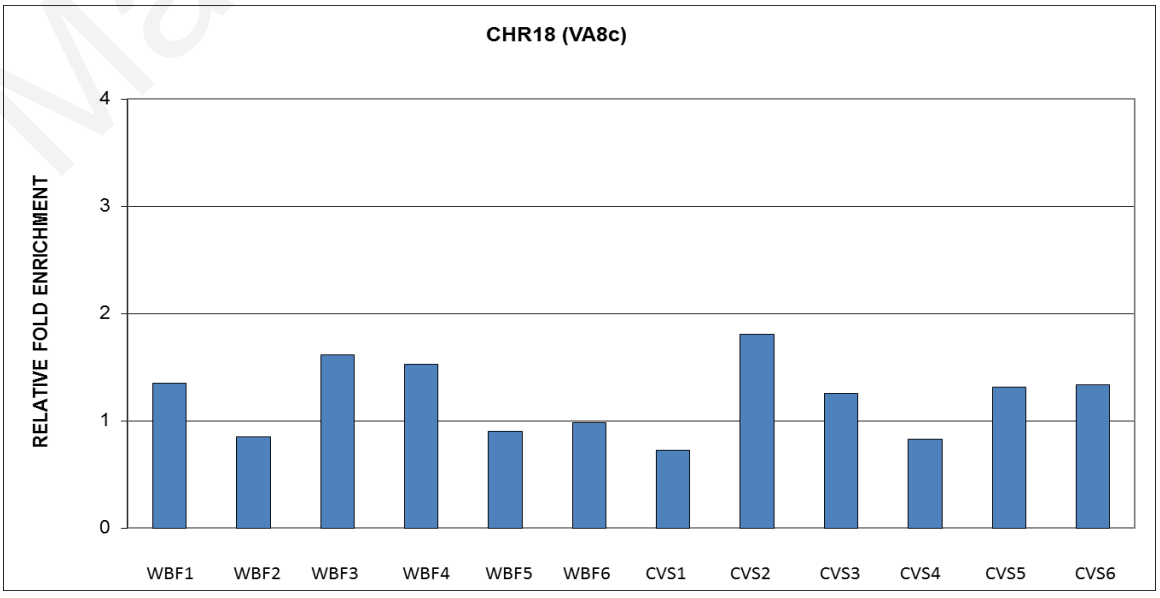


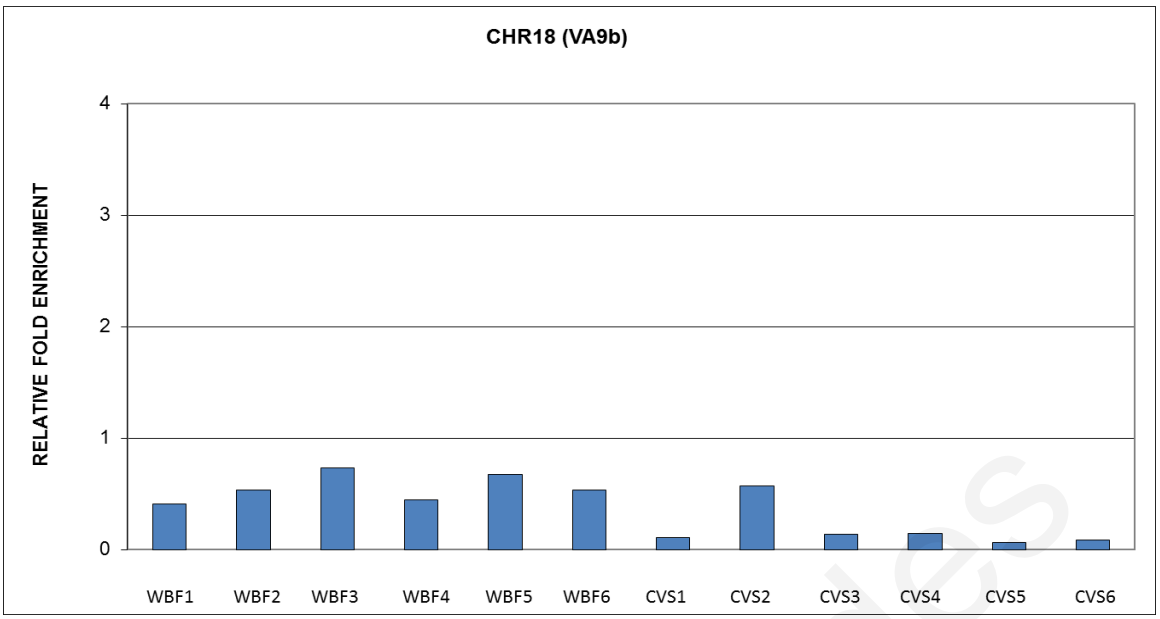


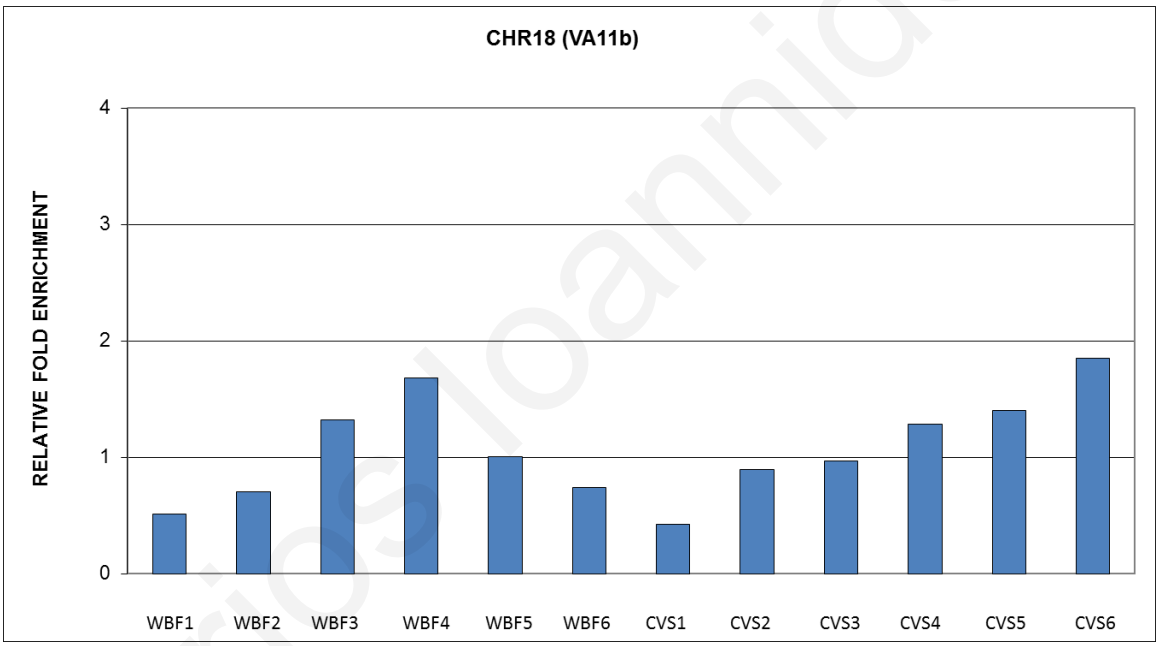
B. Chromosome 18

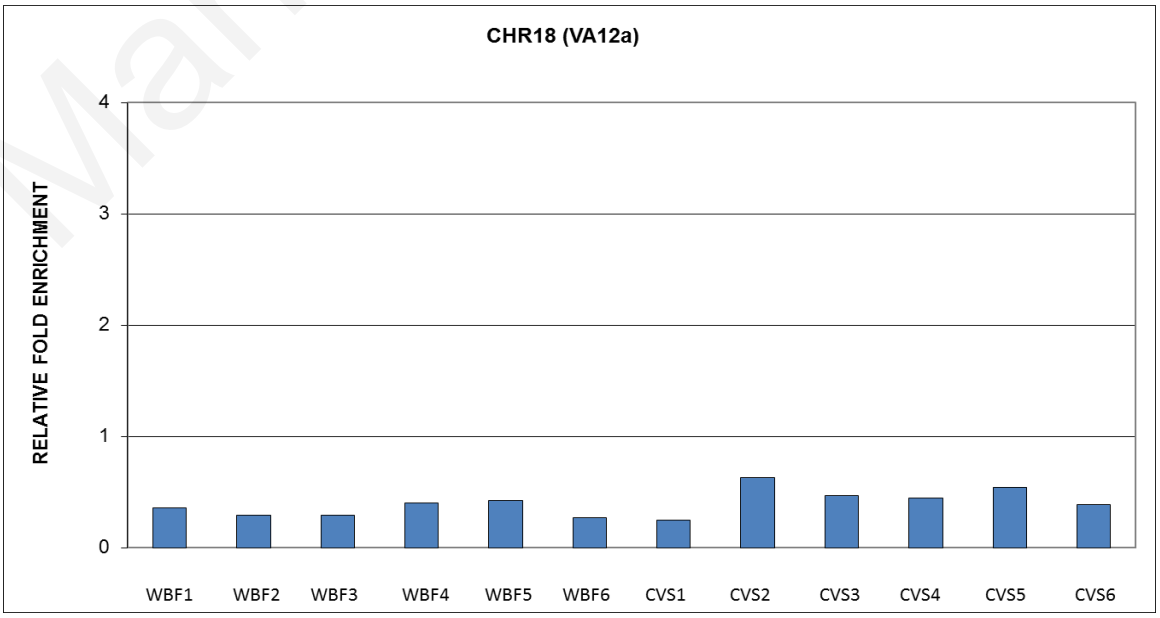


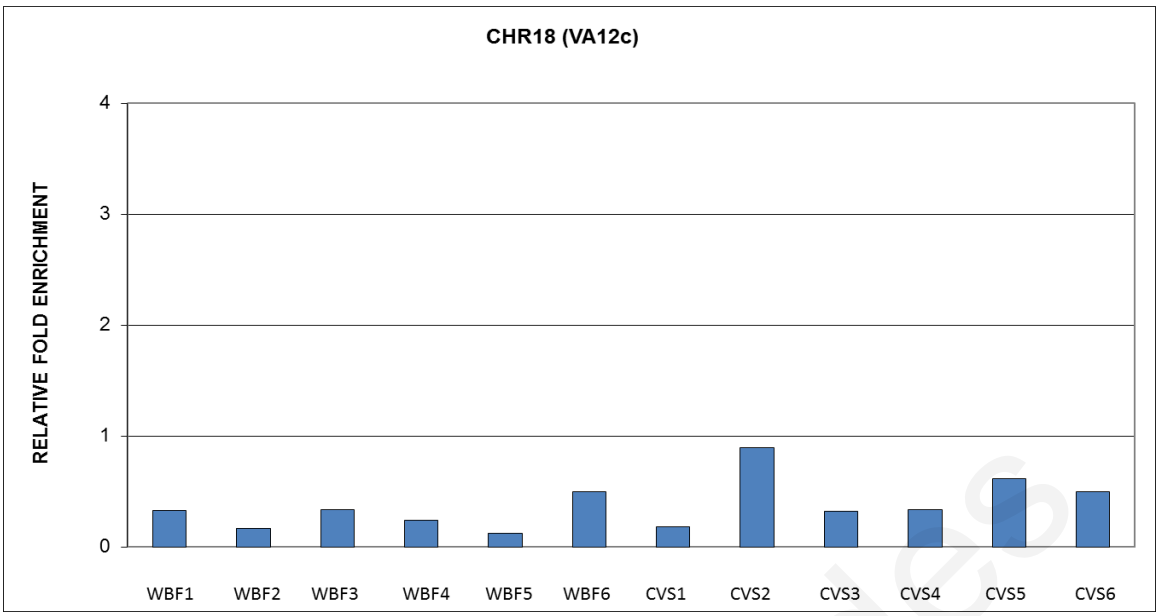


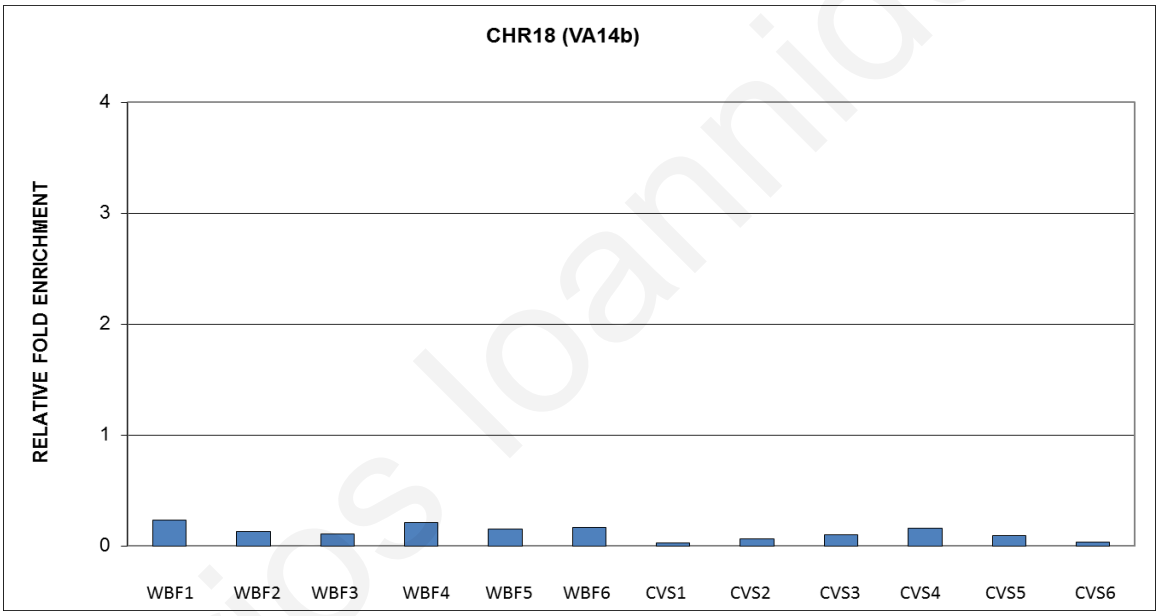


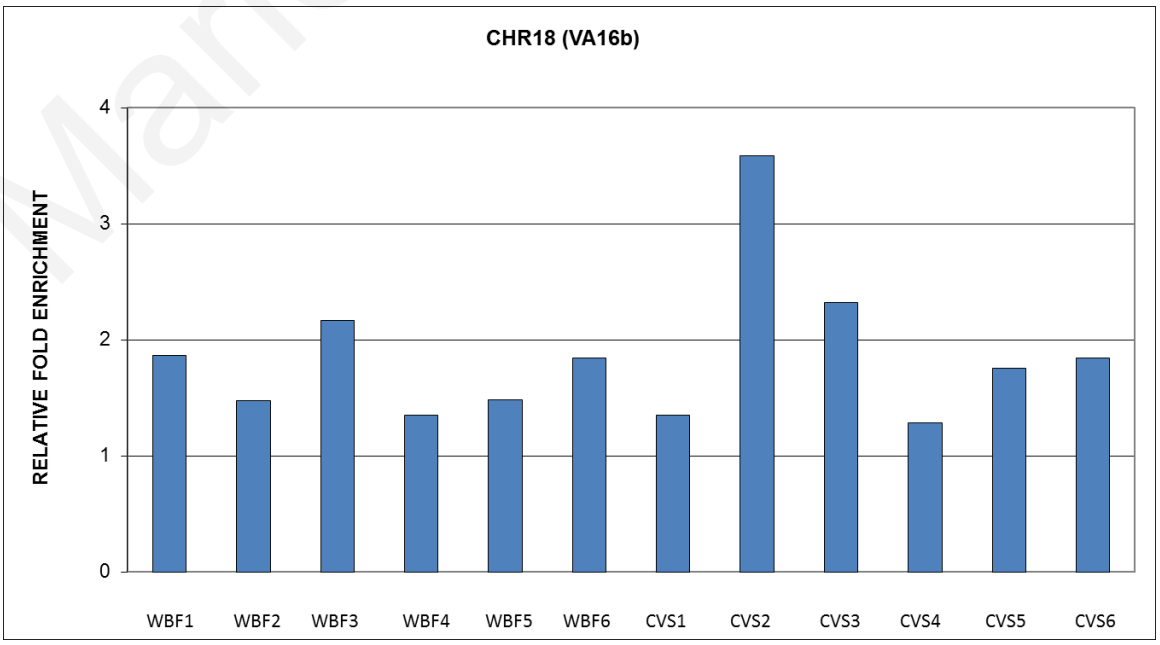


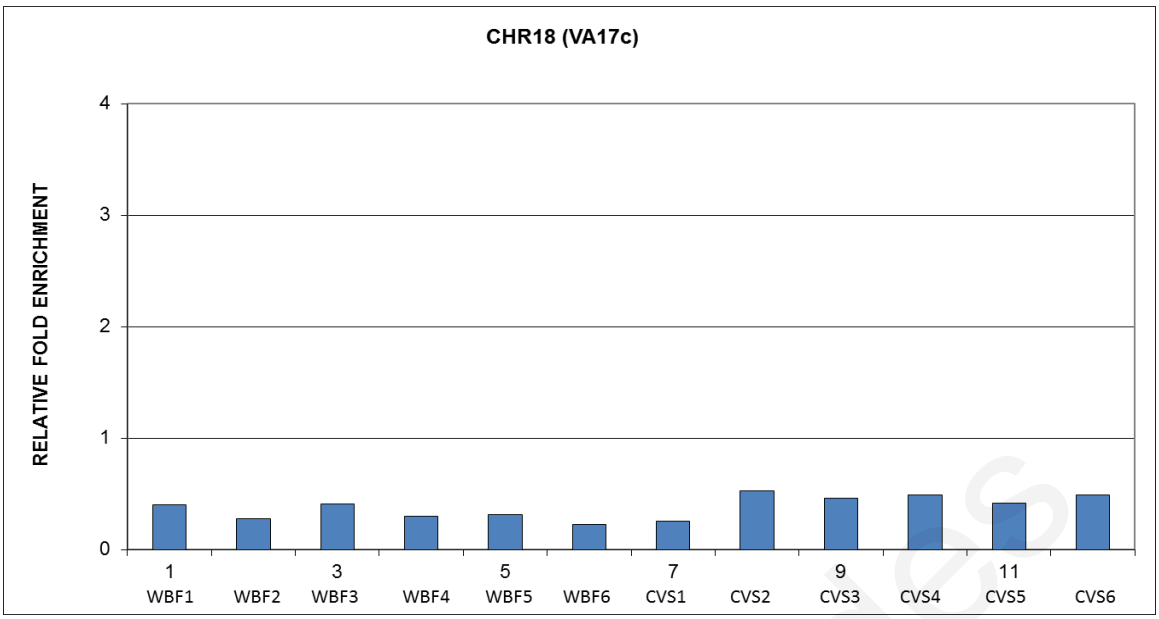


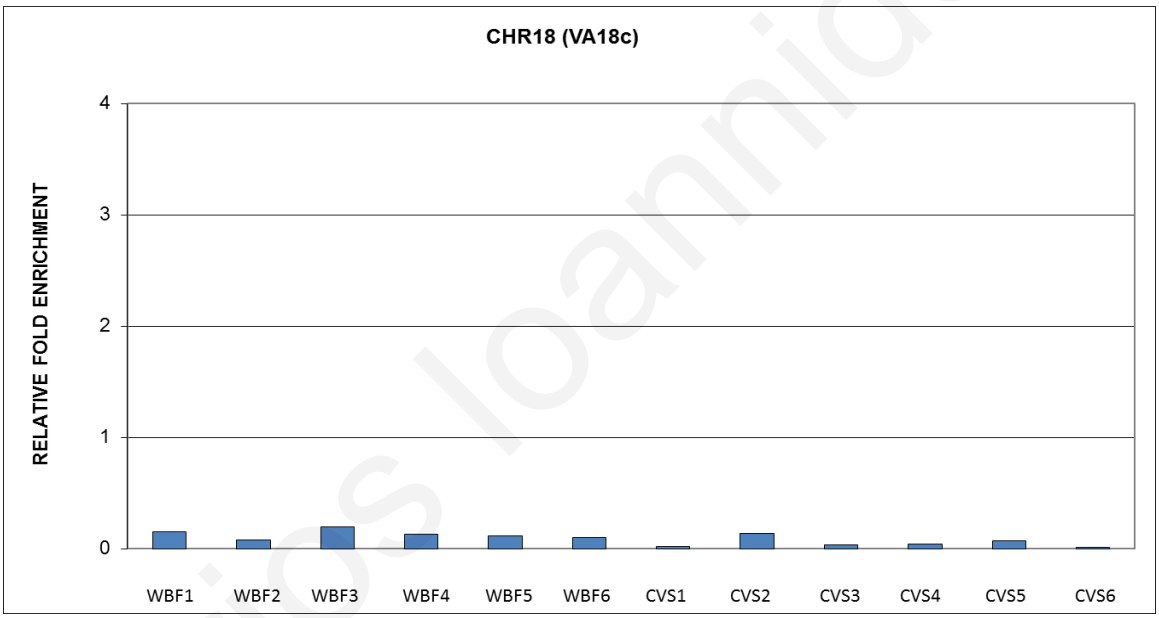


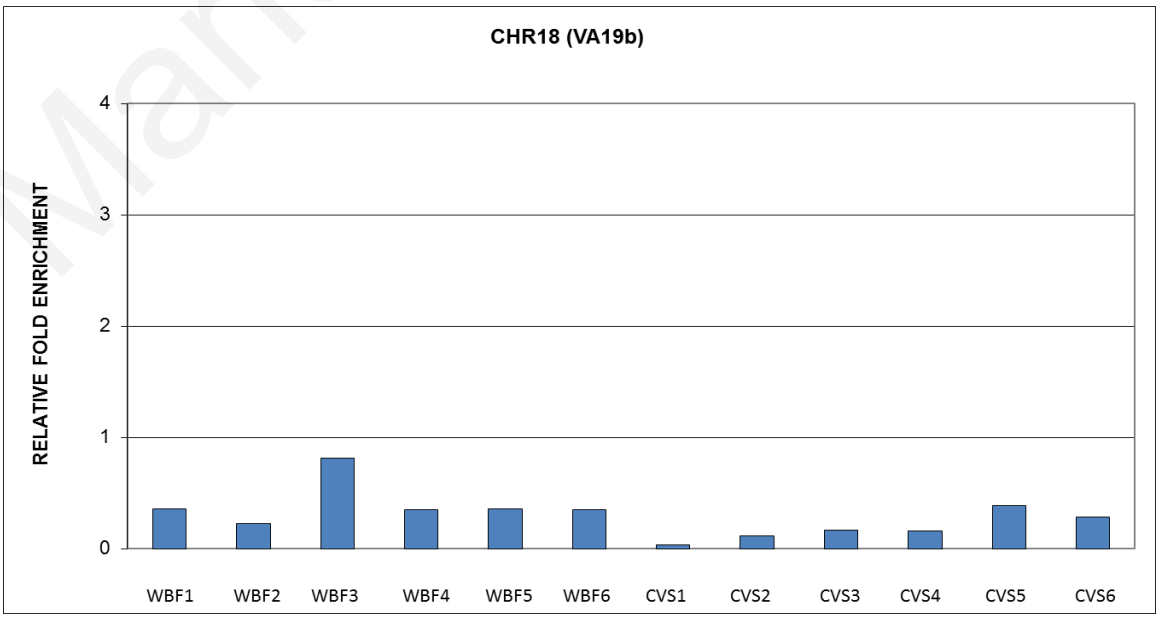


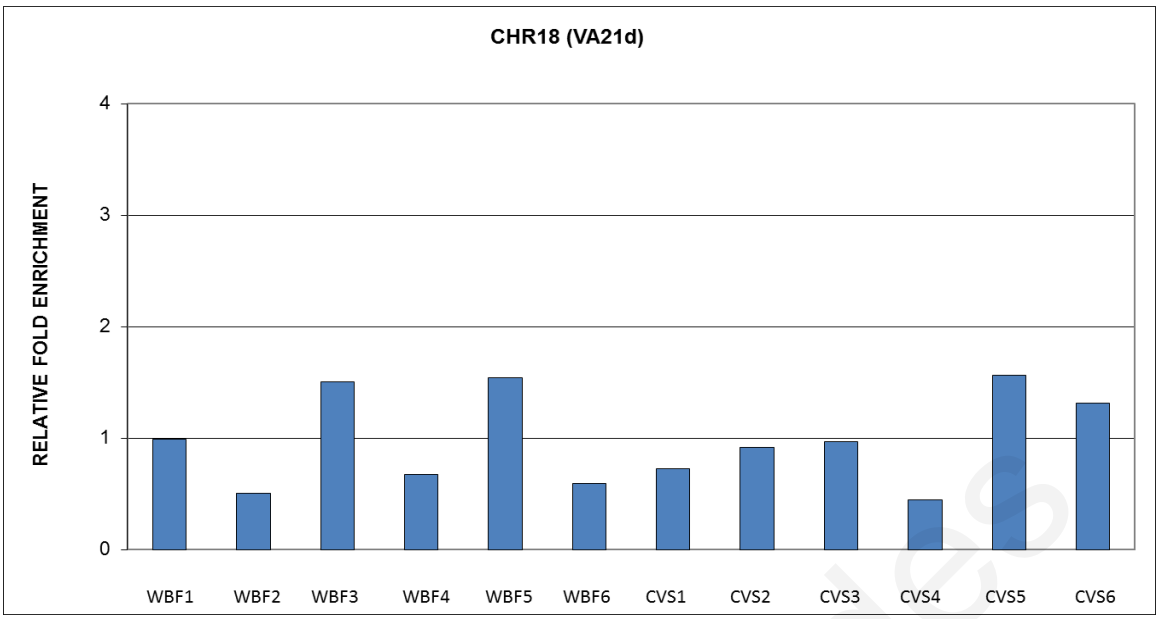


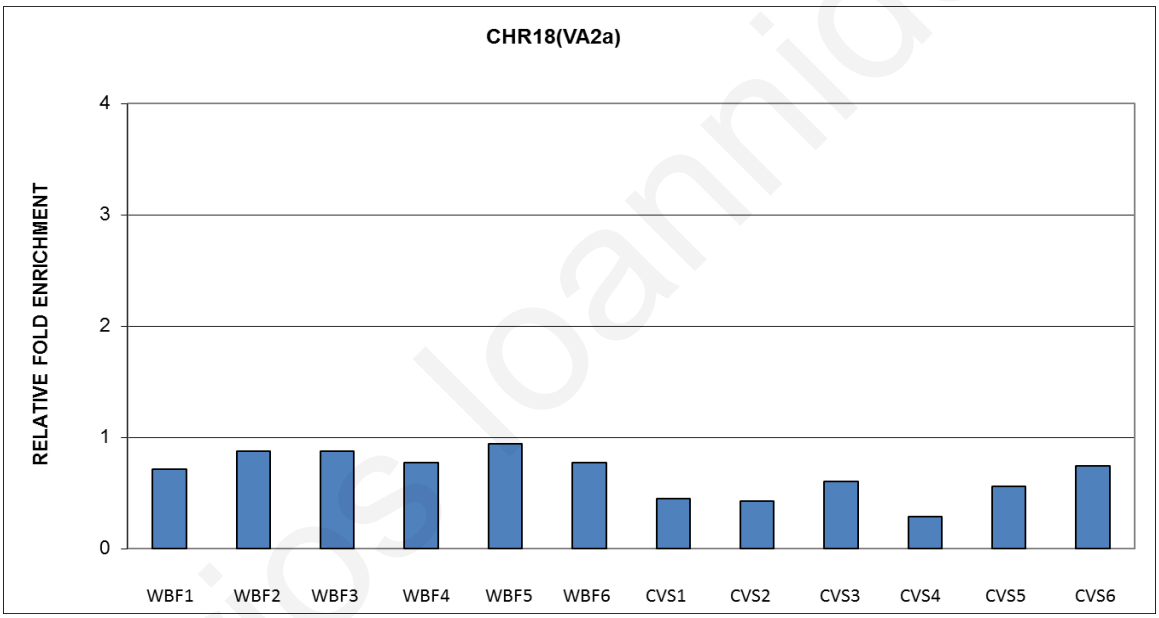


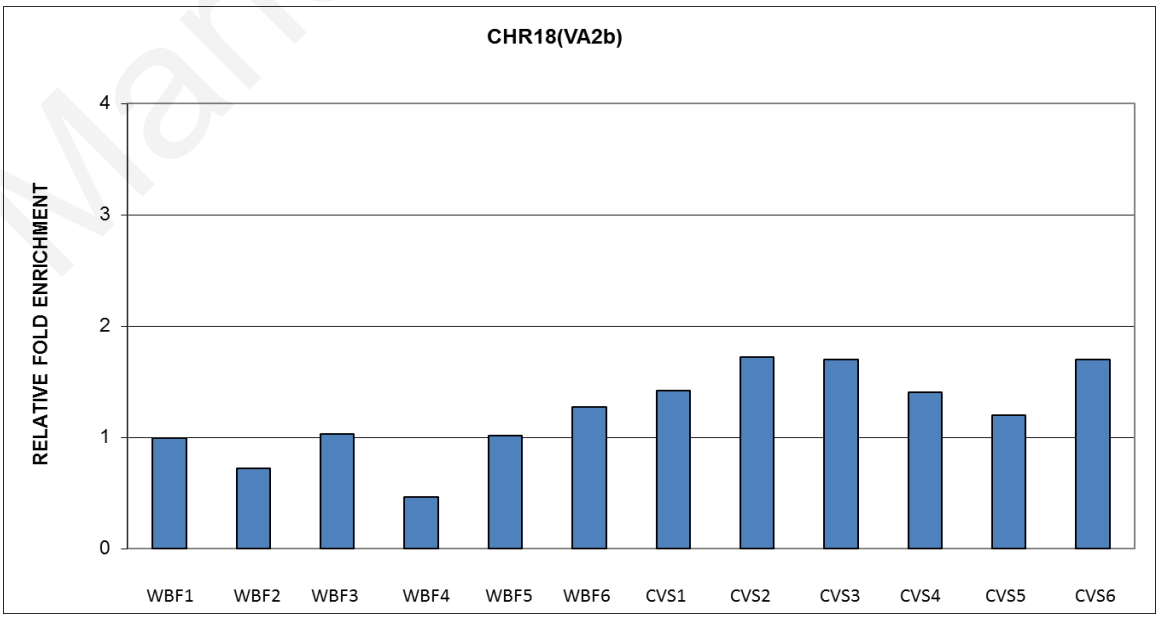


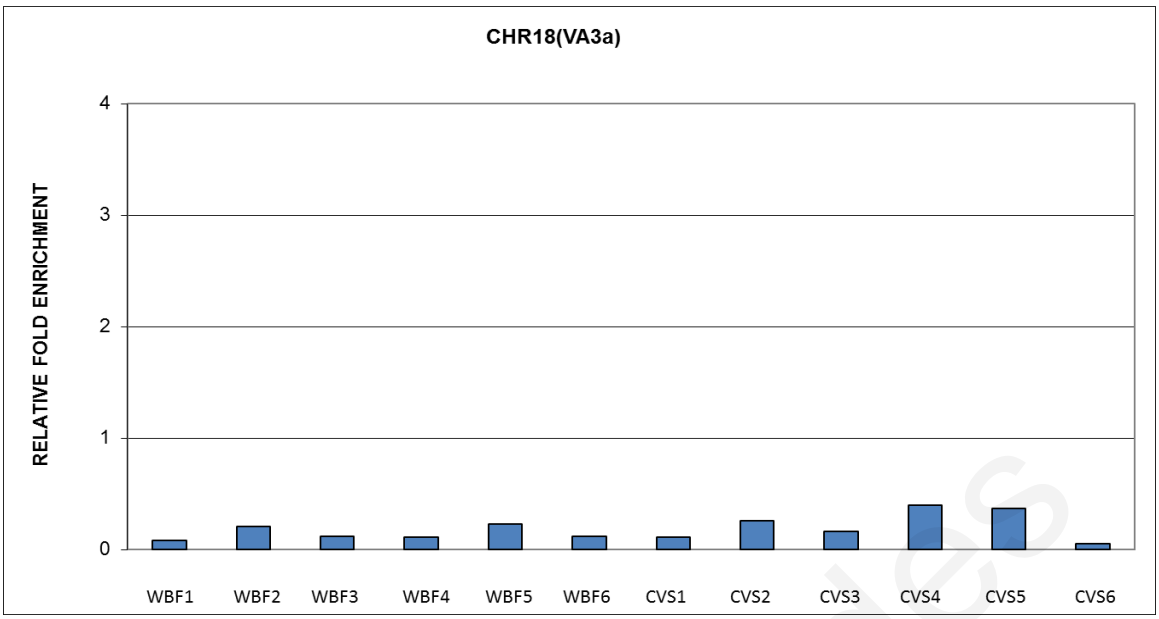


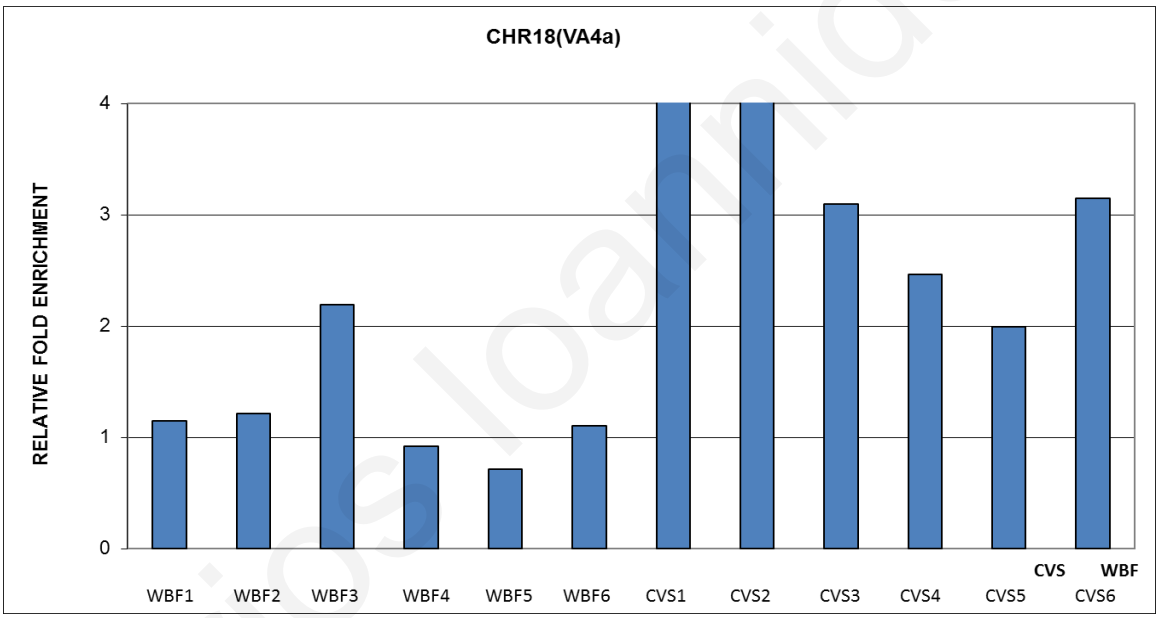


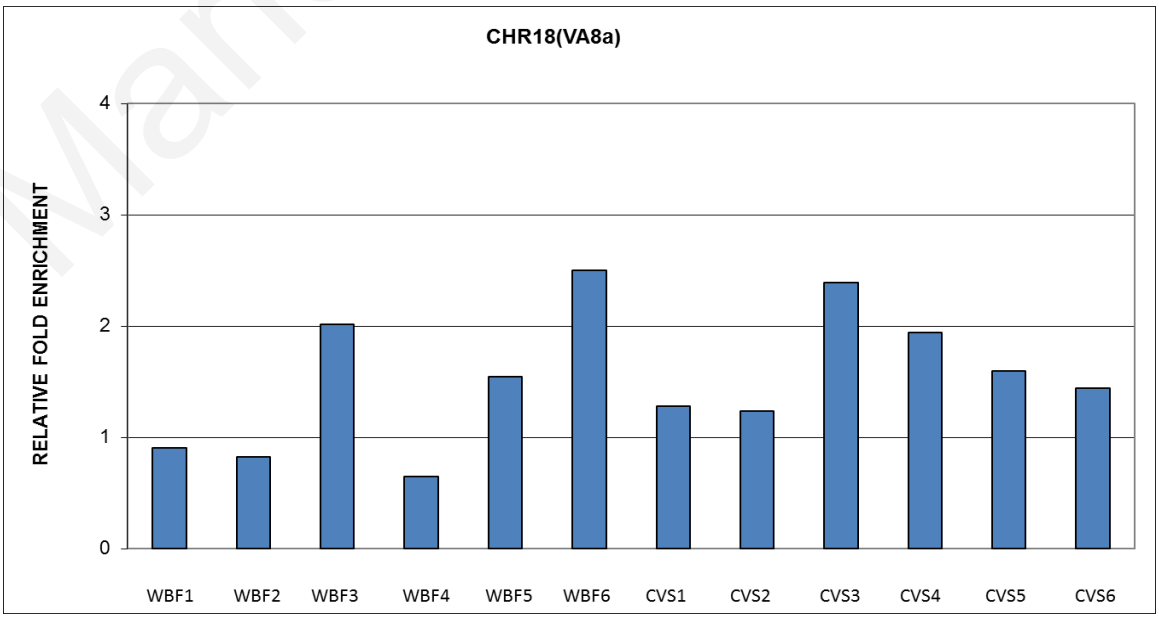


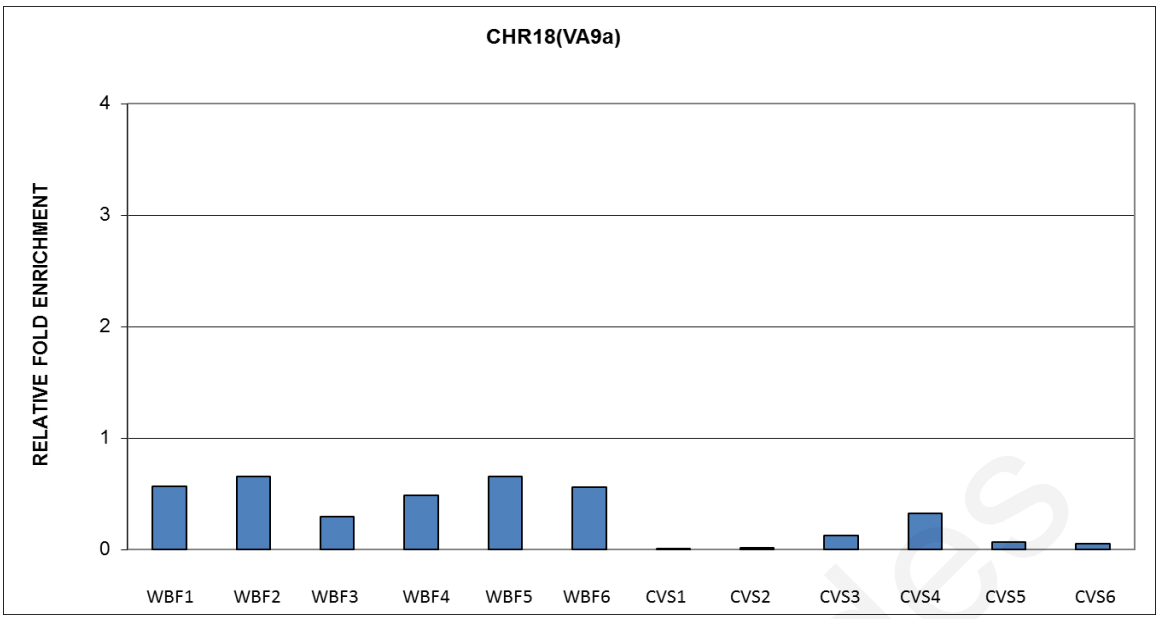


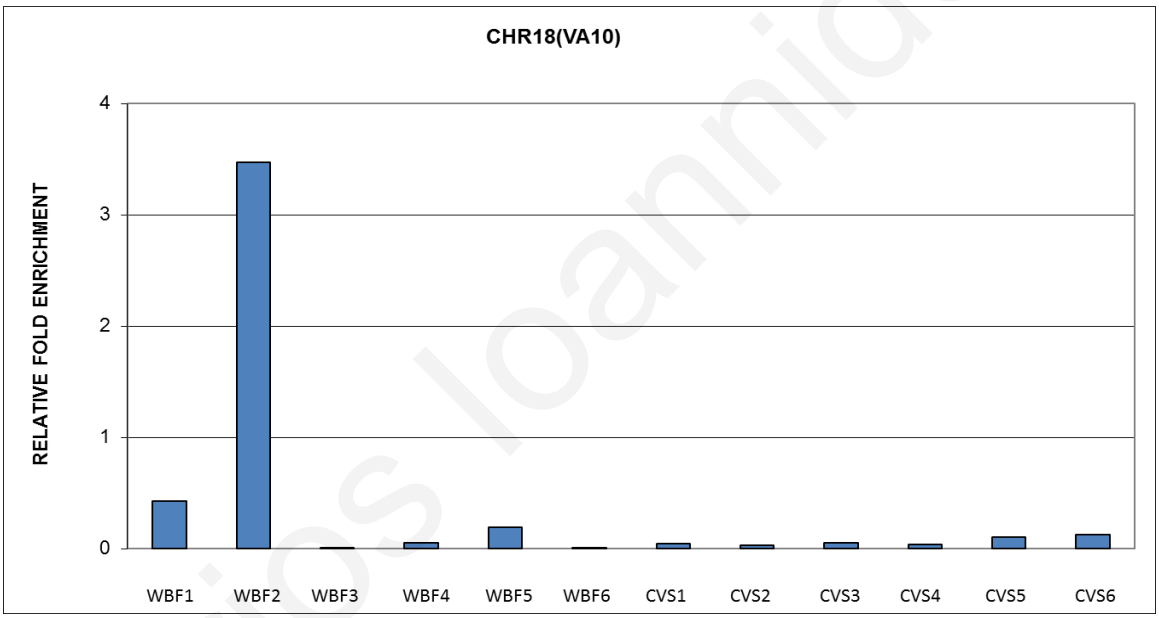


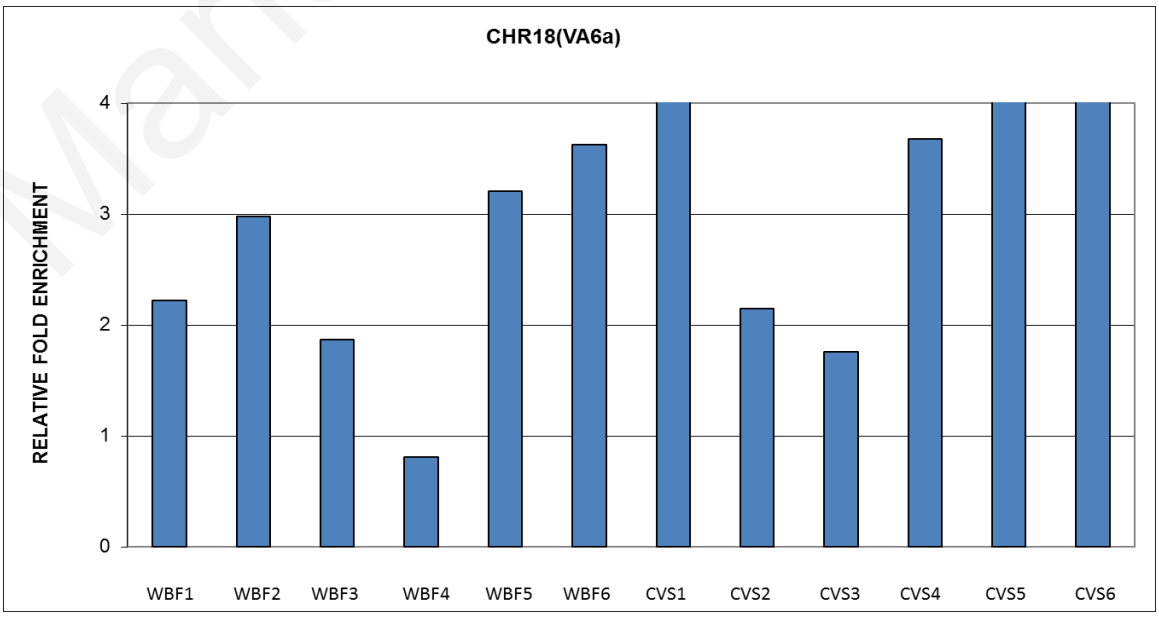


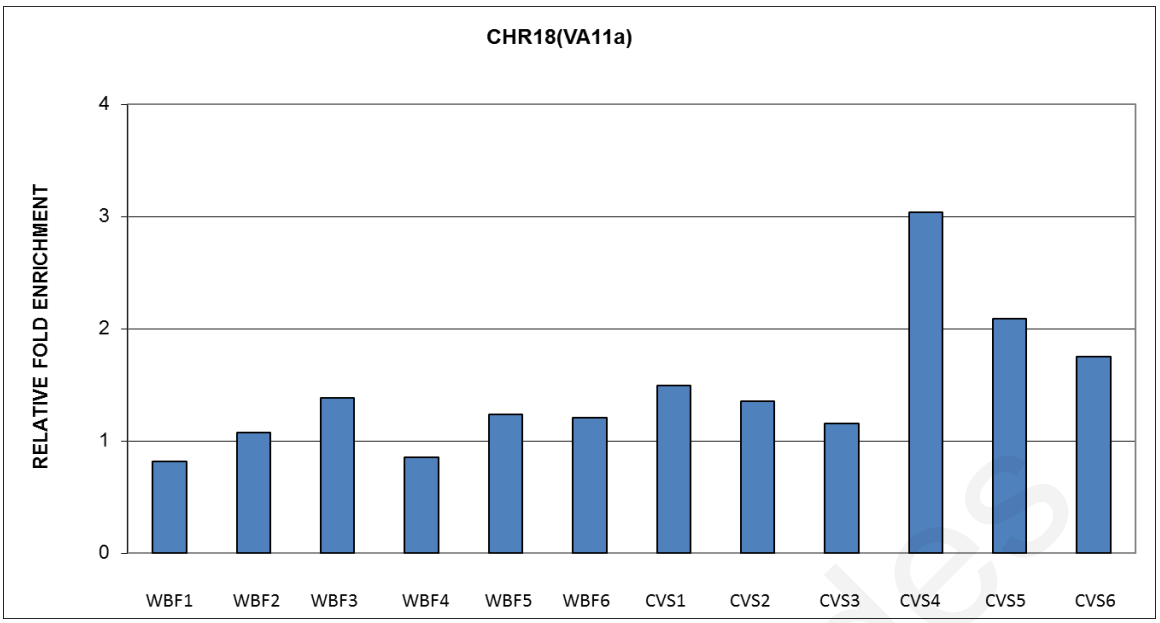


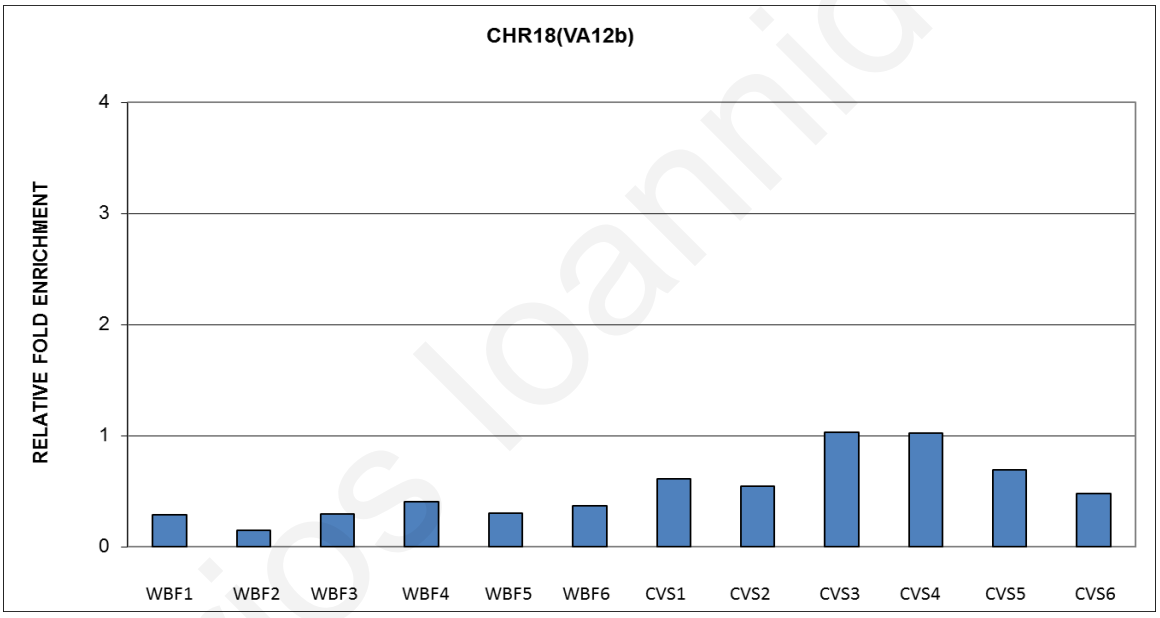


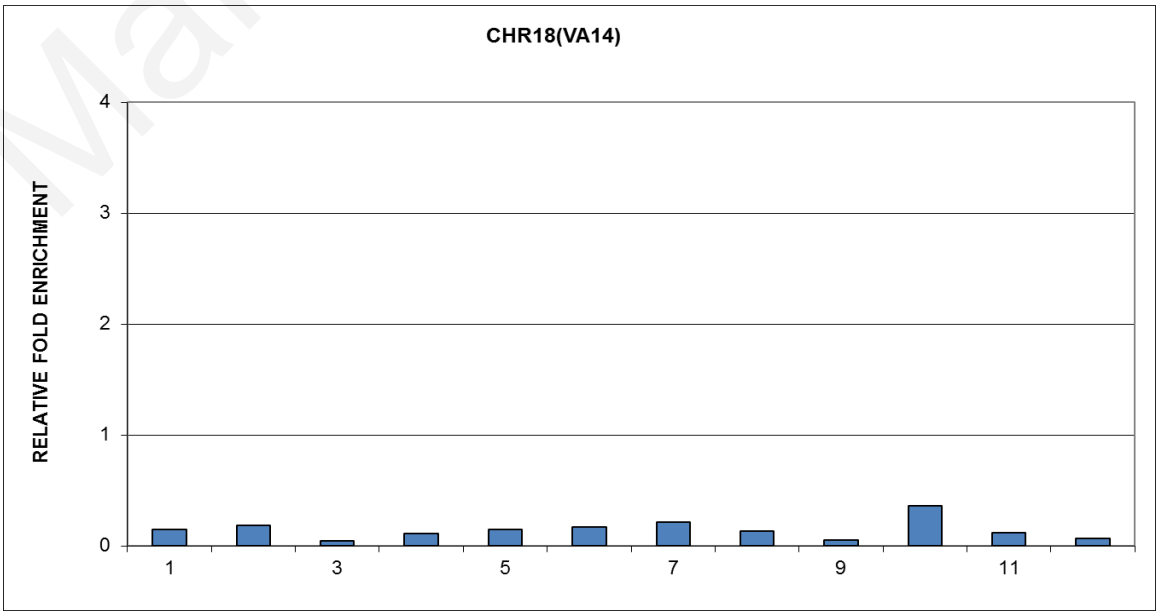


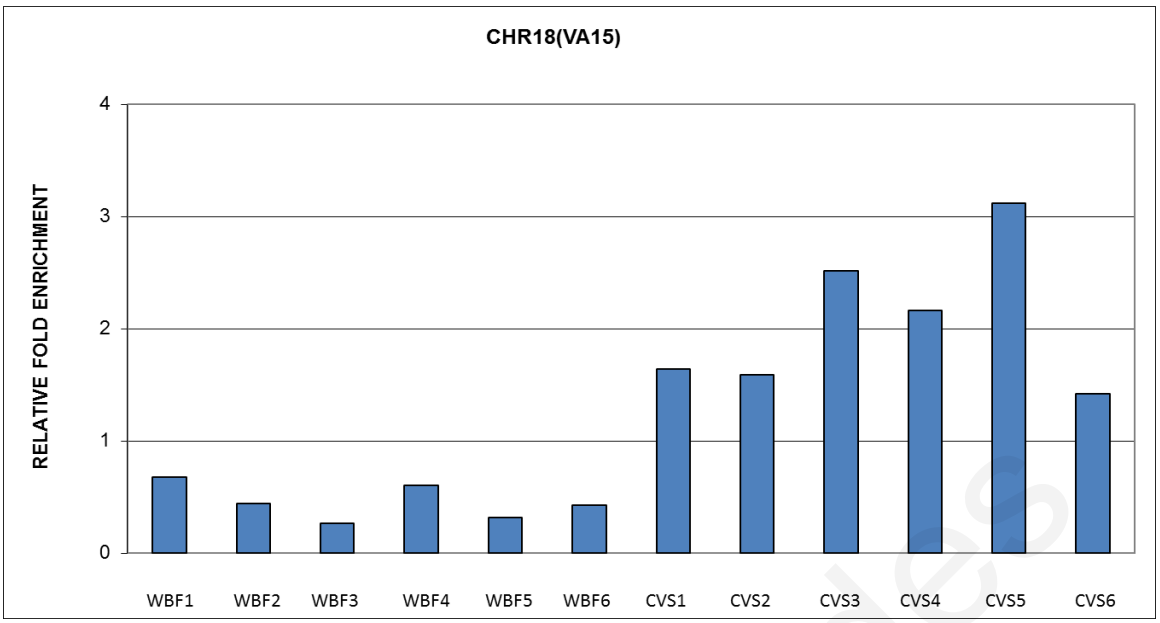


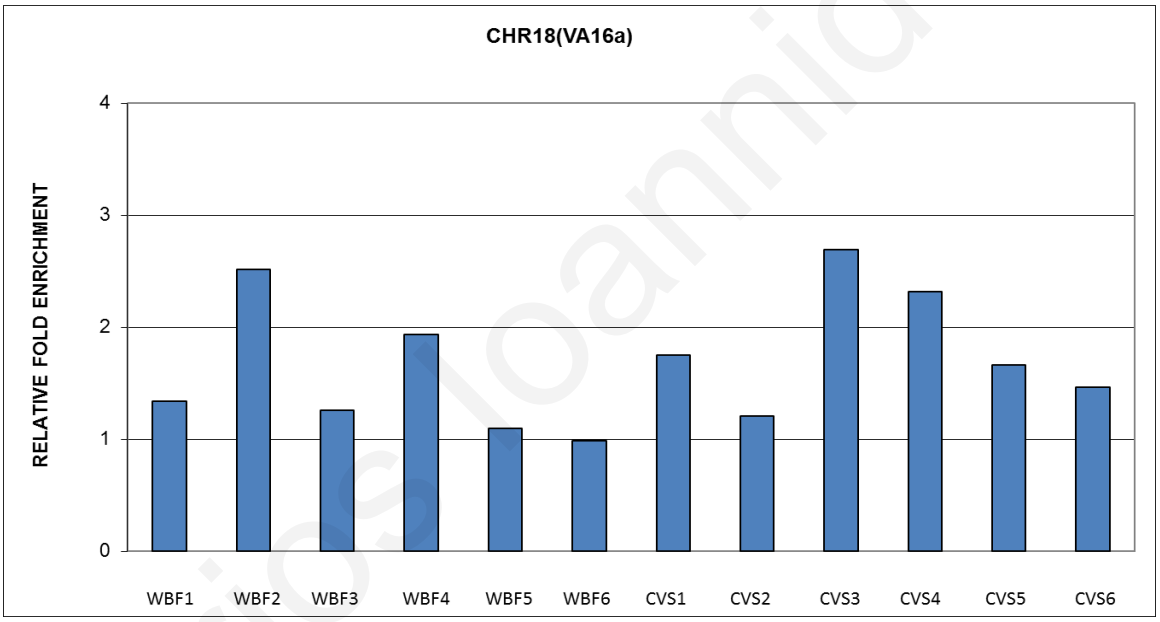


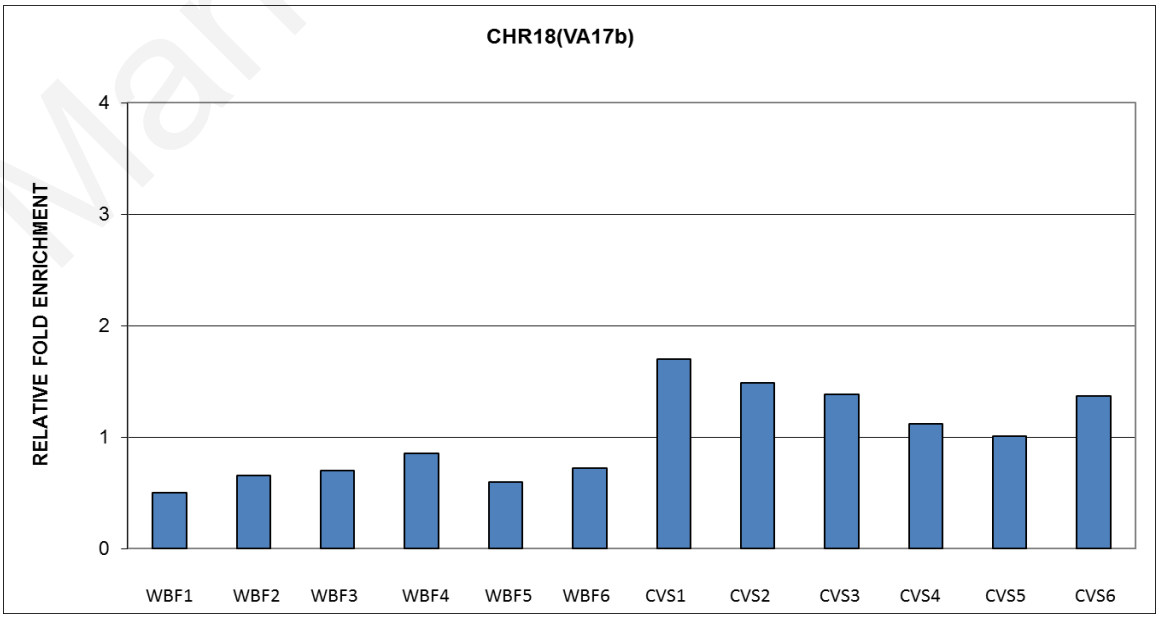


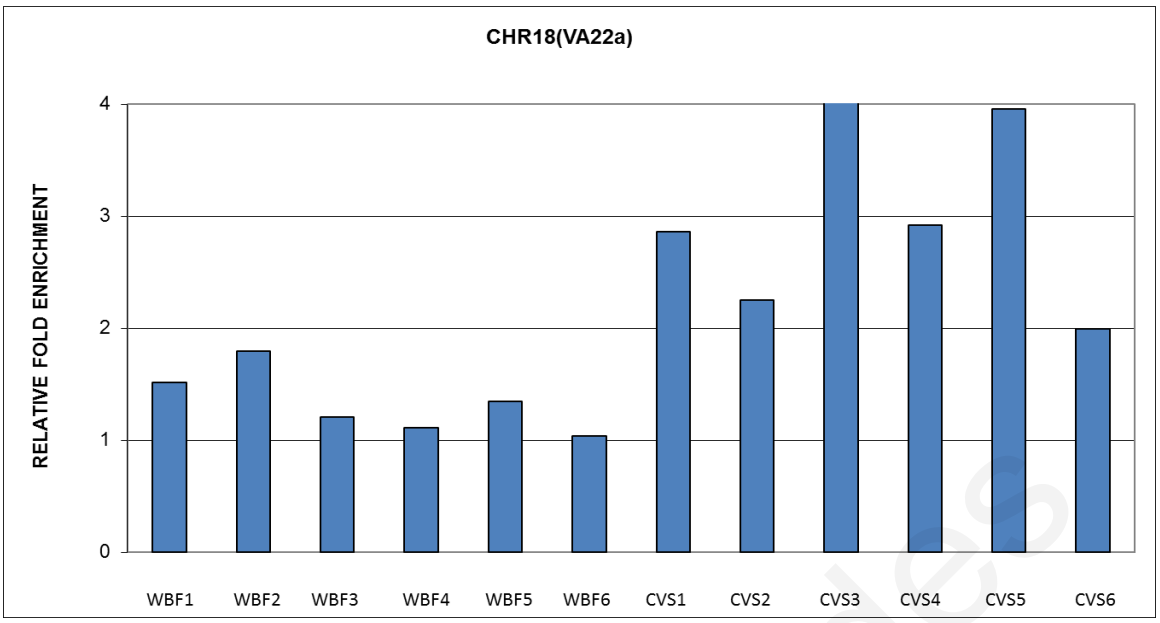


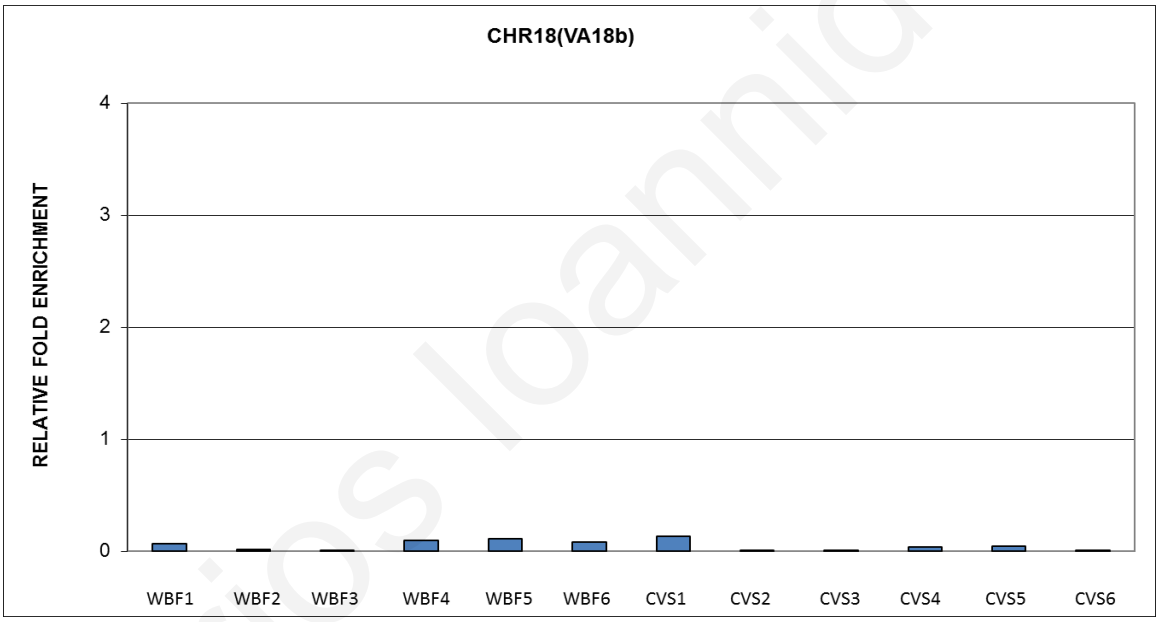


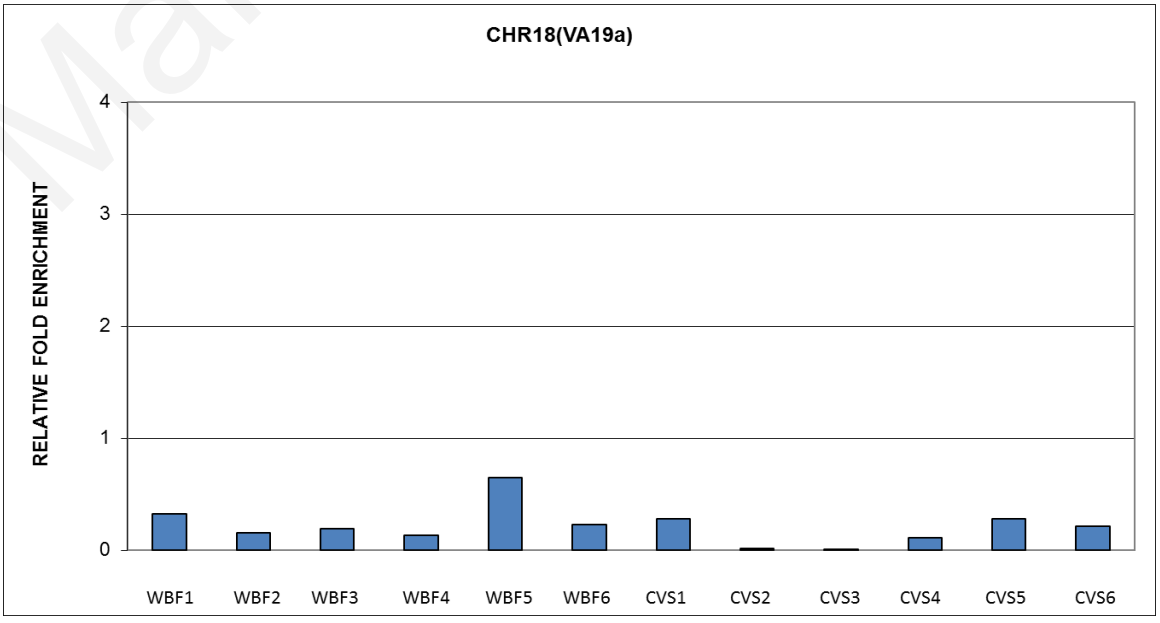


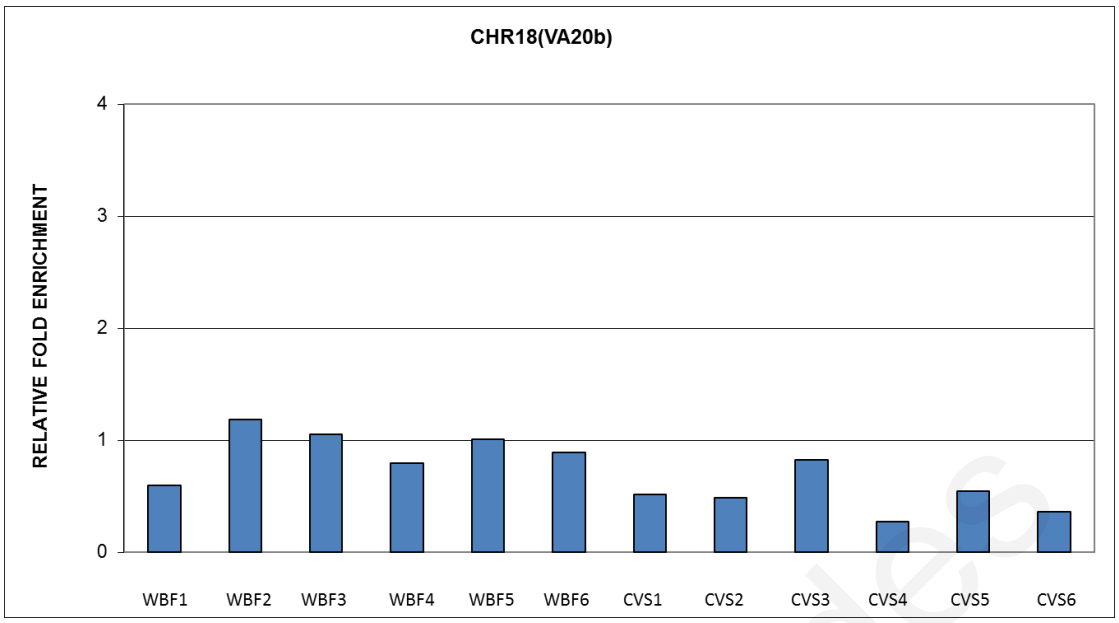


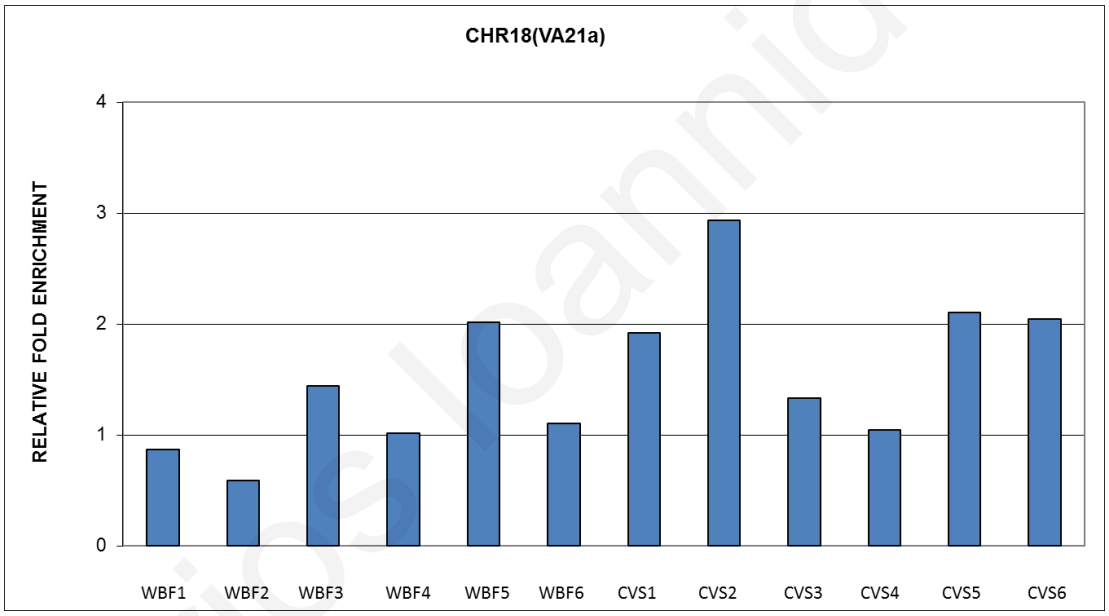


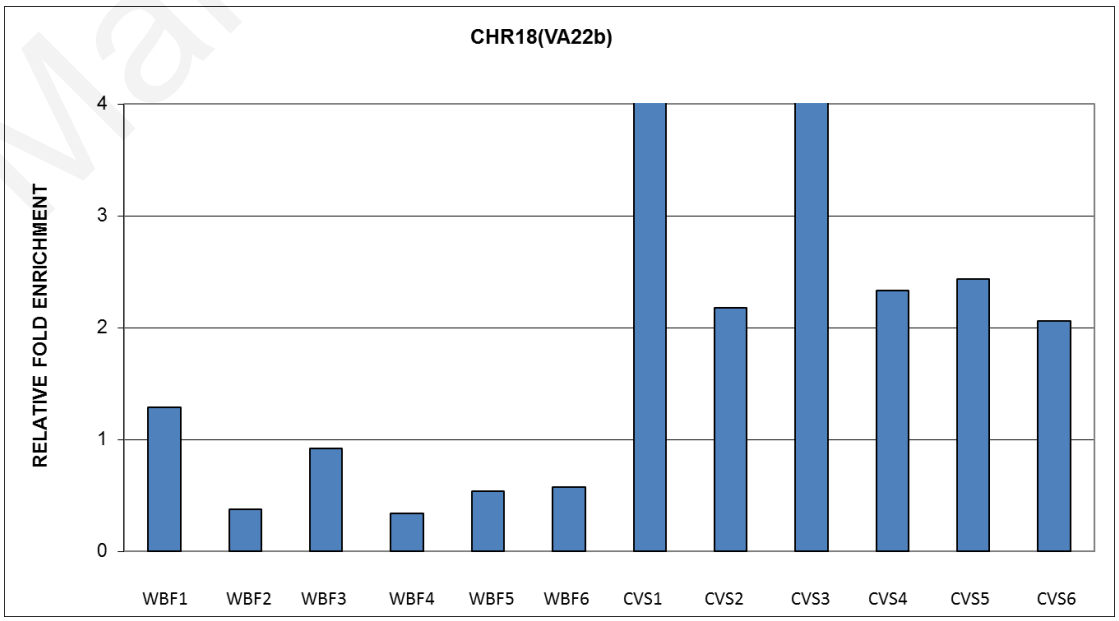


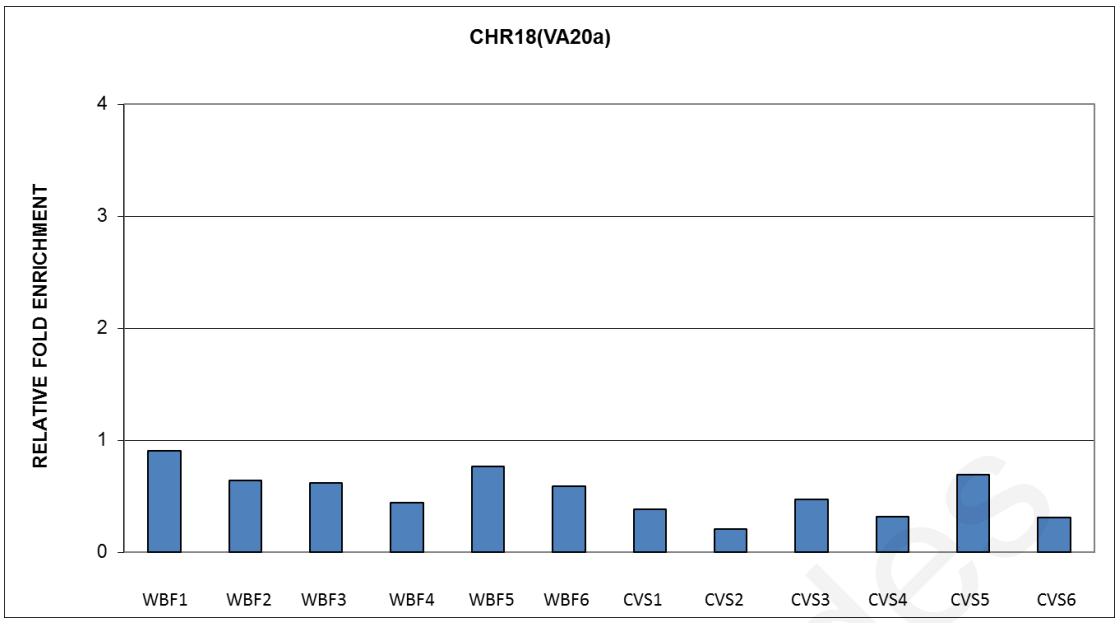


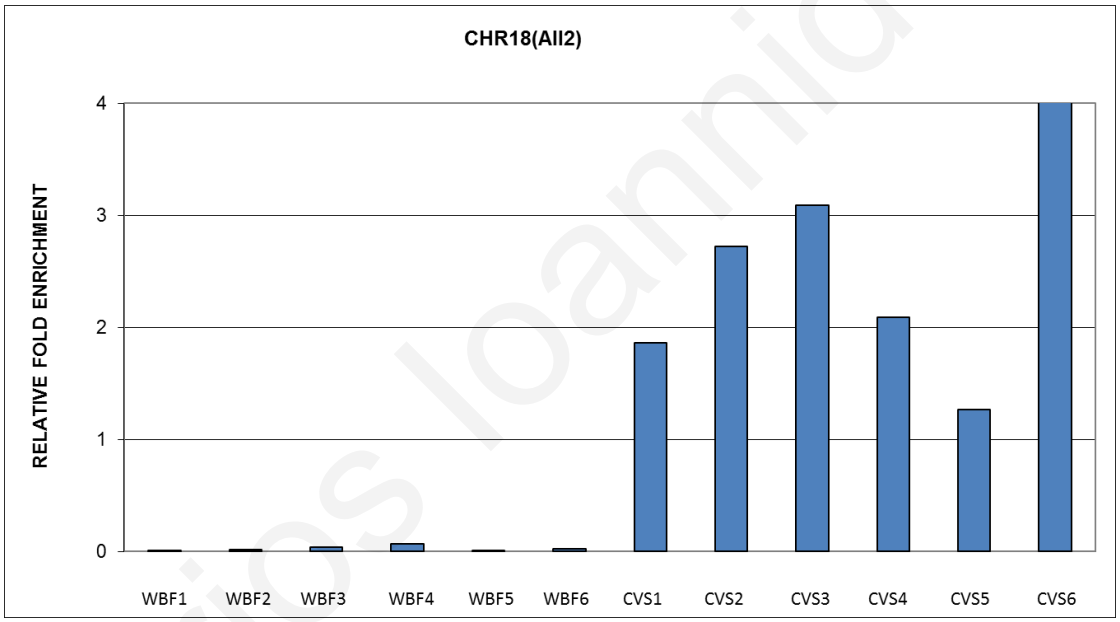


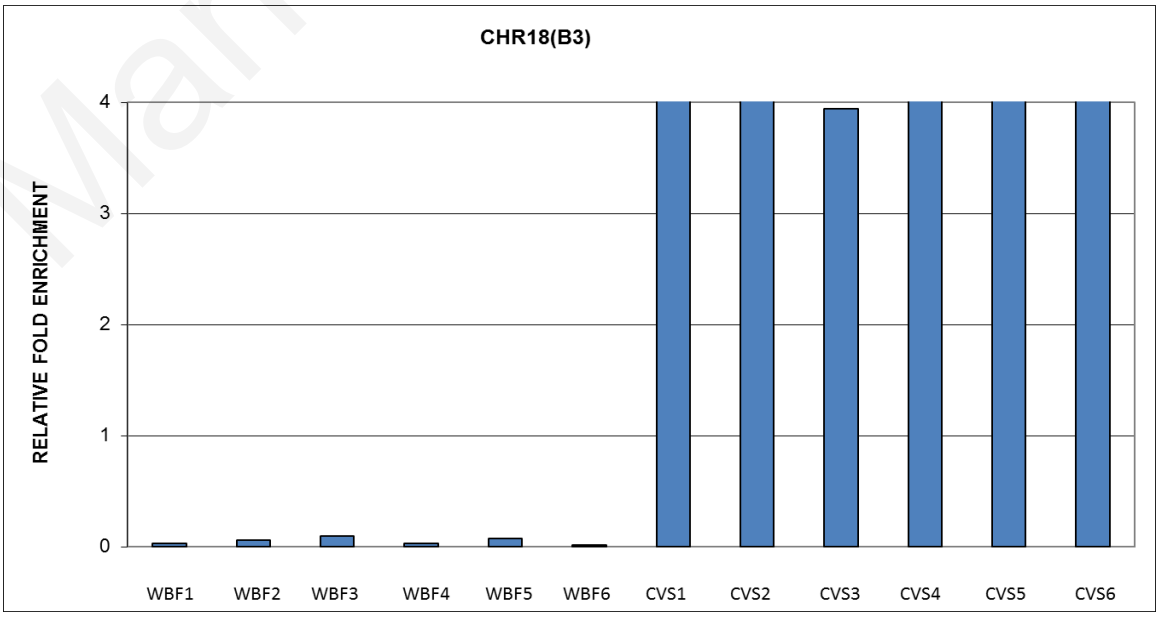


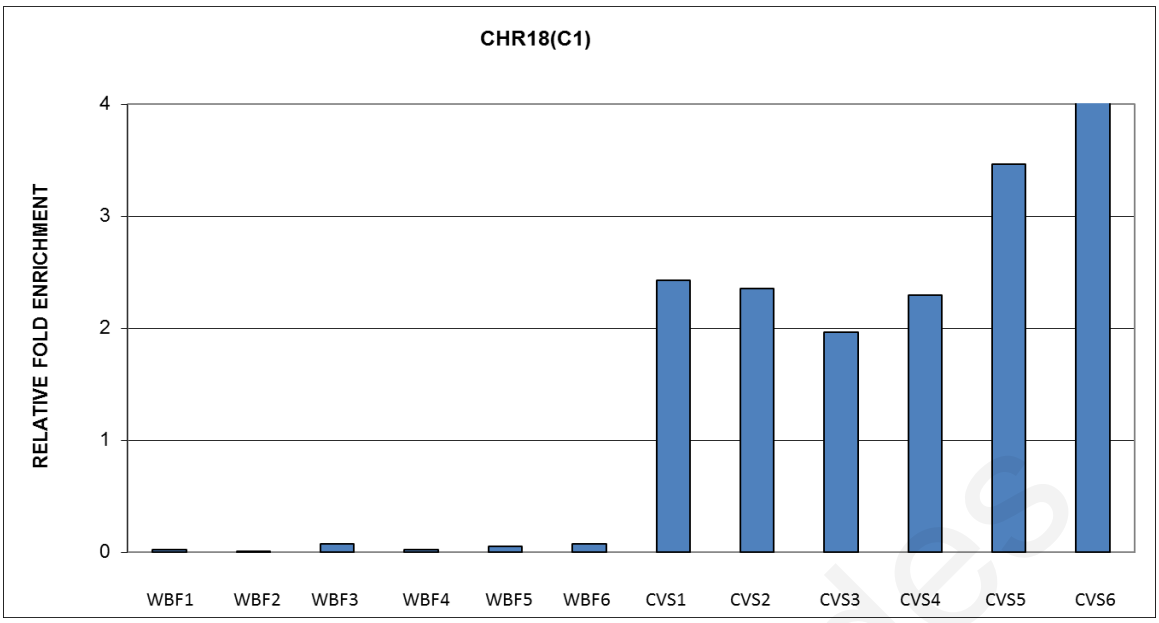




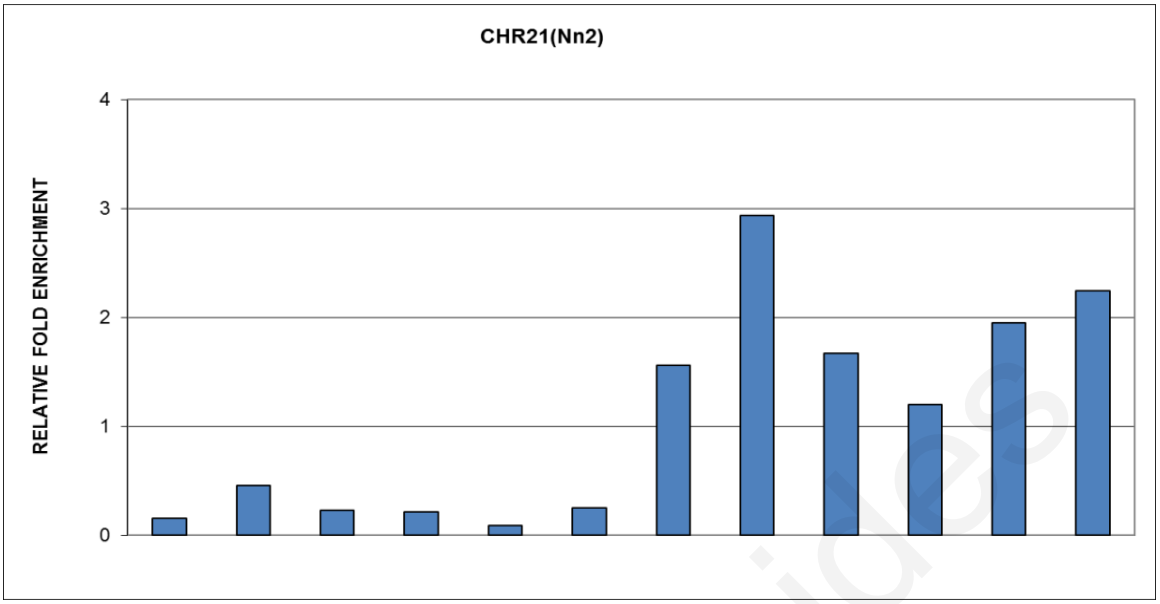


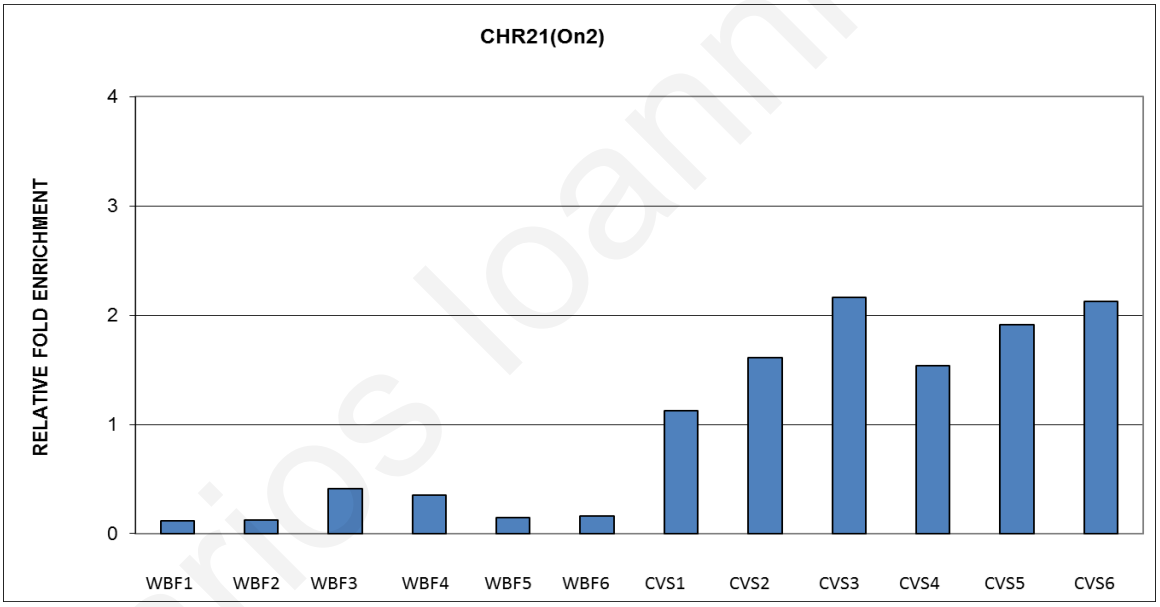


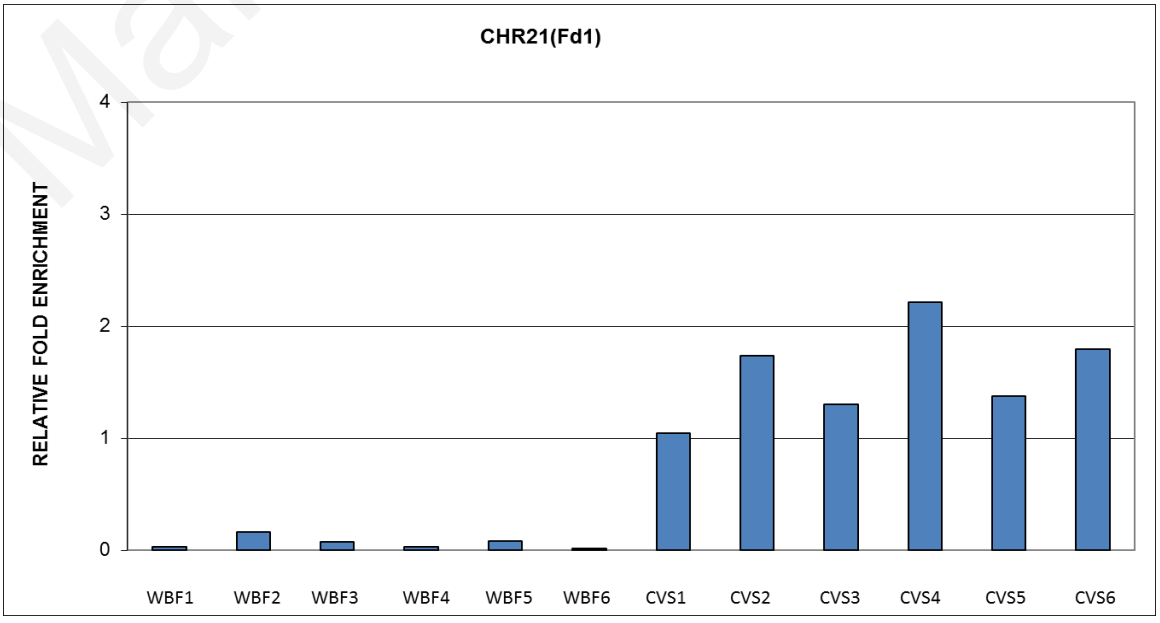


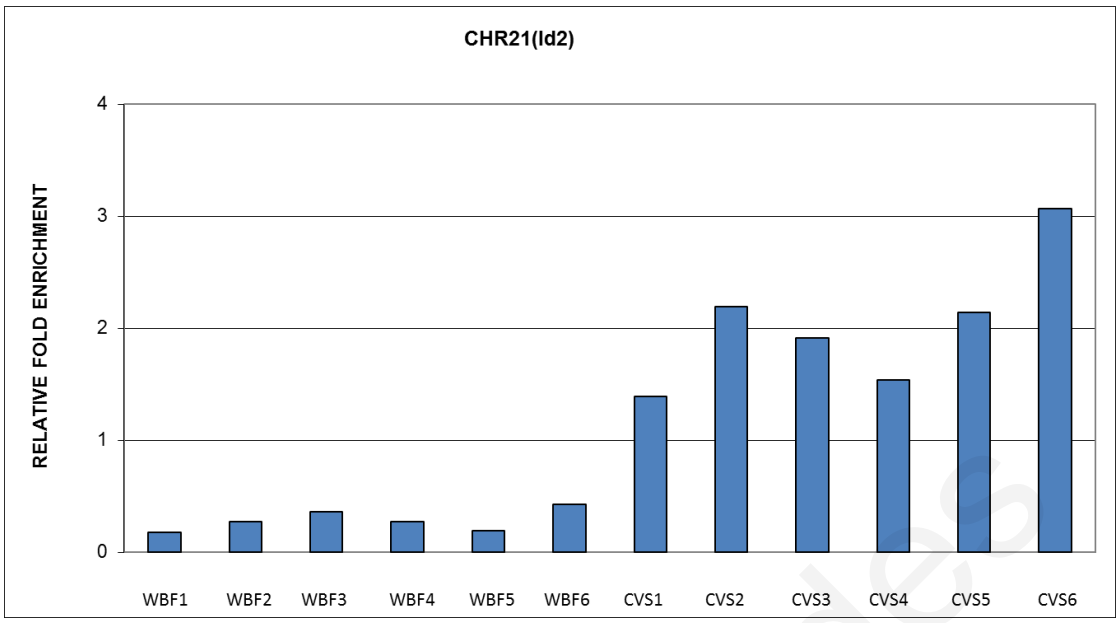


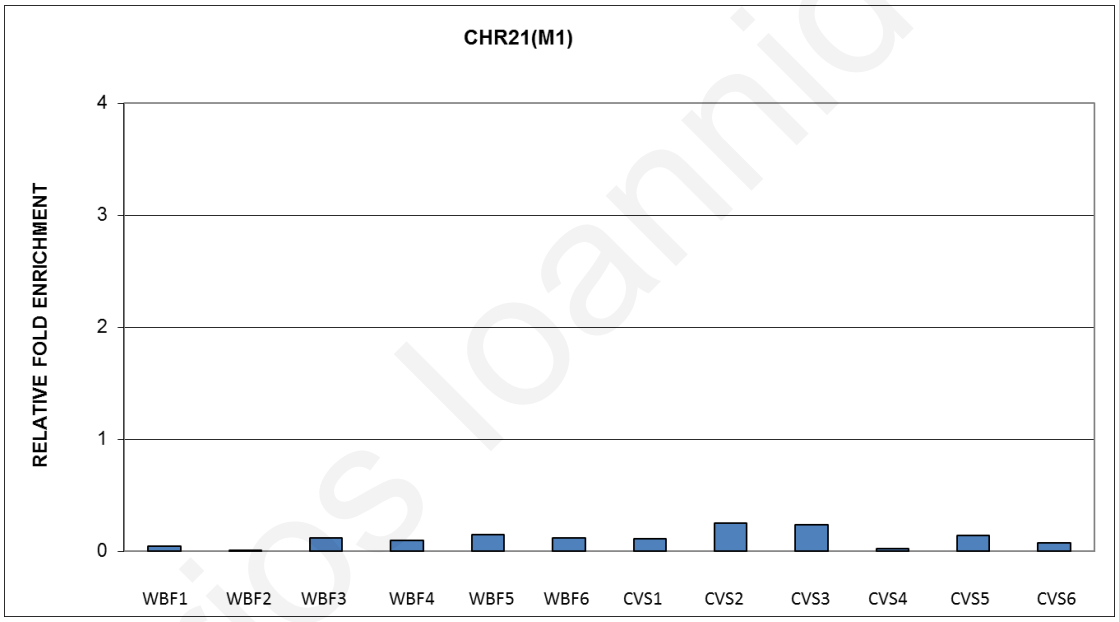
C. Chromosome 21

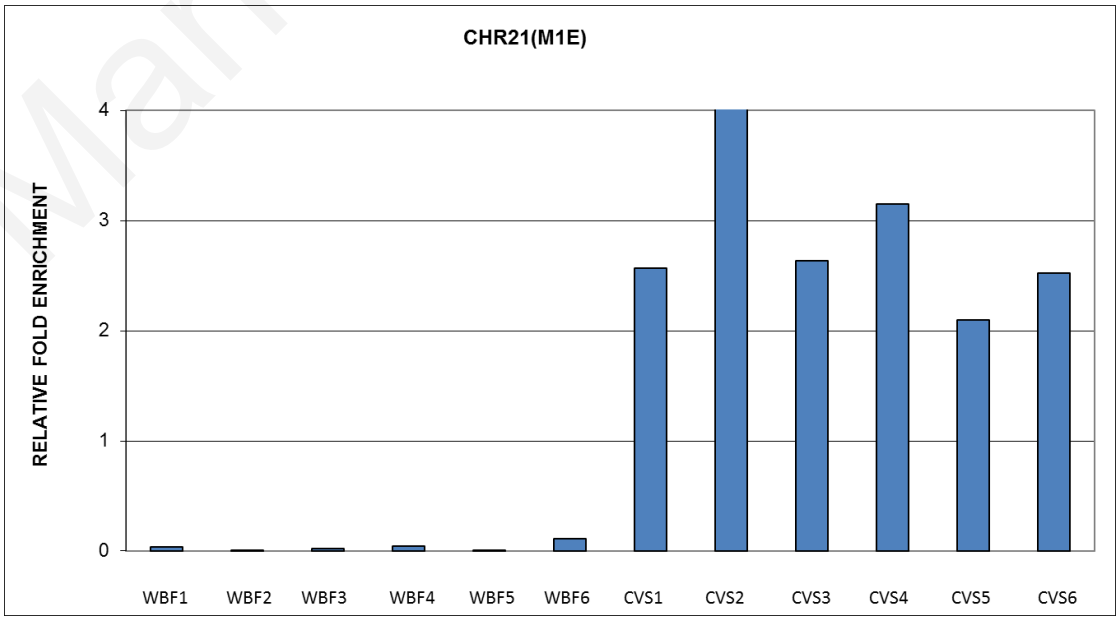


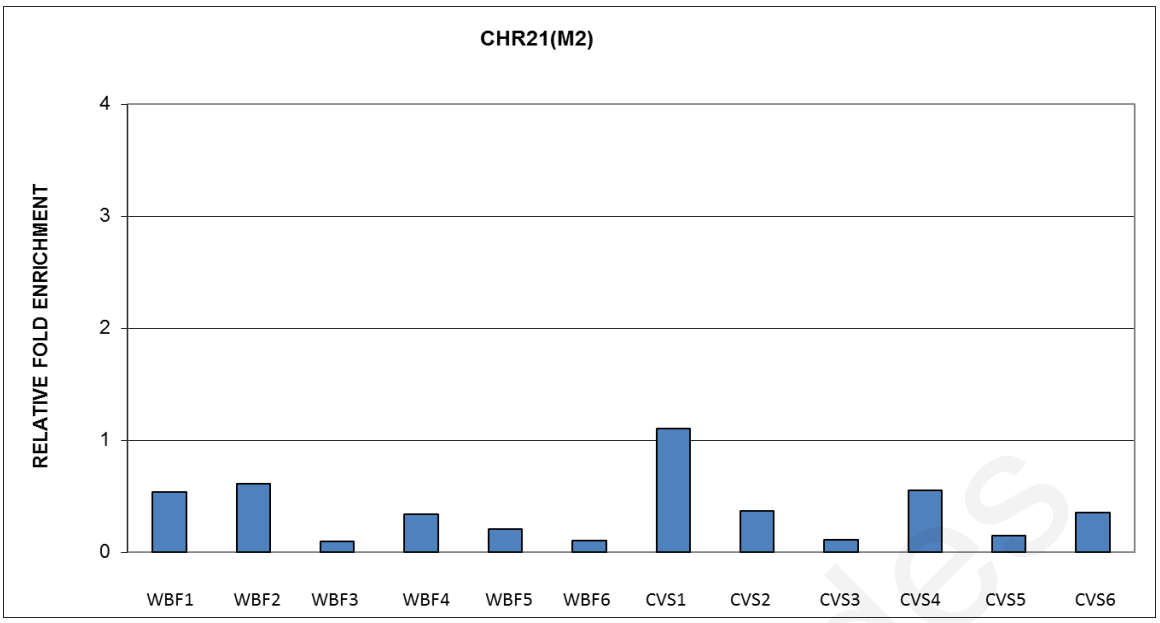


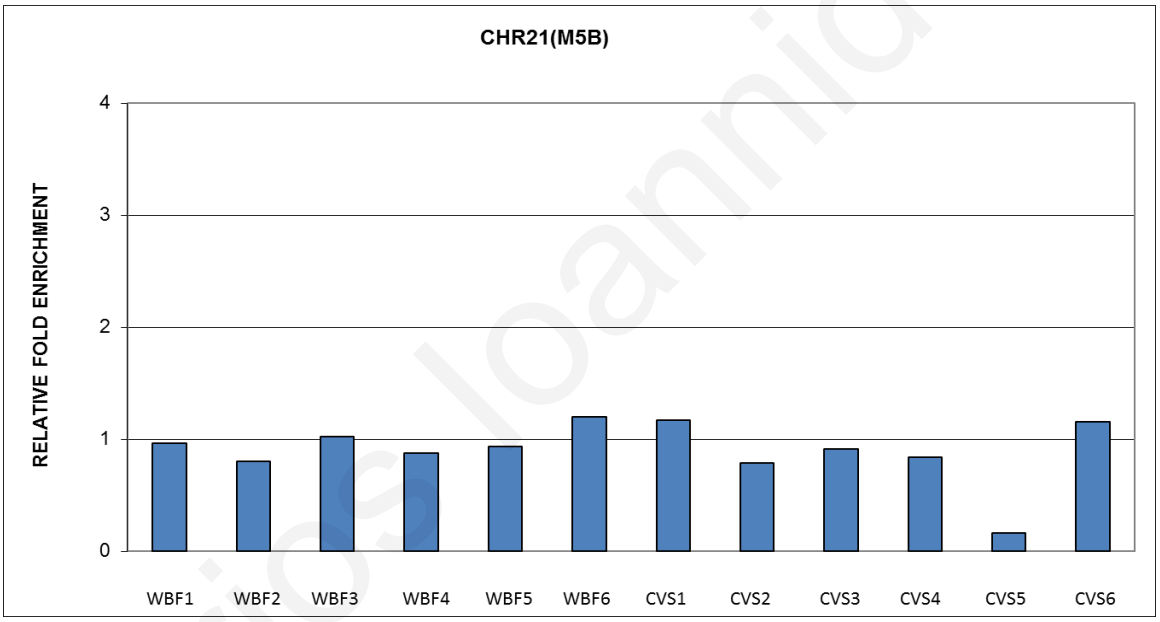


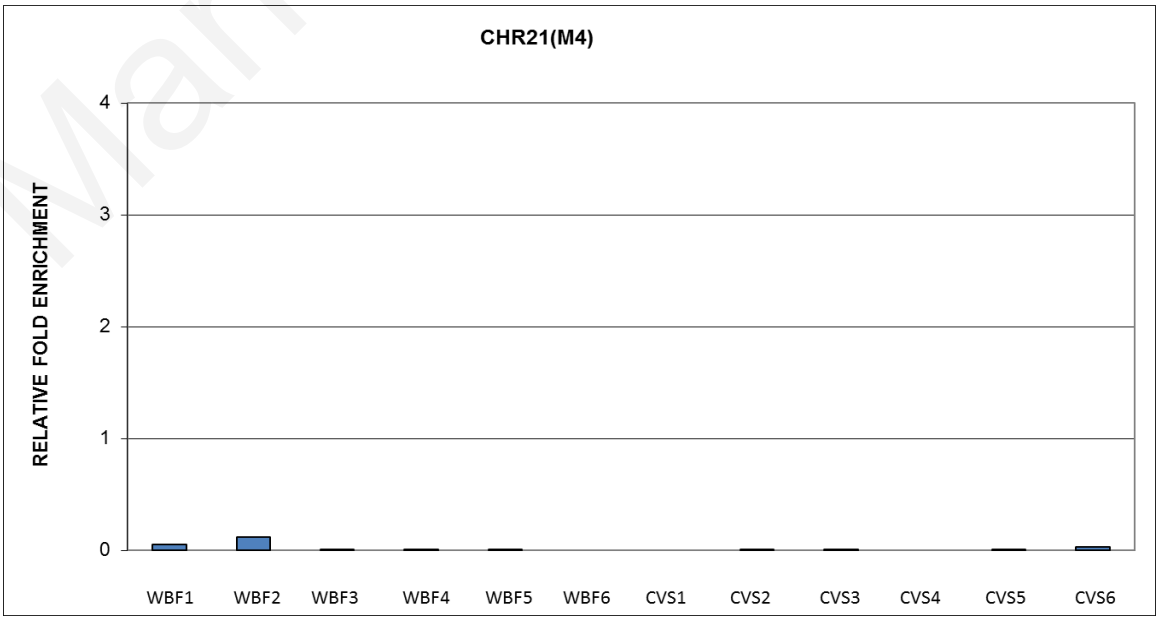


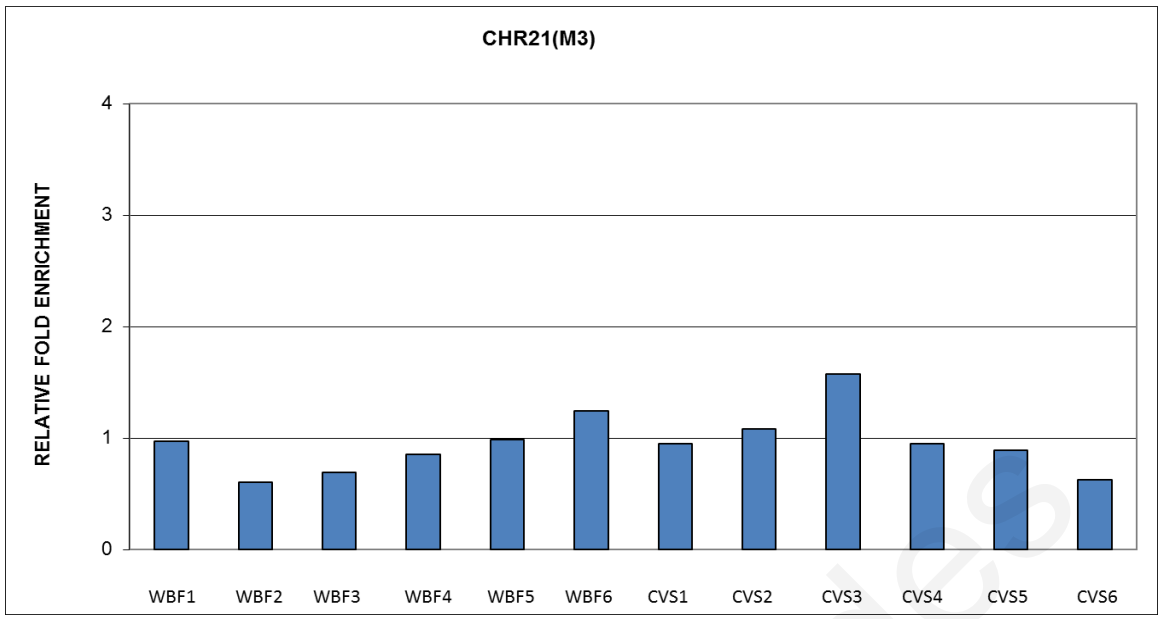


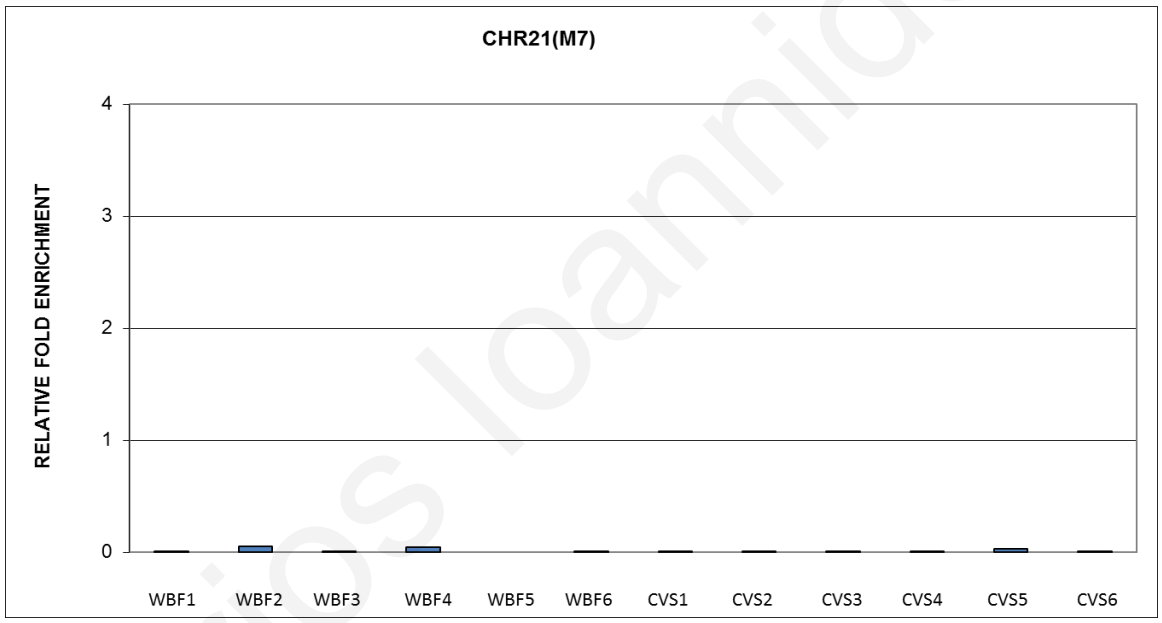


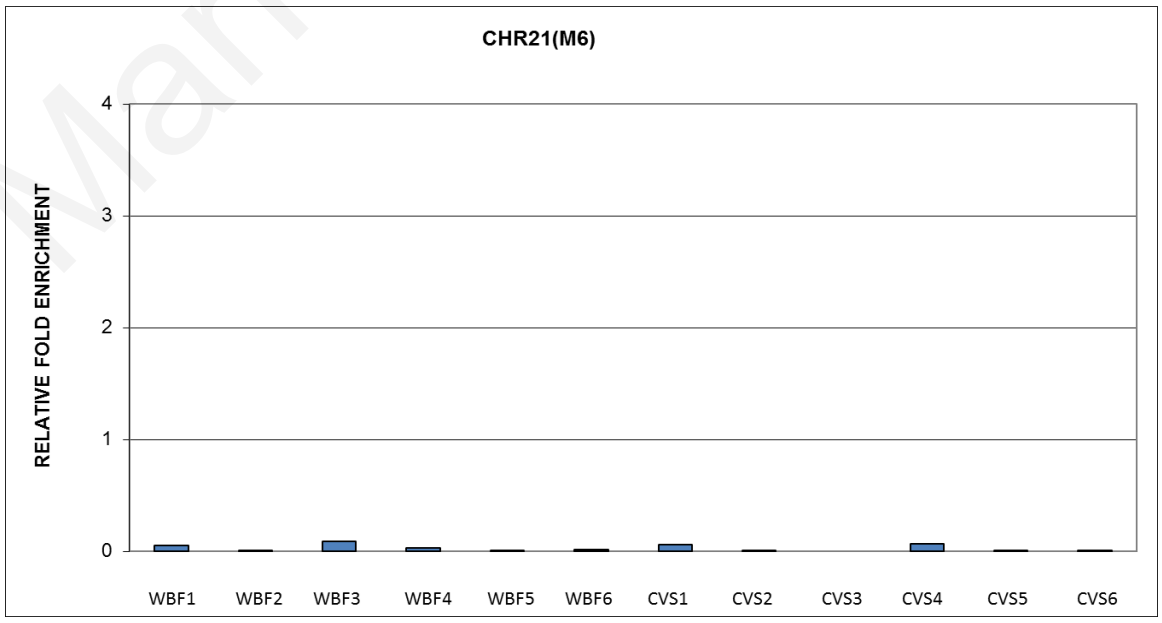


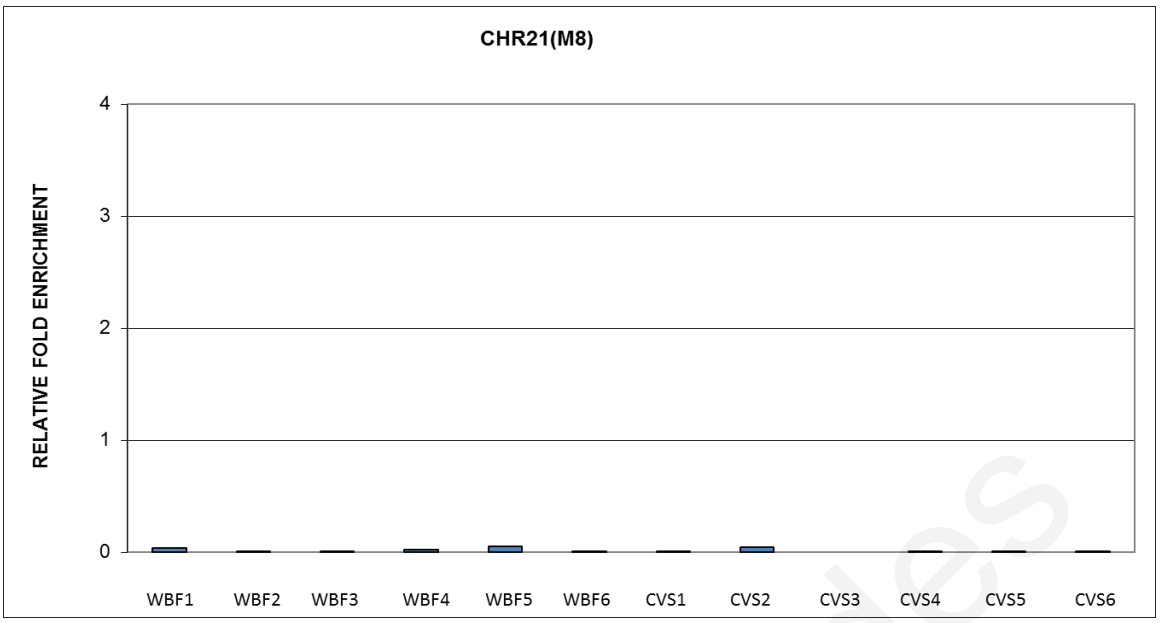


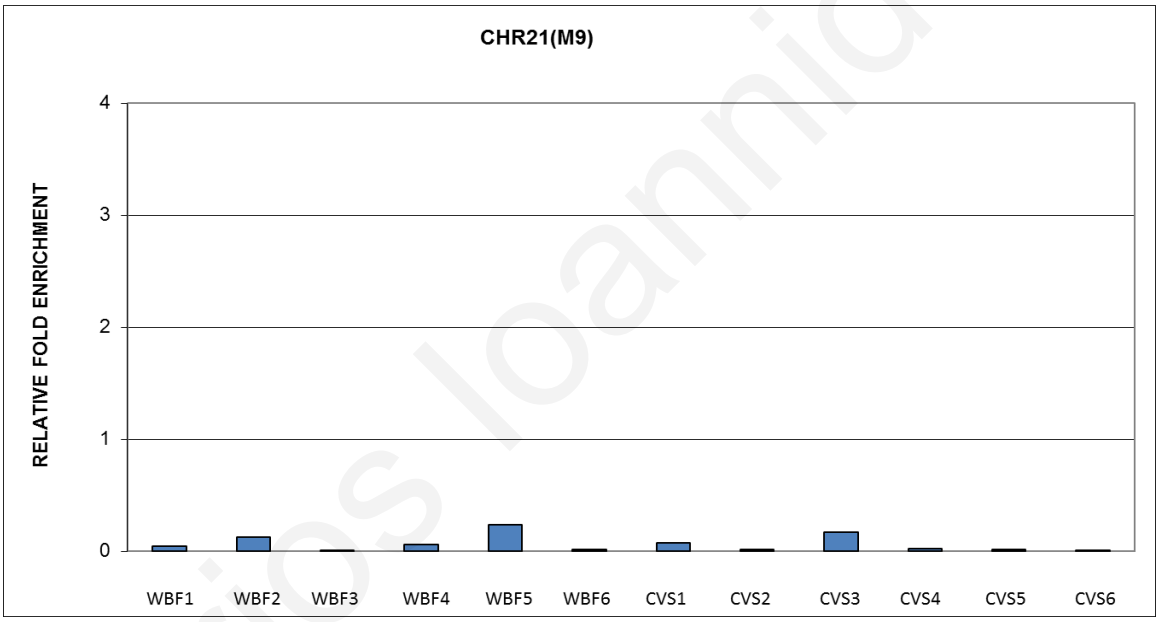


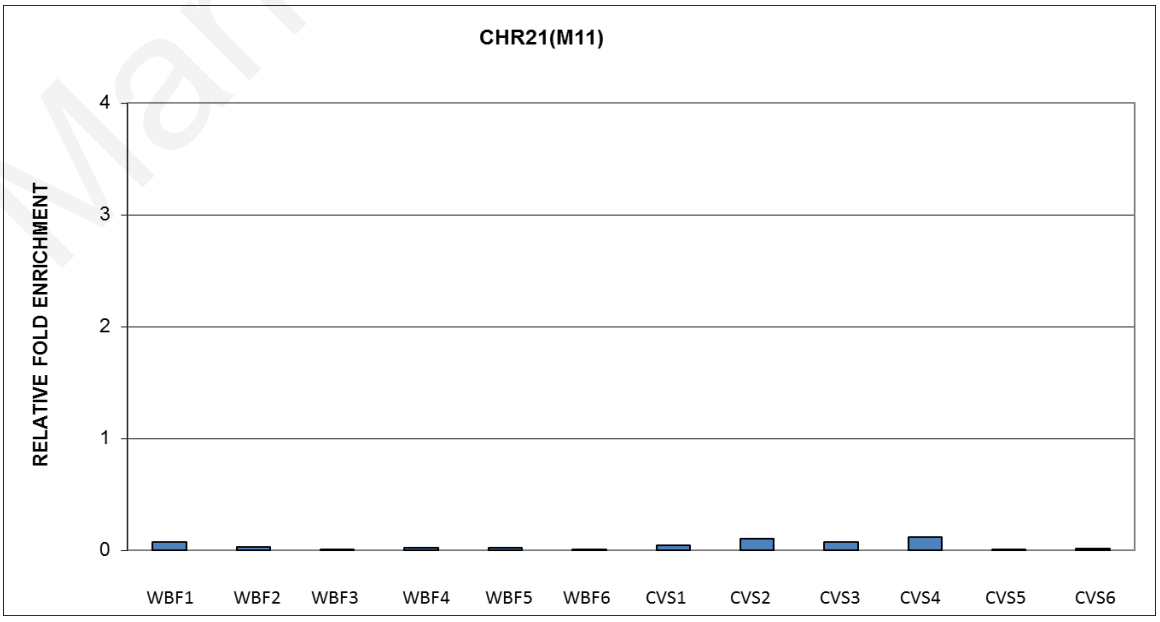


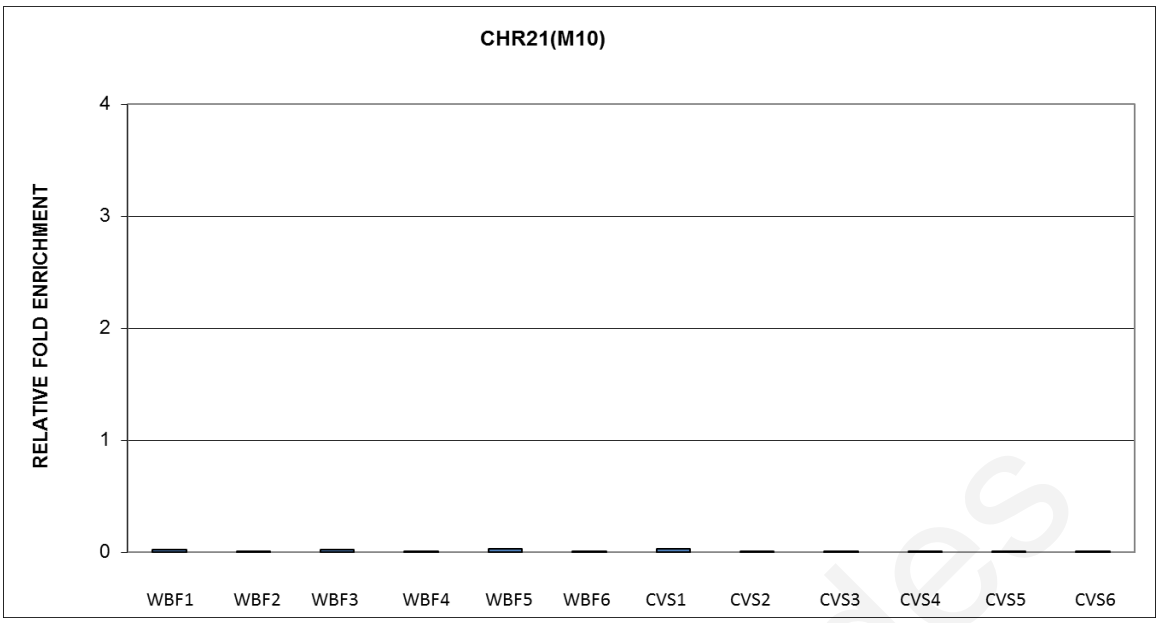


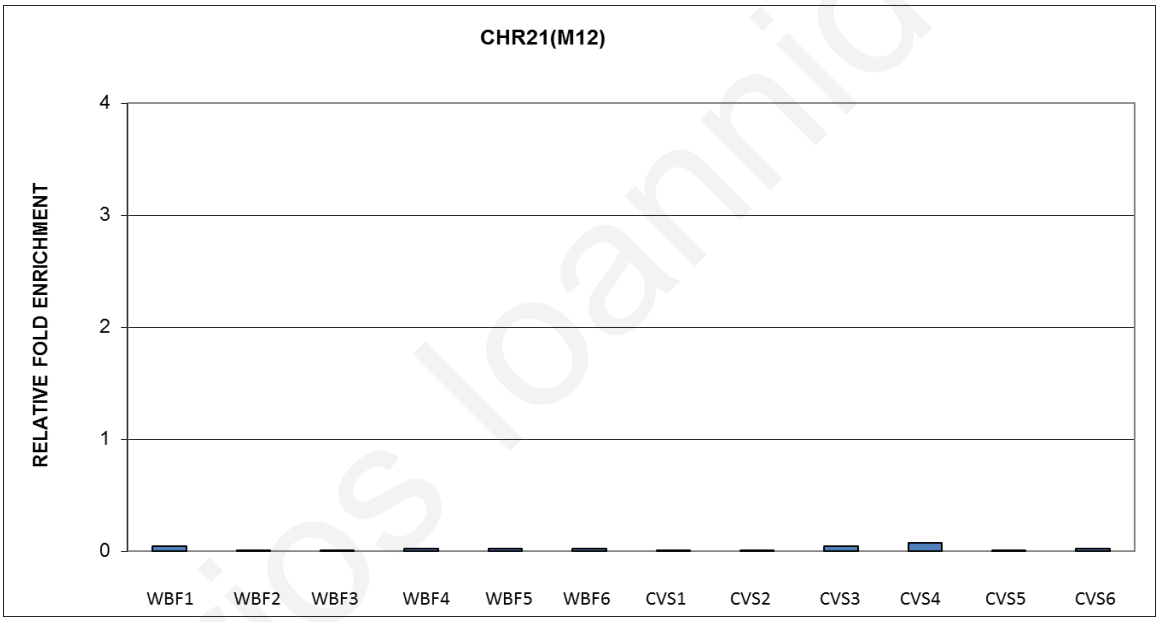


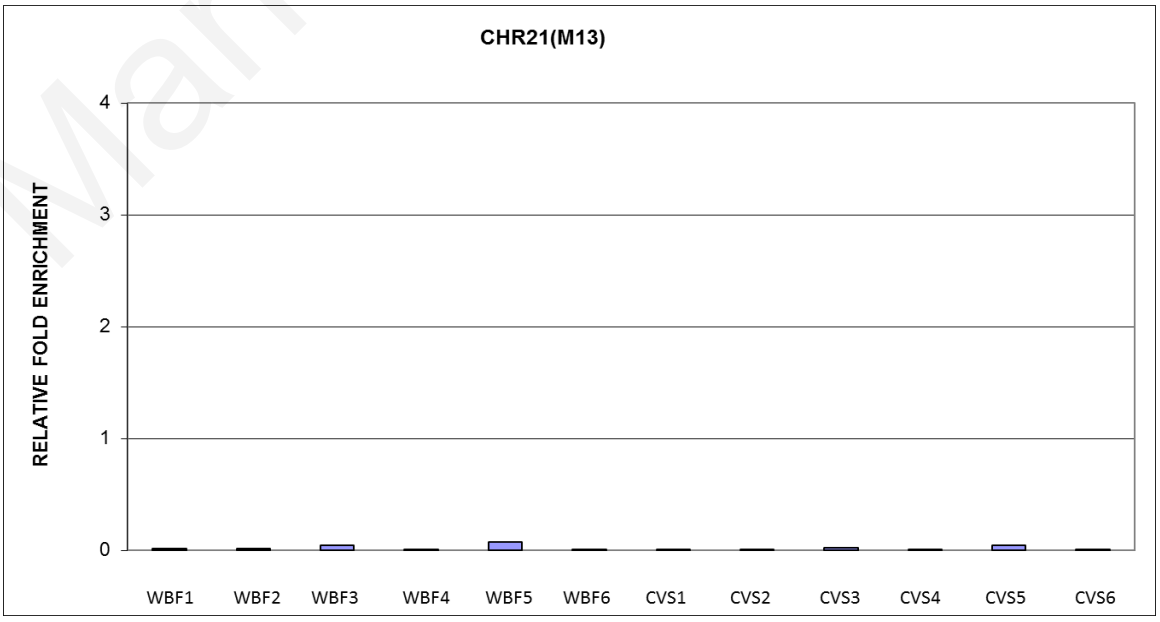


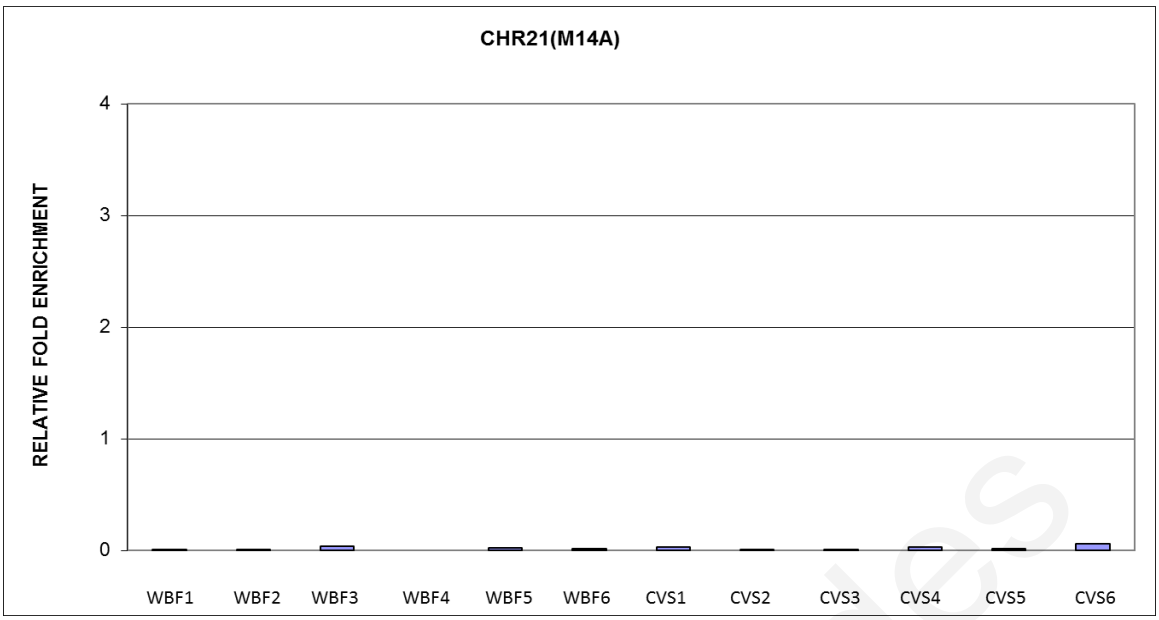


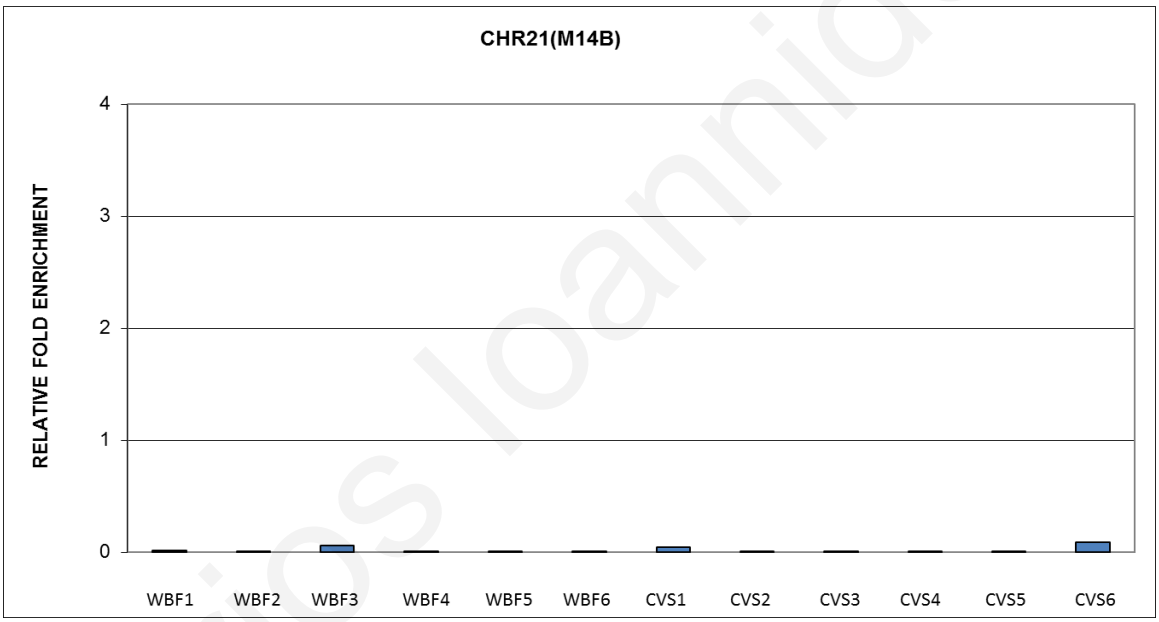


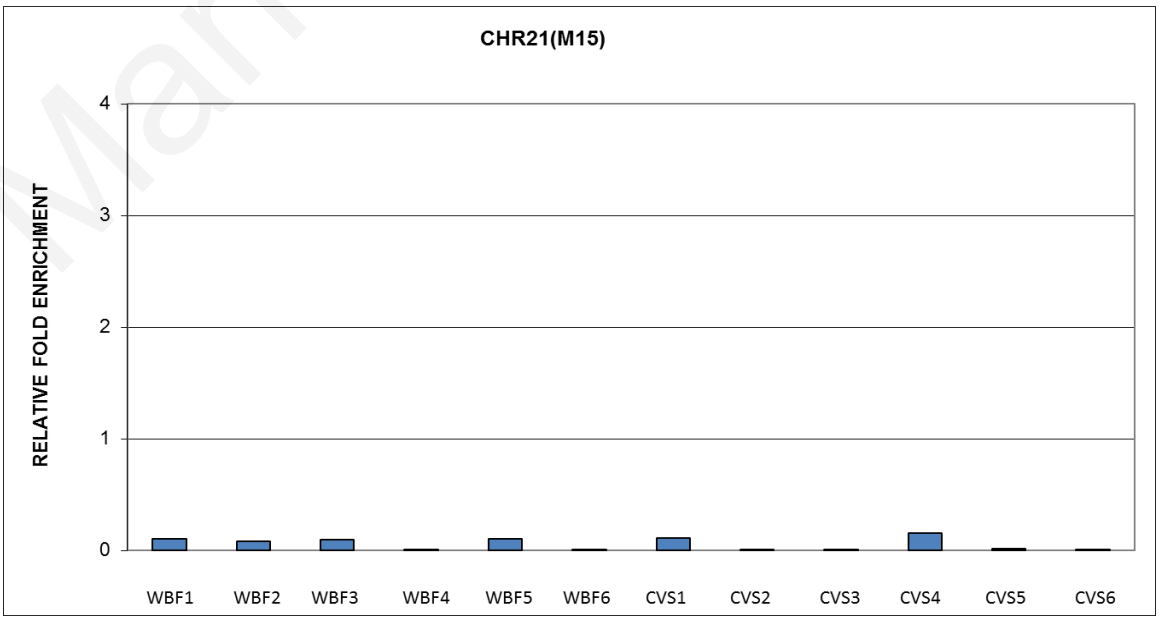


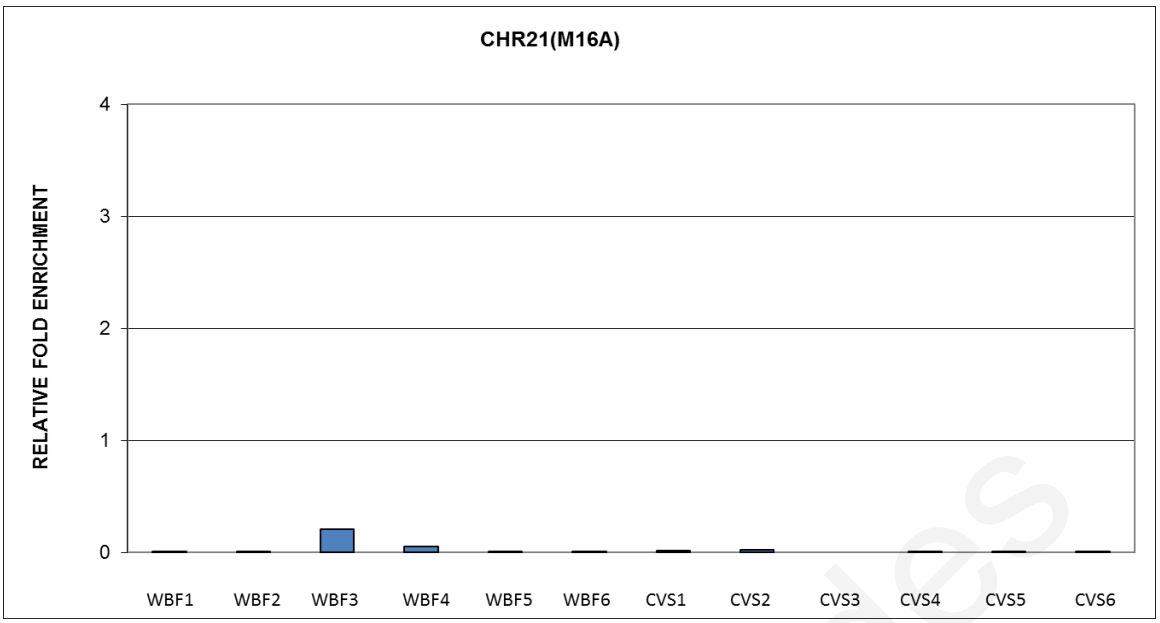


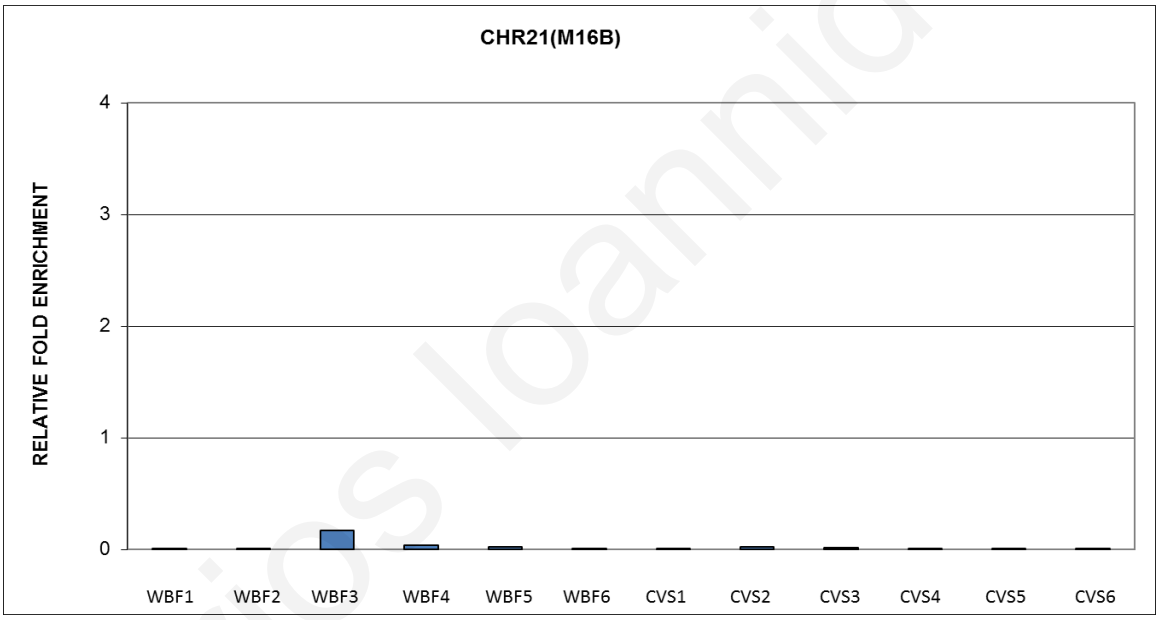


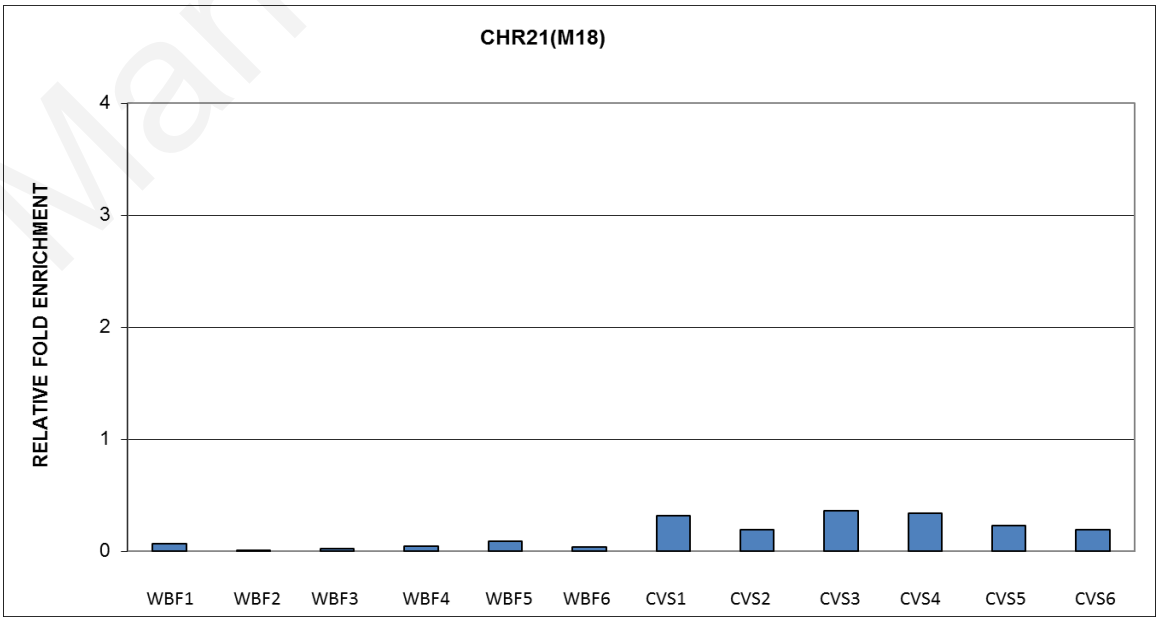


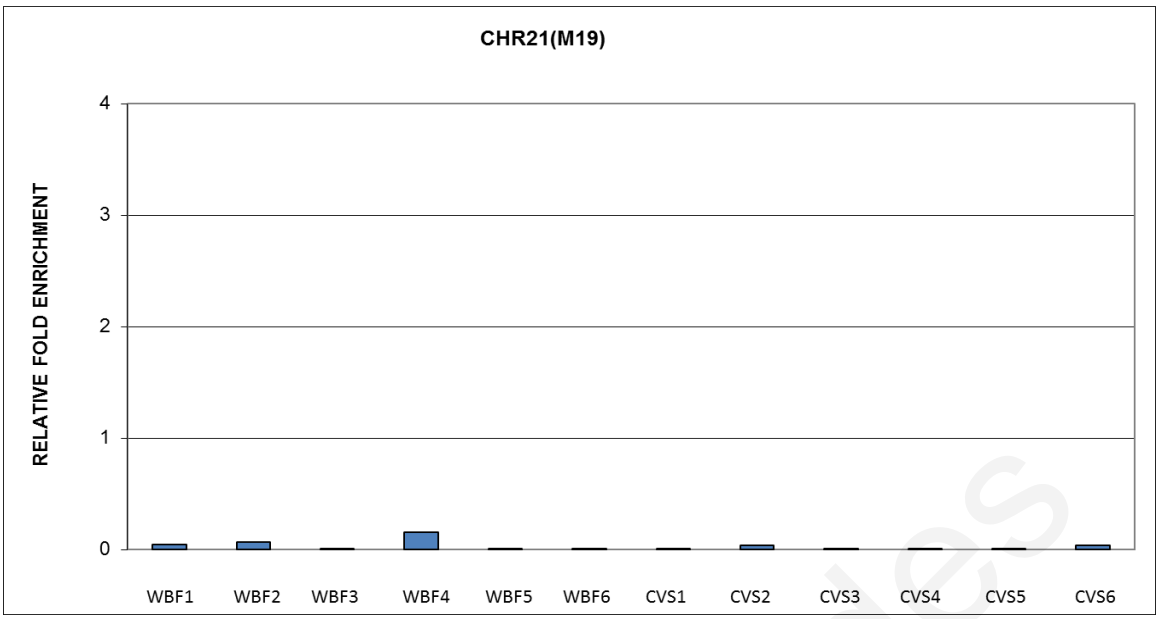


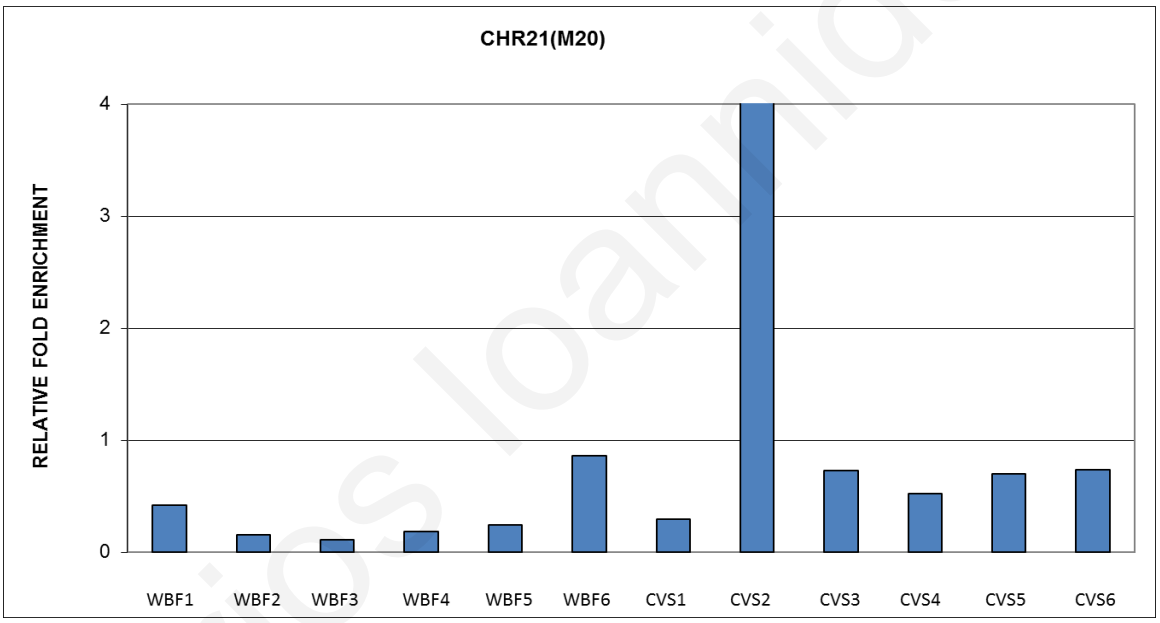


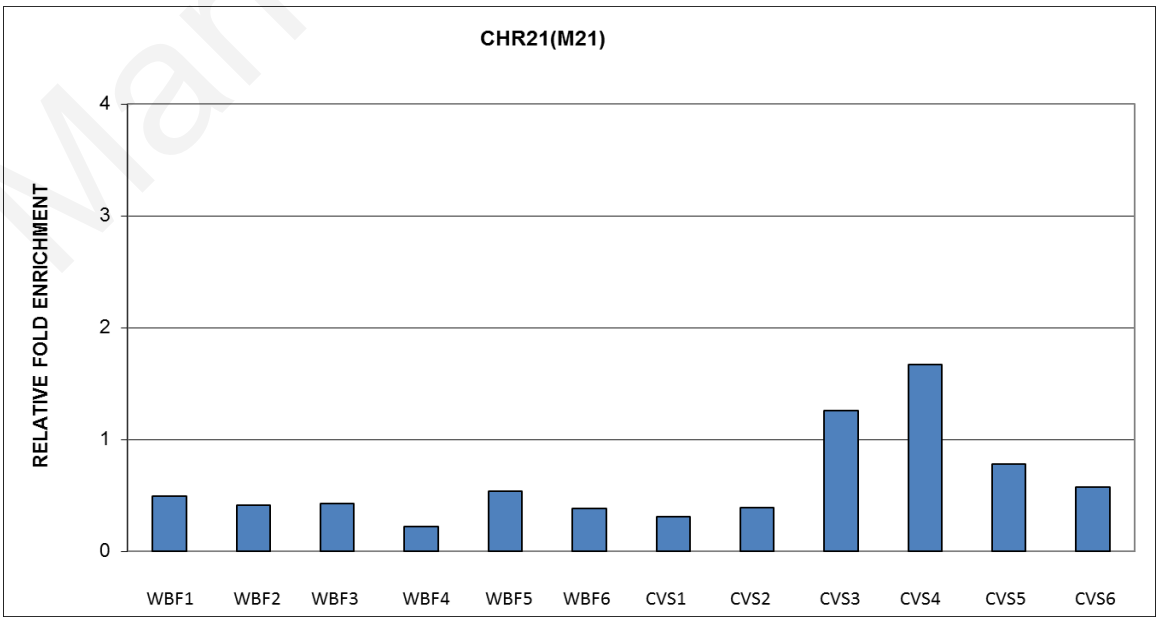


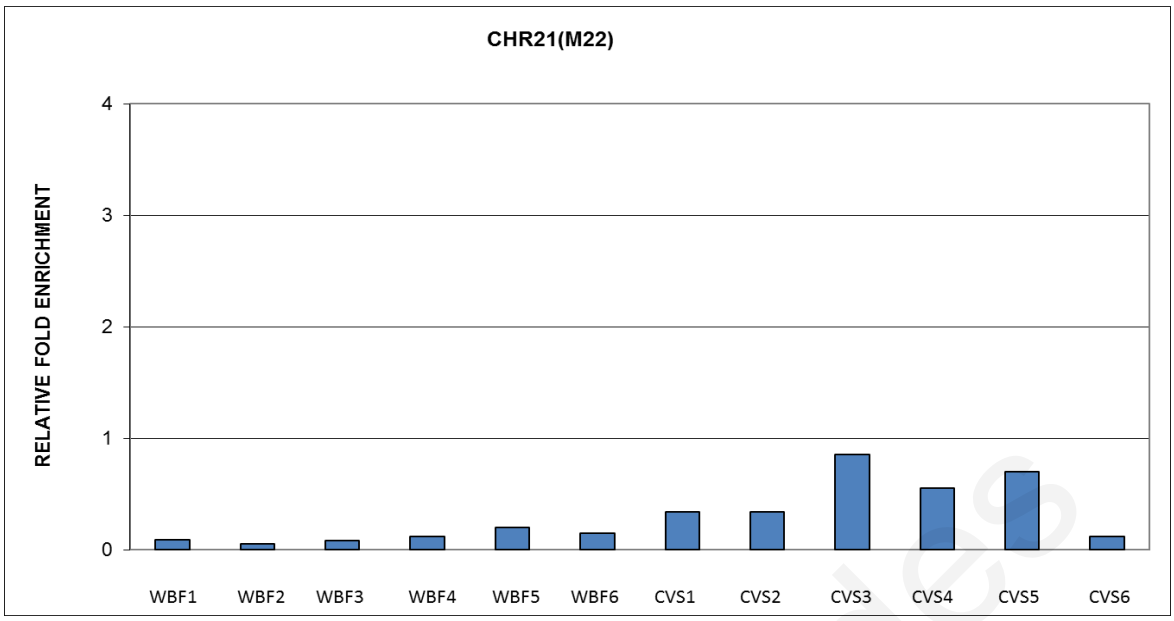


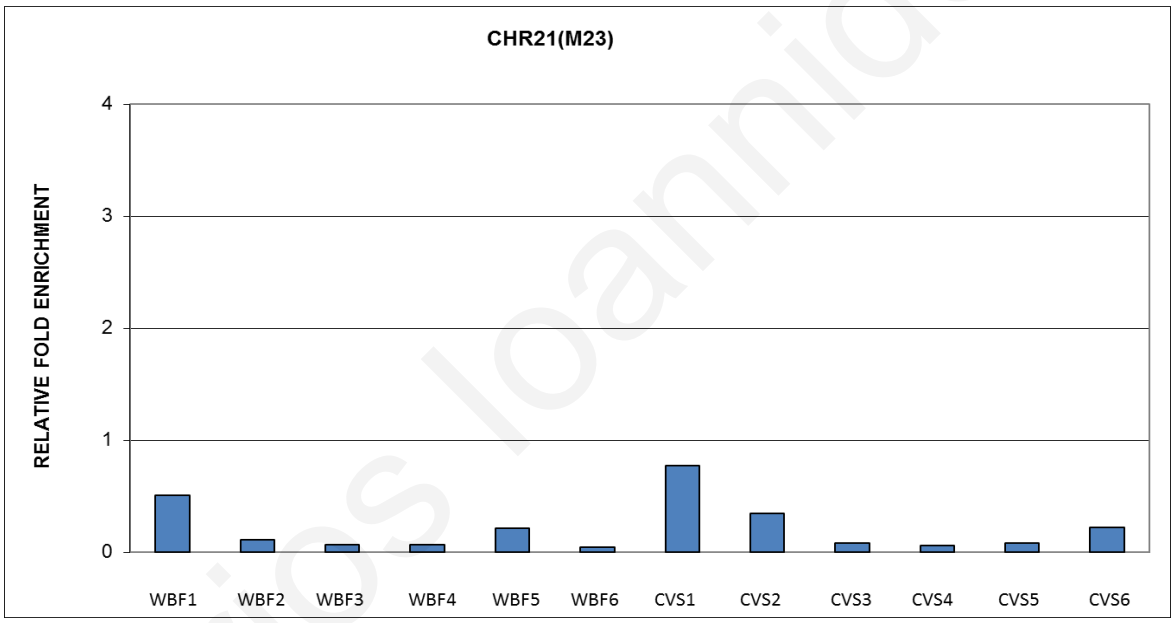


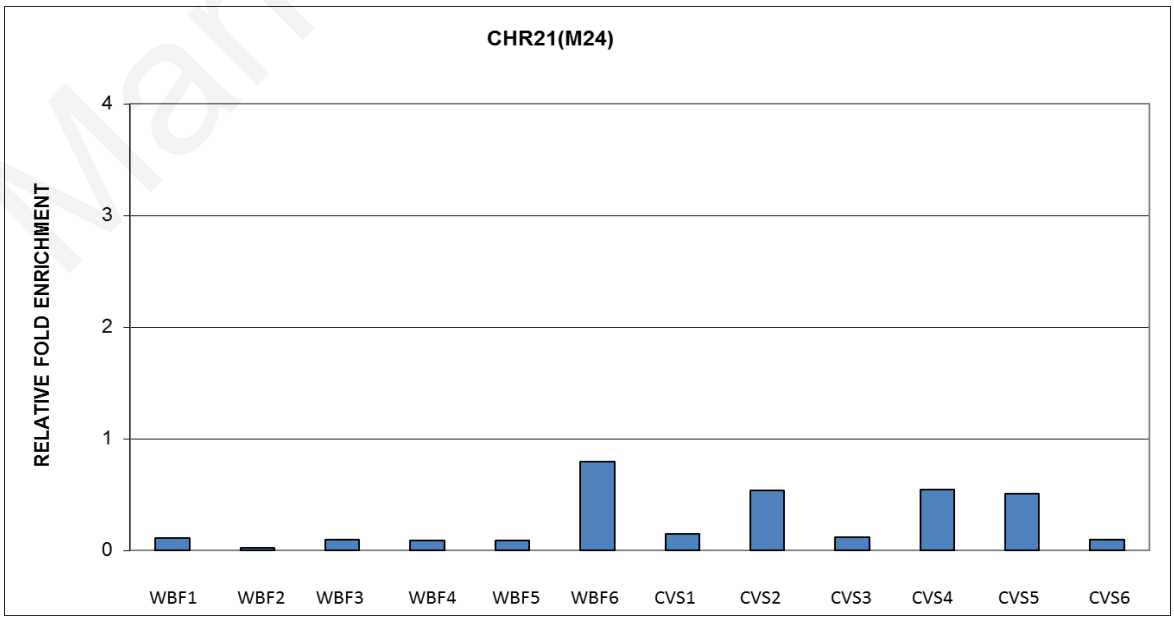


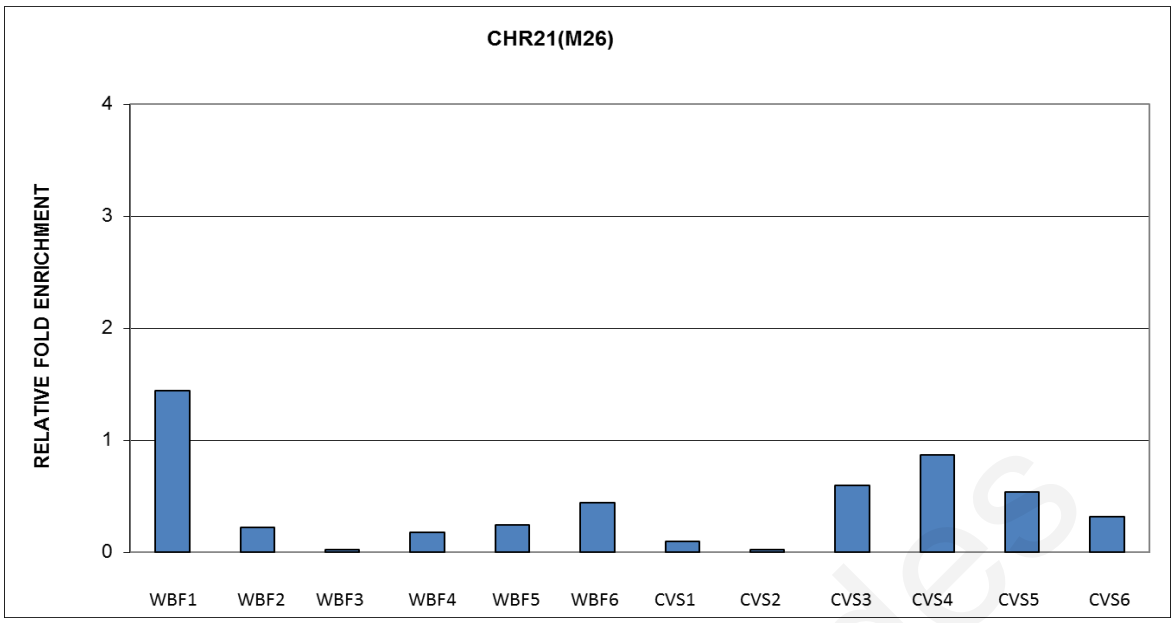


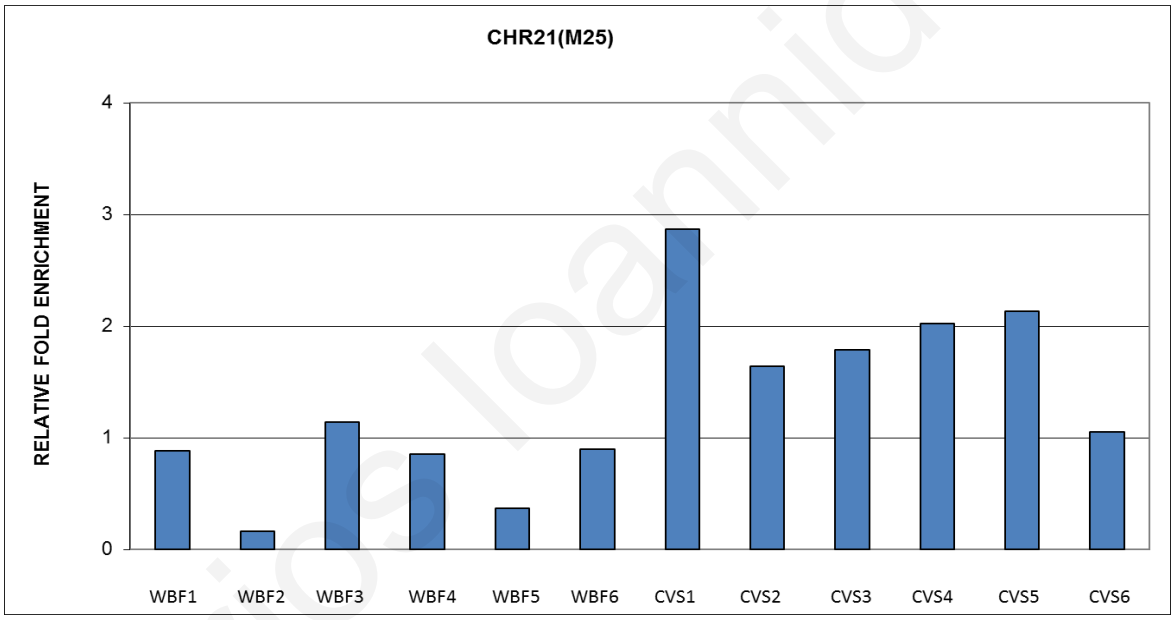


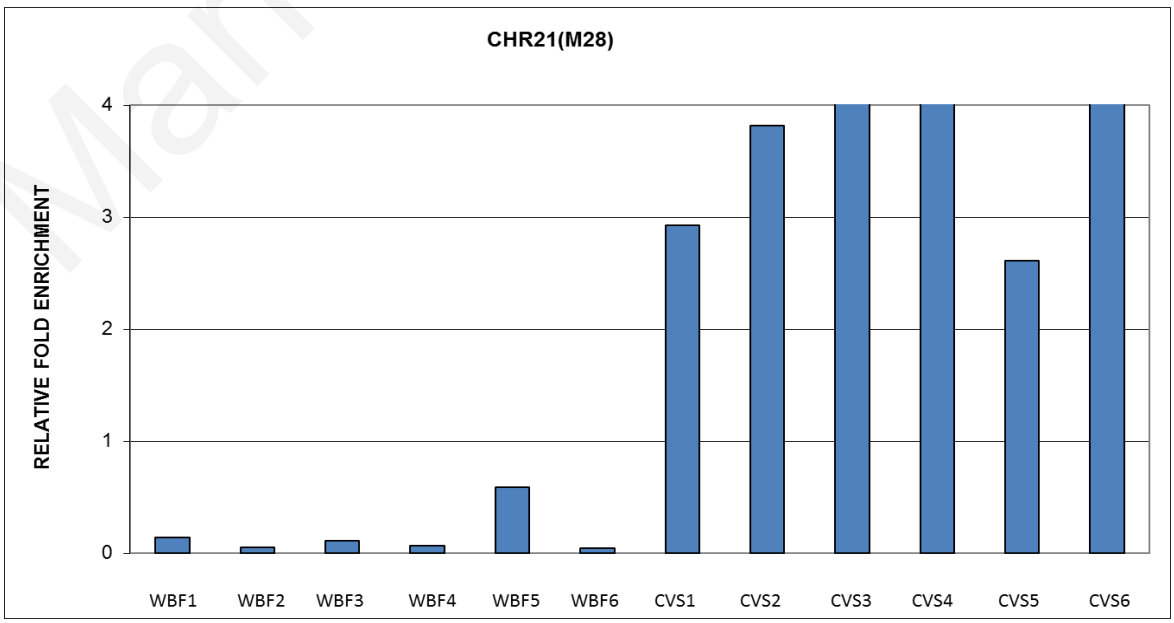


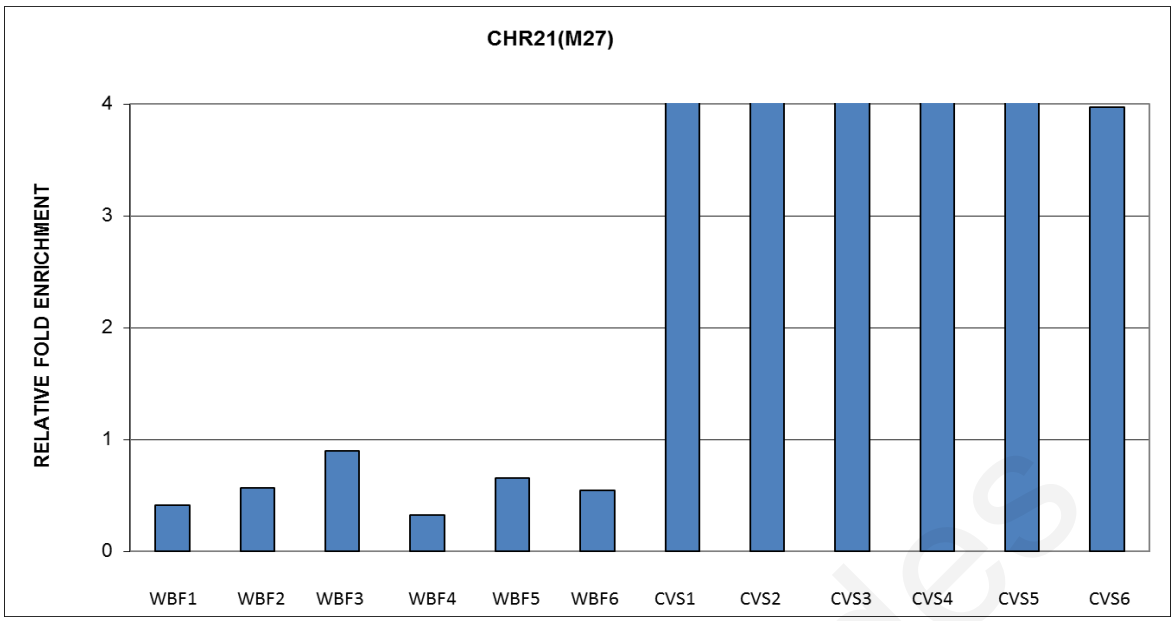


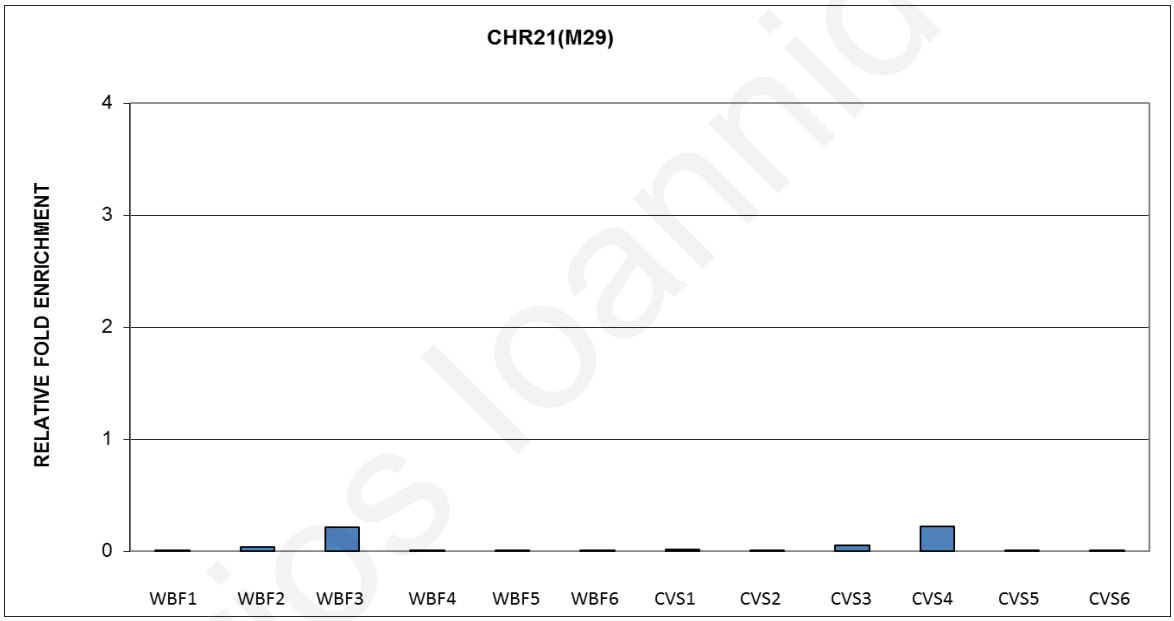


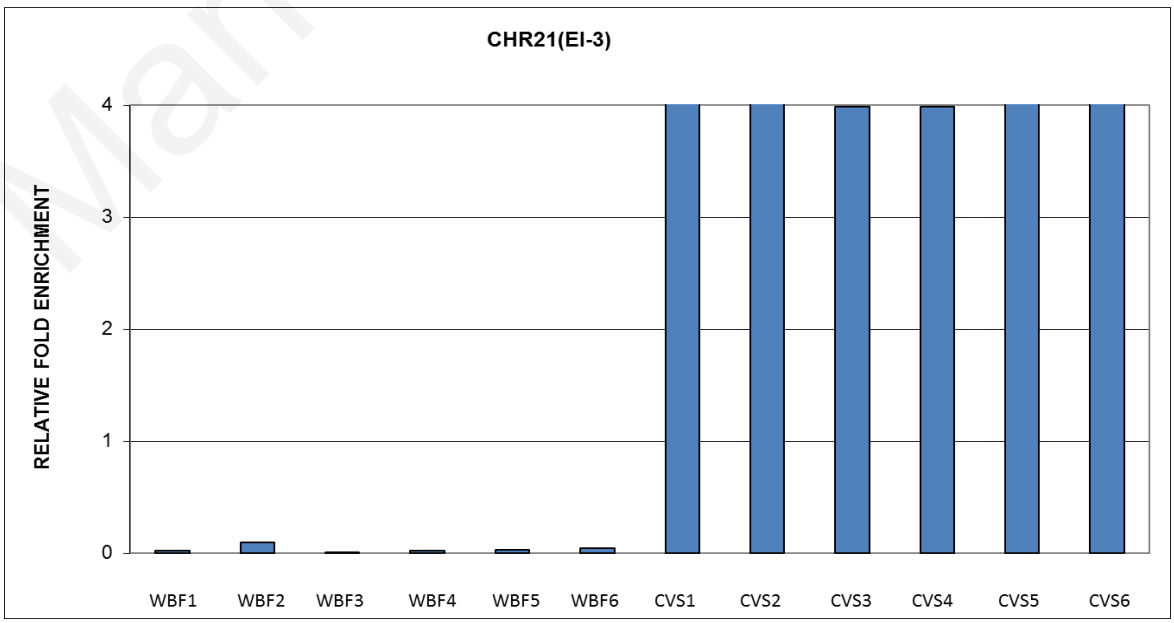


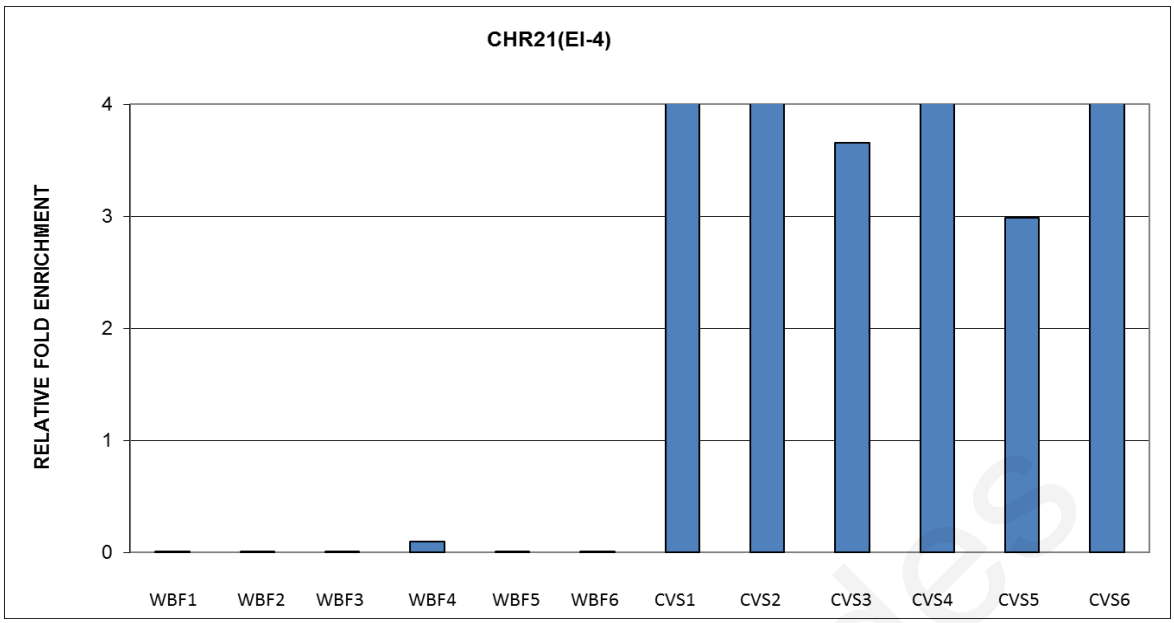








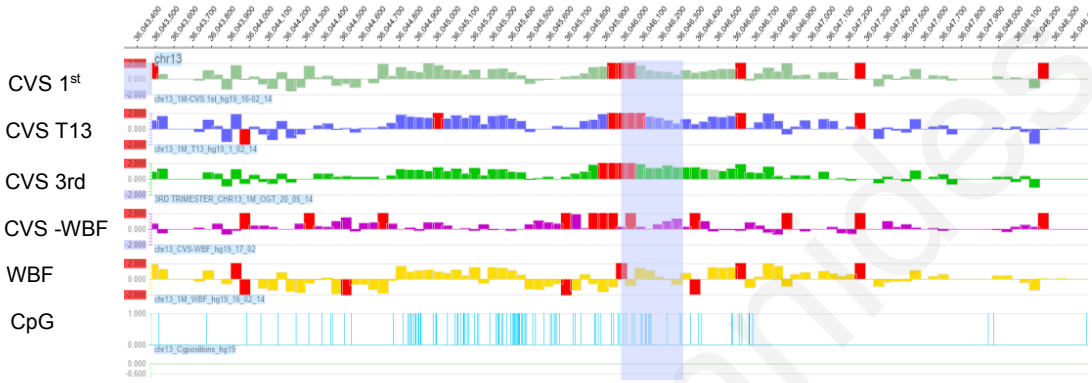




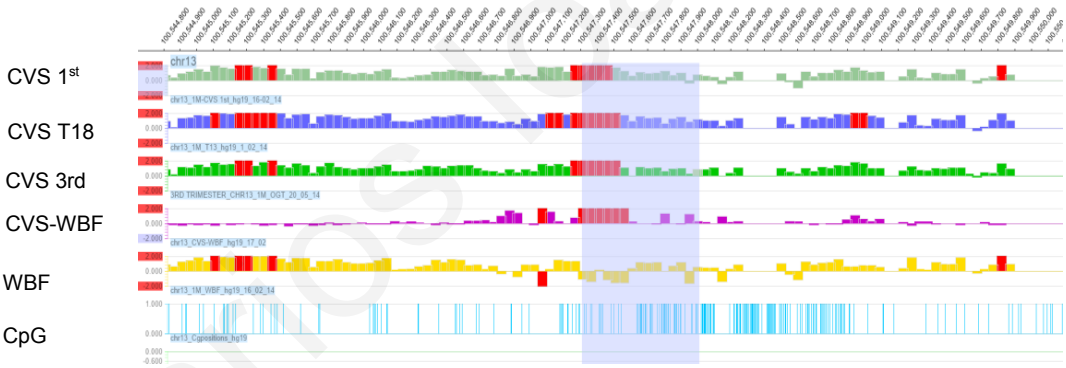
Appendix II

Confirmation of the methylation status between CVS and WBF of control DMRs (SERBIN, A, C, D, J, EI, EII) and DMRs that were confirmed in the first stage of the study using the previous aCGH data for chromosomes 13, 18 and 21.

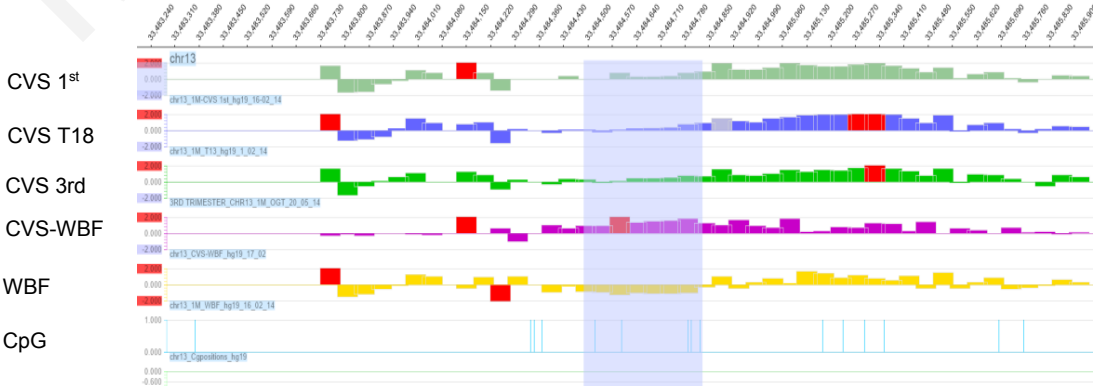
A. Chromosome 13



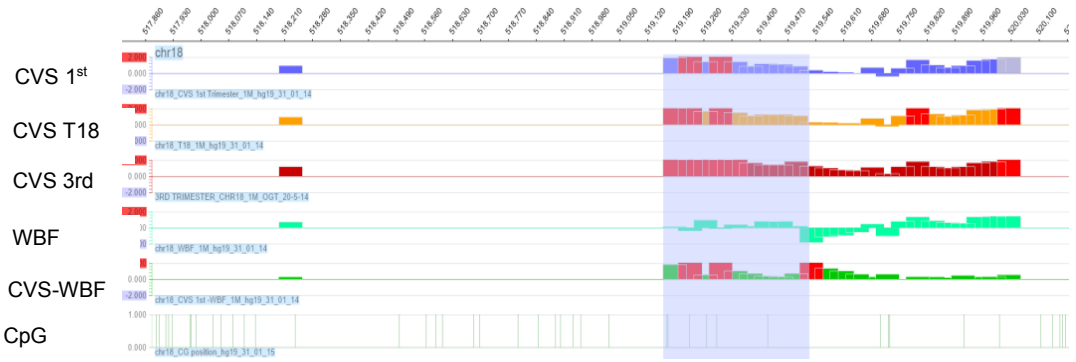
AV2



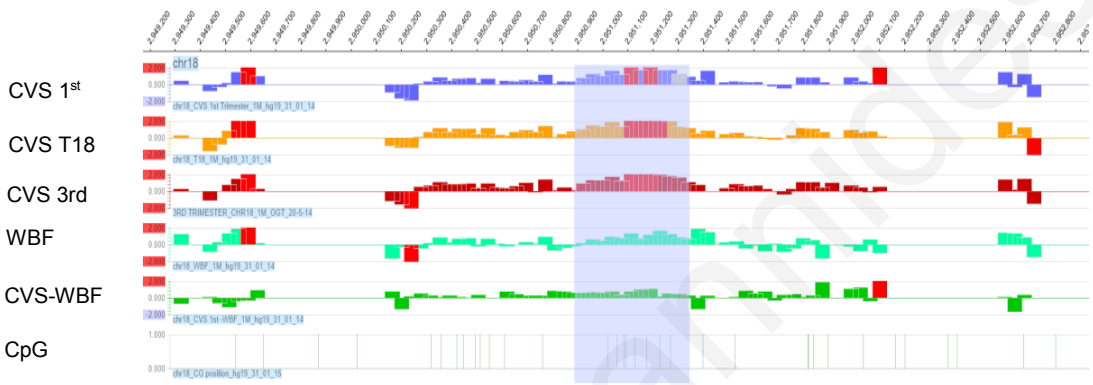
AV5



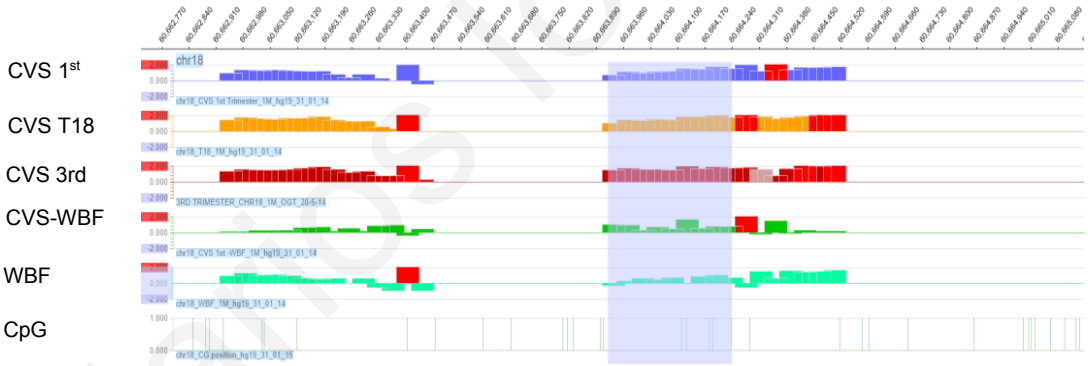
R2



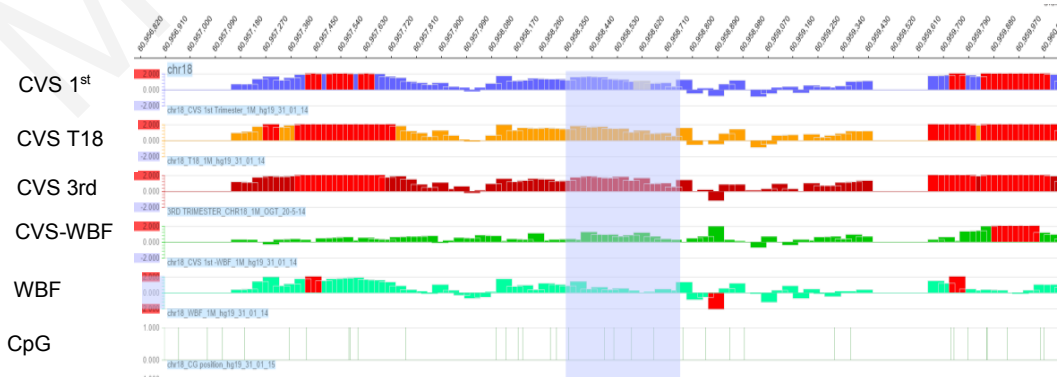
VA1



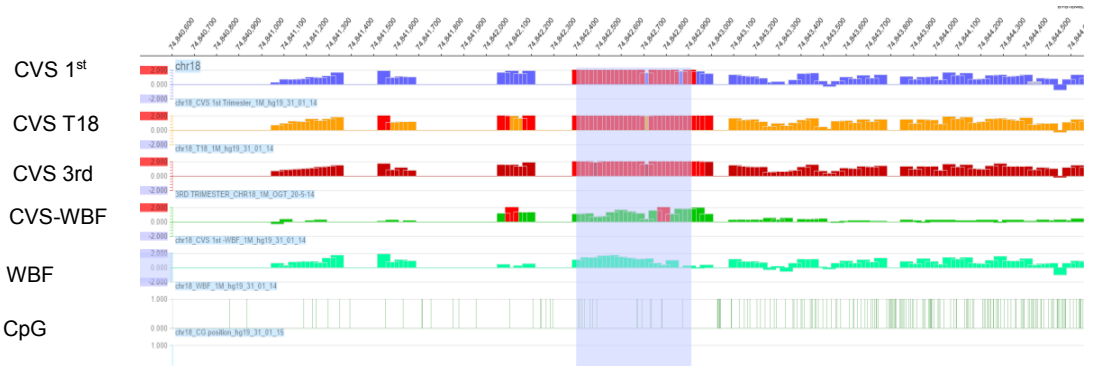
VA4



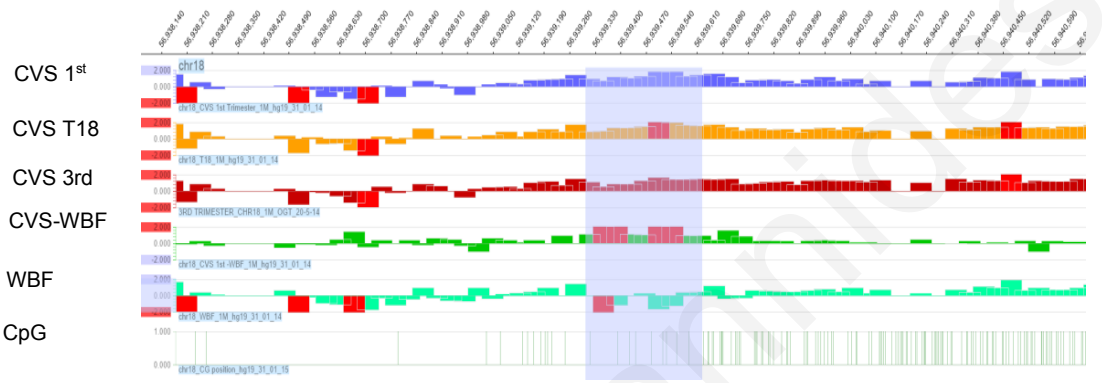
VA15



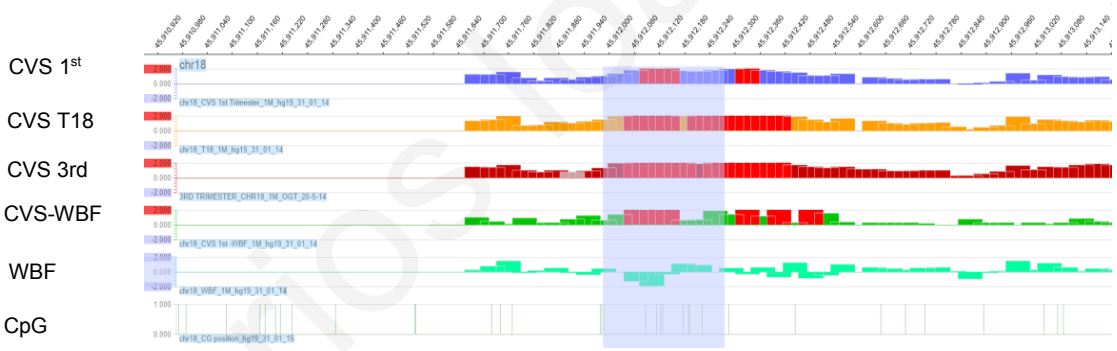
VA17



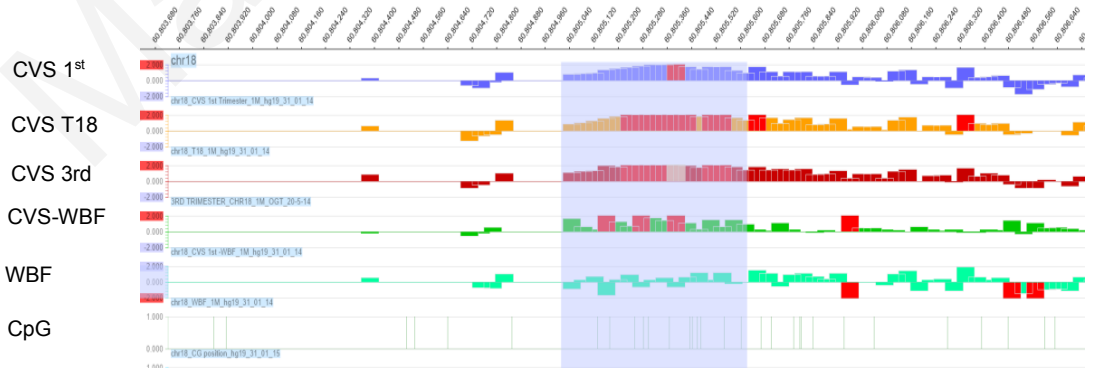
VA22



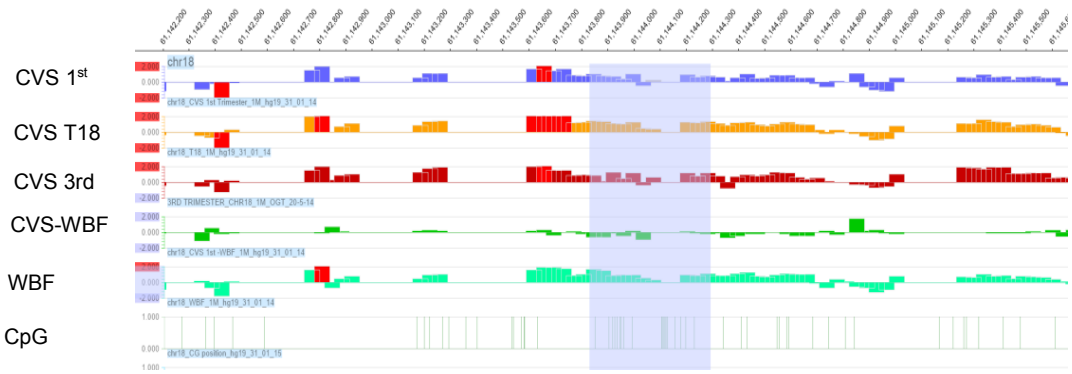
AII



B

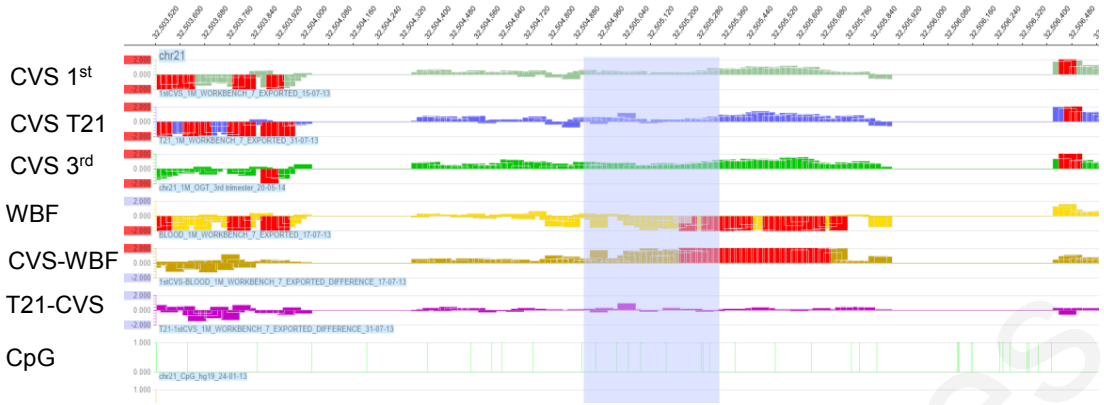


C

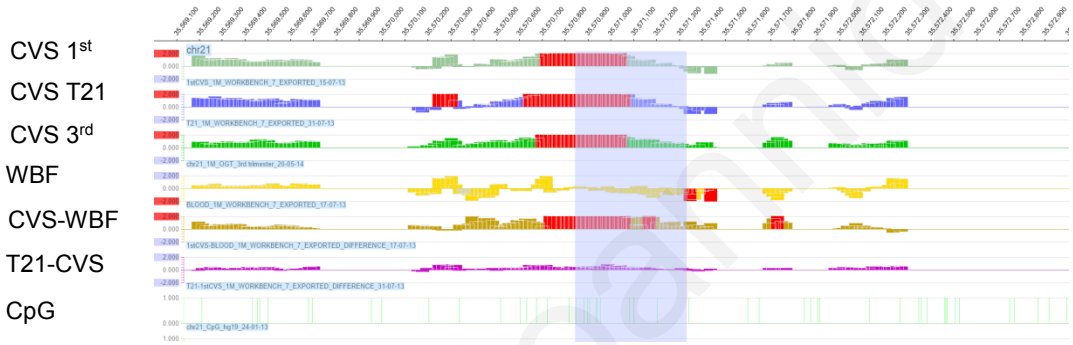


SERPINB5 (Control region- Opposite methylation status as indicated by negative difference between CVS-WBF)

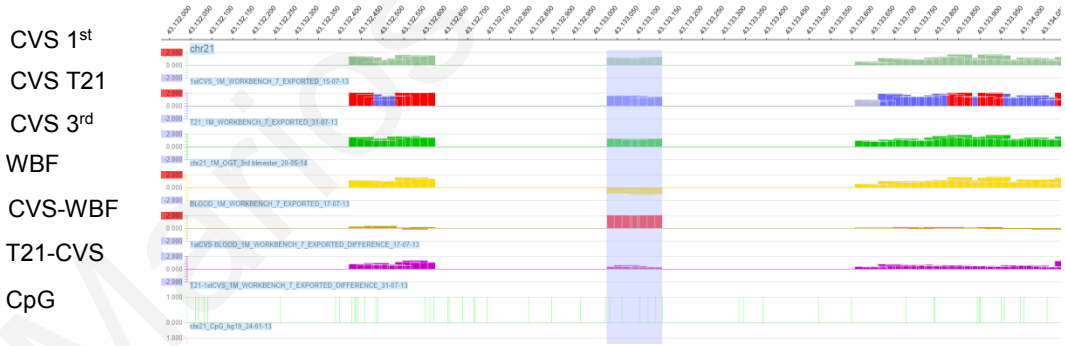
C. Chromosome 21



Nn



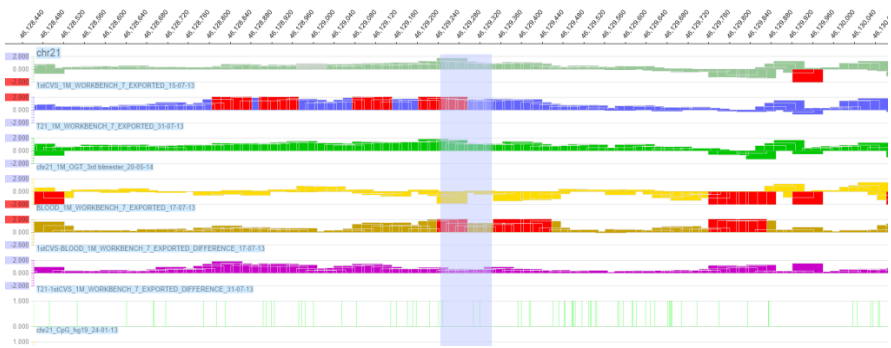
On



Fd

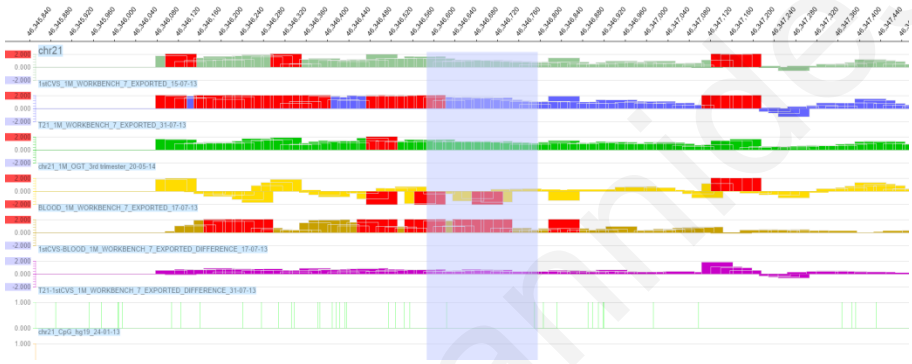
Id

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG



M1E

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG

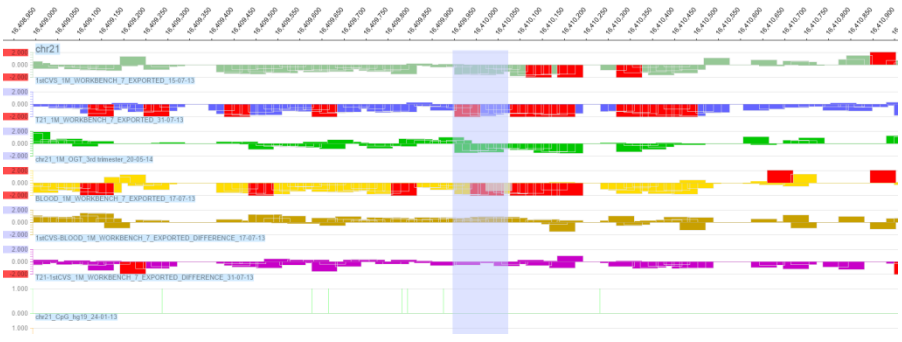


M28

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG

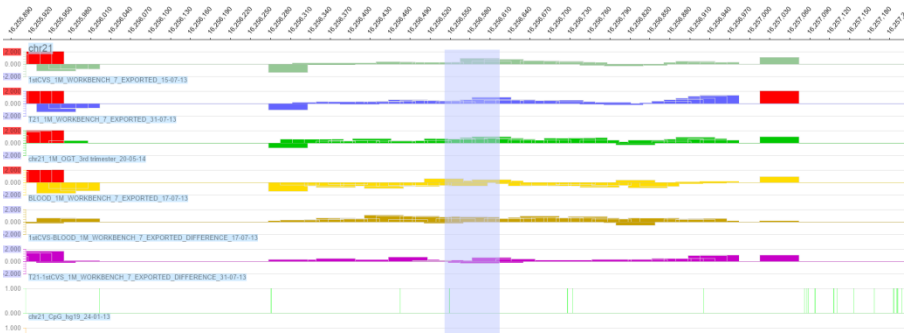
M27

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG



M18

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG



M20

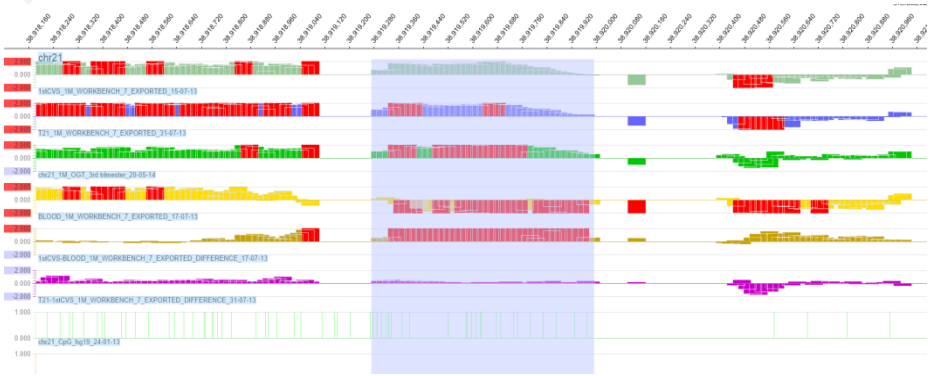
CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG

M25

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG

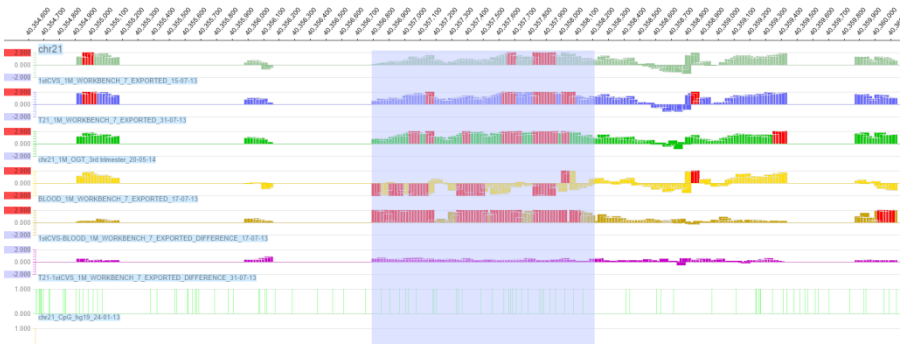
C

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG



J

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG



A

CVS 1st
CVS T21
CVS 3rd
WBF
CVS-WBF
T21-CVS
CpG

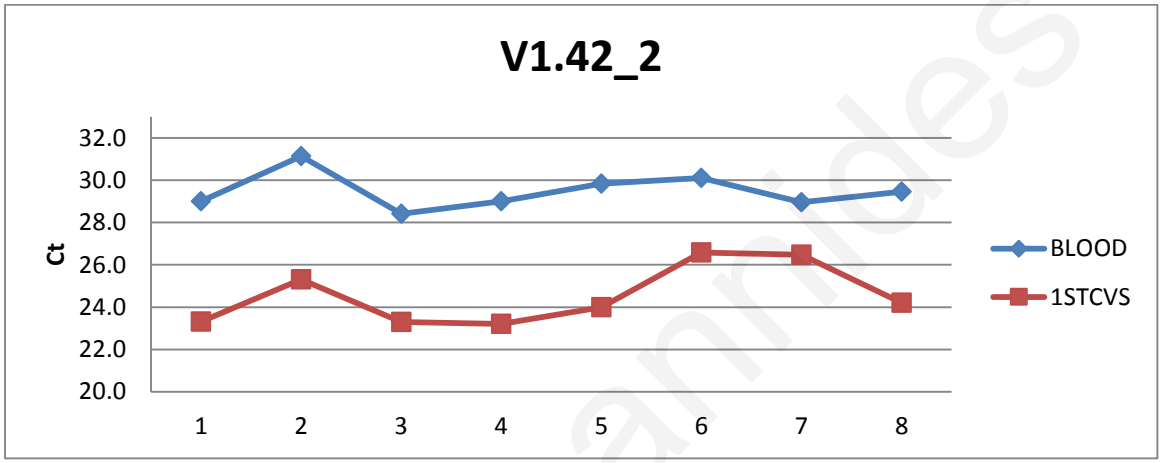
D

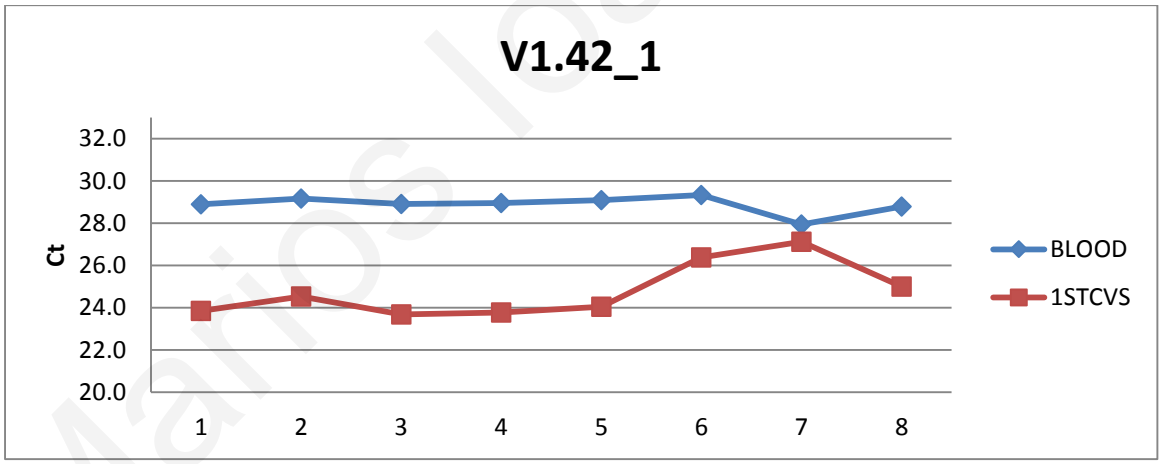
Appendix III

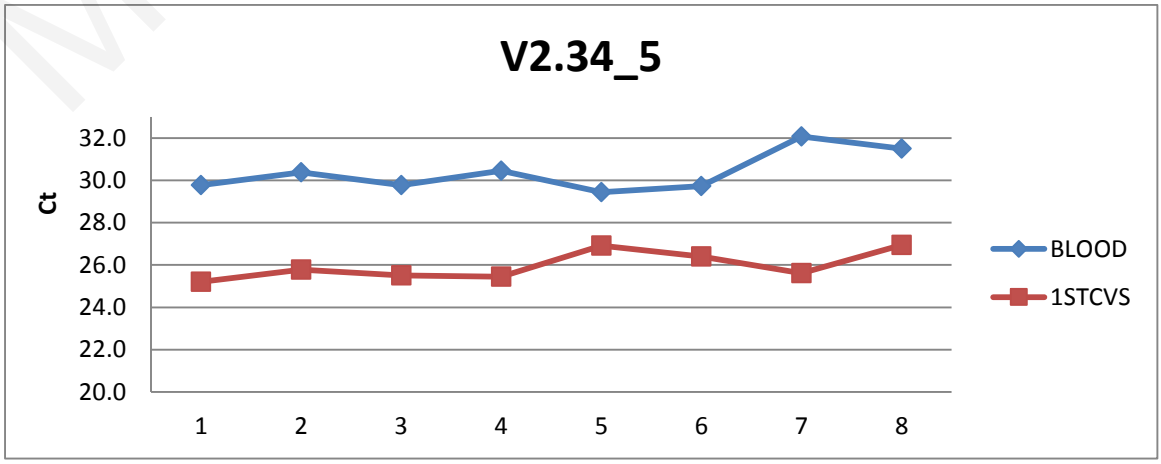
qPCR confirmation of DMRs using the 1M whole genome custom array chip for the chromosomes 13, 18, 21 and X

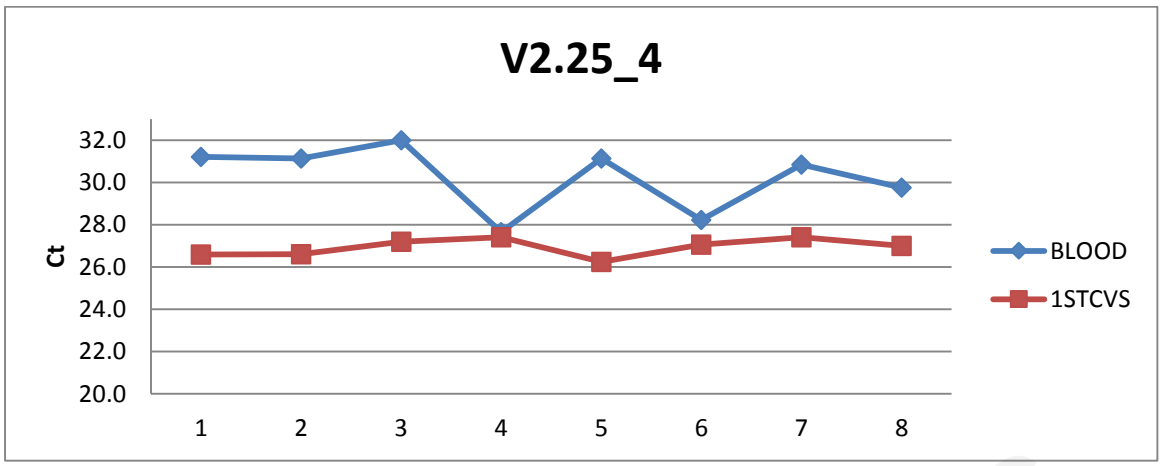
A. Chromosome 13

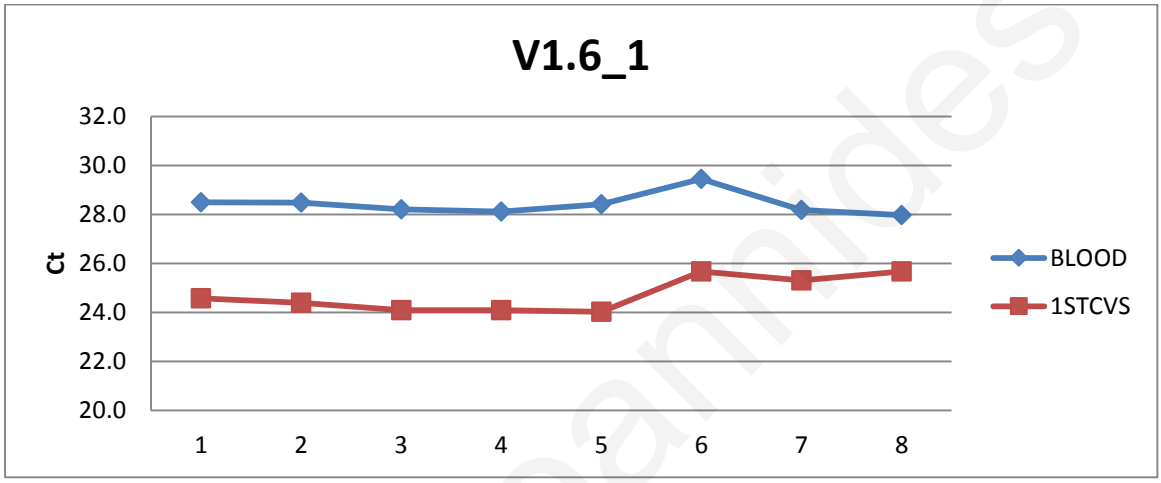
1. Bad DMRs

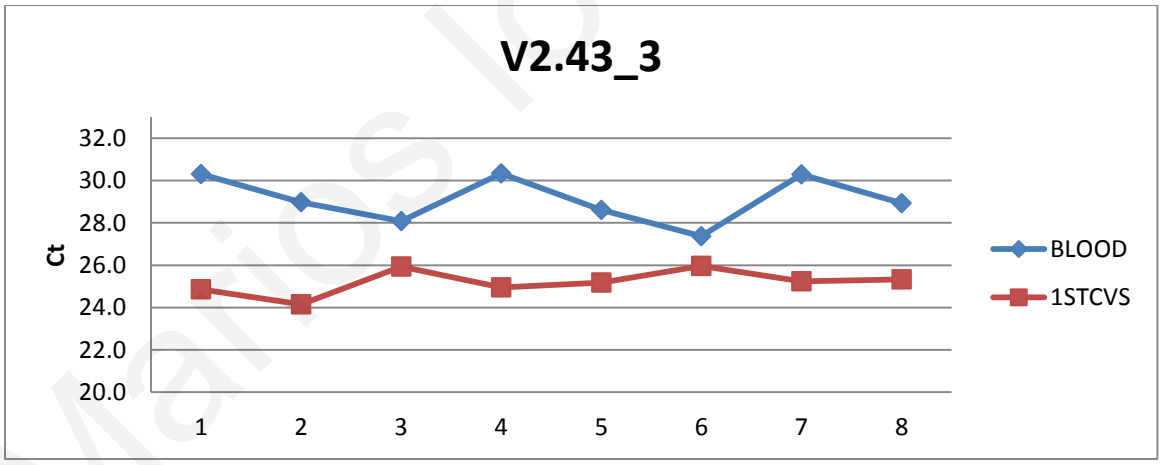


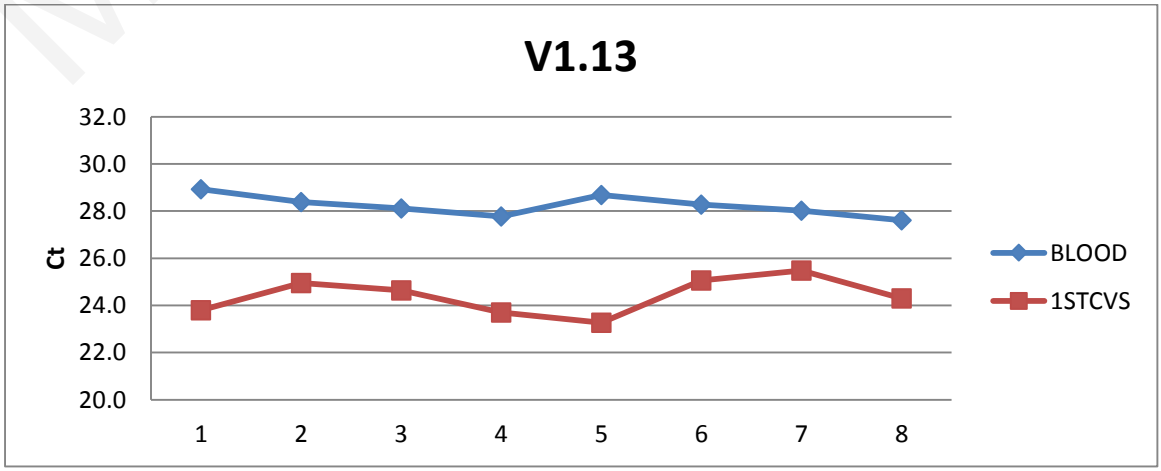


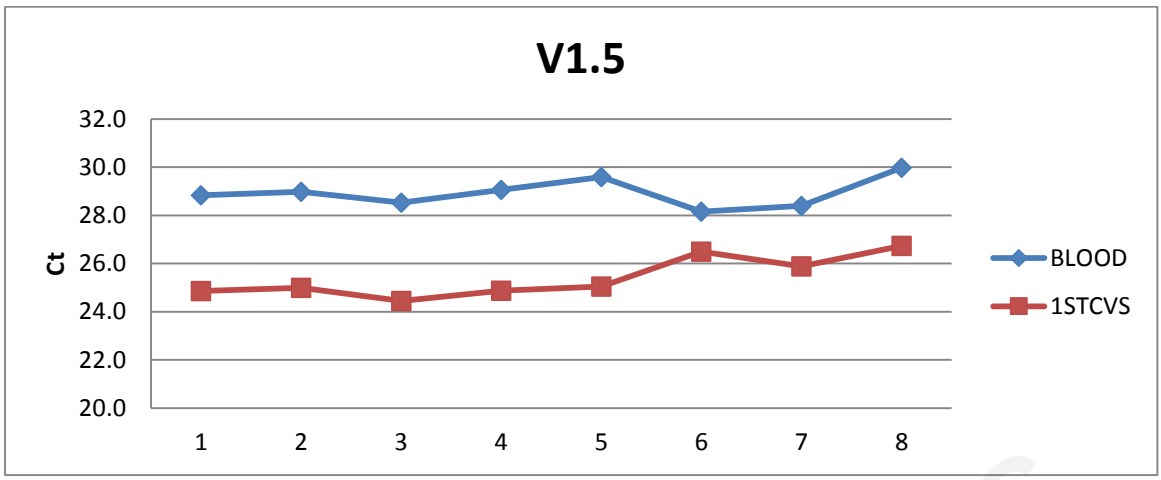


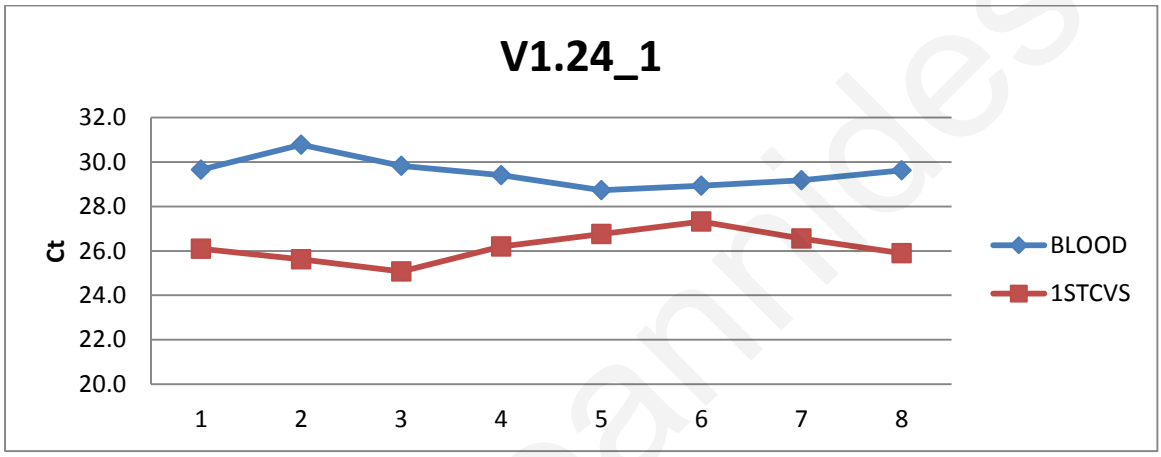


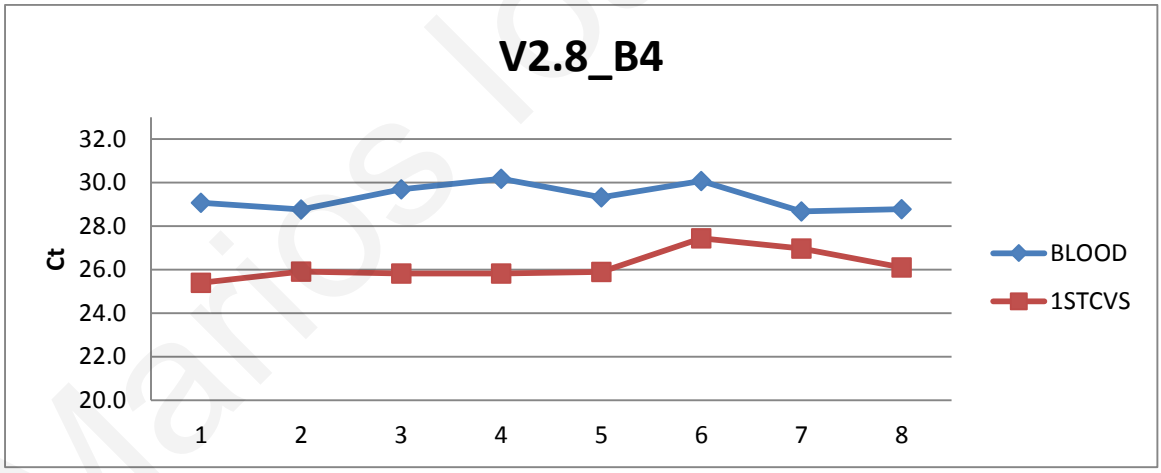


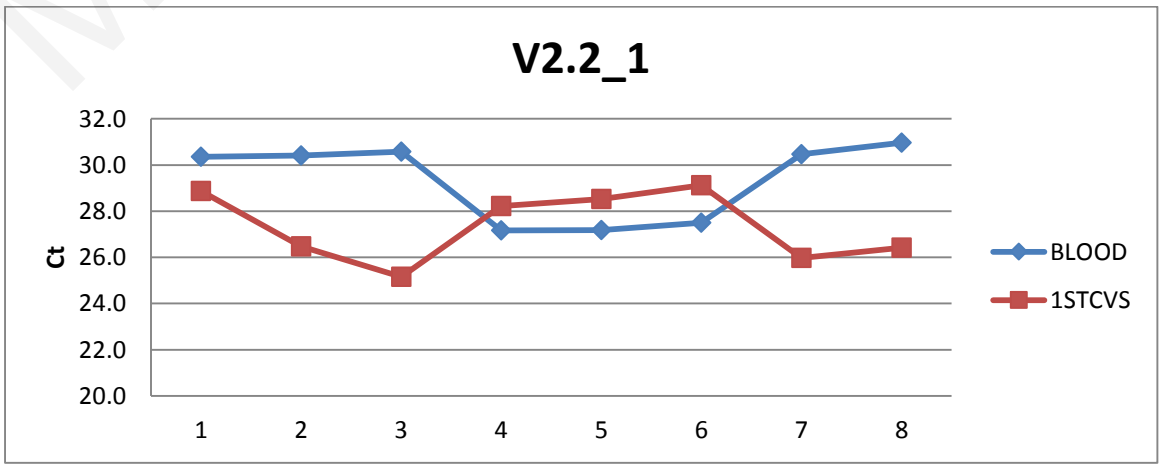


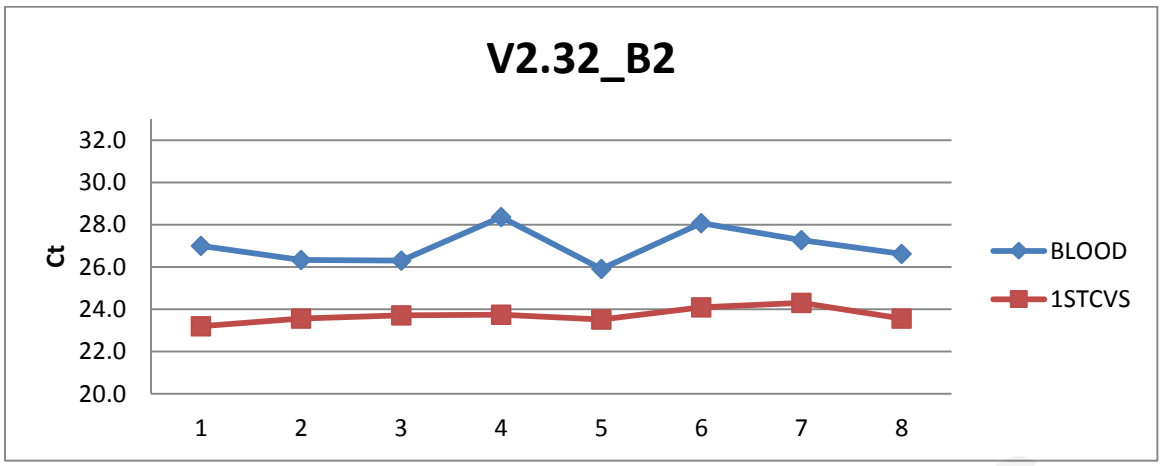


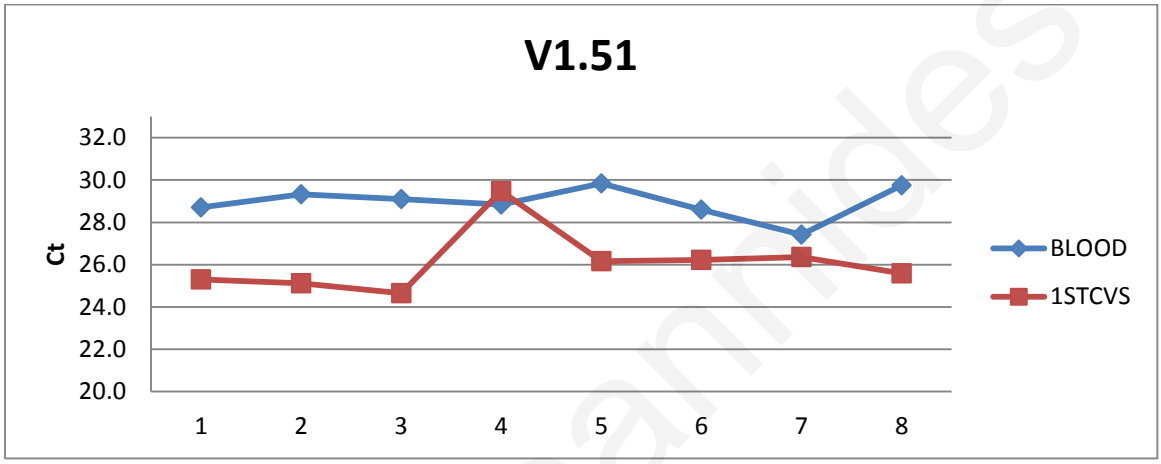


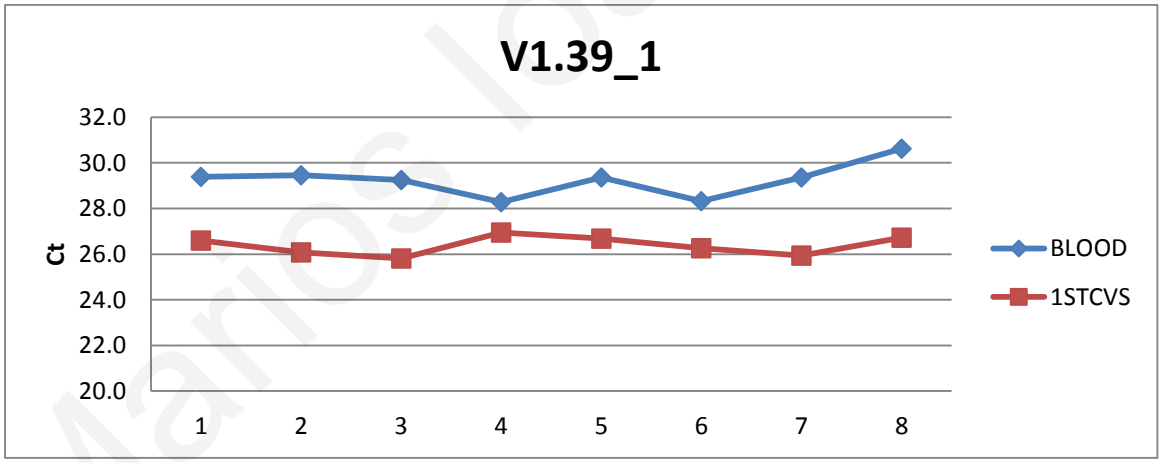


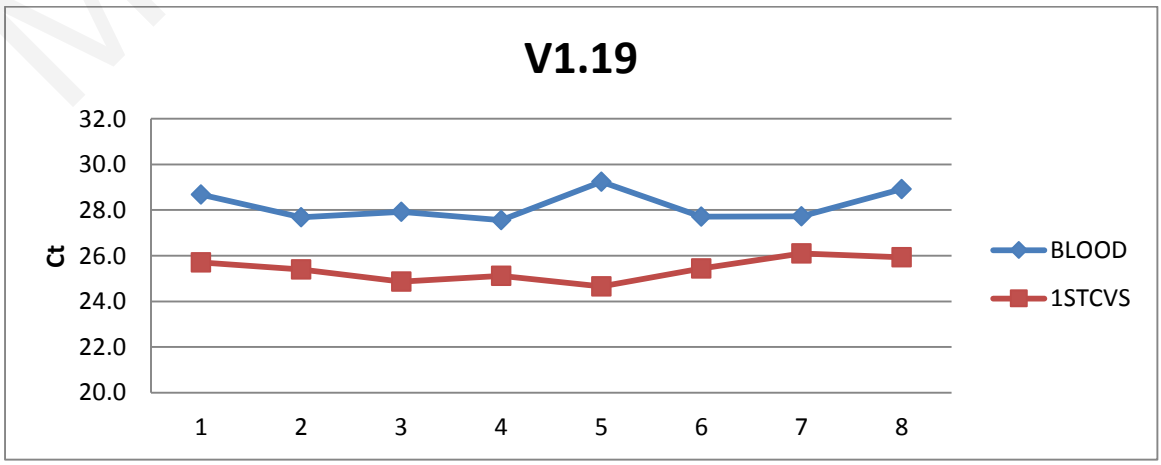


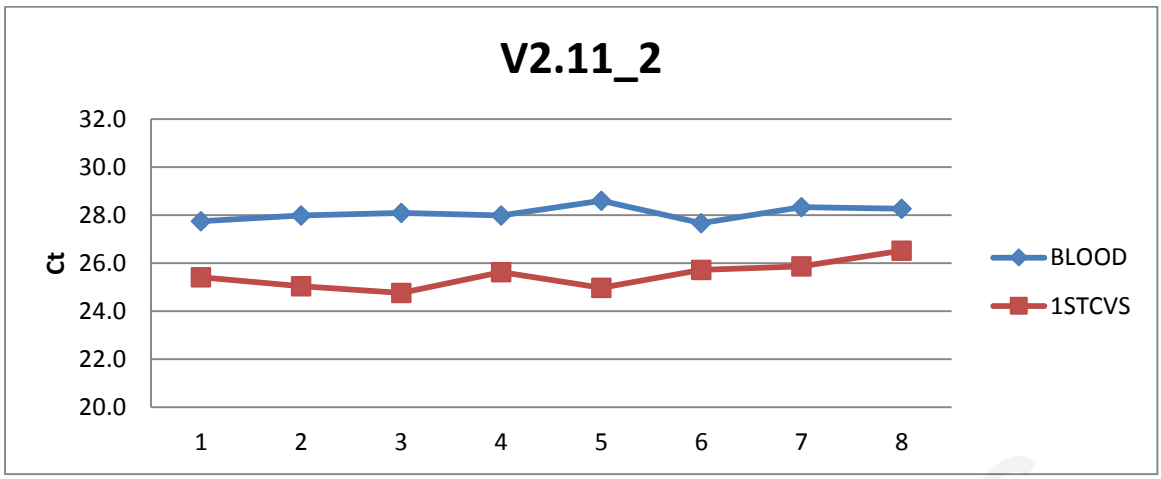


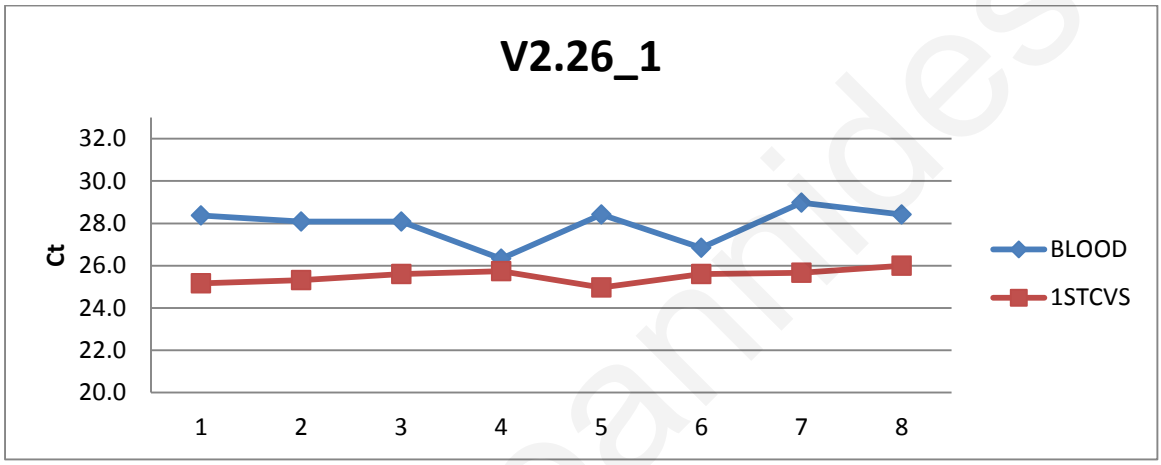


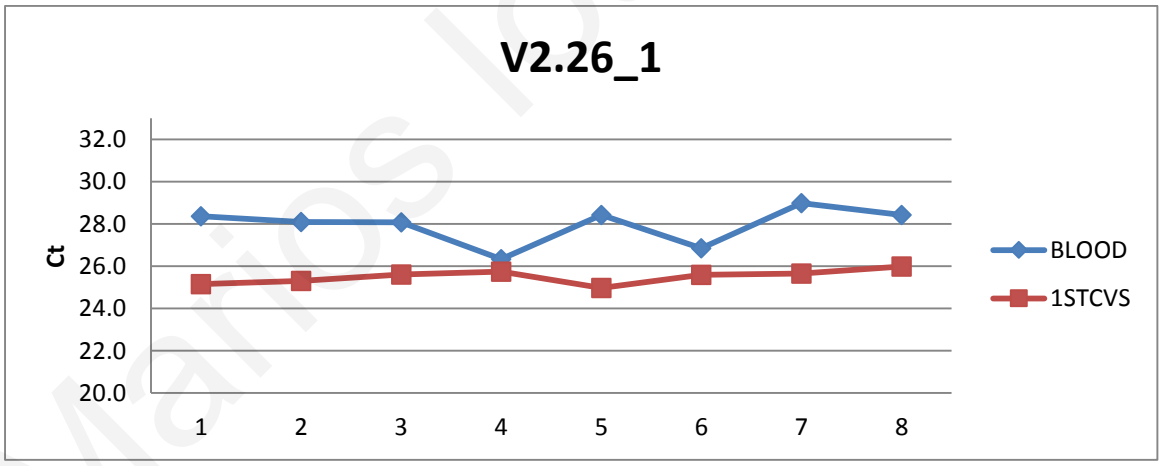


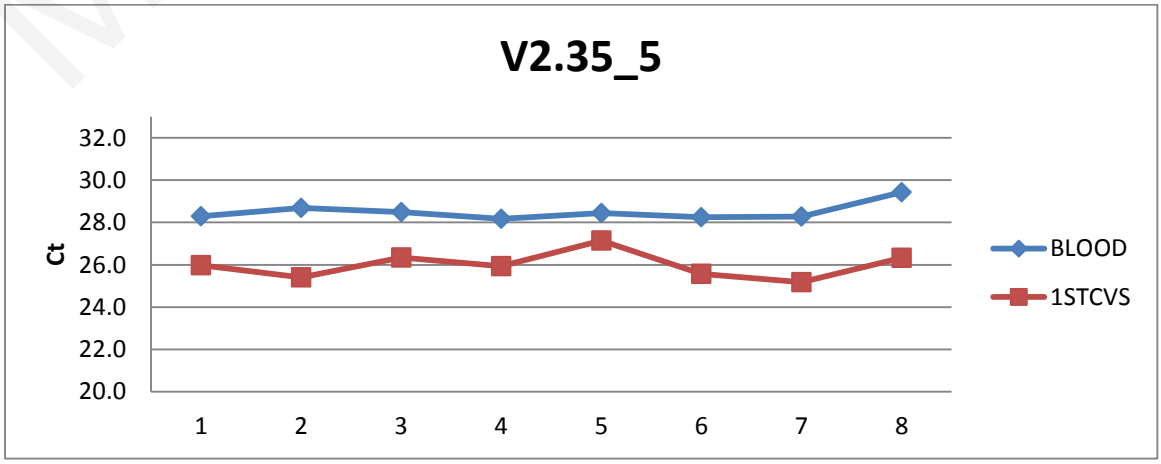


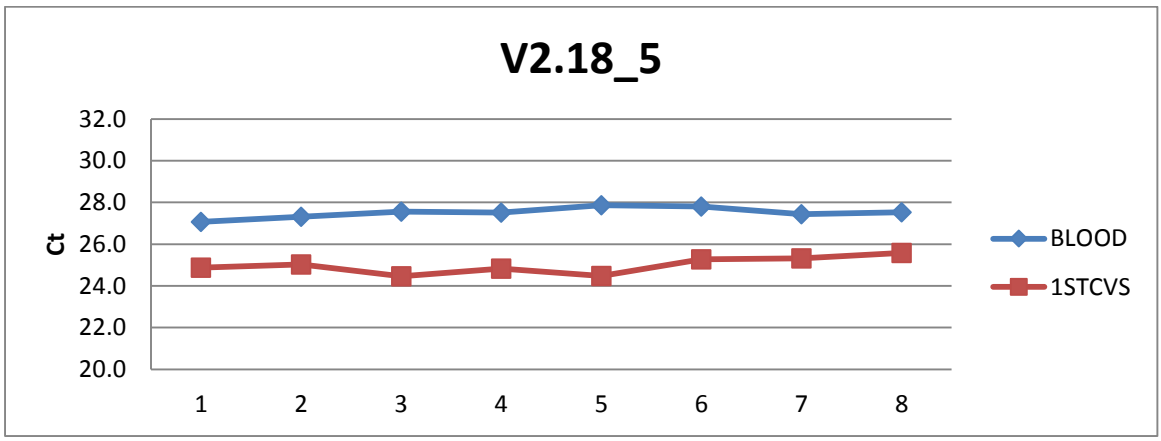


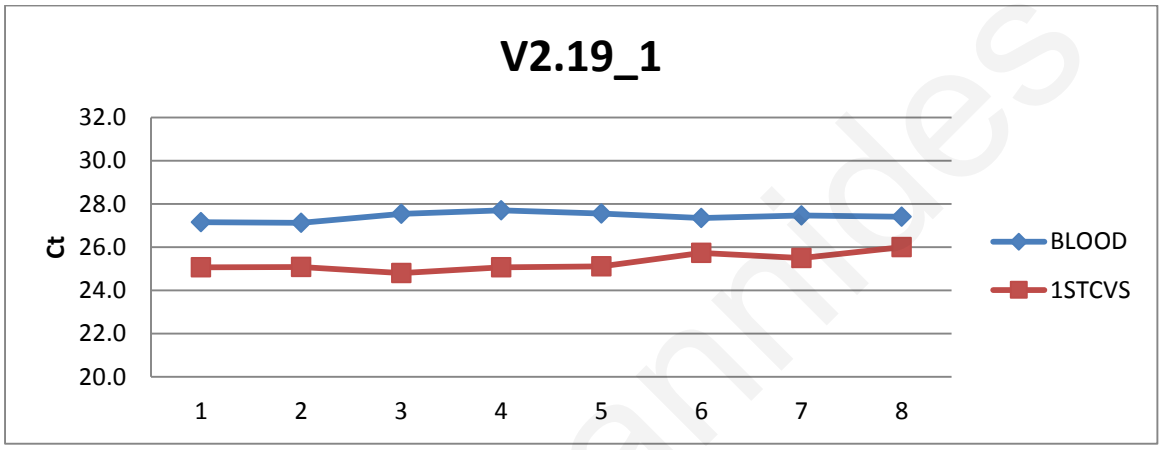


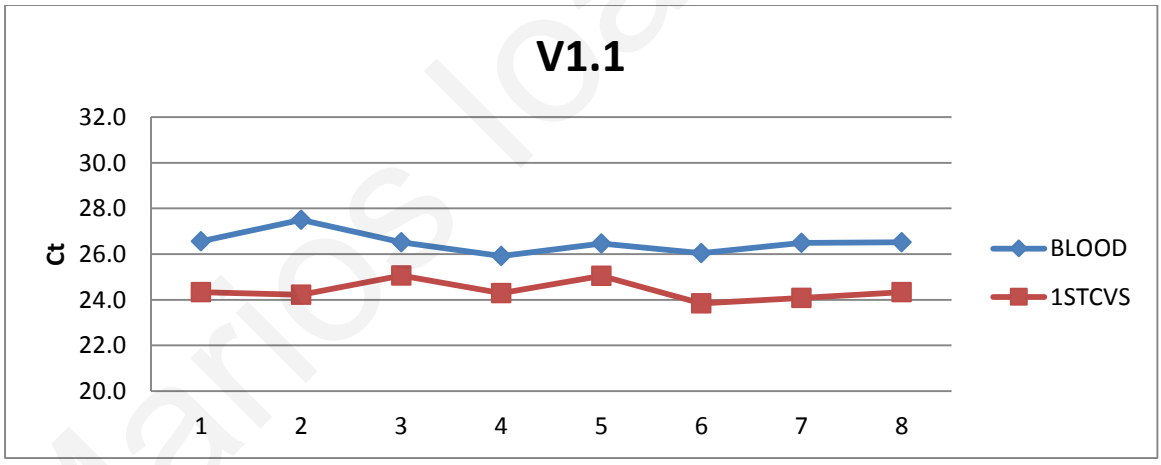


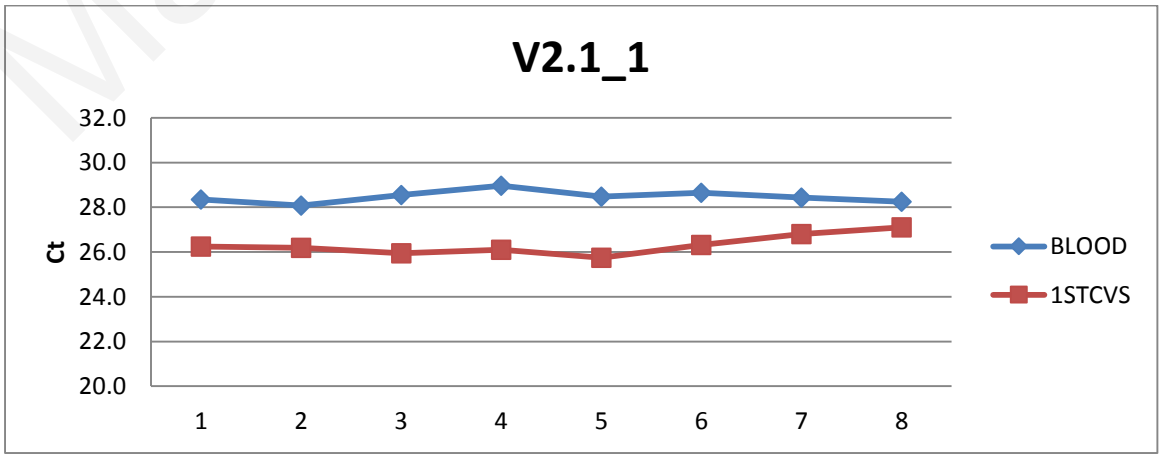


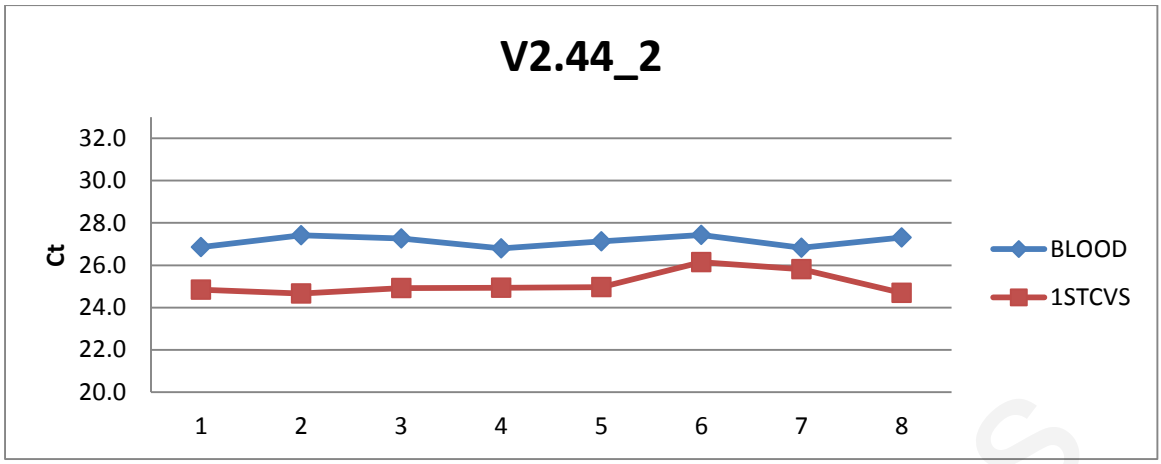


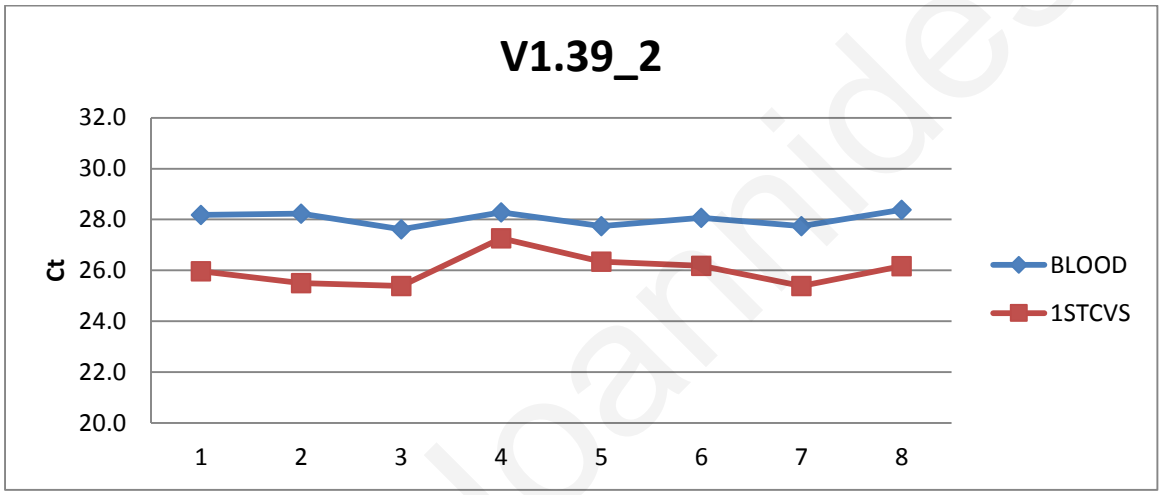


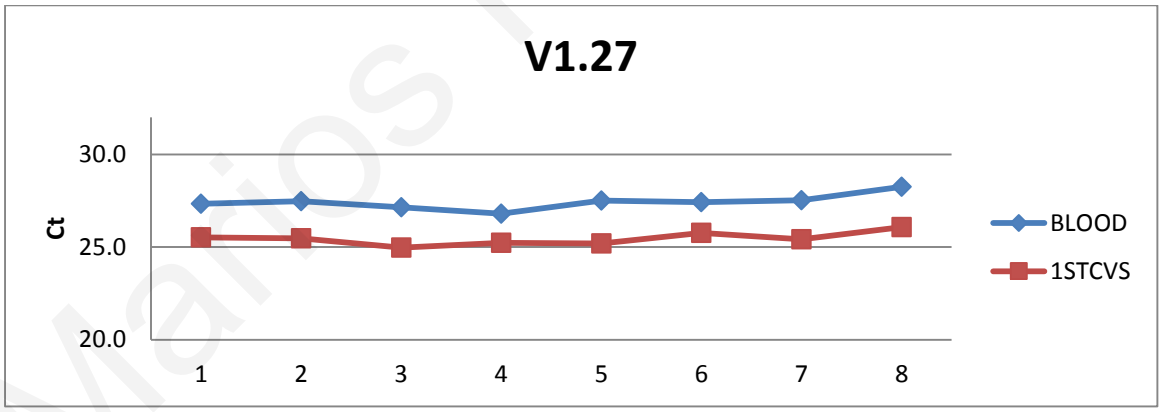


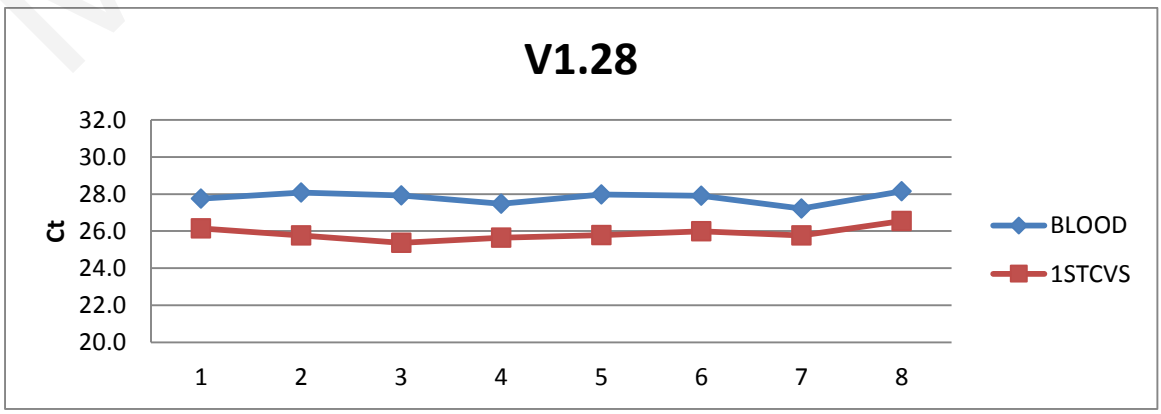


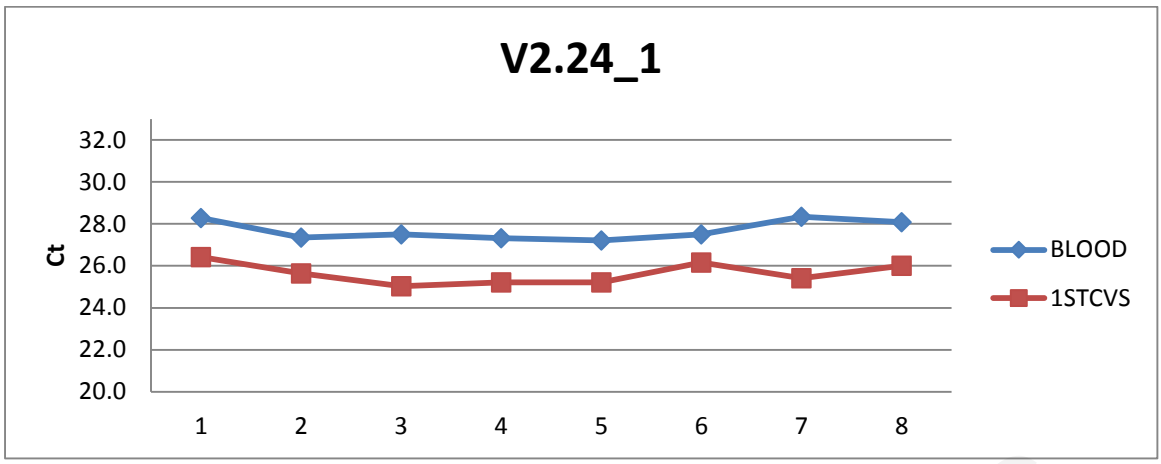


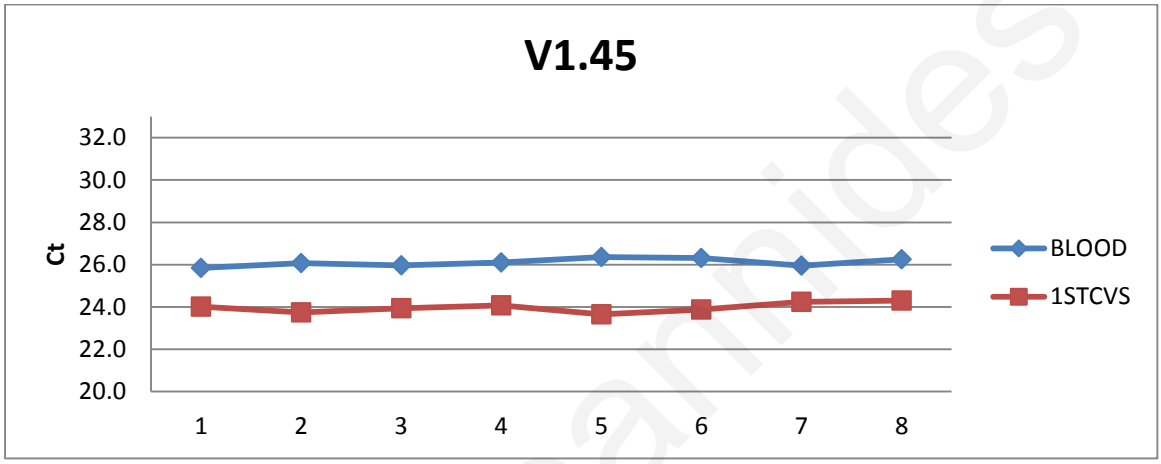


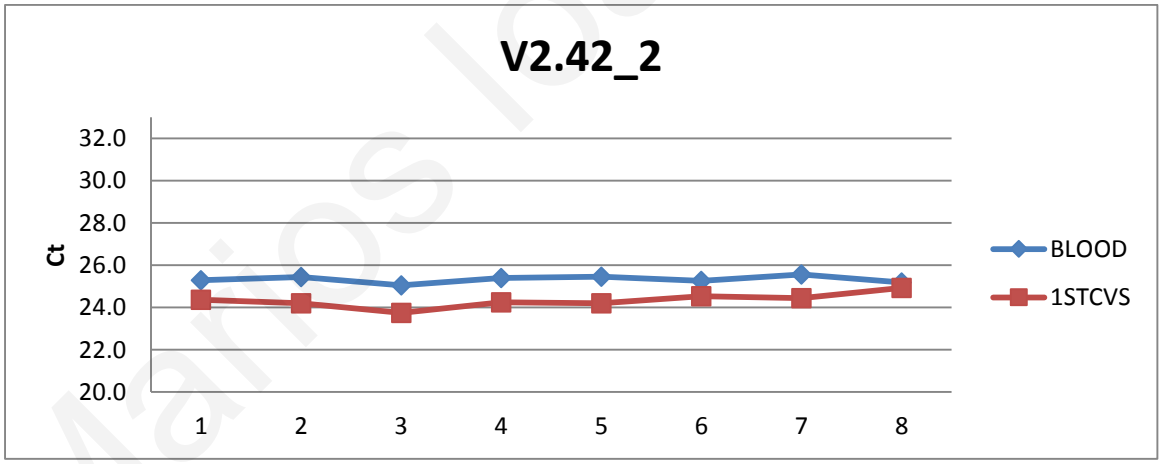


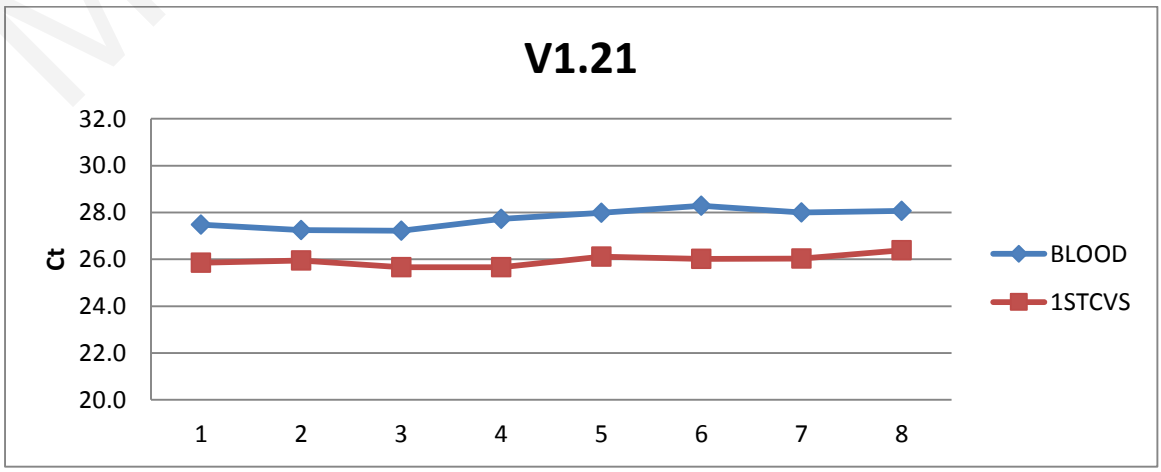


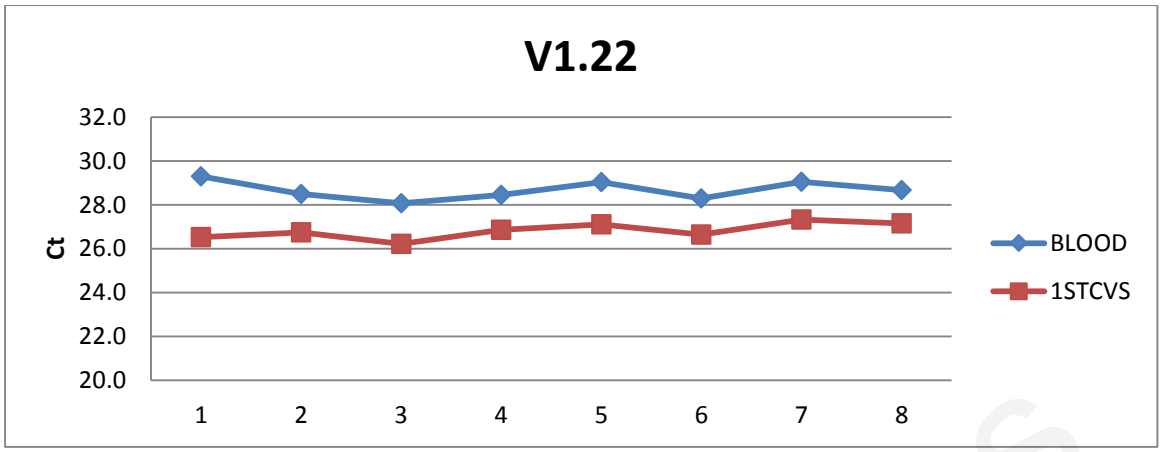


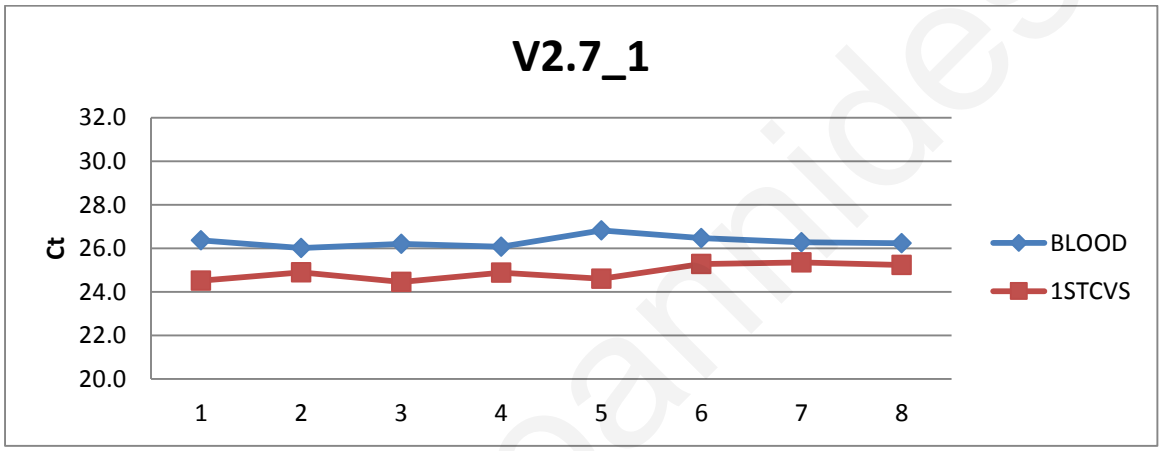


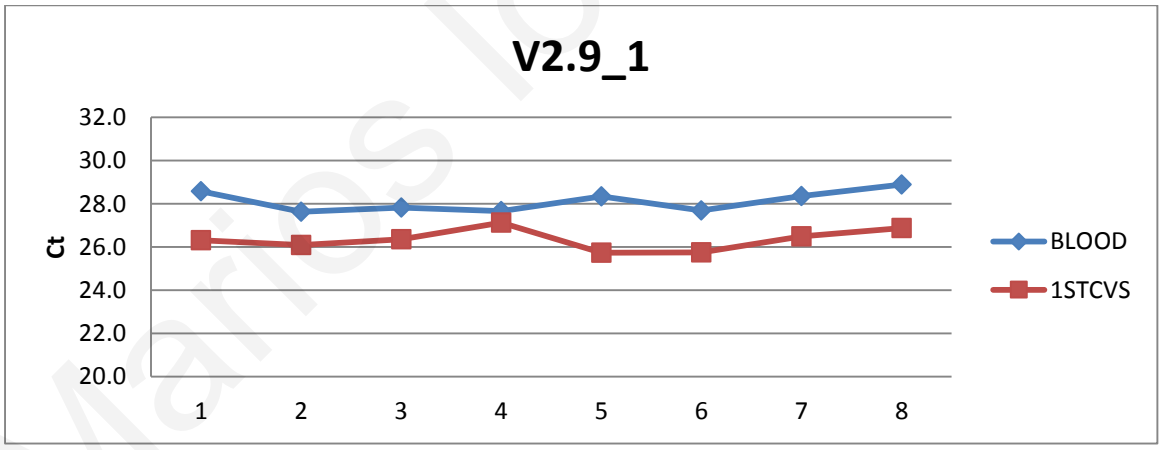


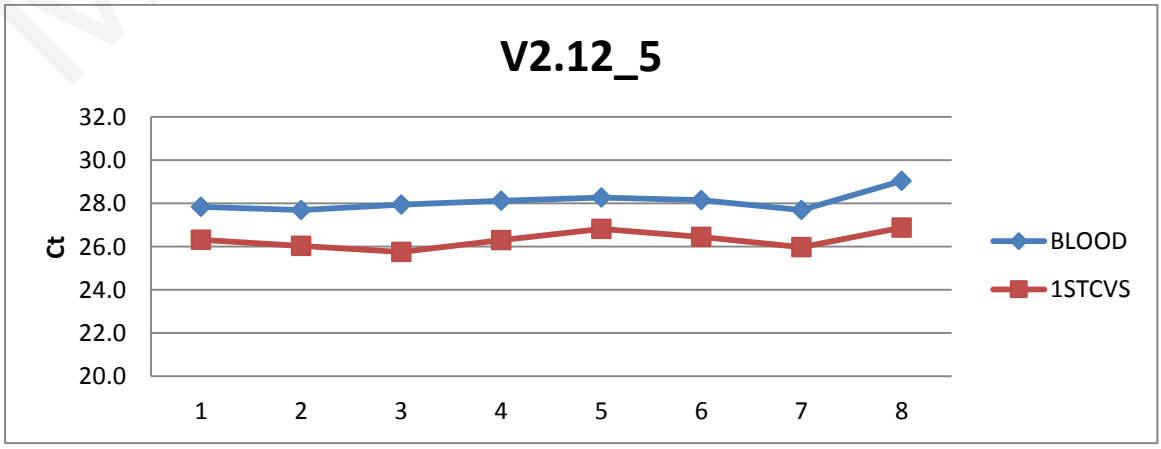


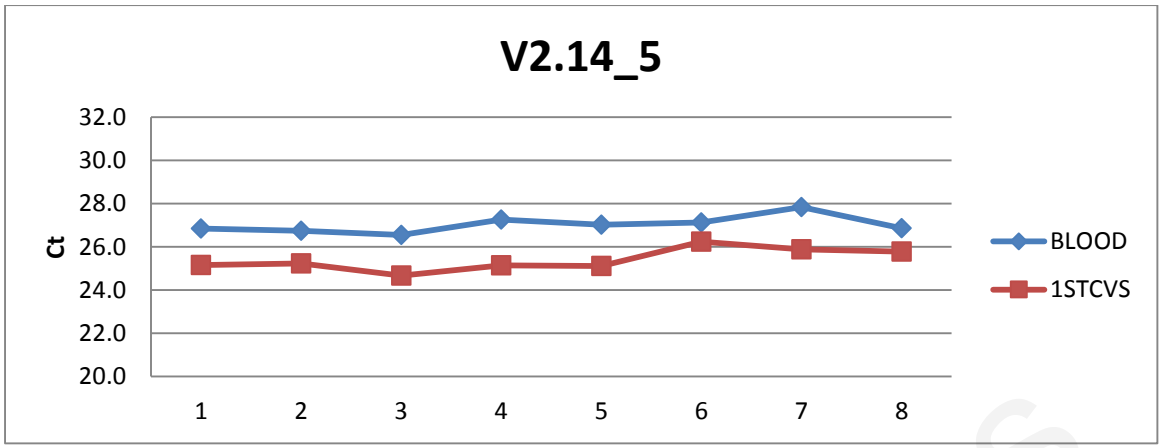


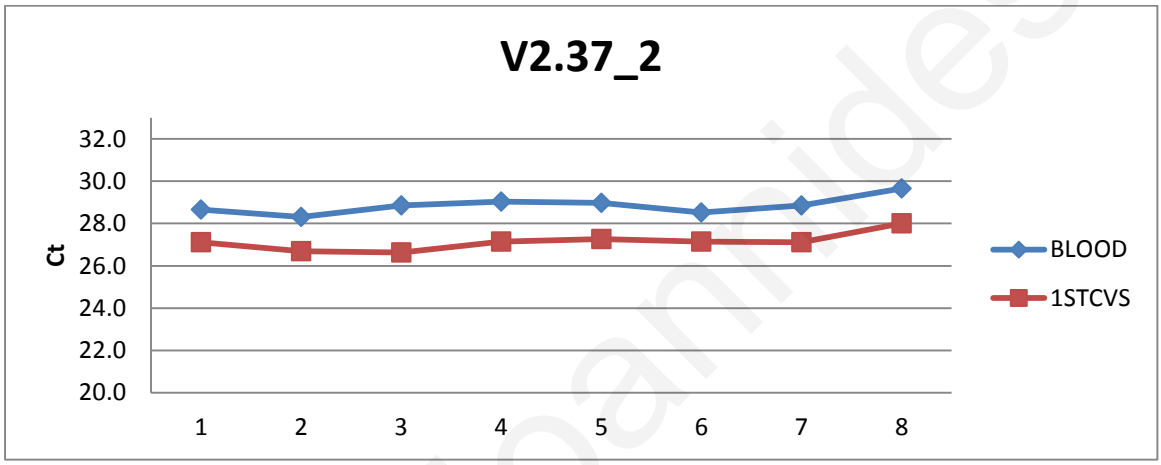


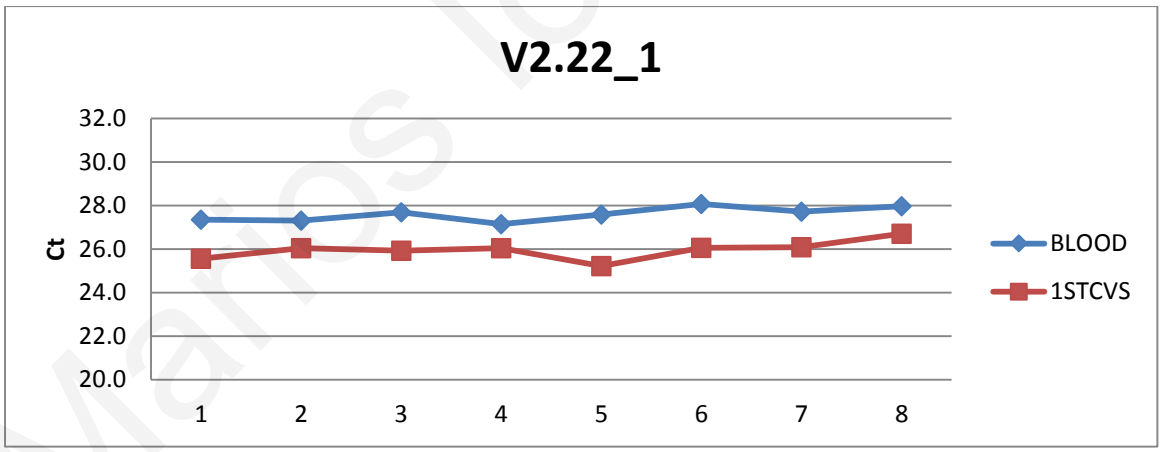


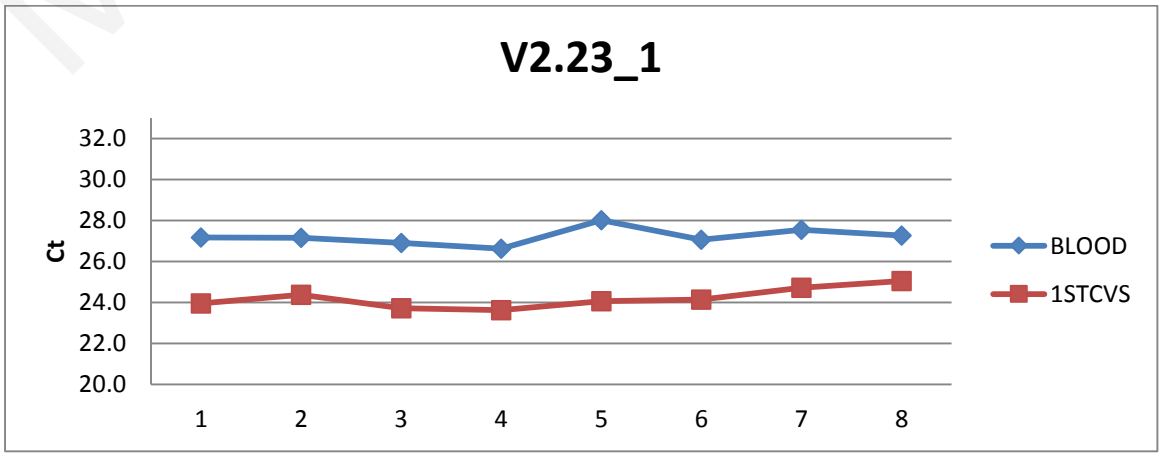


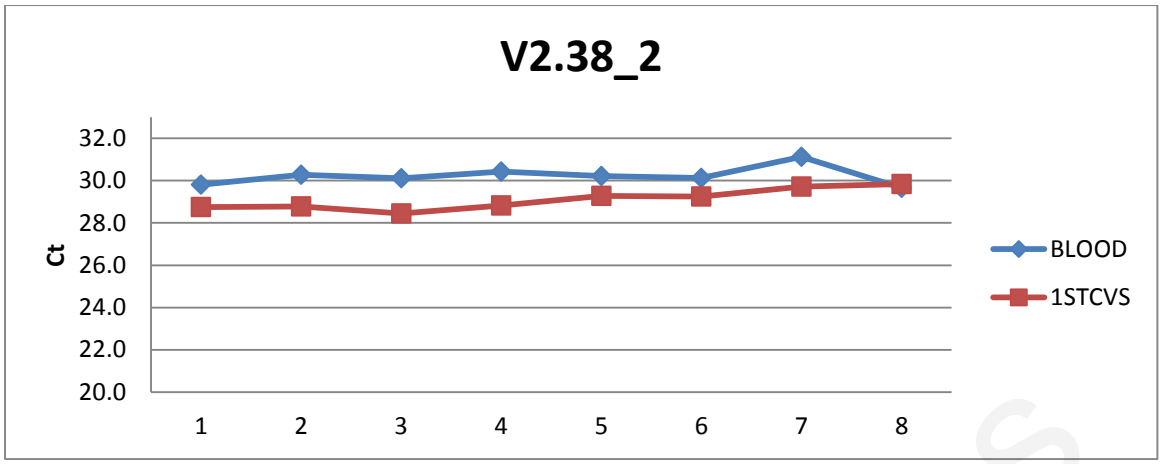


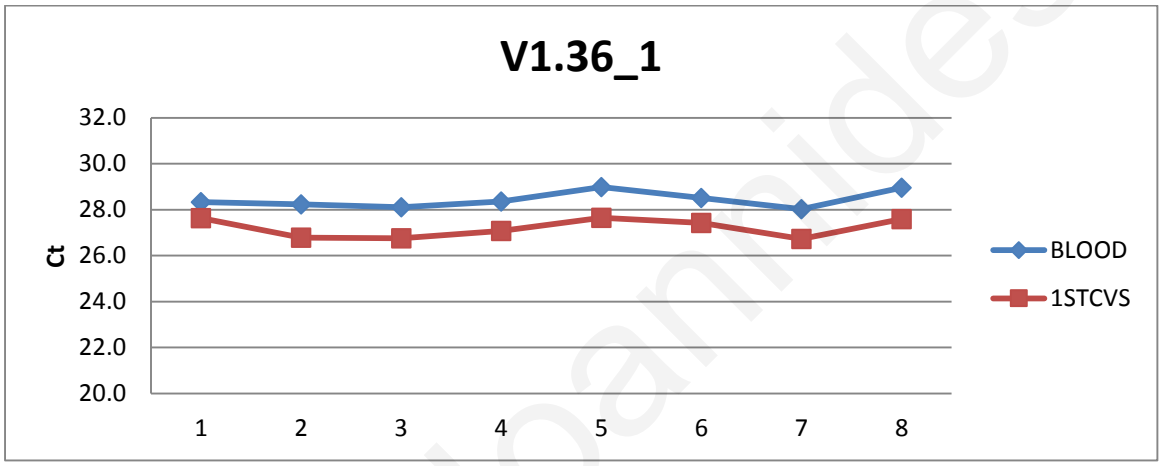


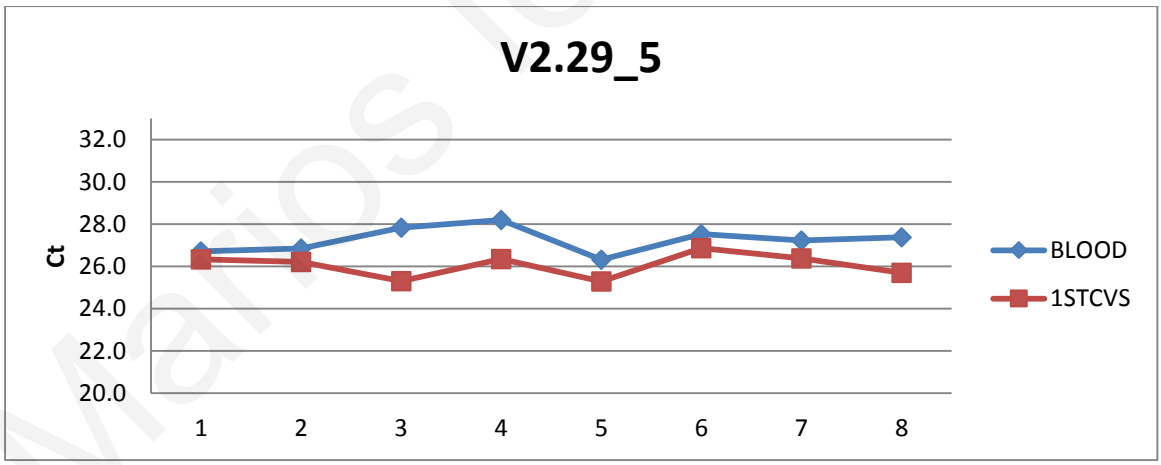


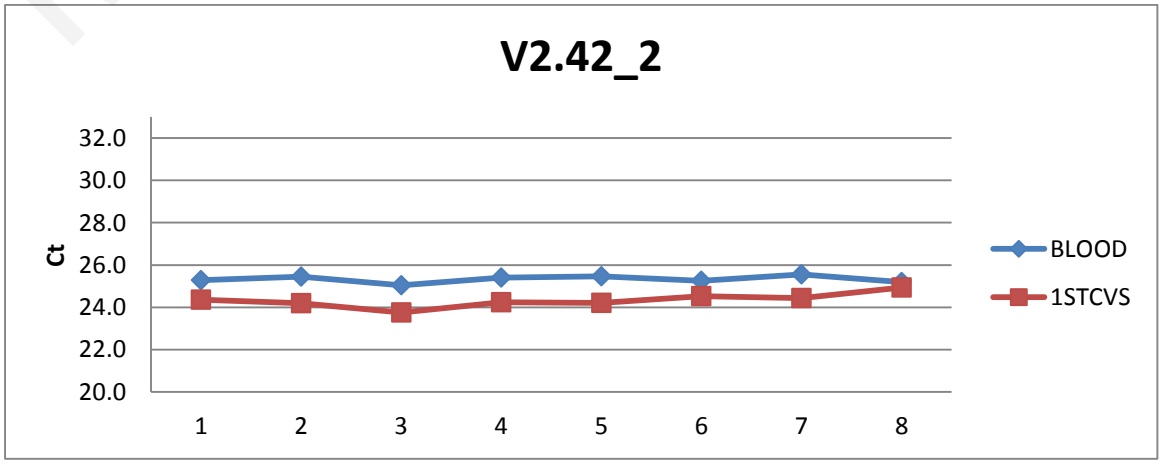


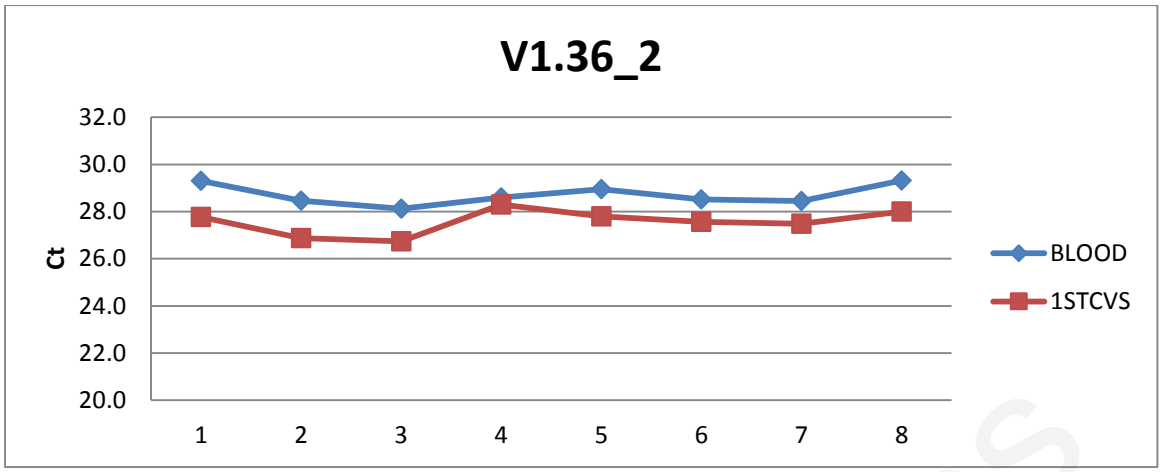


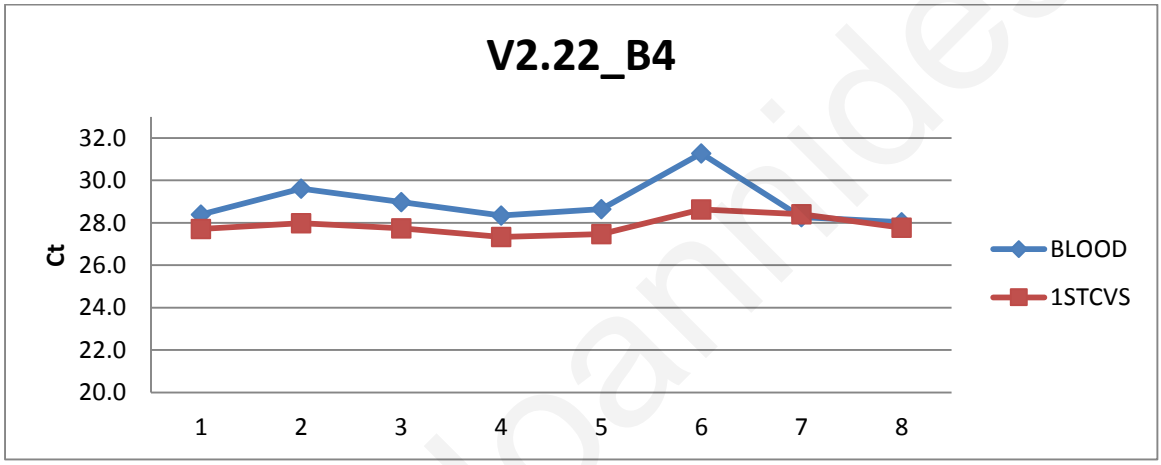




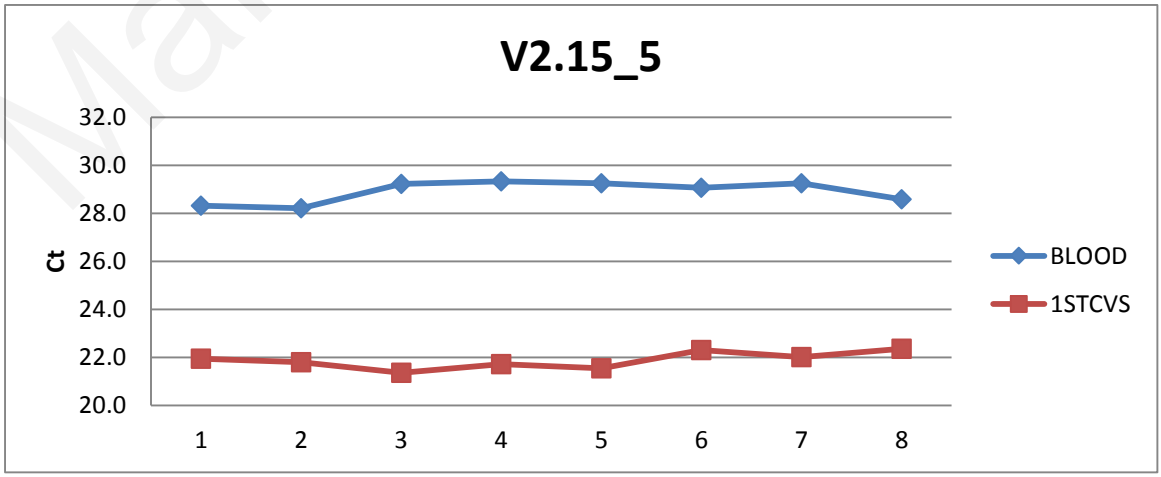


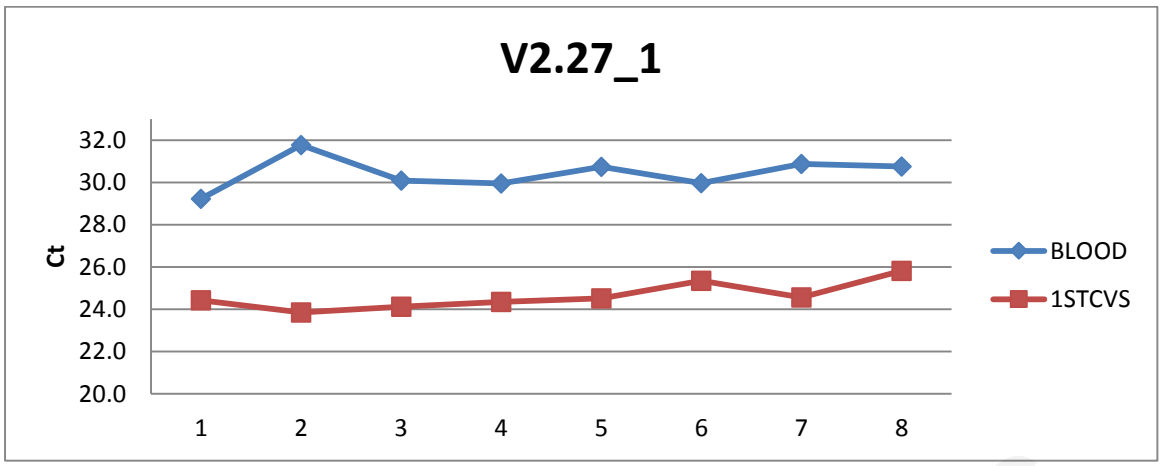


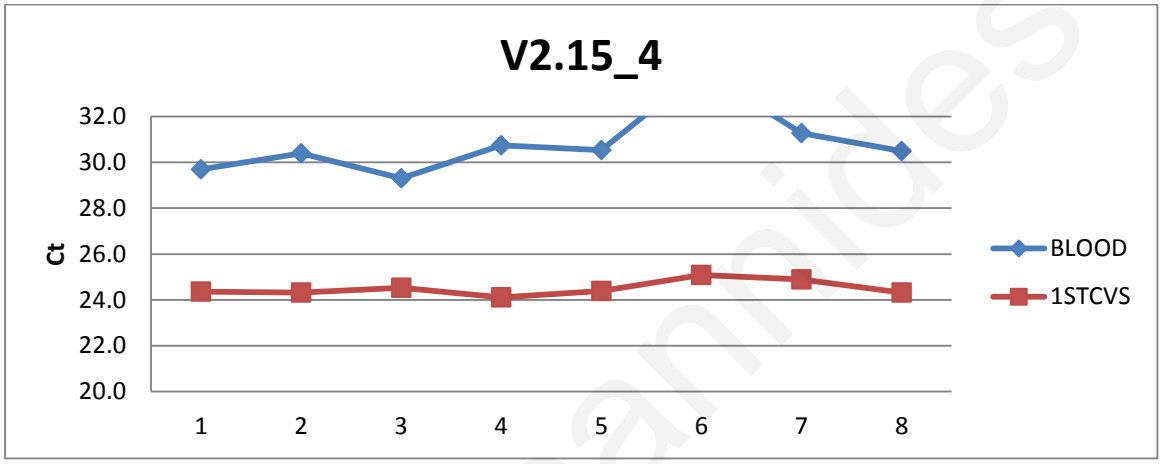


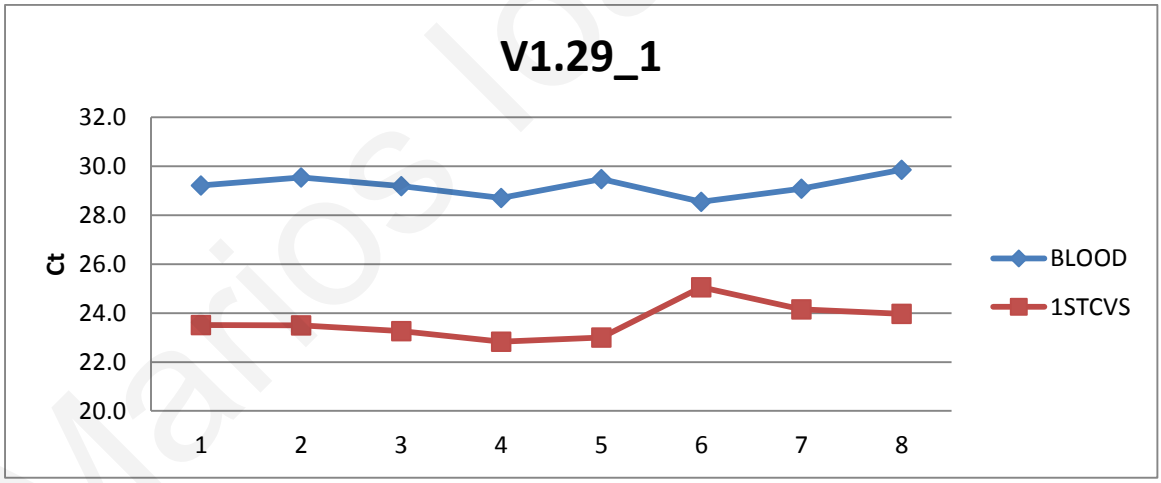


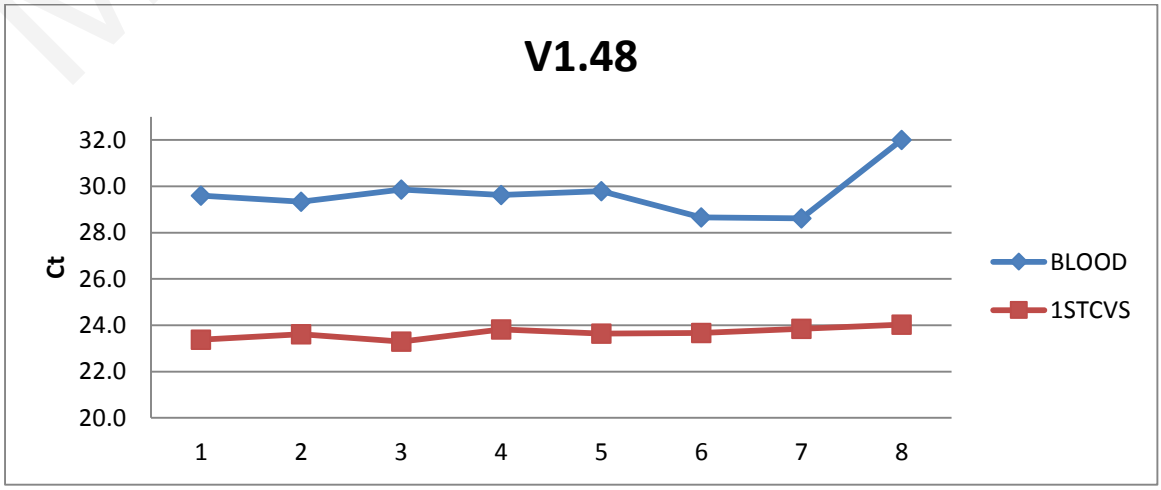
2. ***"GOOD DMRs"***

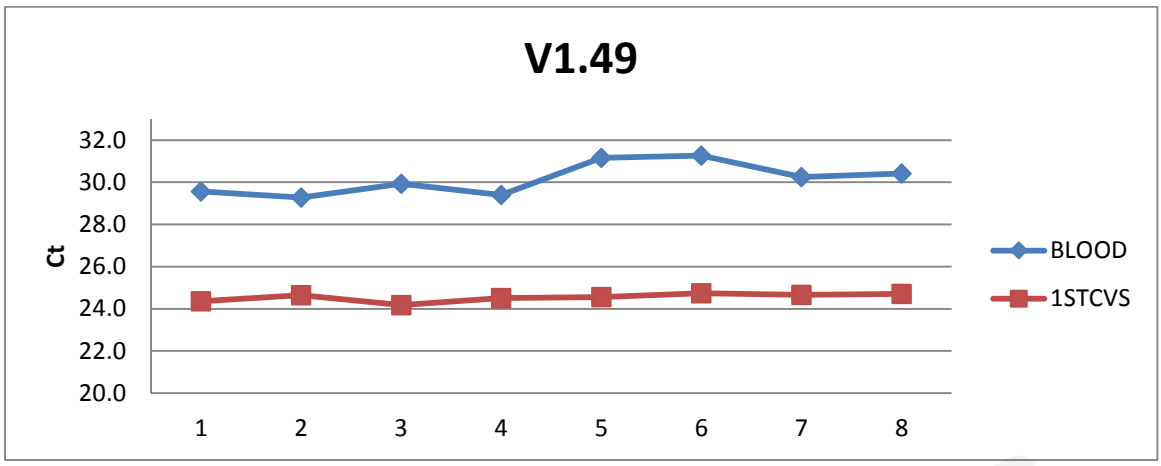


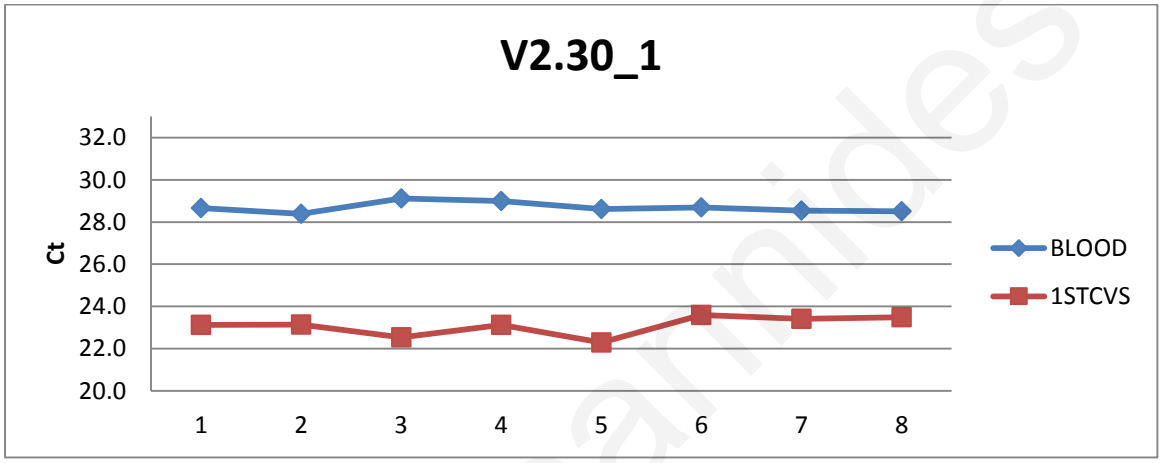


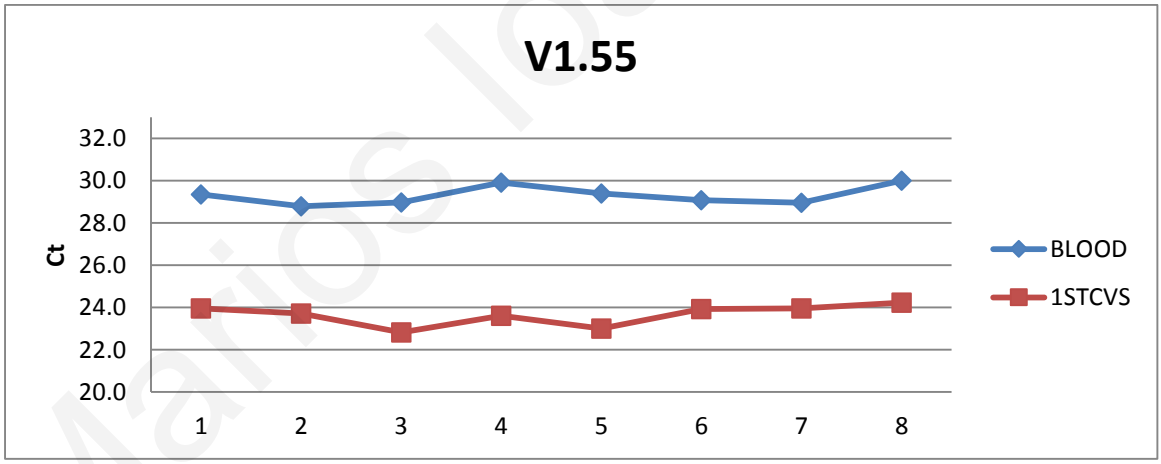


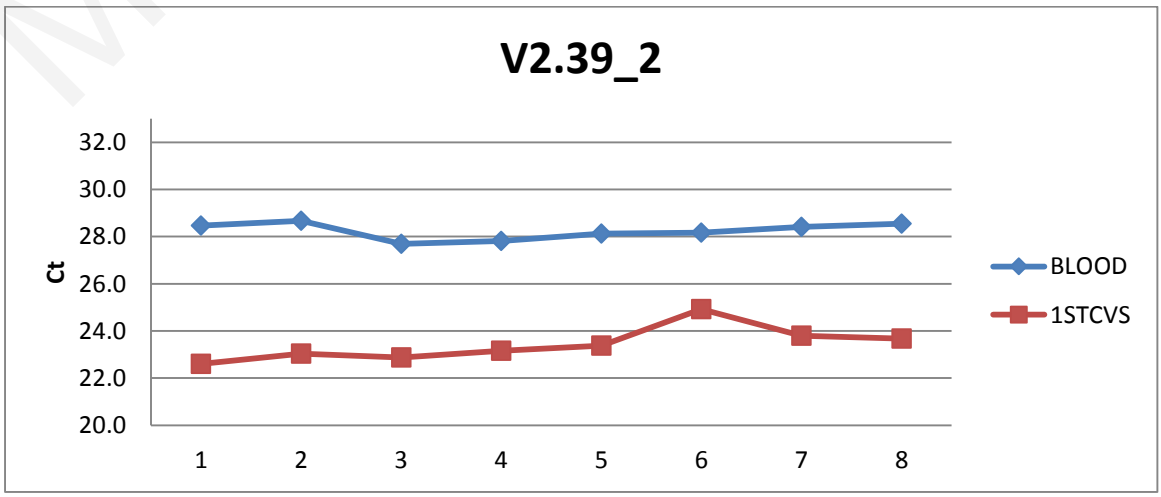


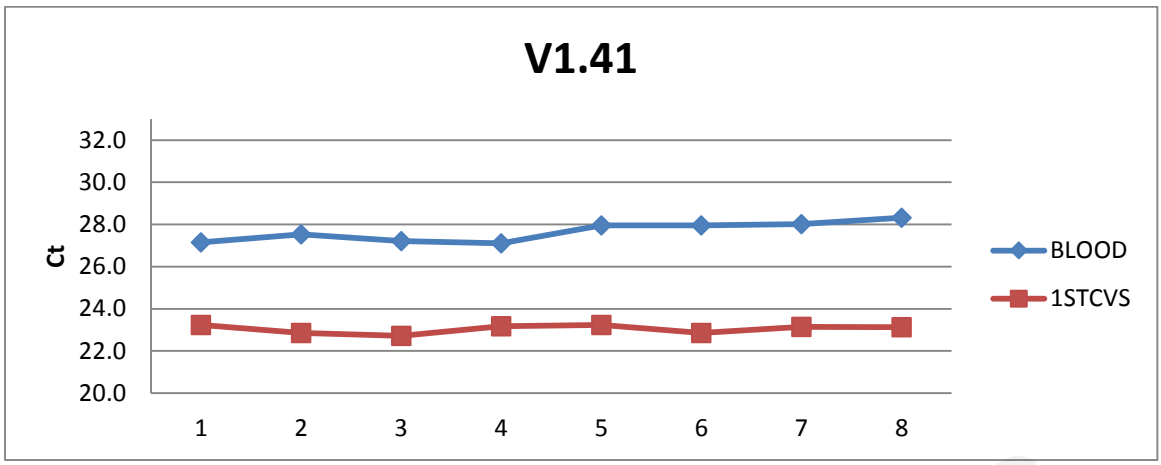


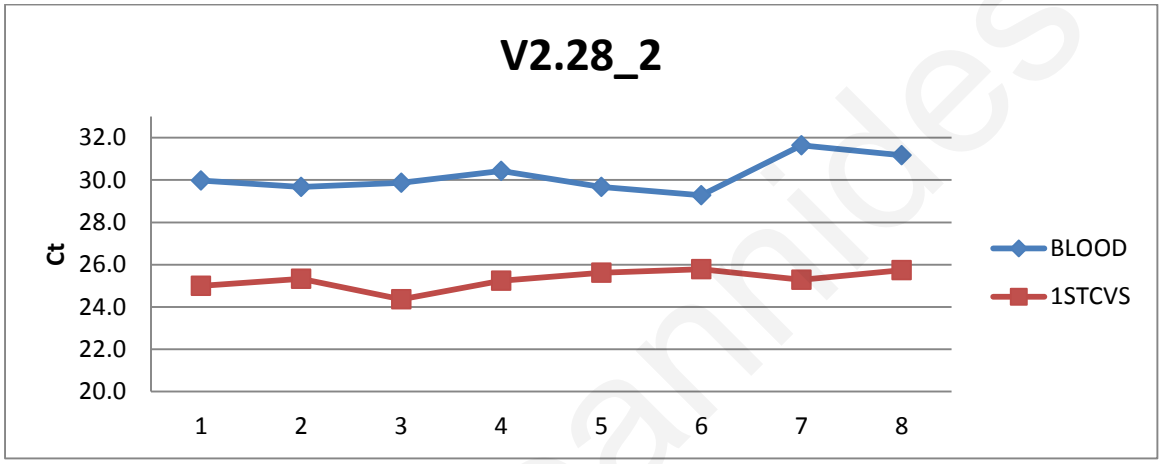


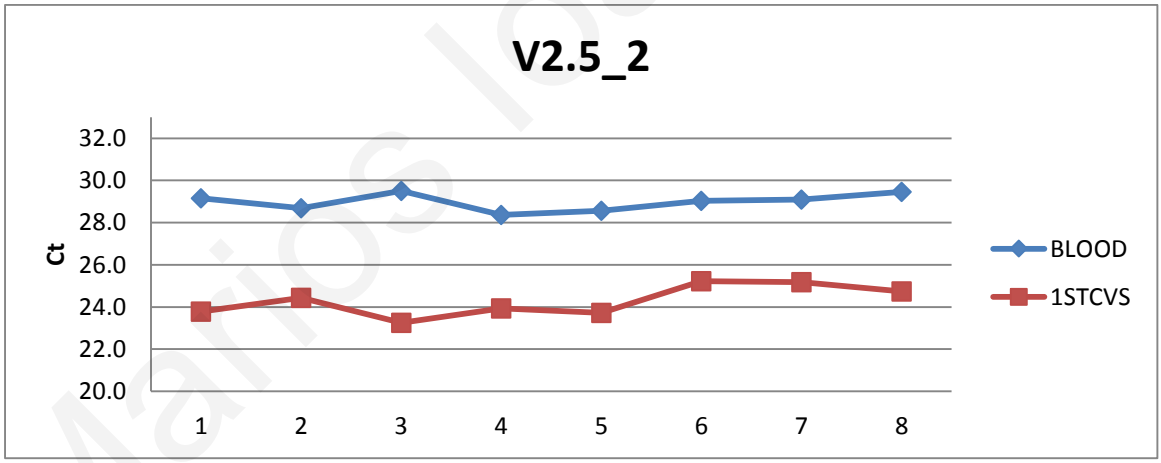


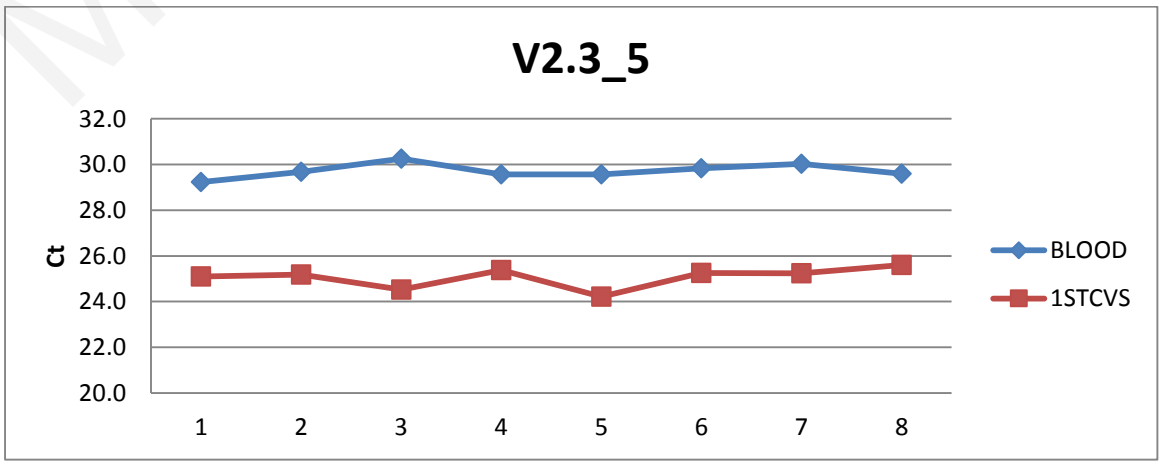


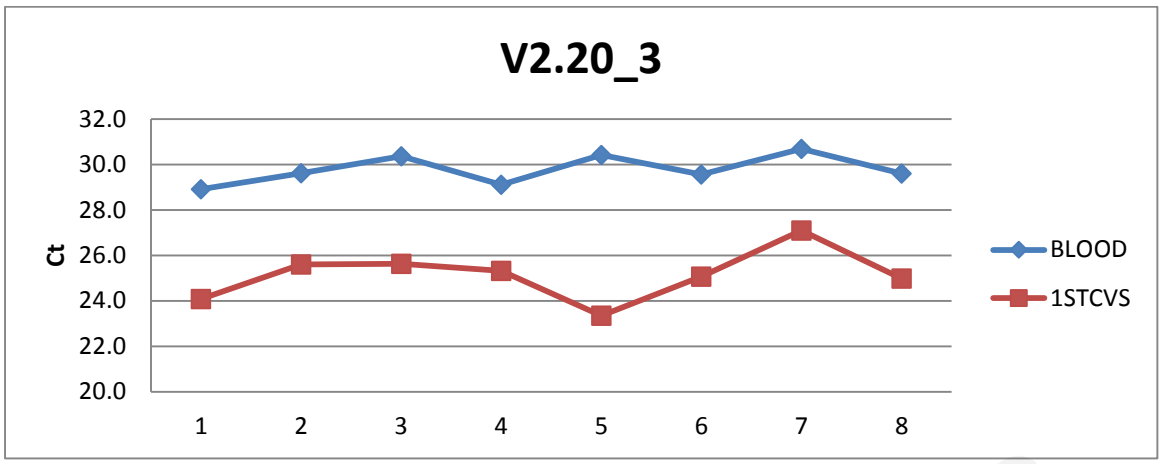


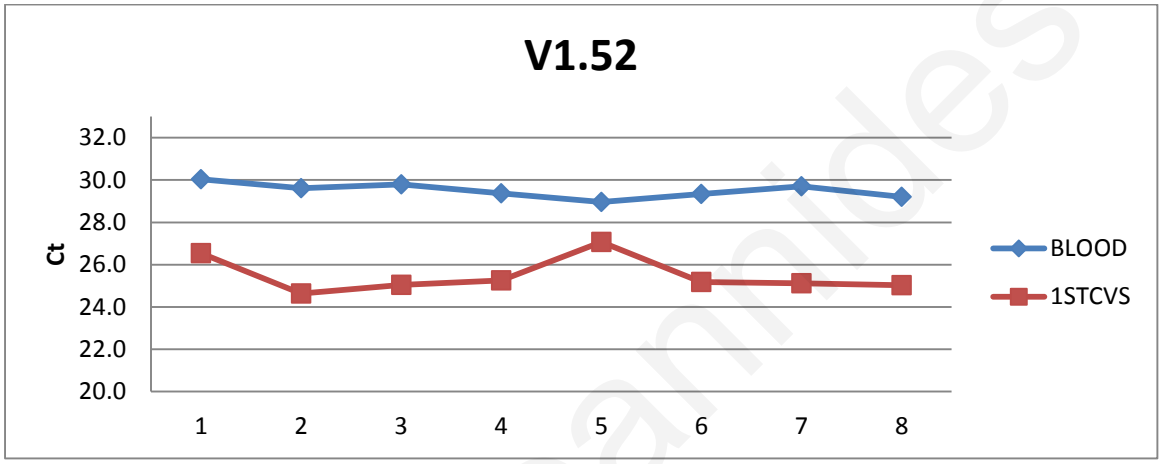


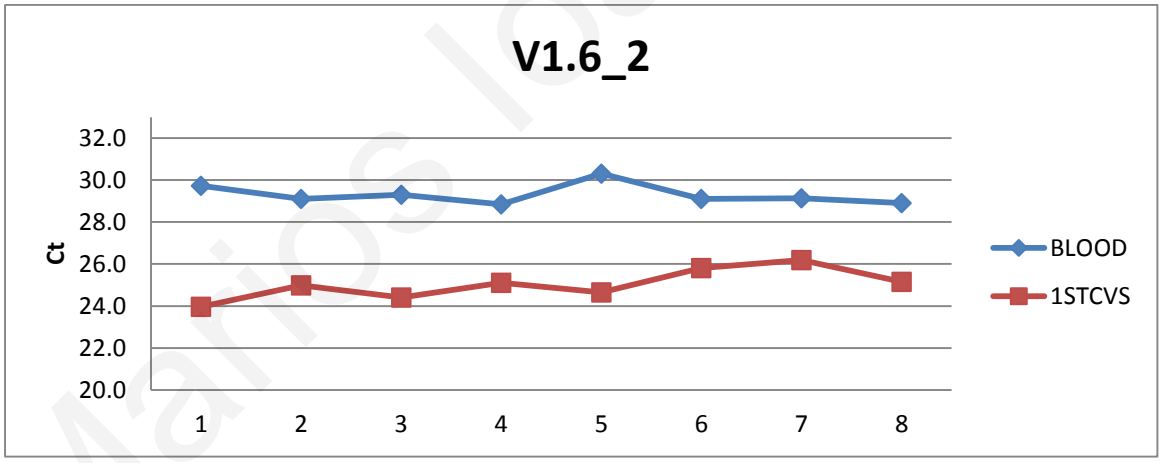


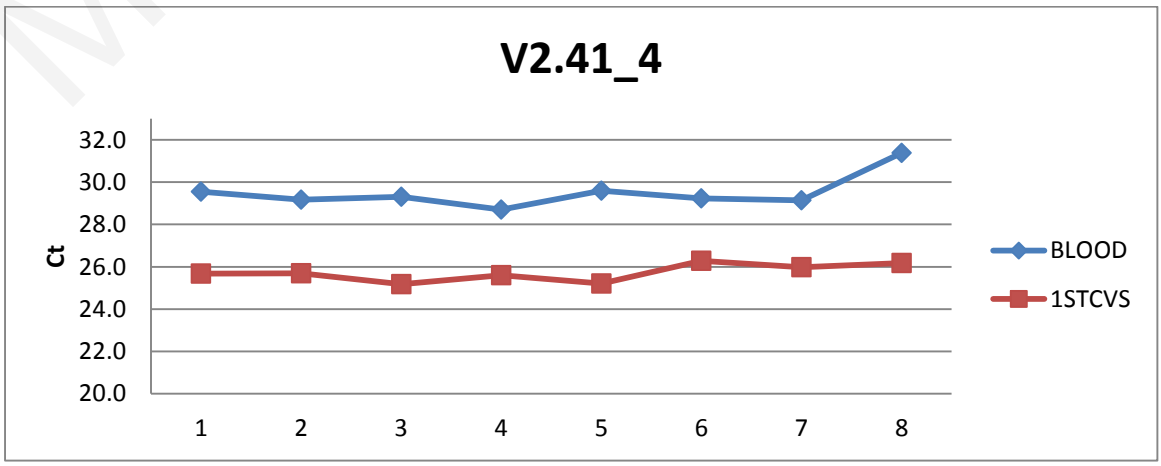


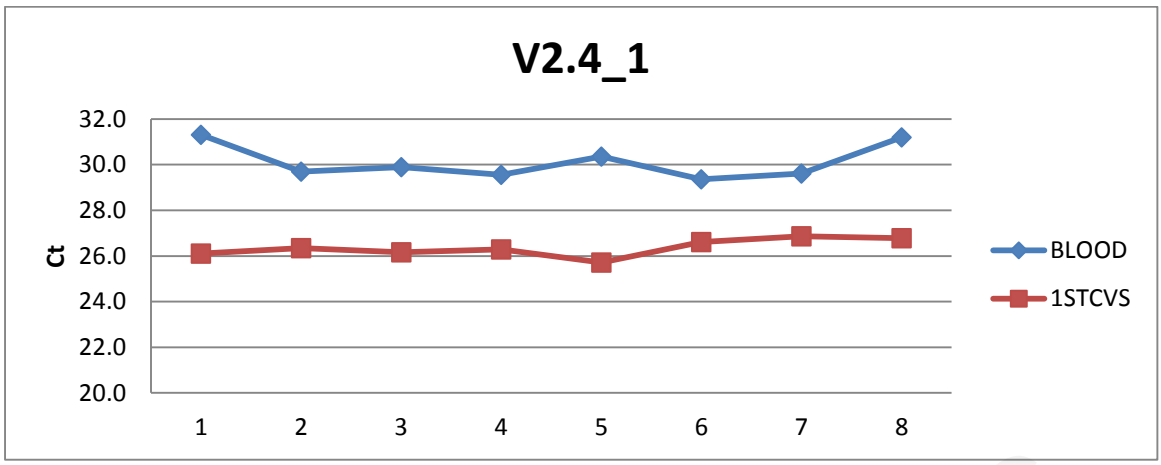


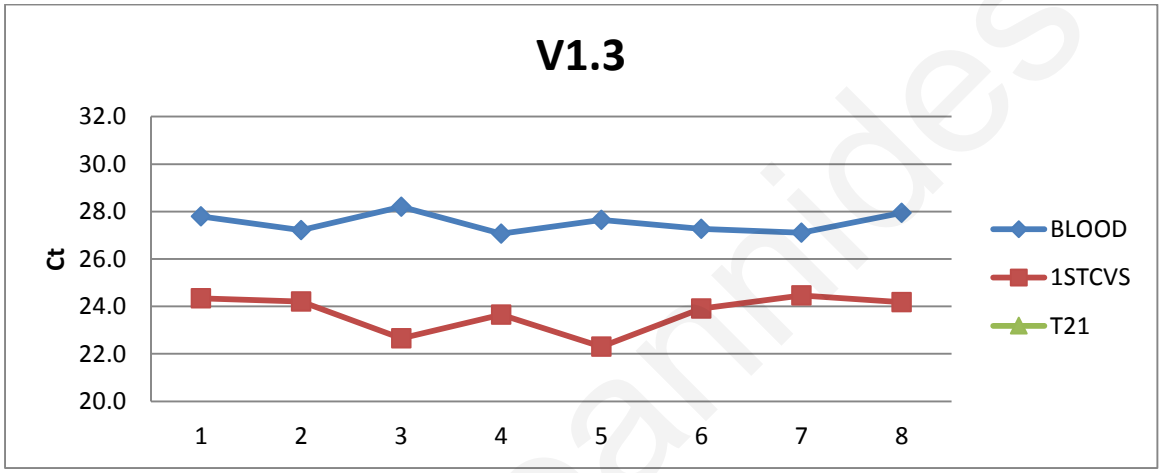


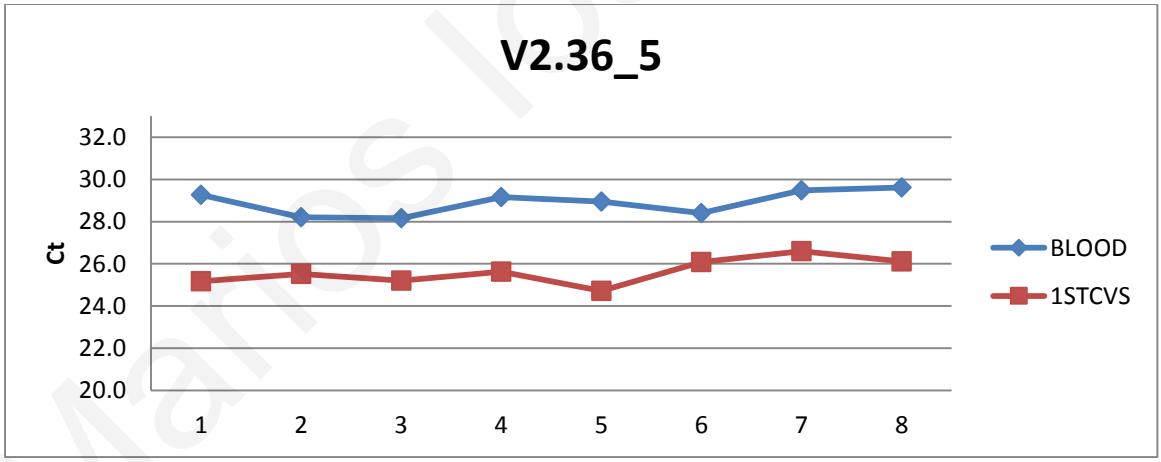


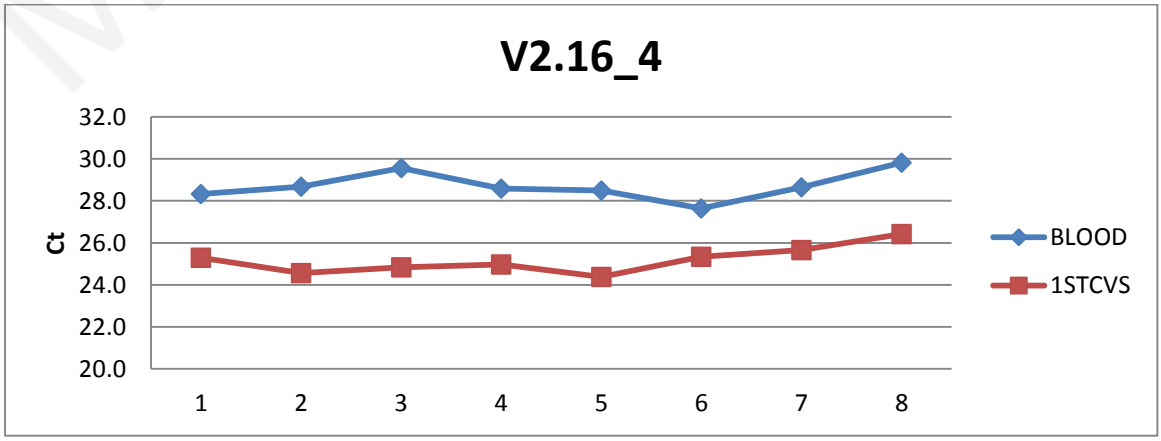


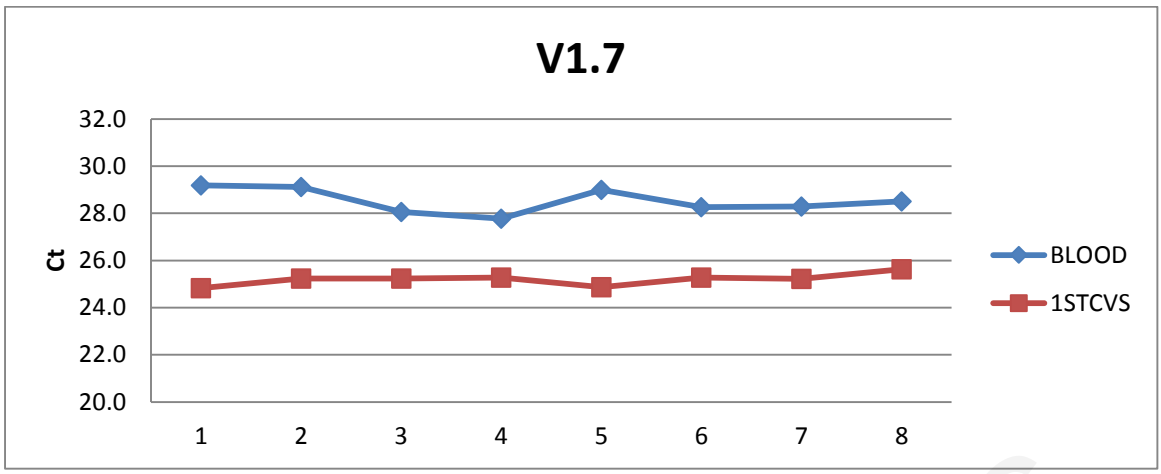


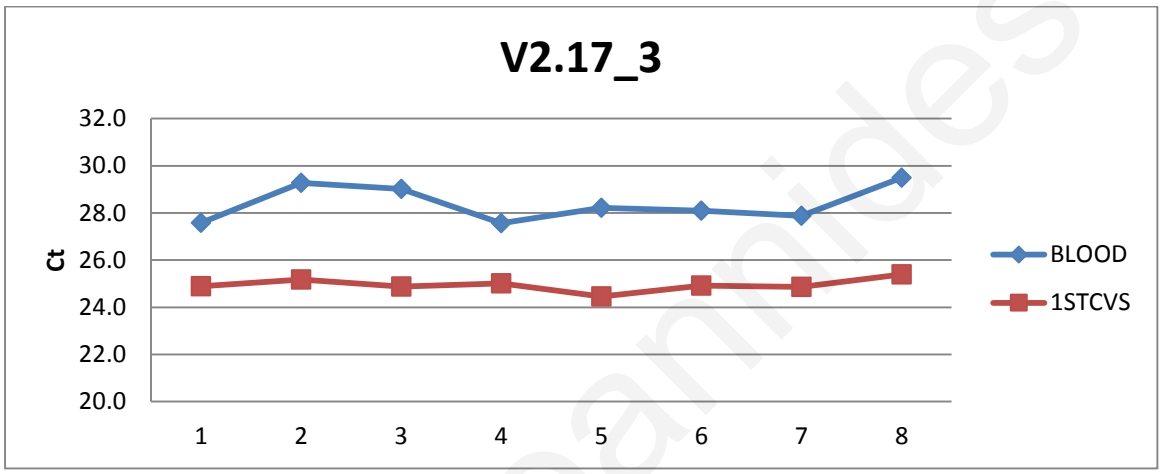


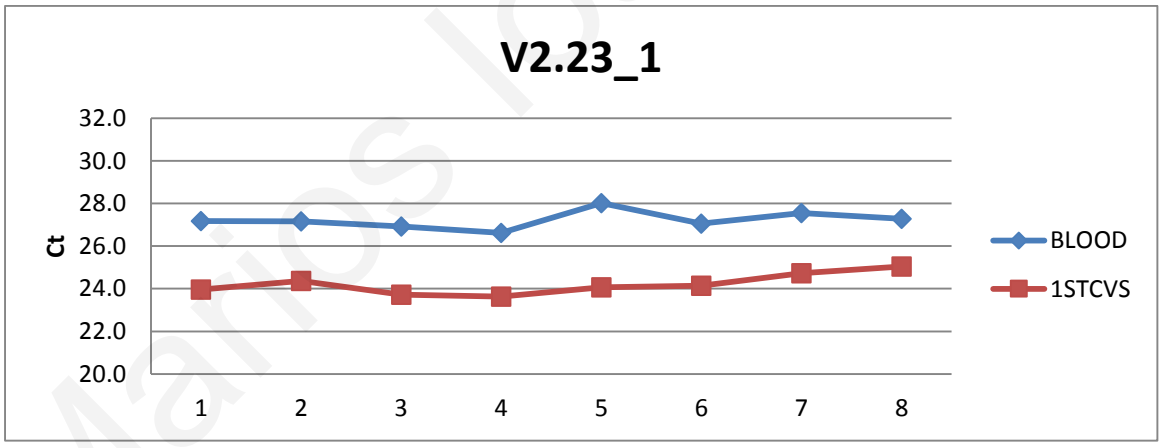


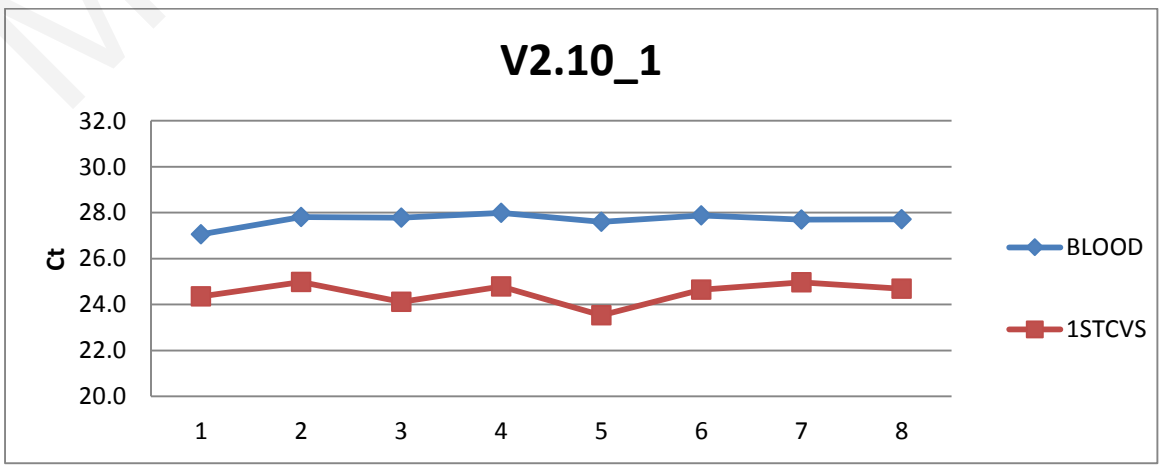


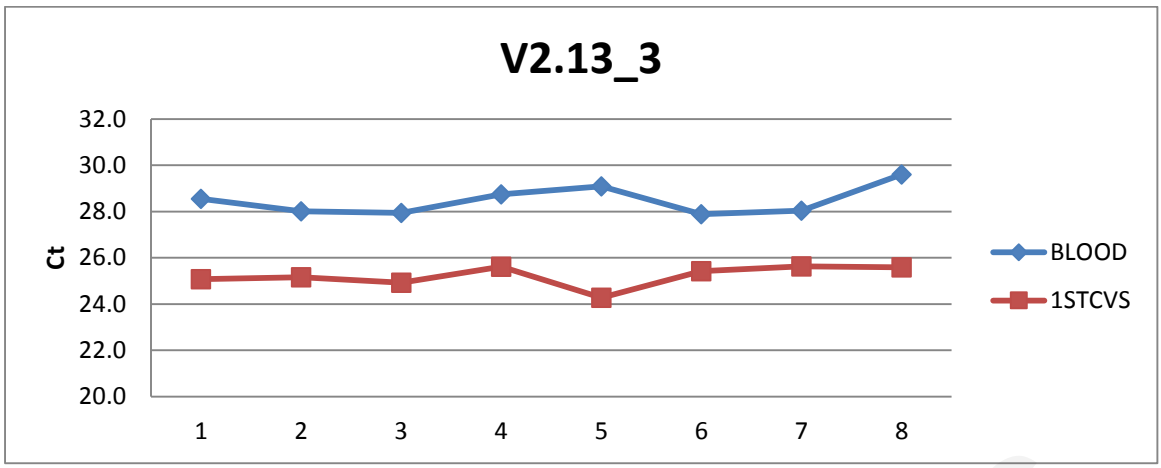


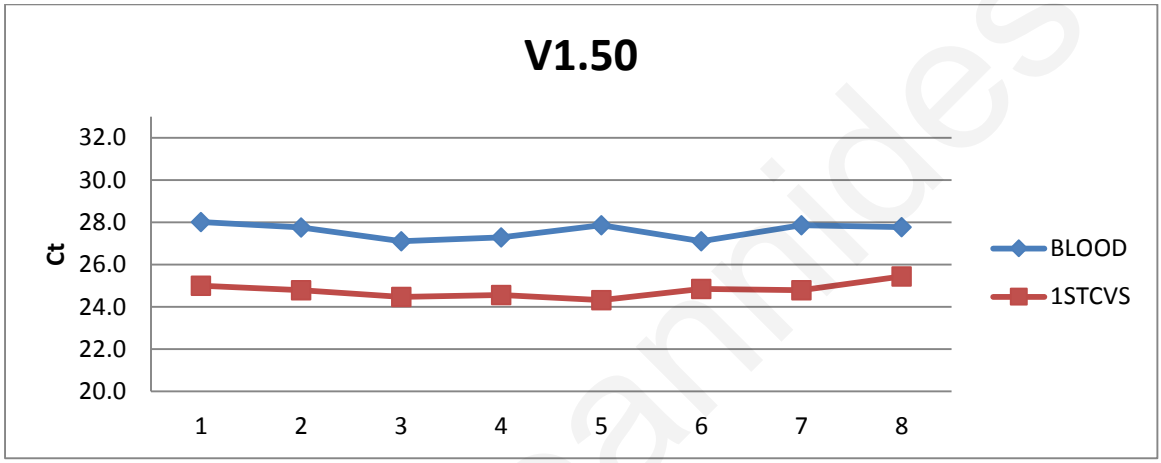






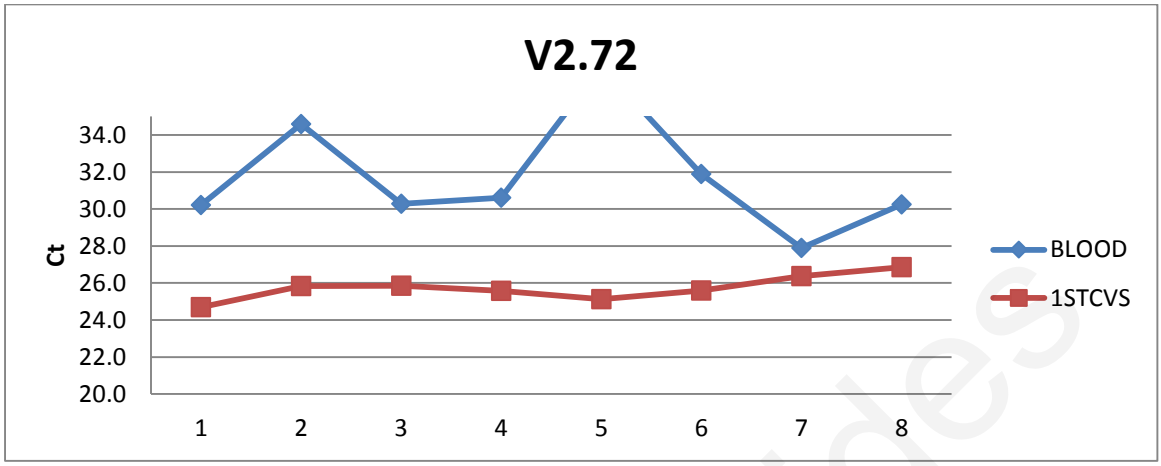


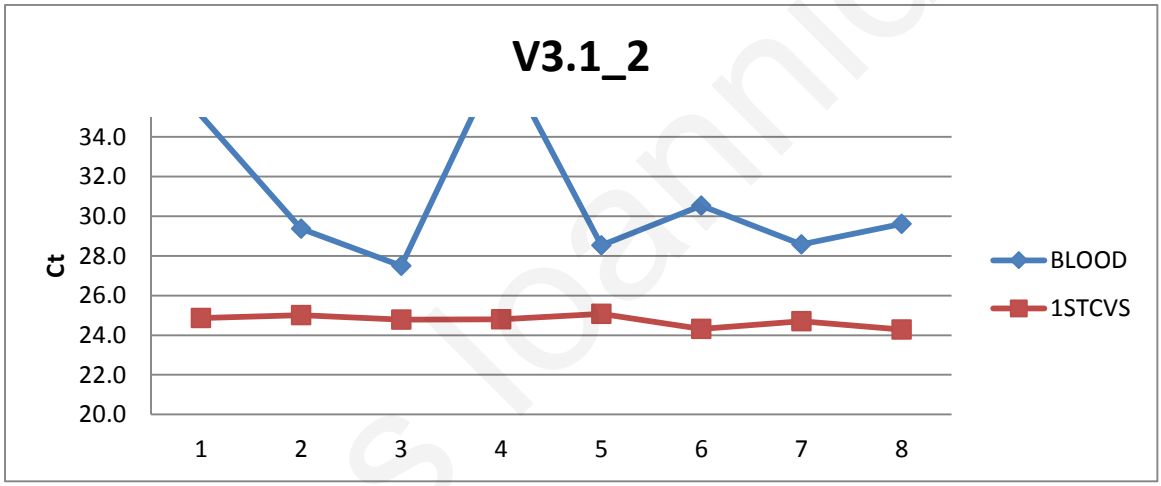


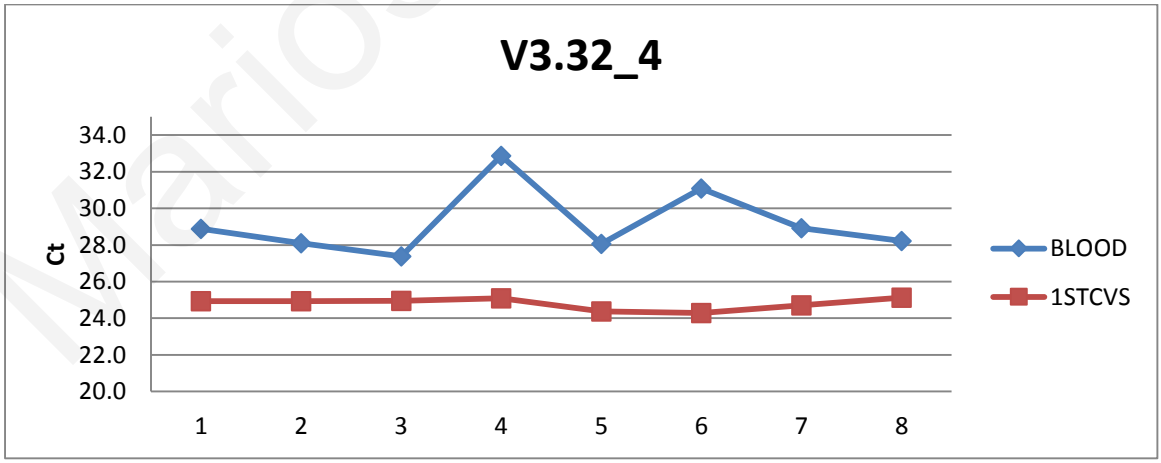


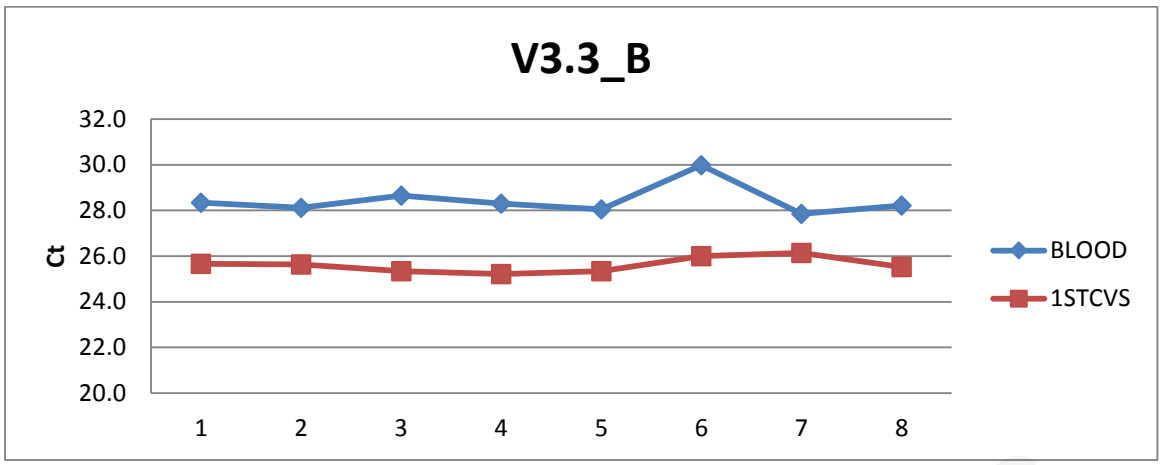
B. Chromosome 18

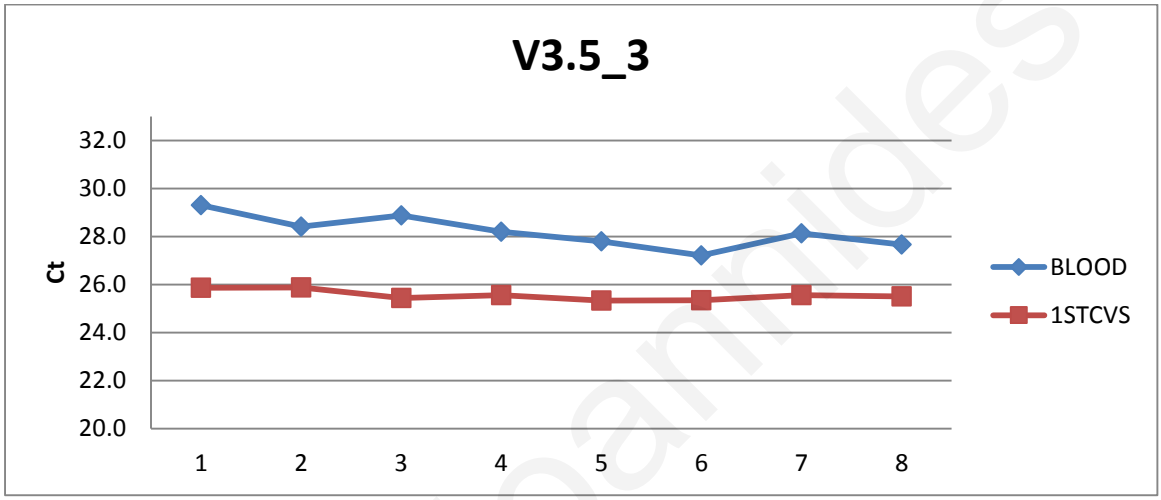
1. "Bad DMRs"

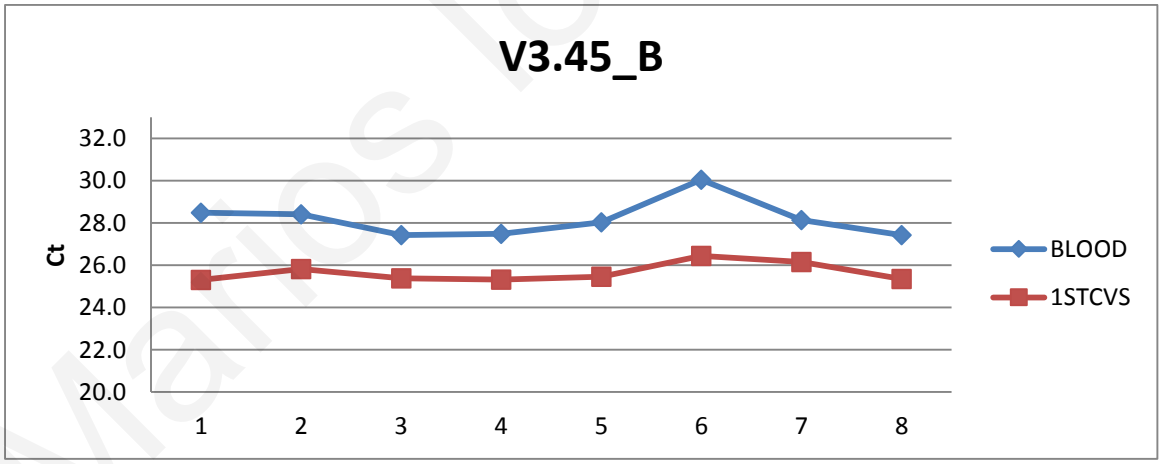


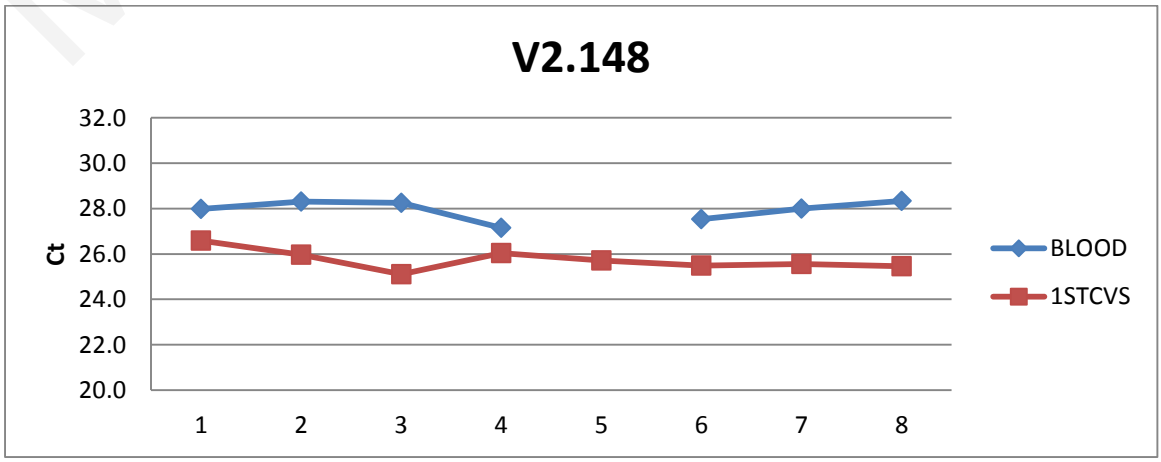


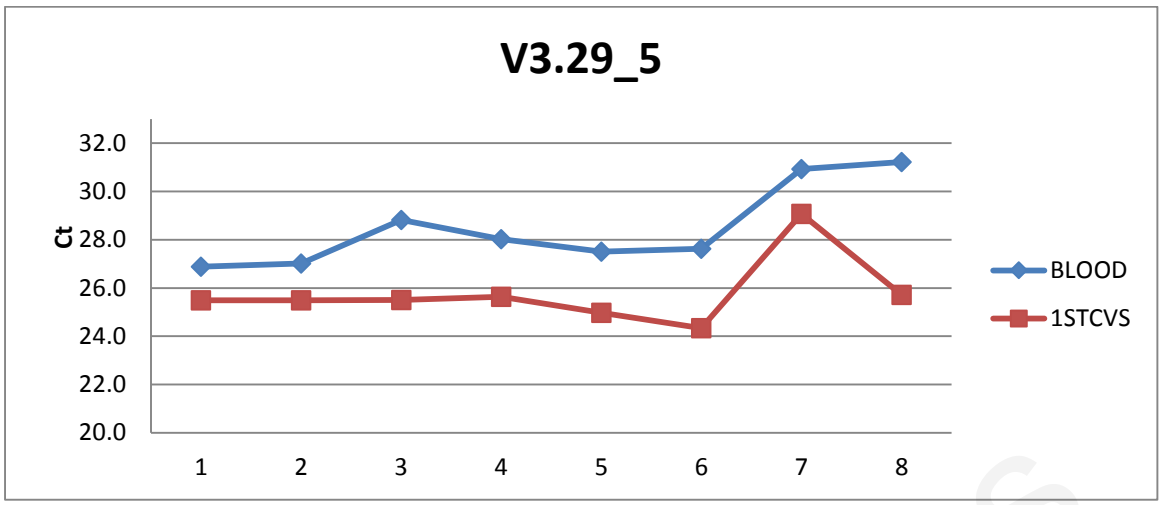


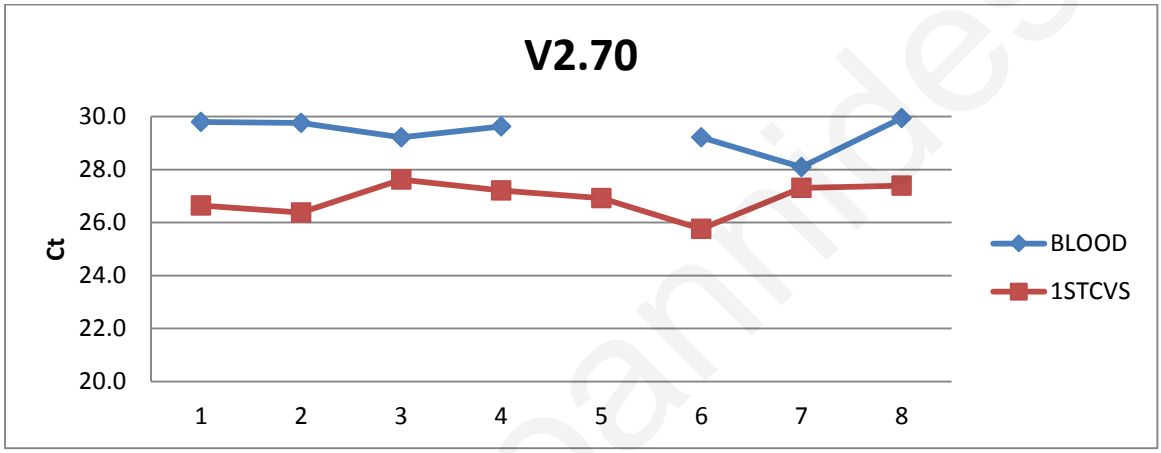


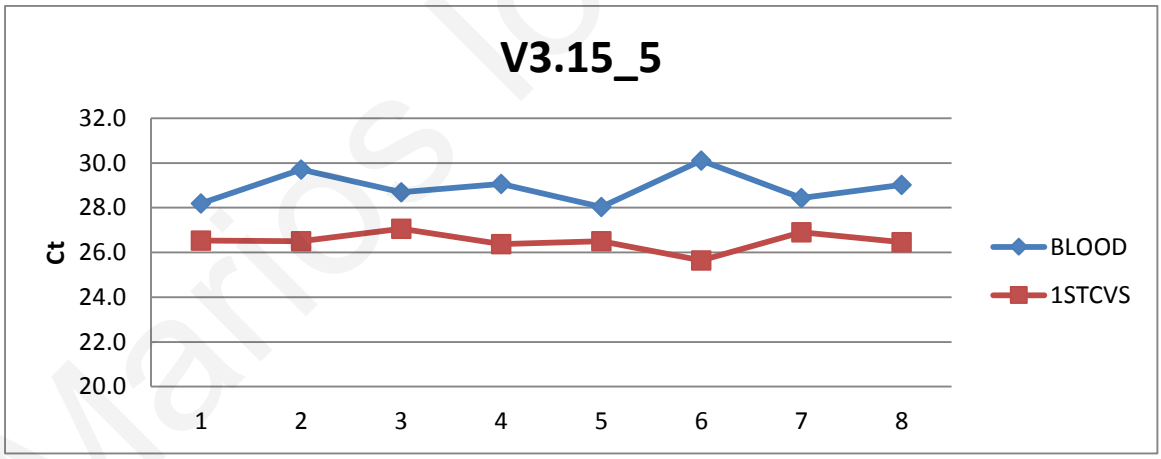


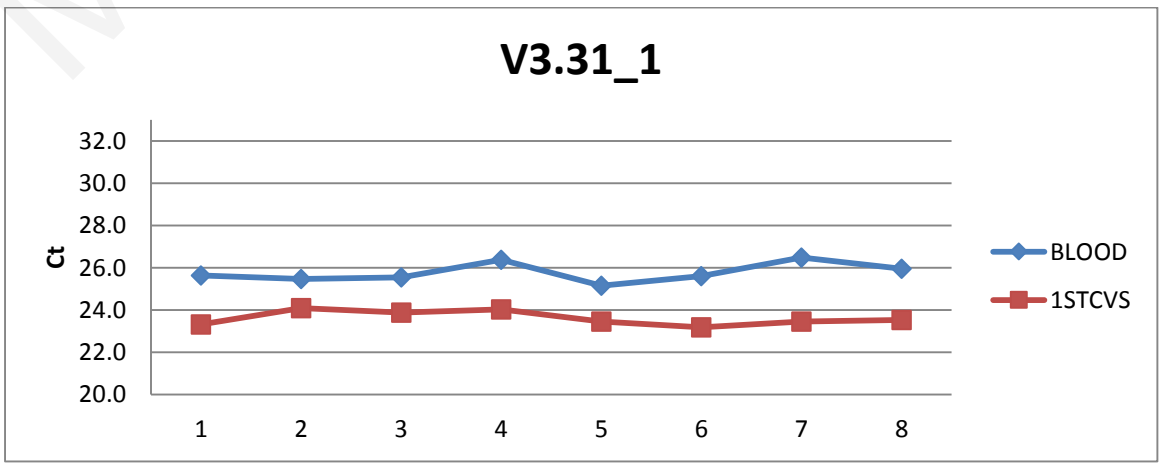


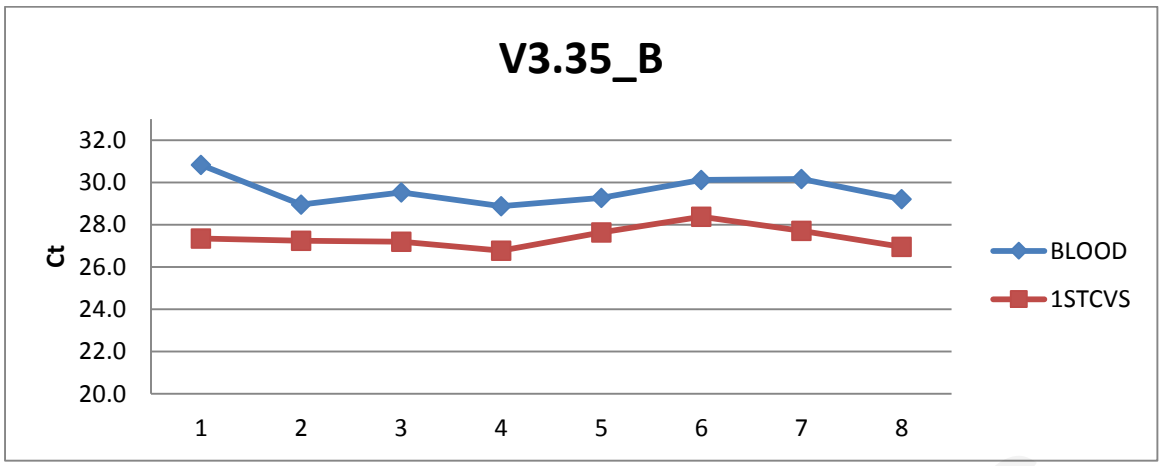


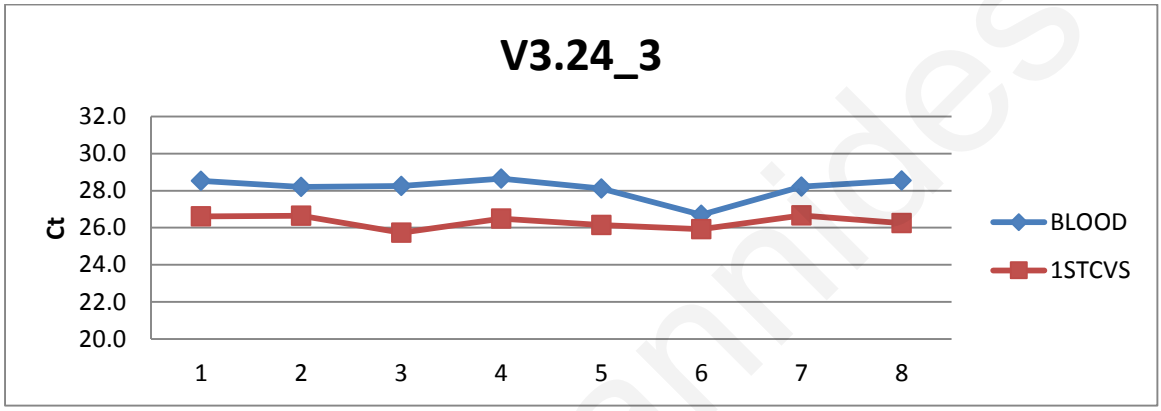


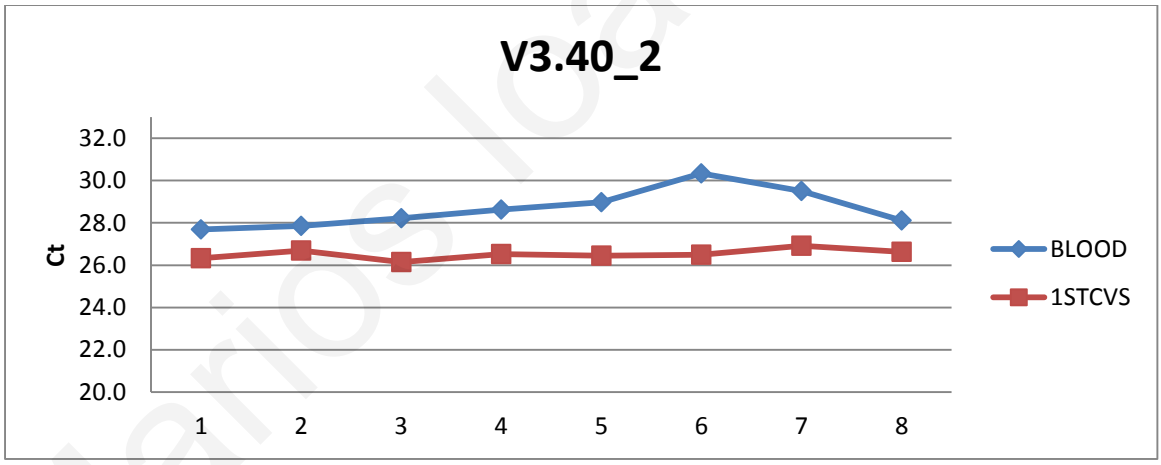


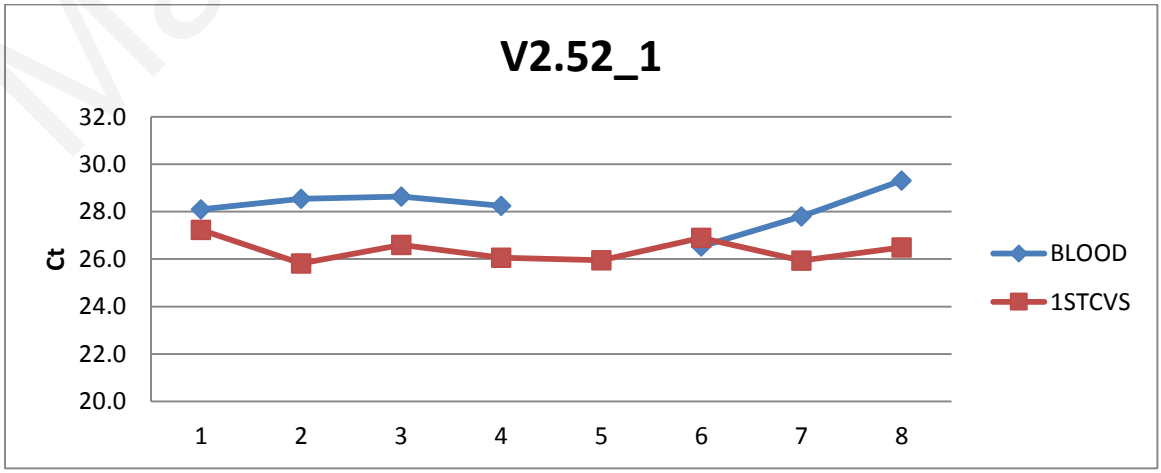


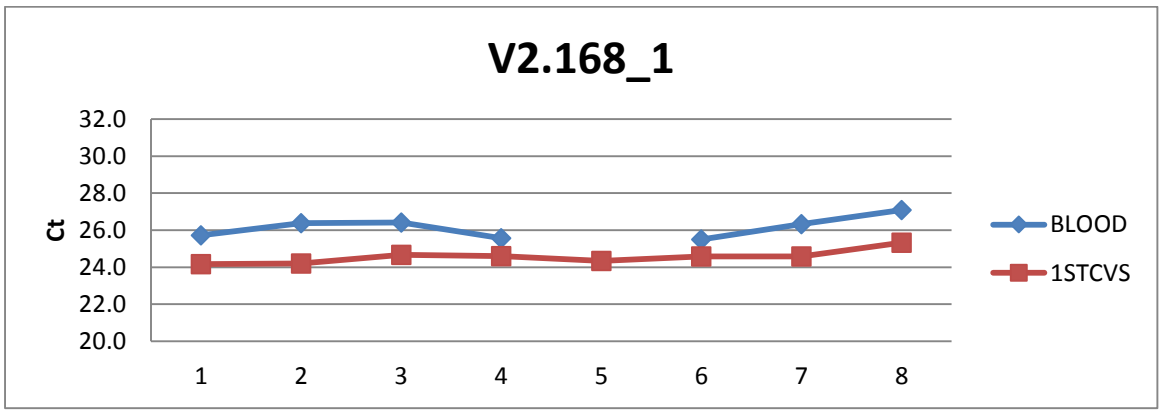


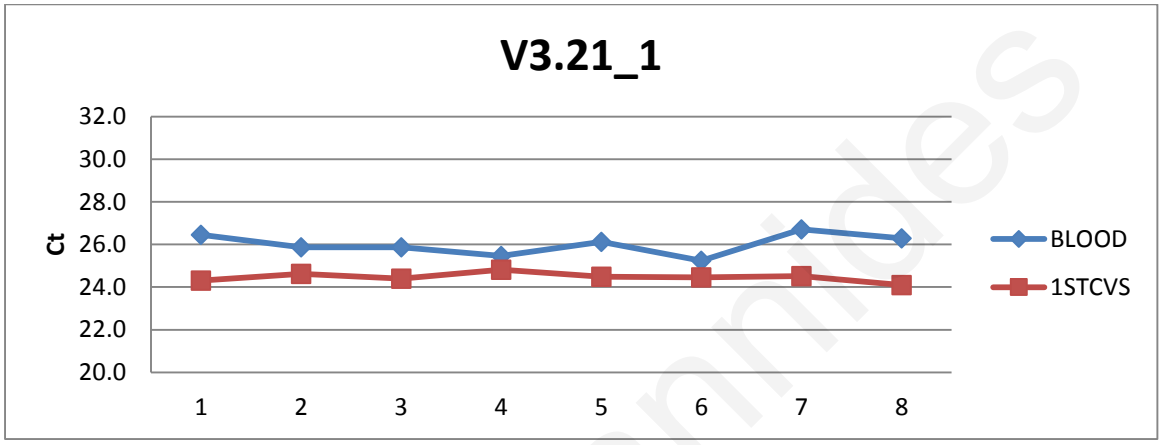


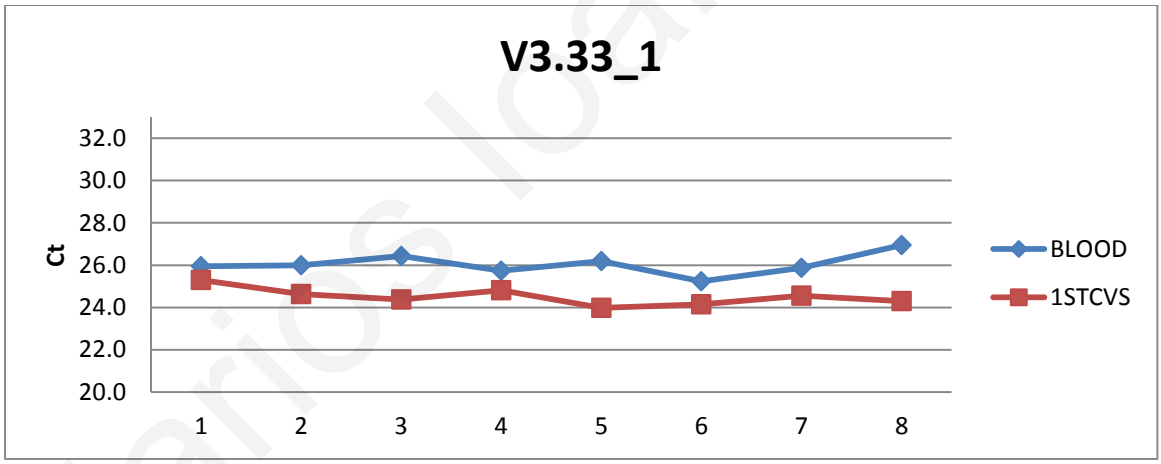


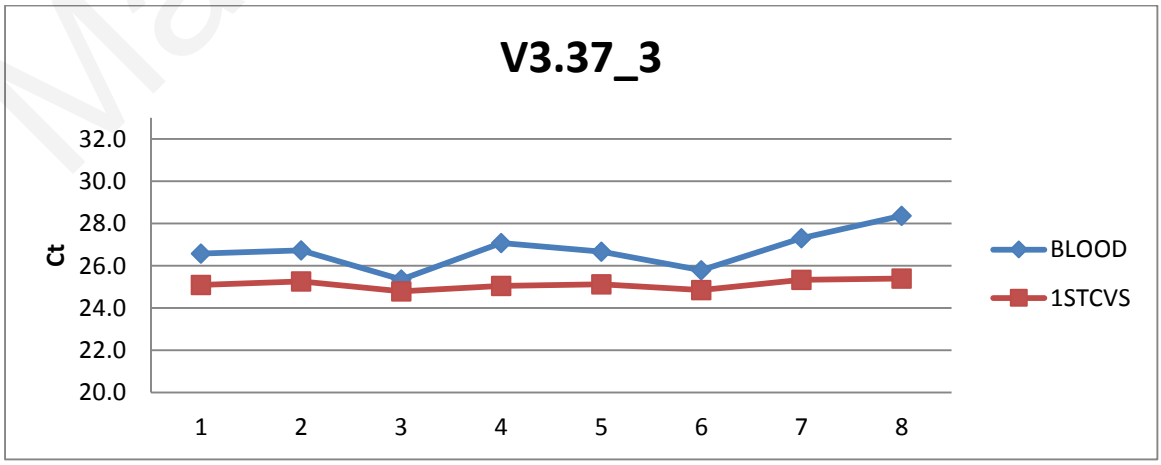


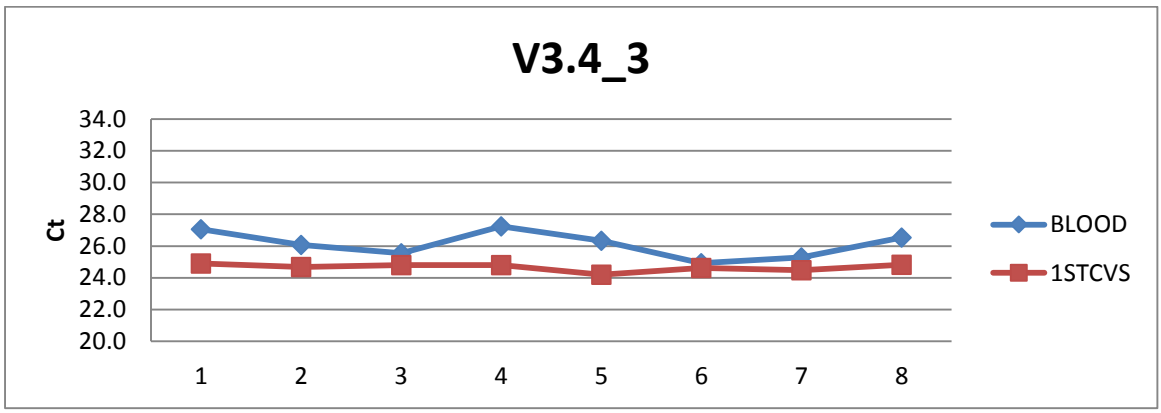


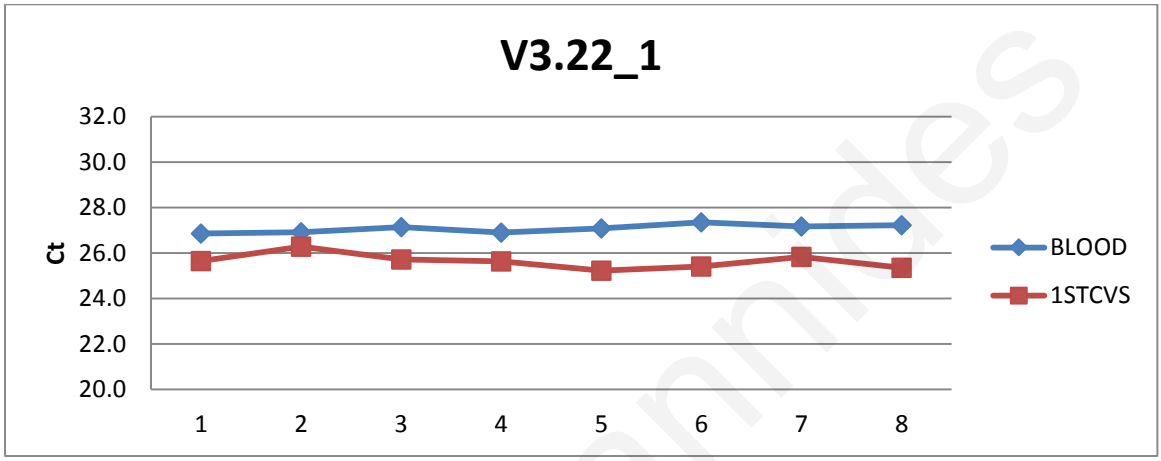


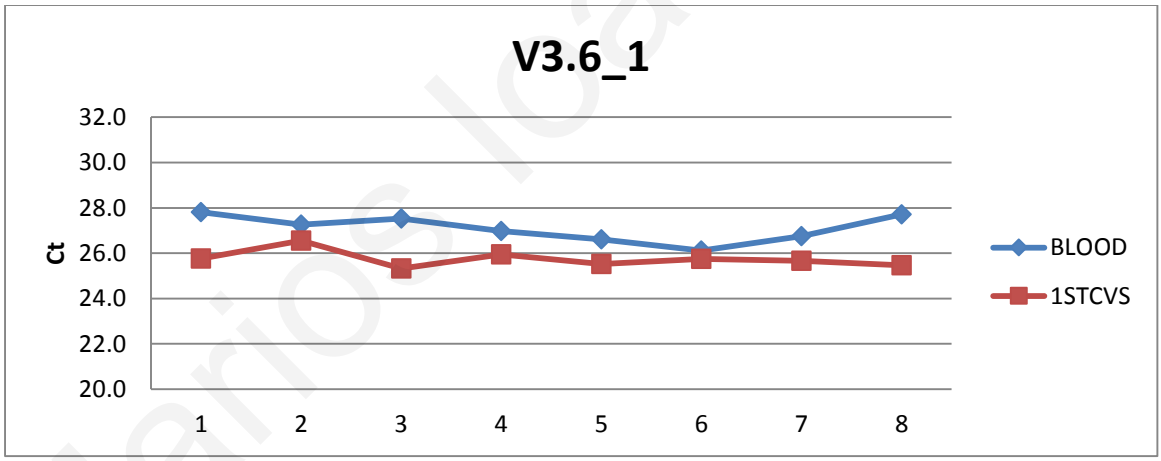


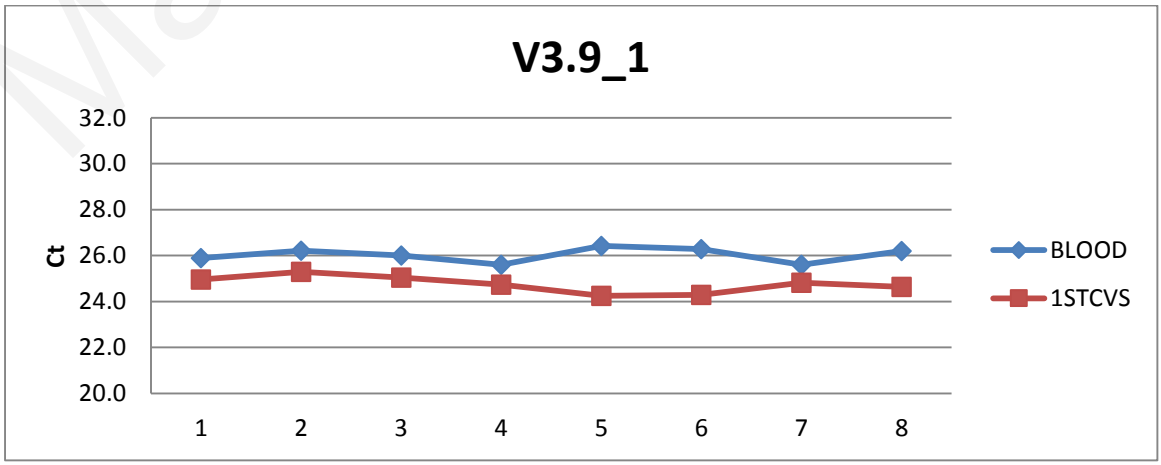


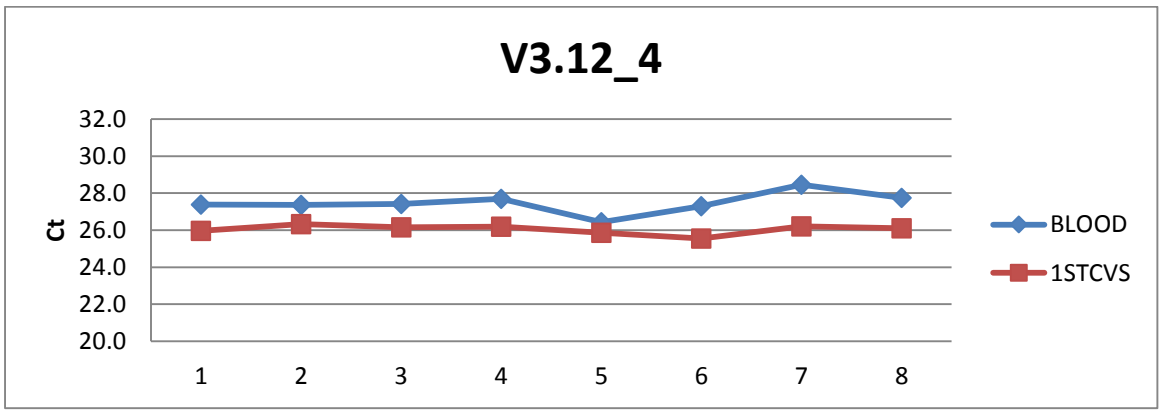


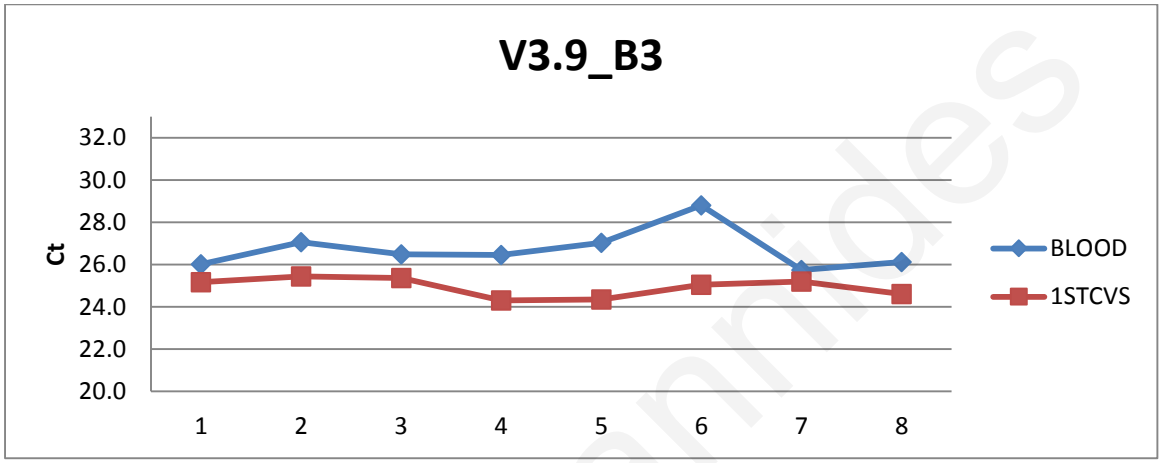


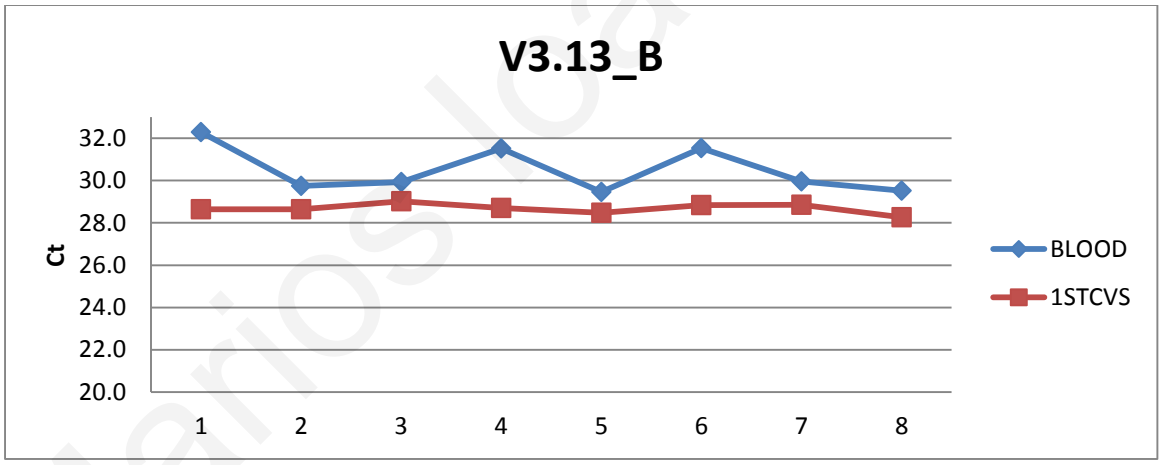


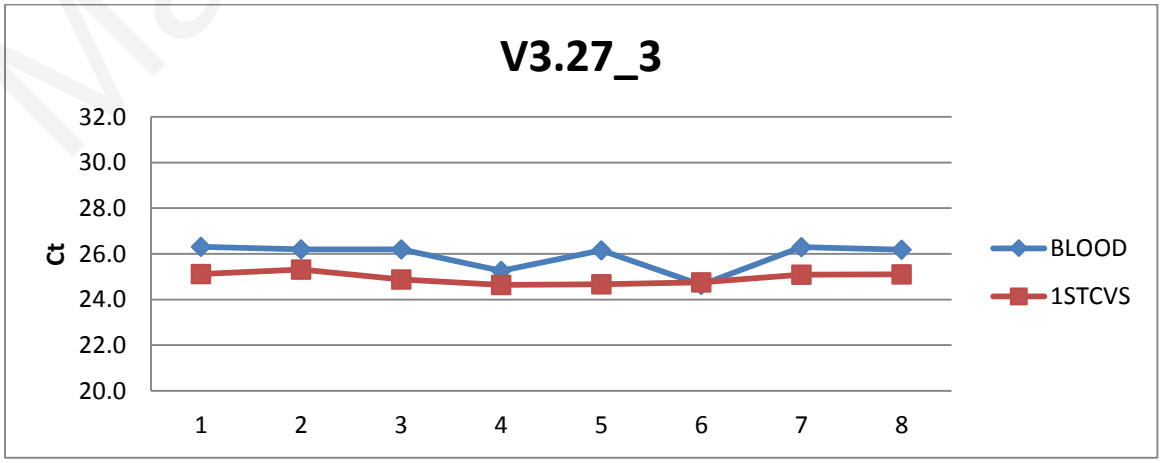


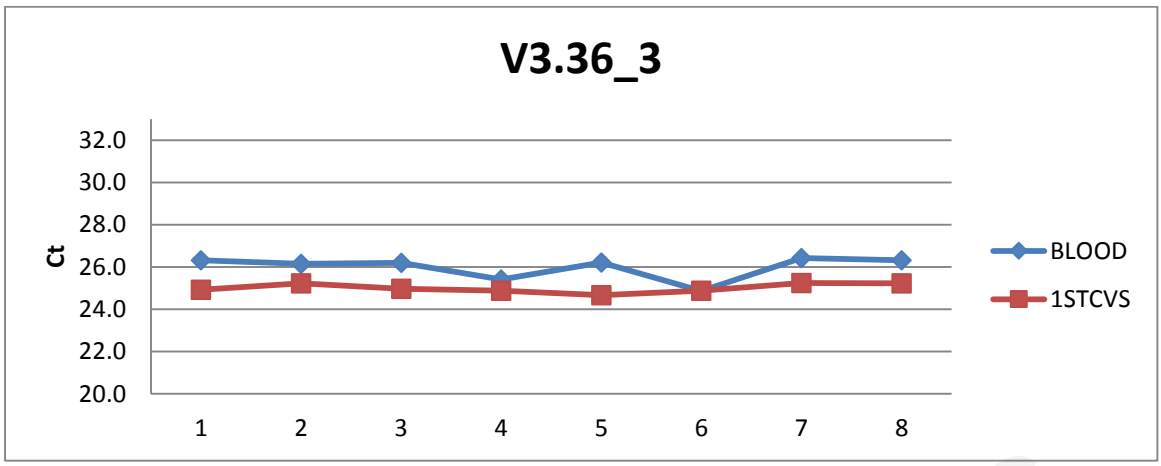


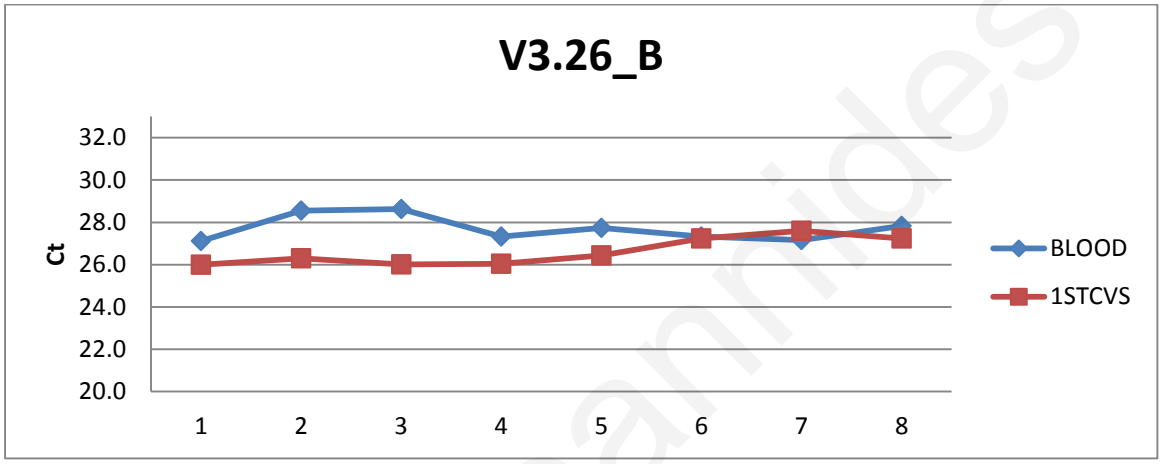


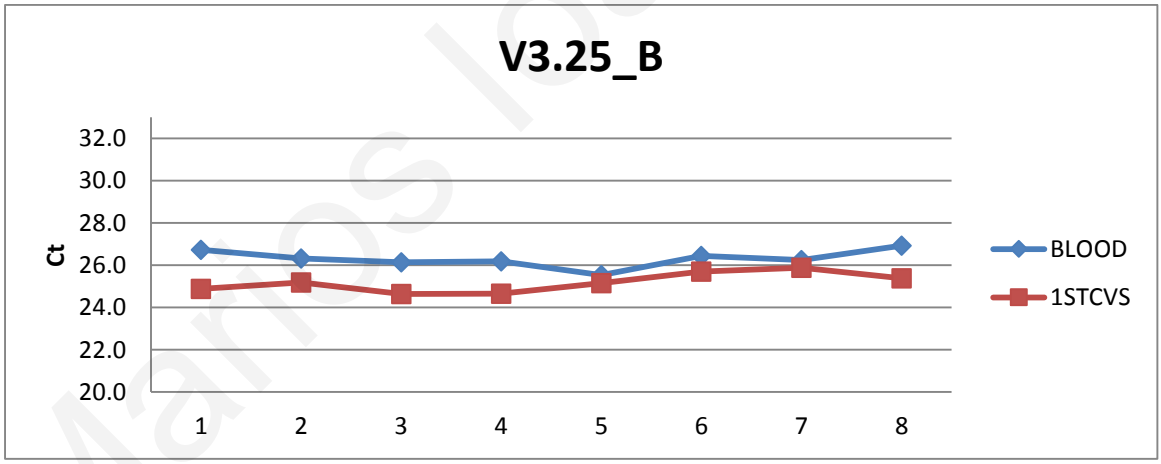


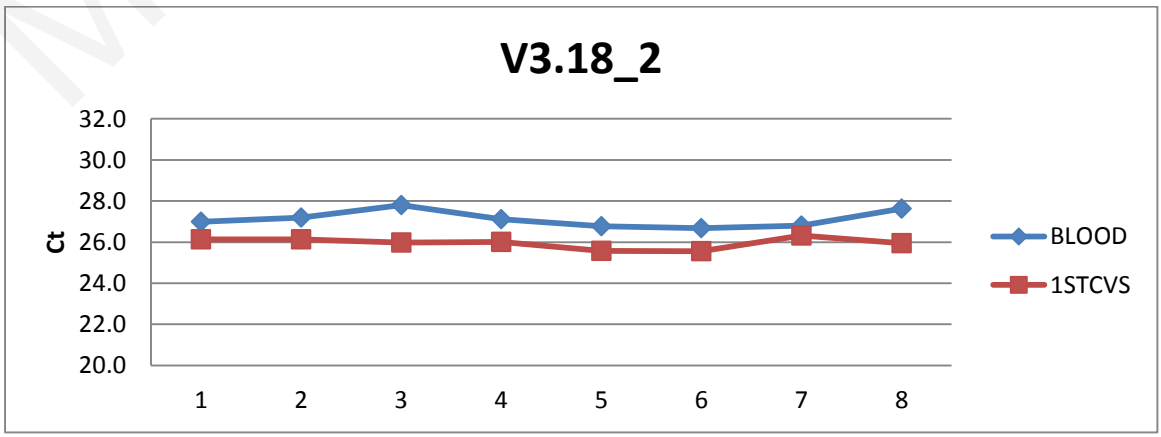


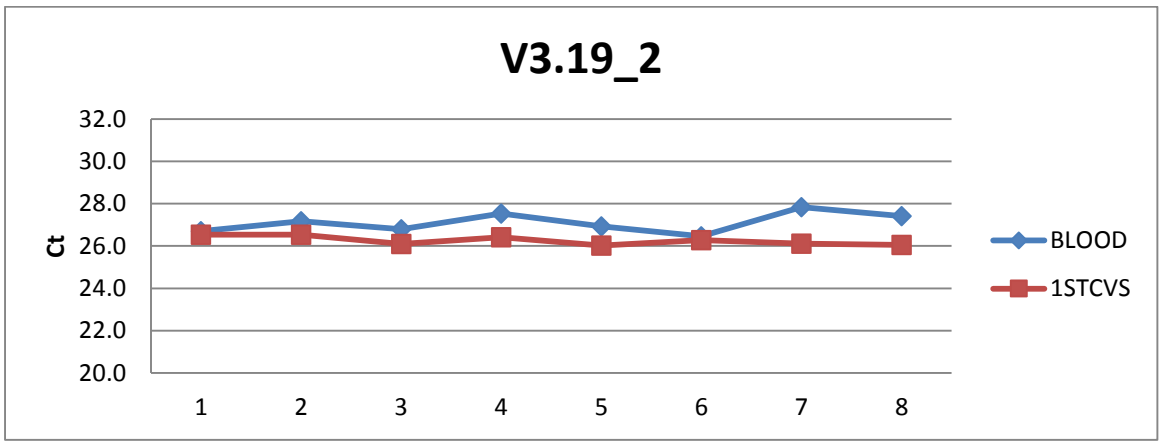


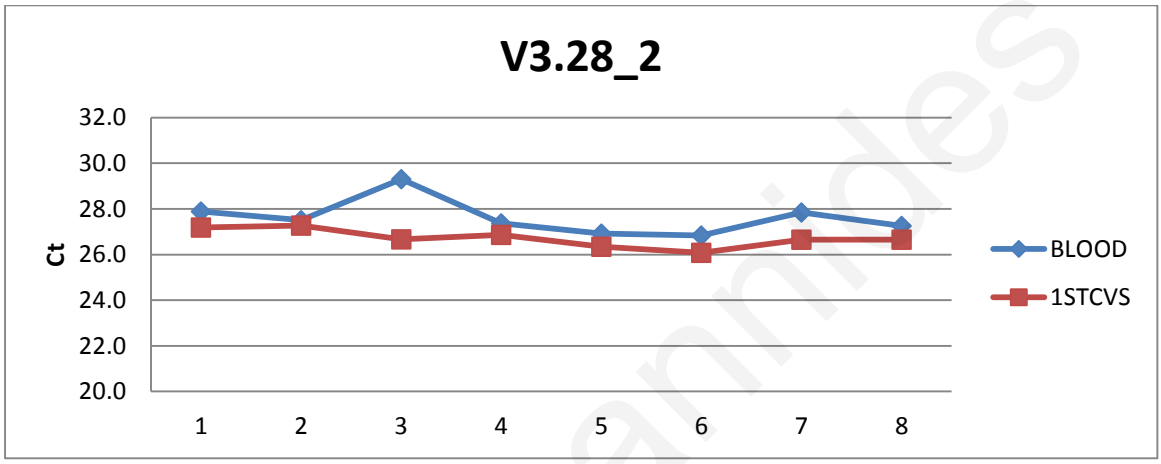


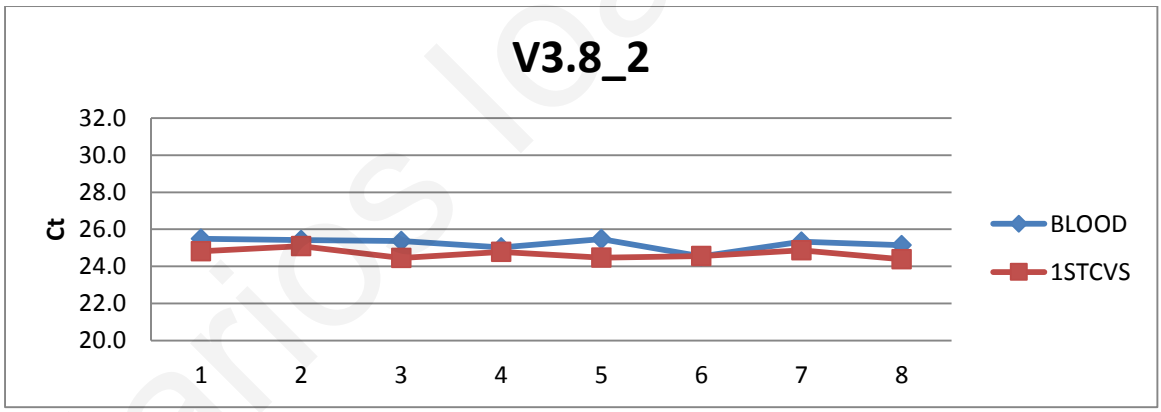


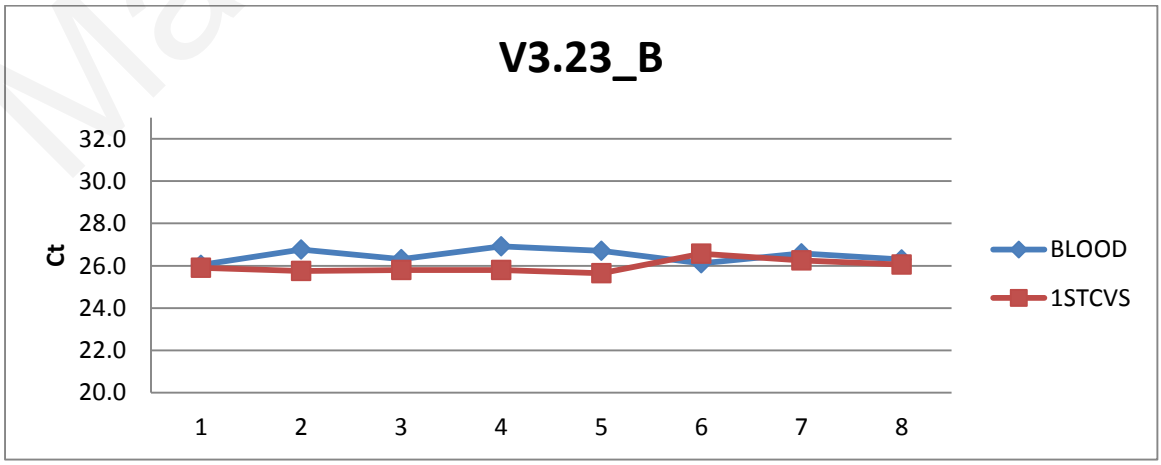


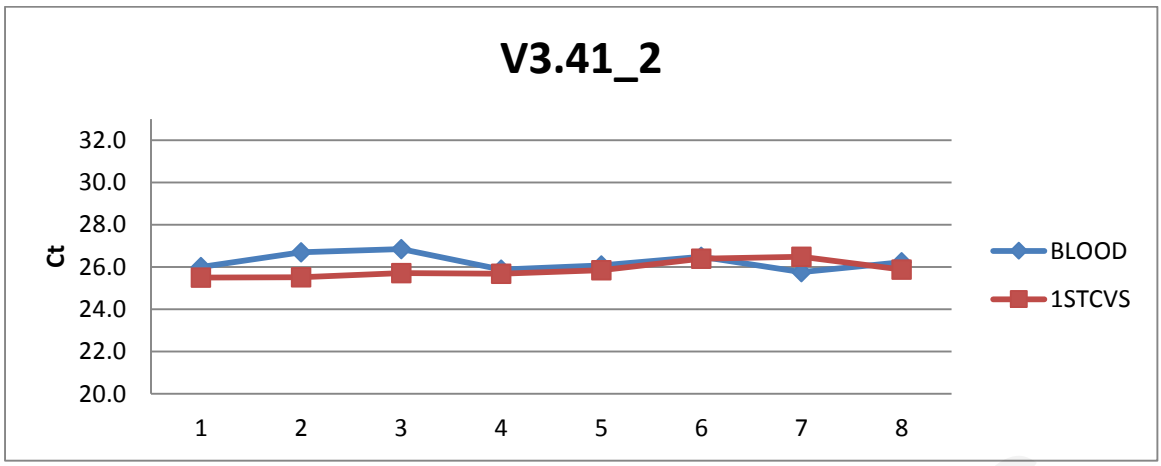






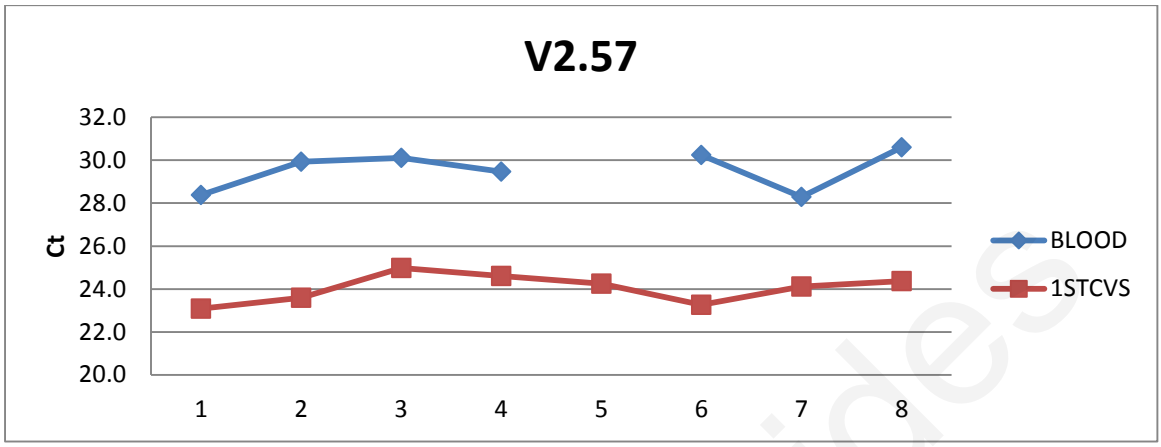


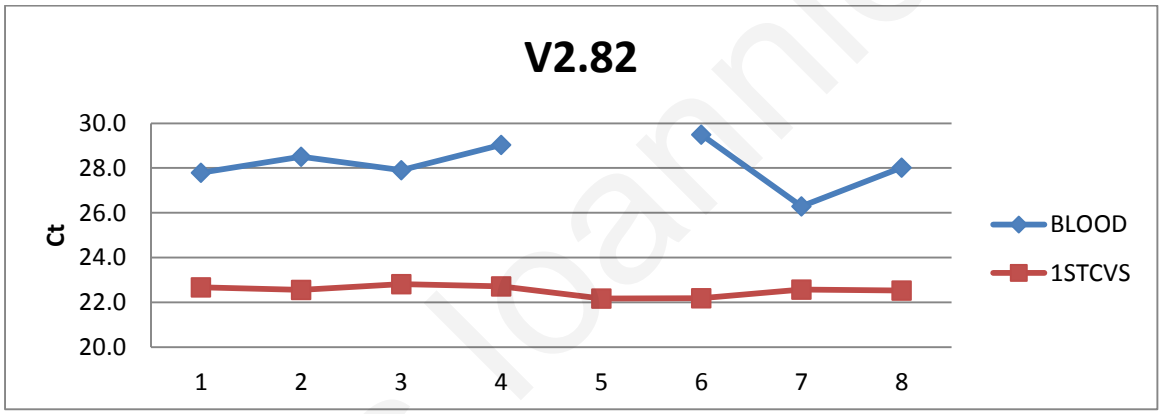


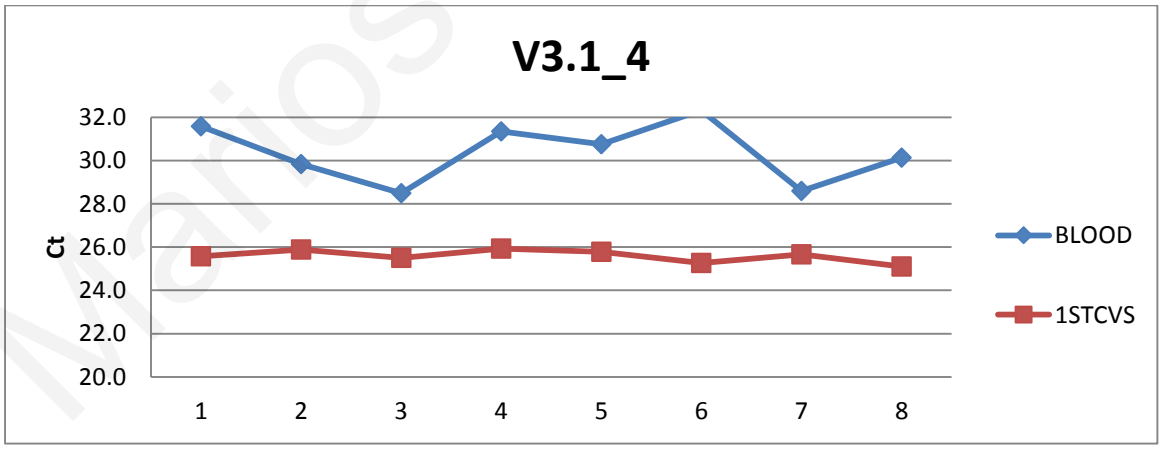


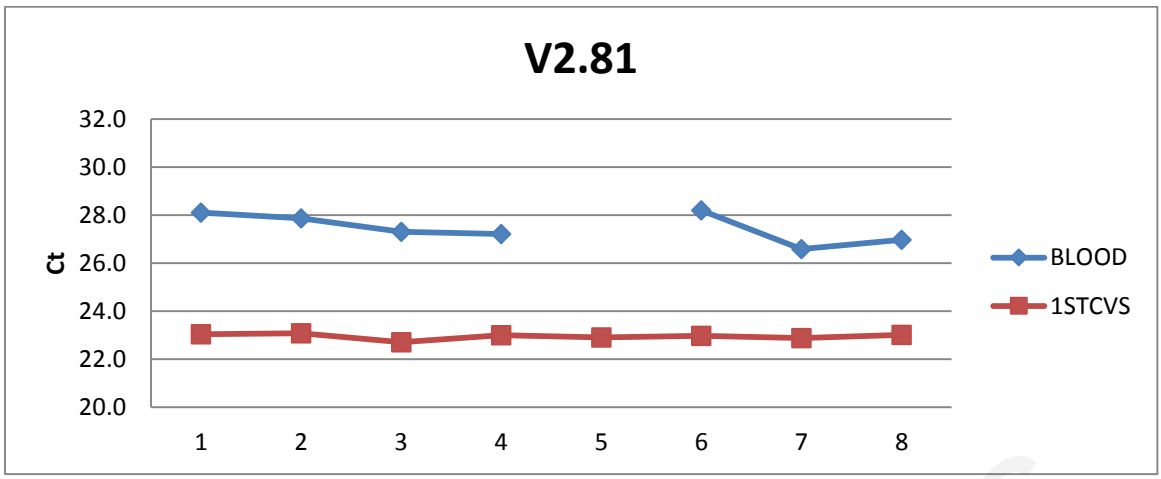
B. Chromosome 18

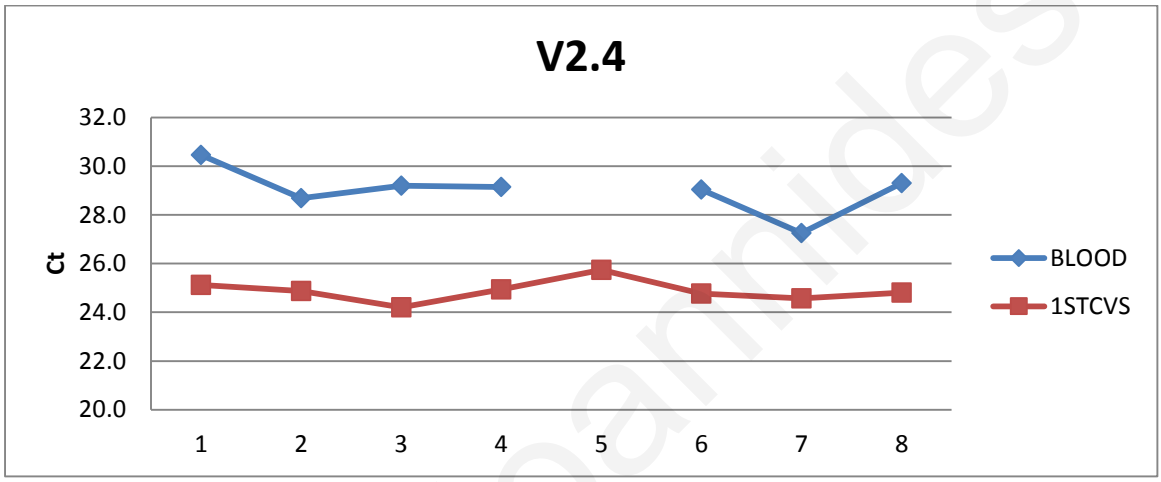
2. "Good DMRs"

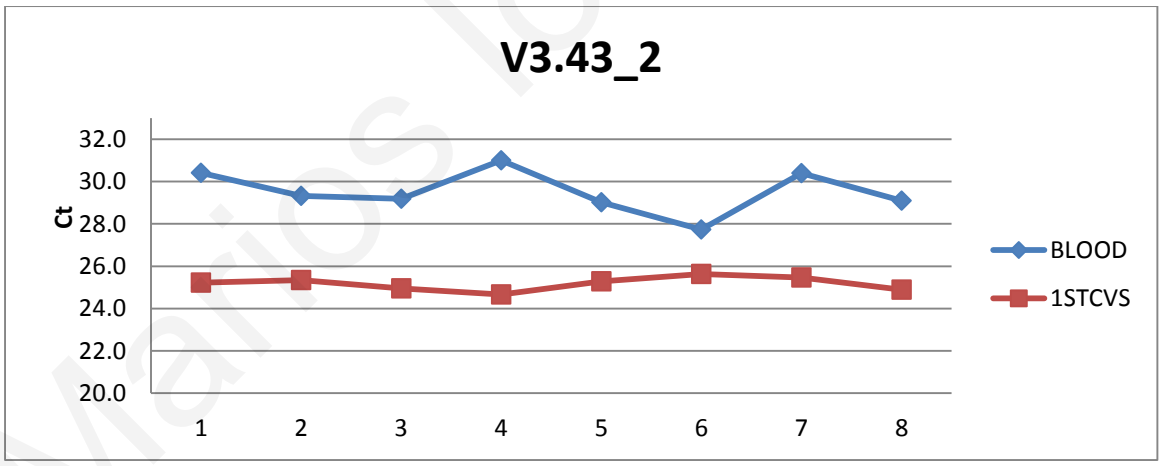


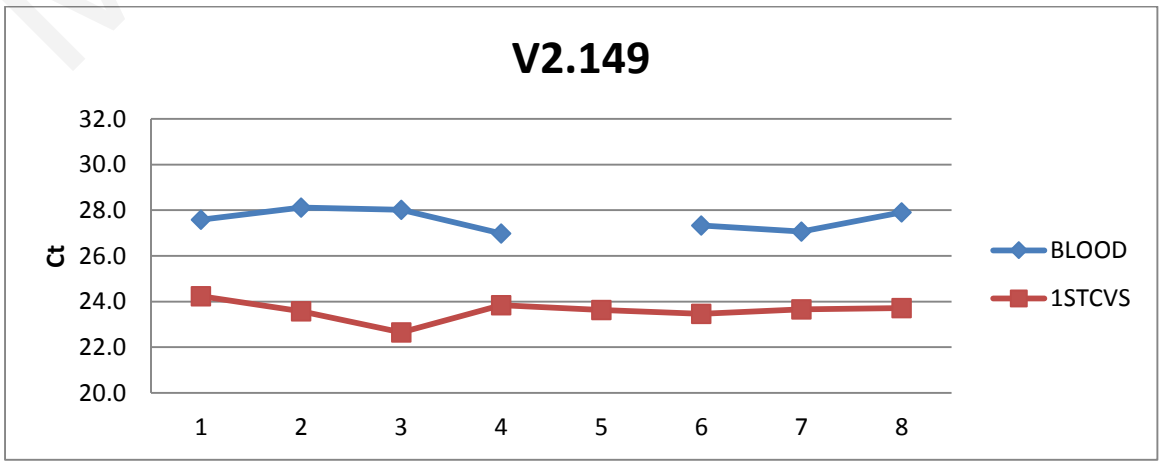


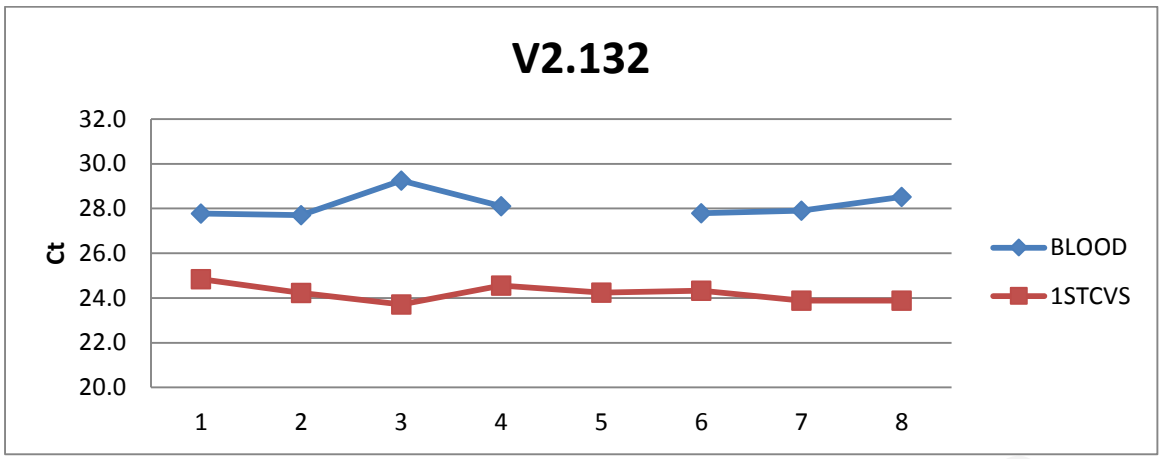


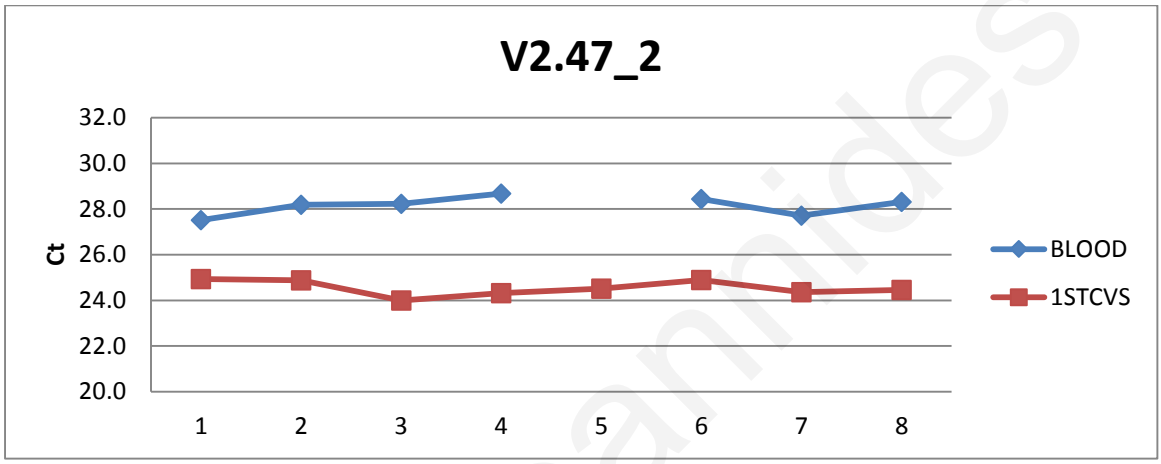


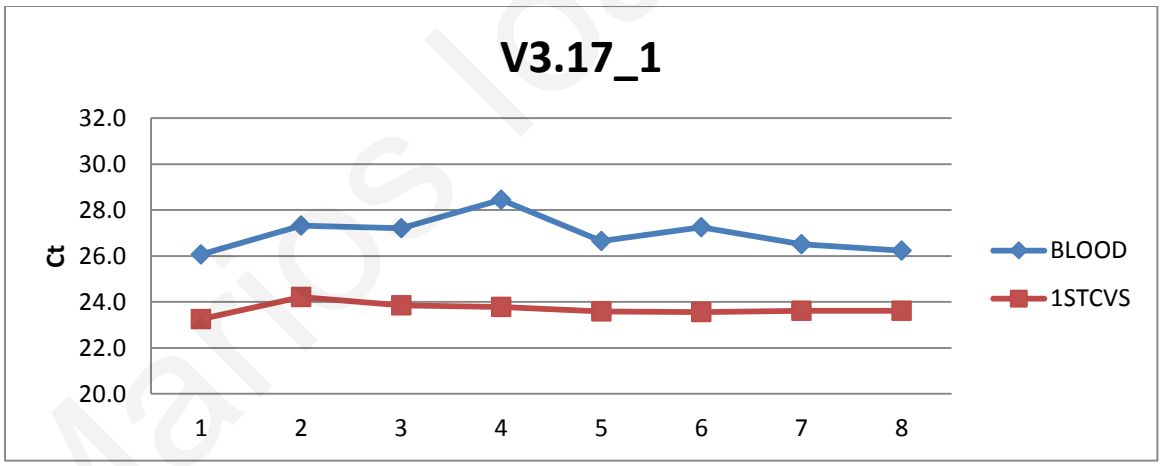


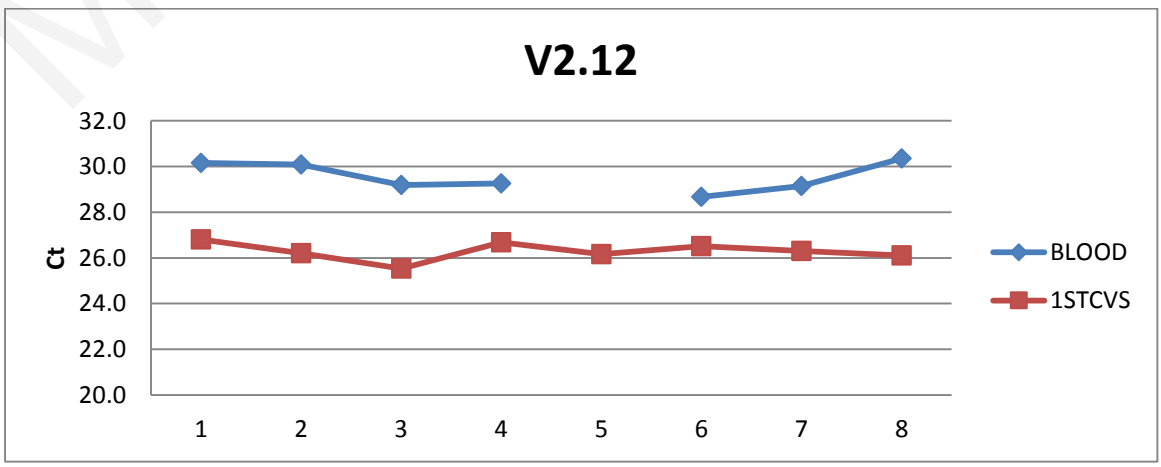


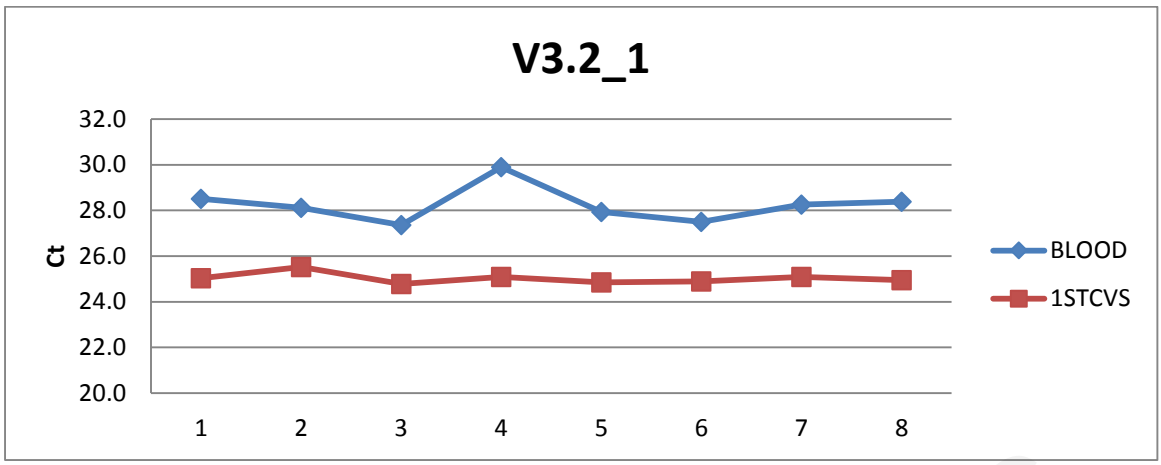


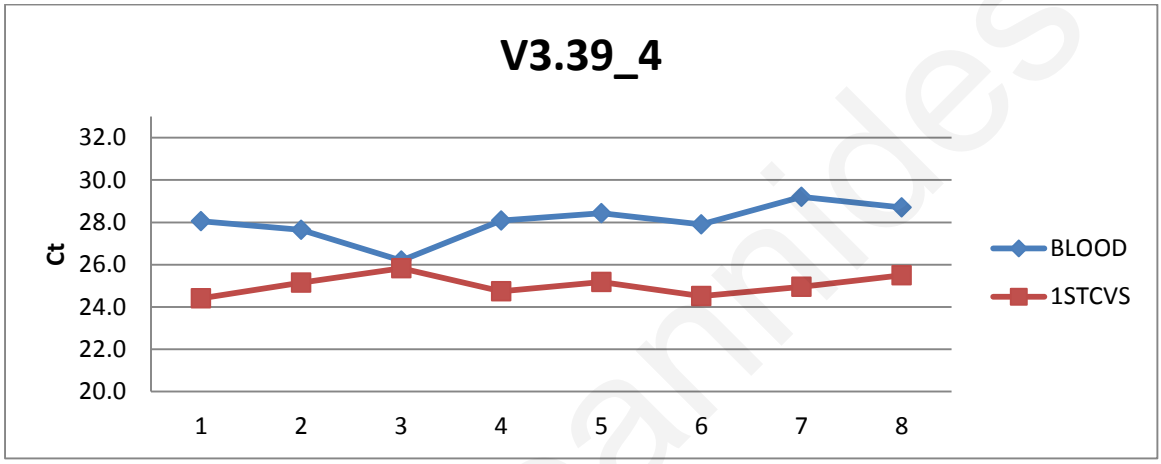


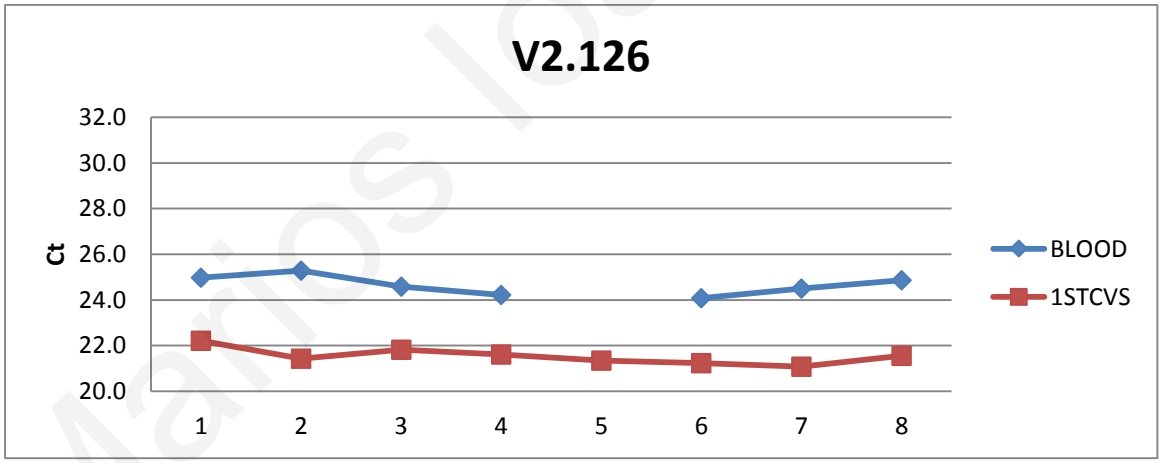


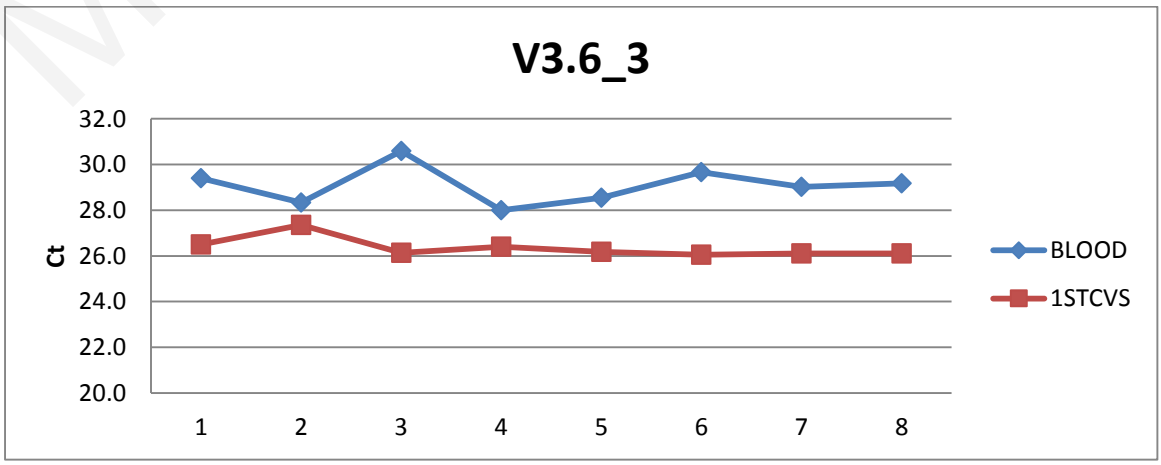


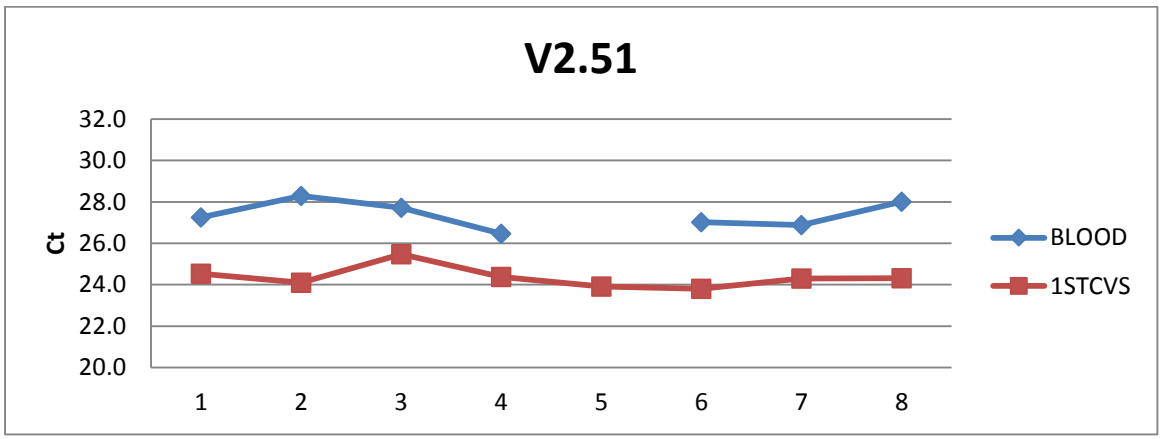


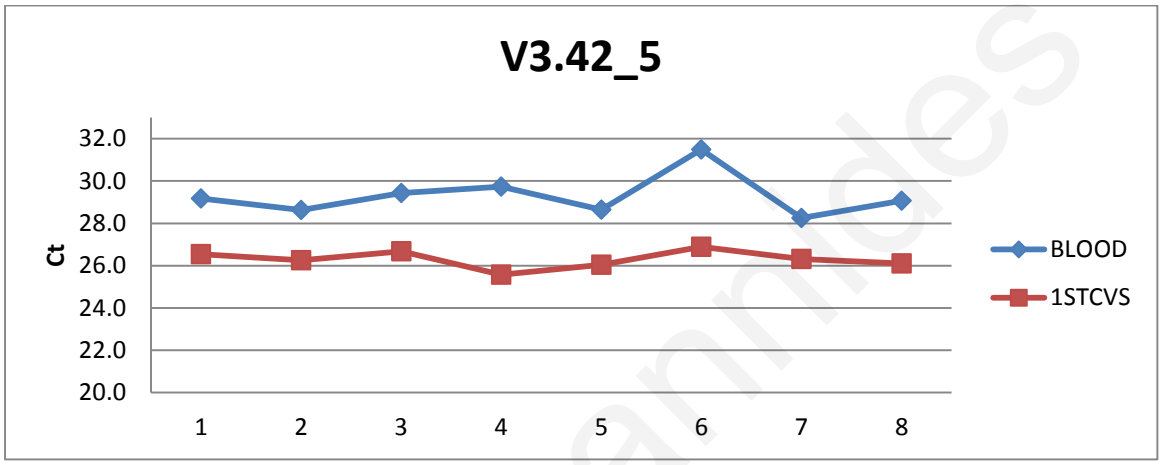


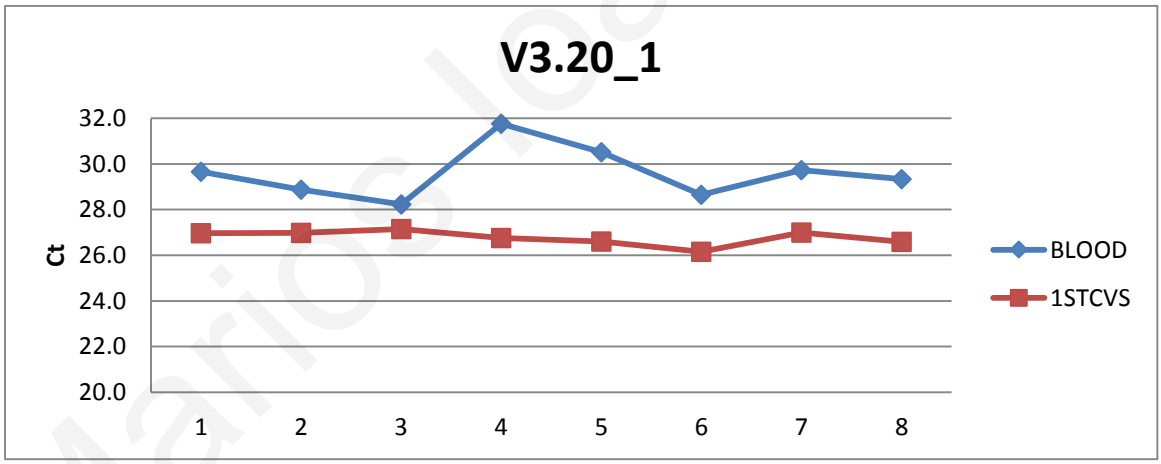


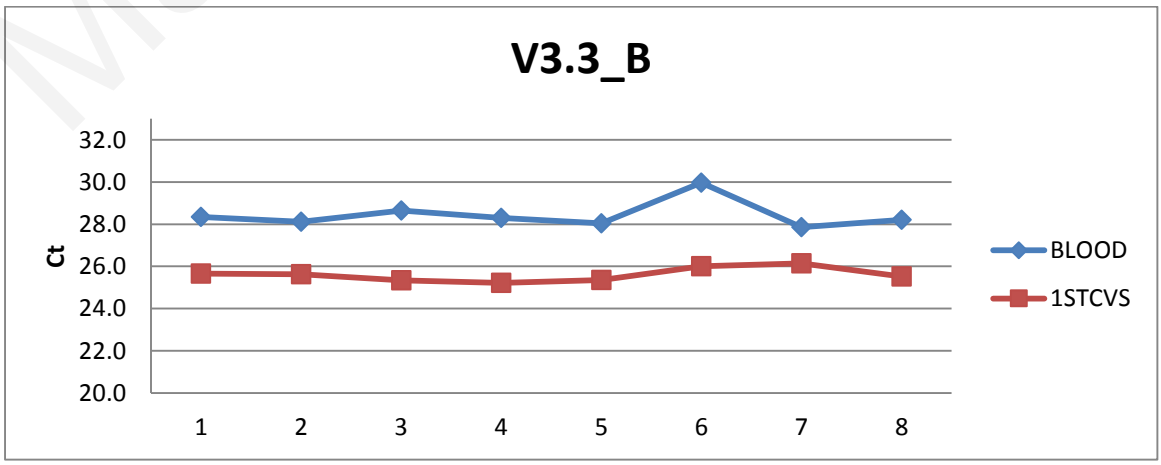






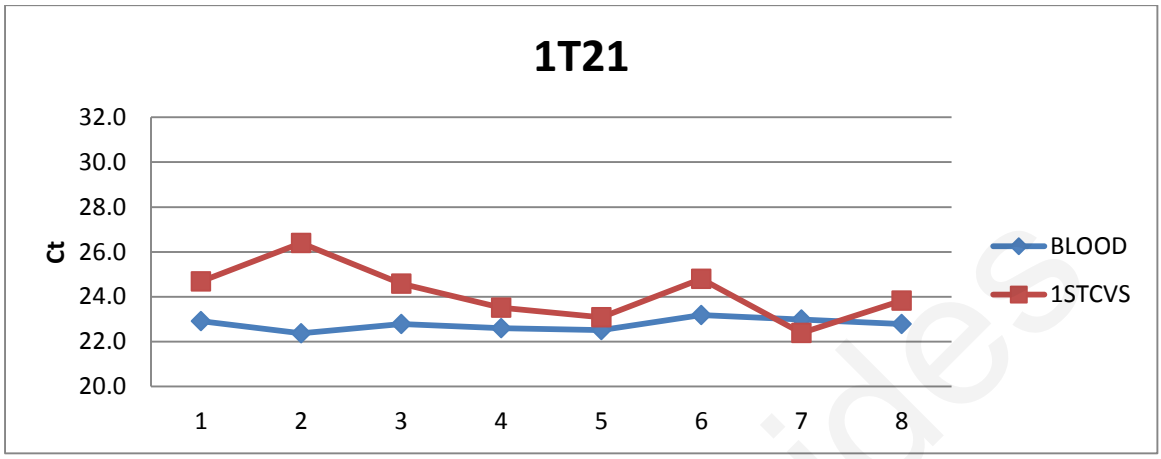


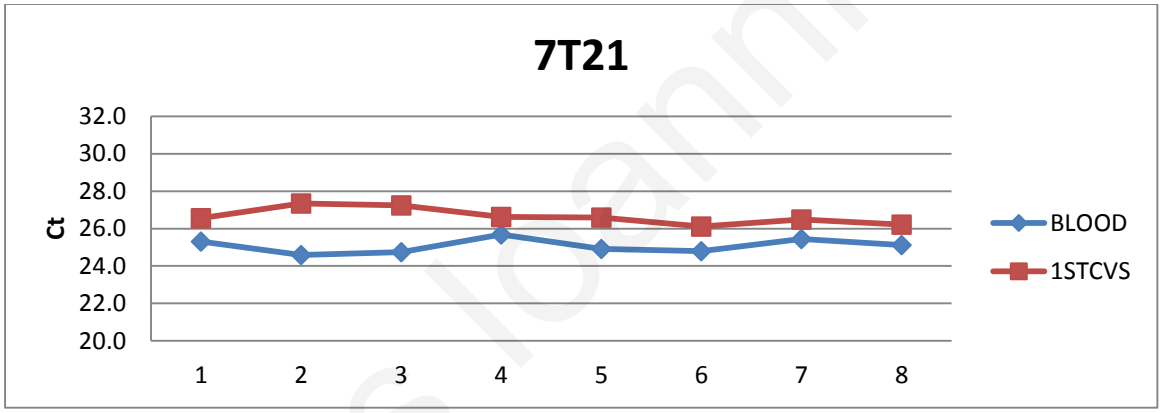


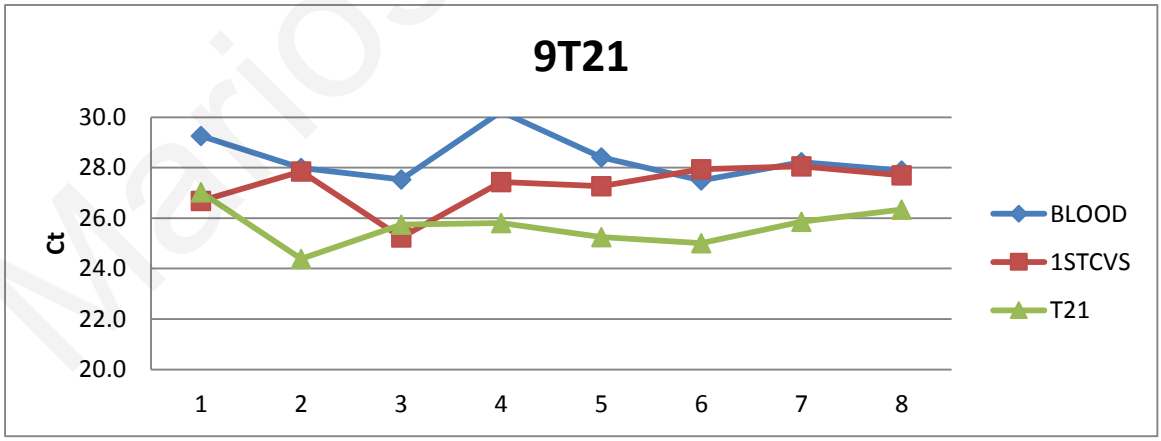


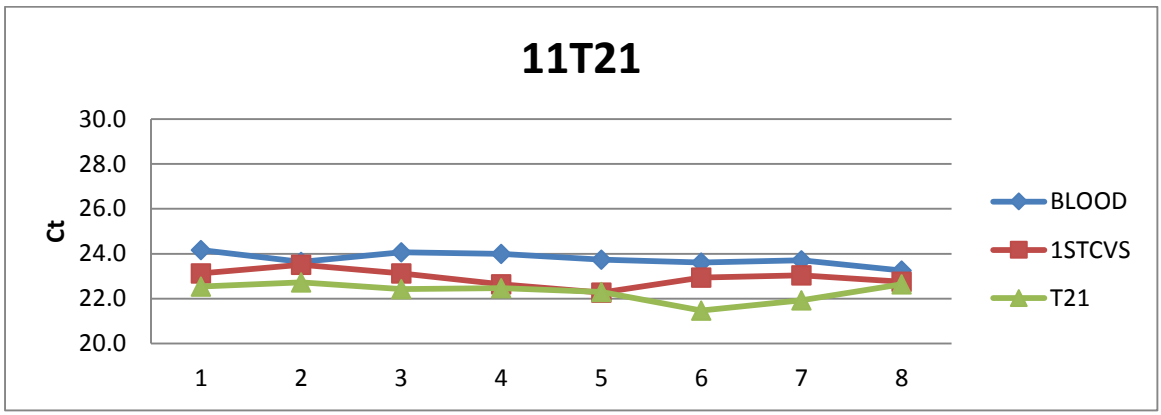
C. Chromosome 21

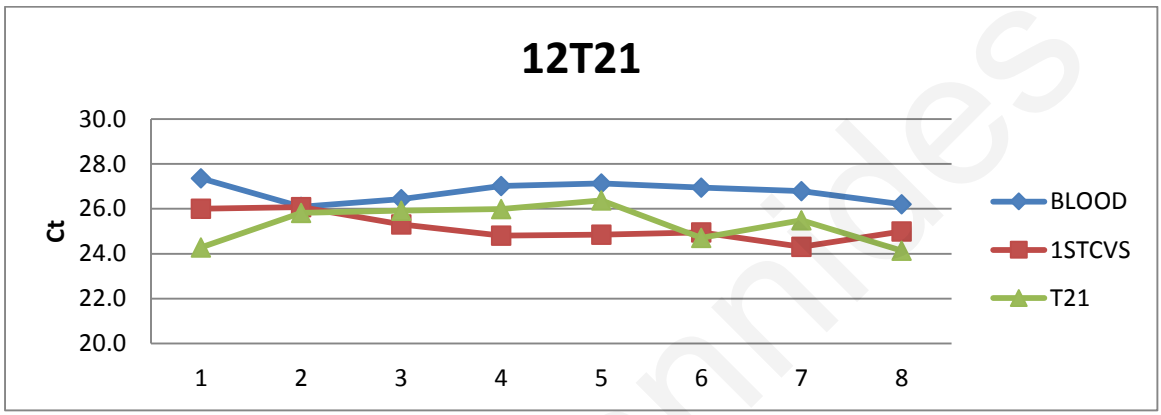
1. "Not DMRs"

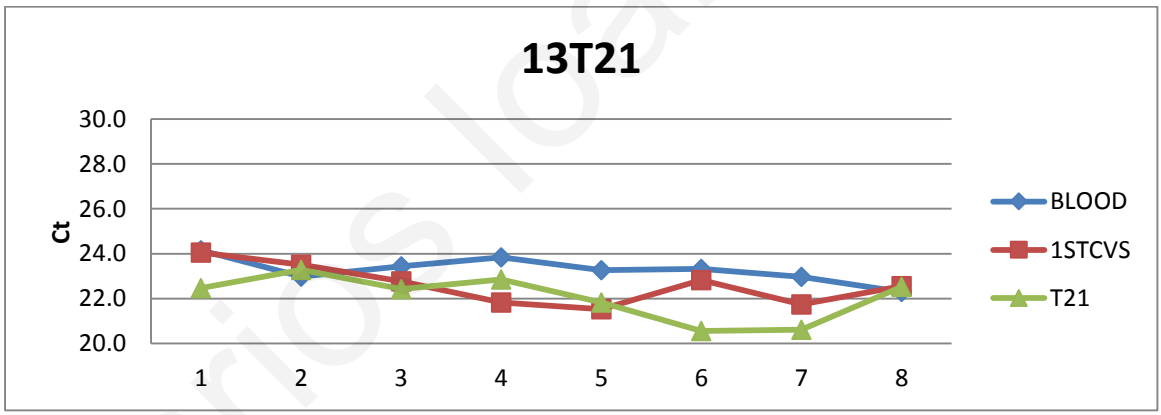


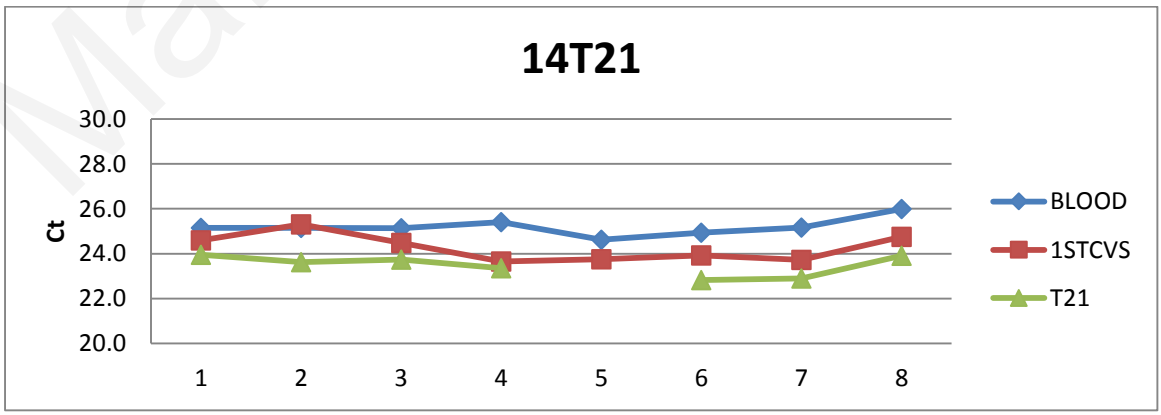


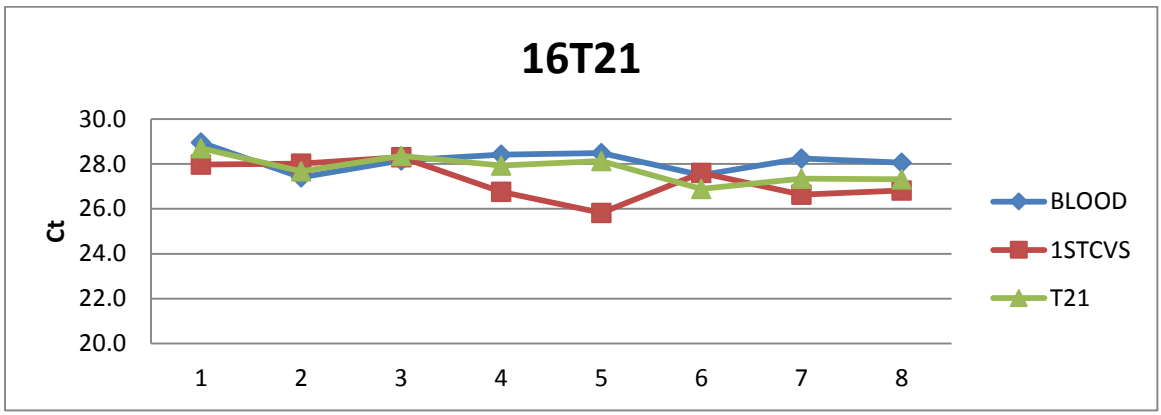


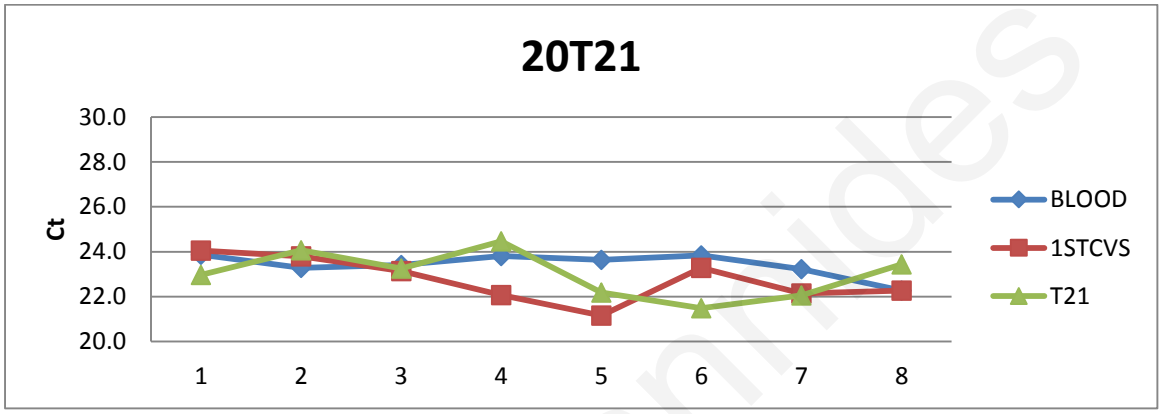


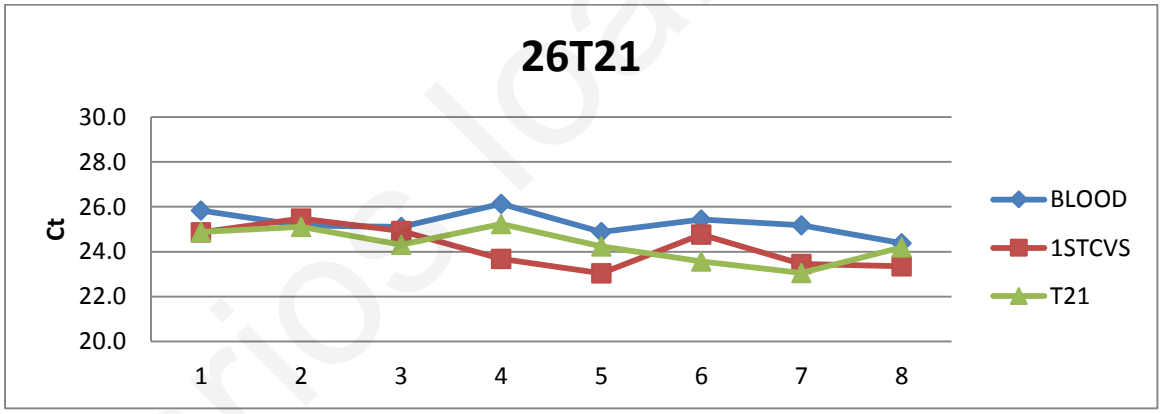


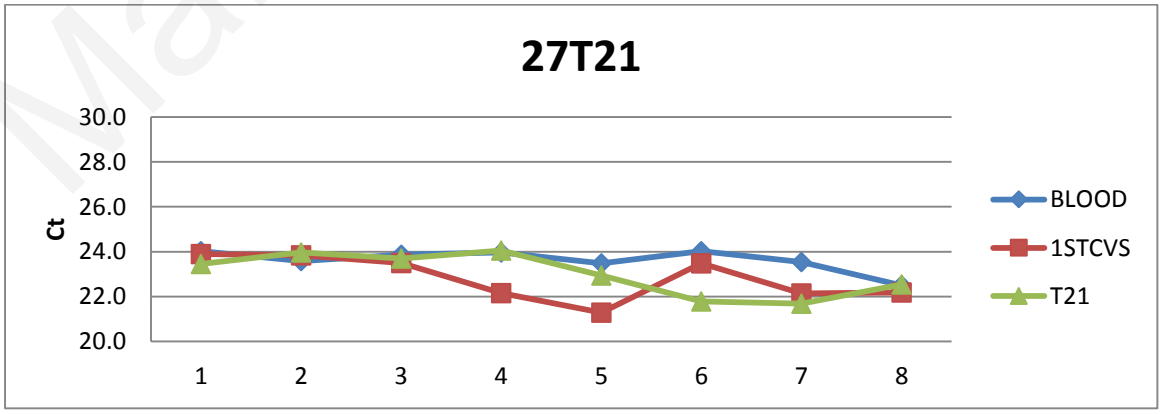


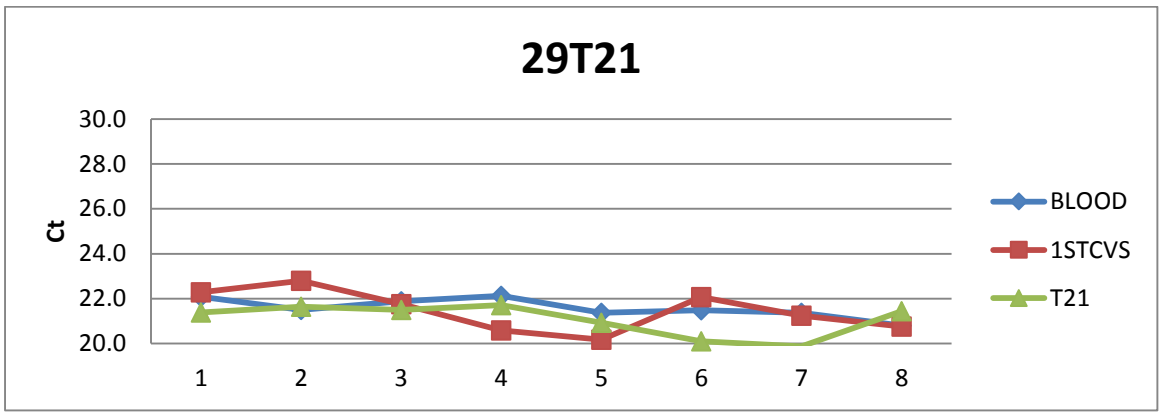


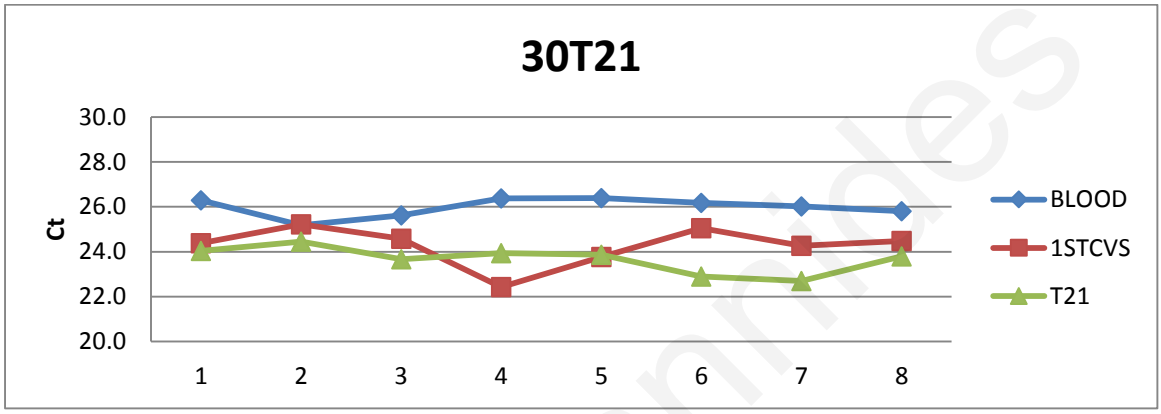


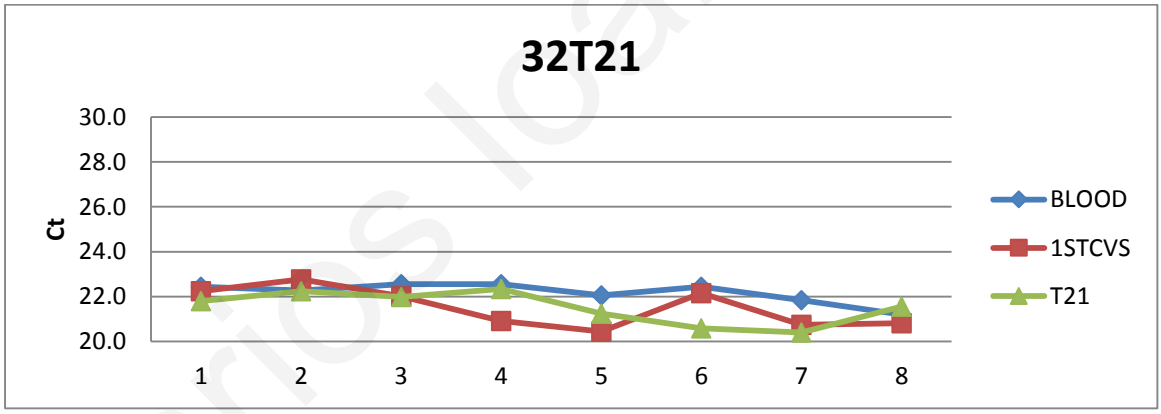


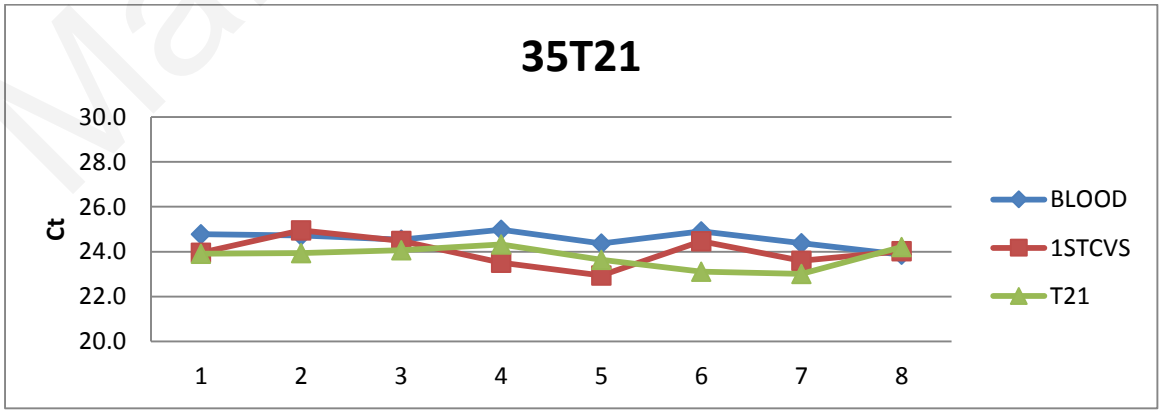


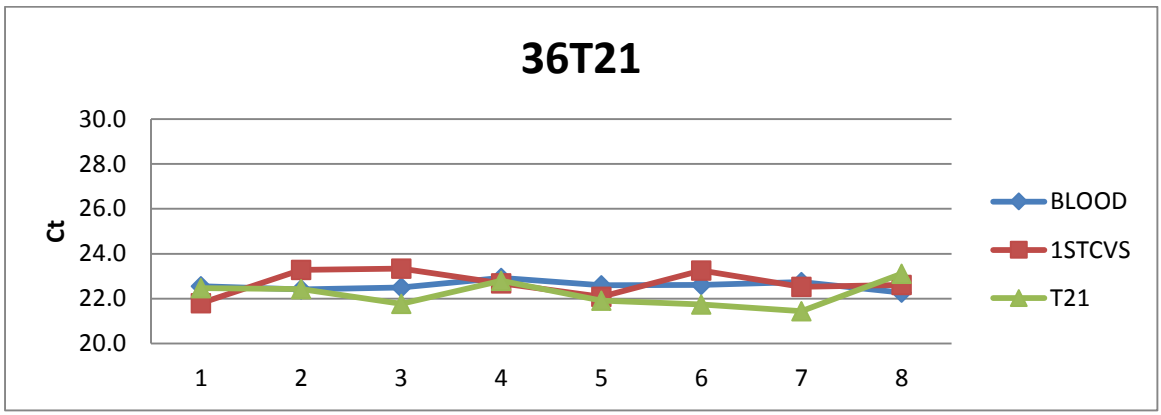


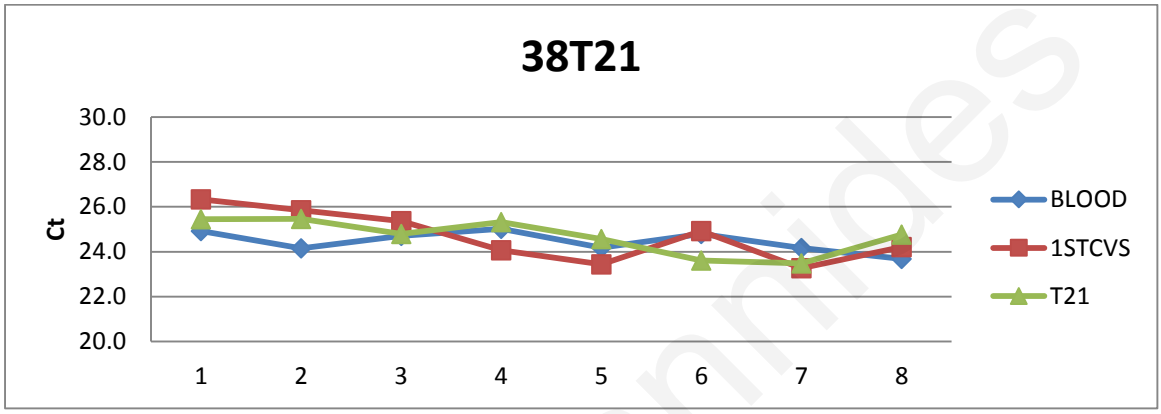


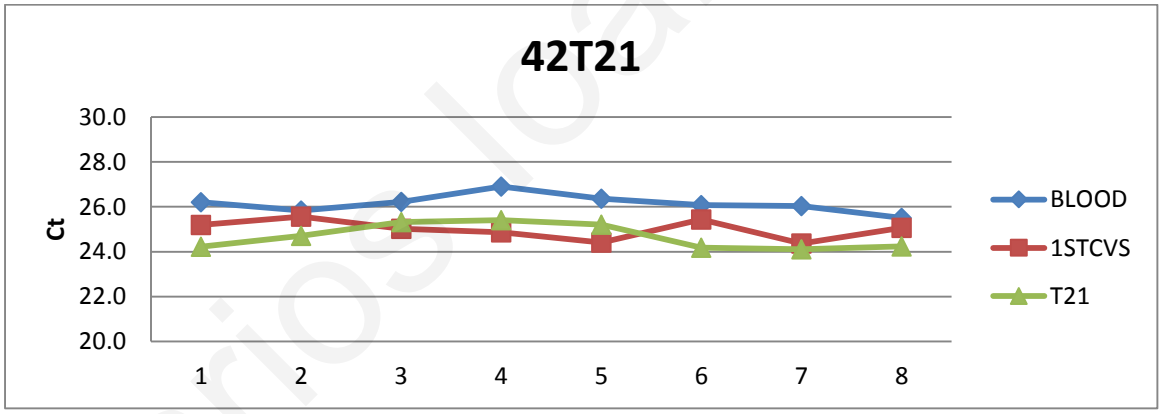


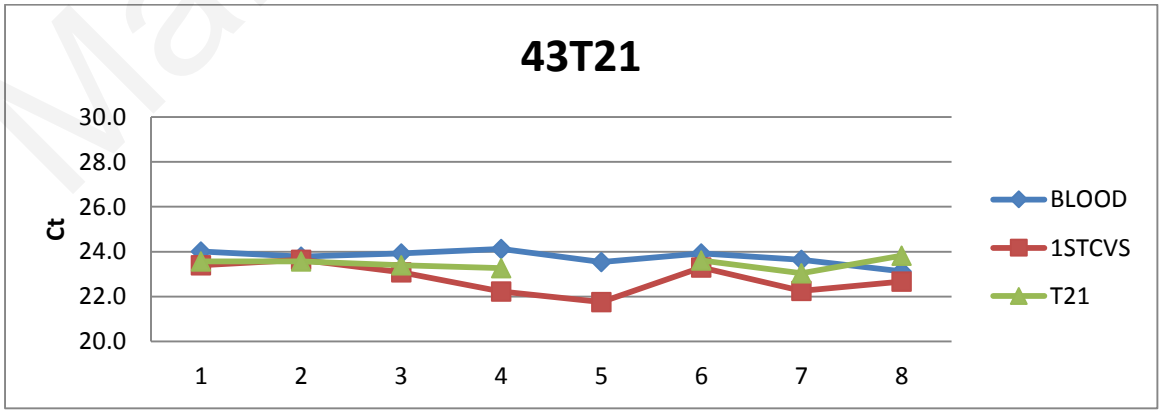


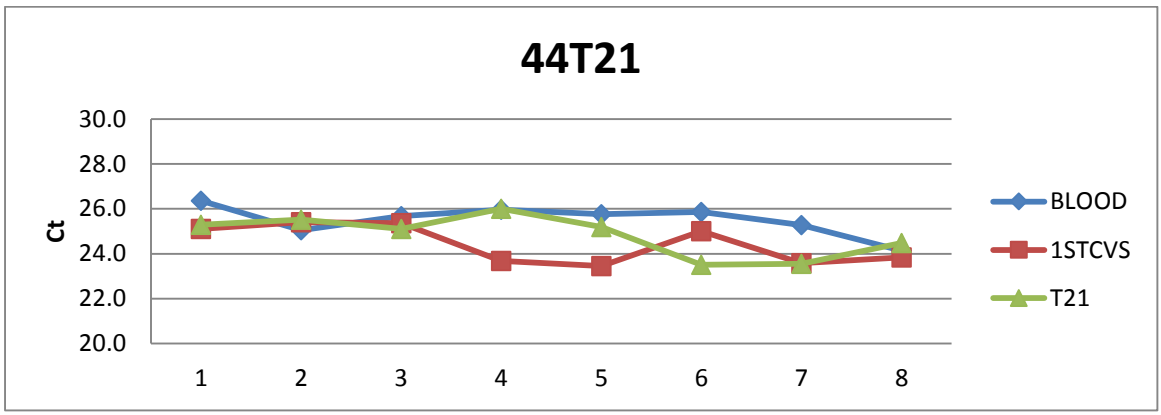


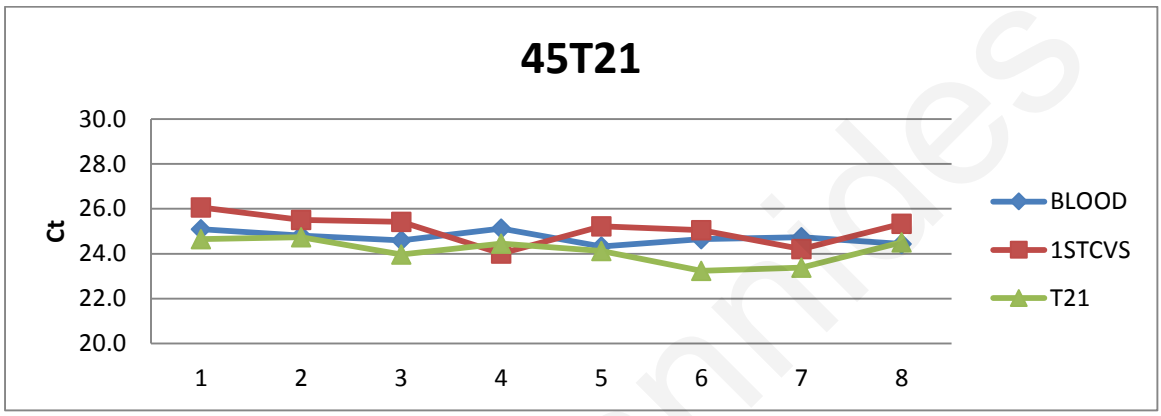


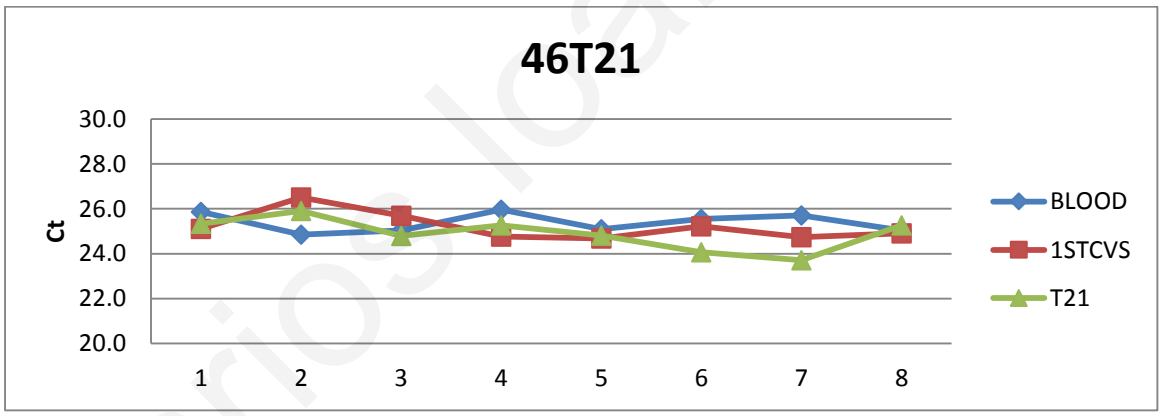


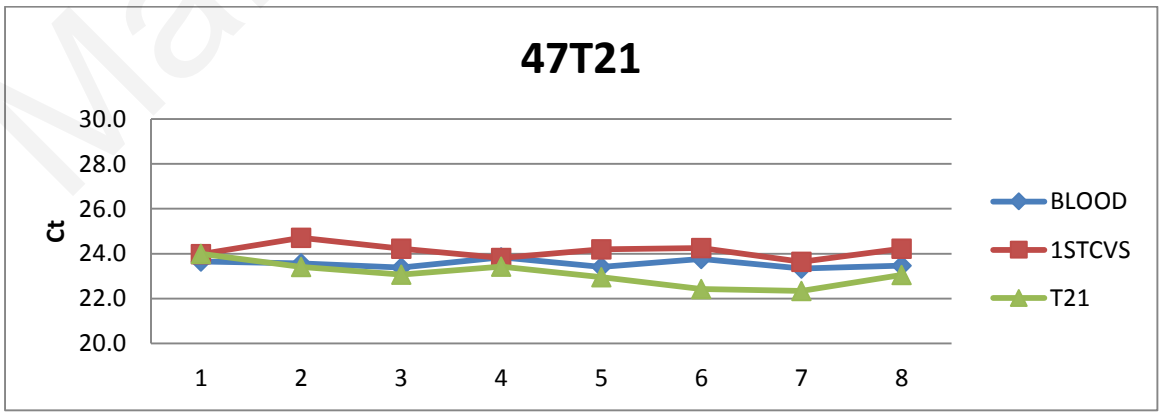


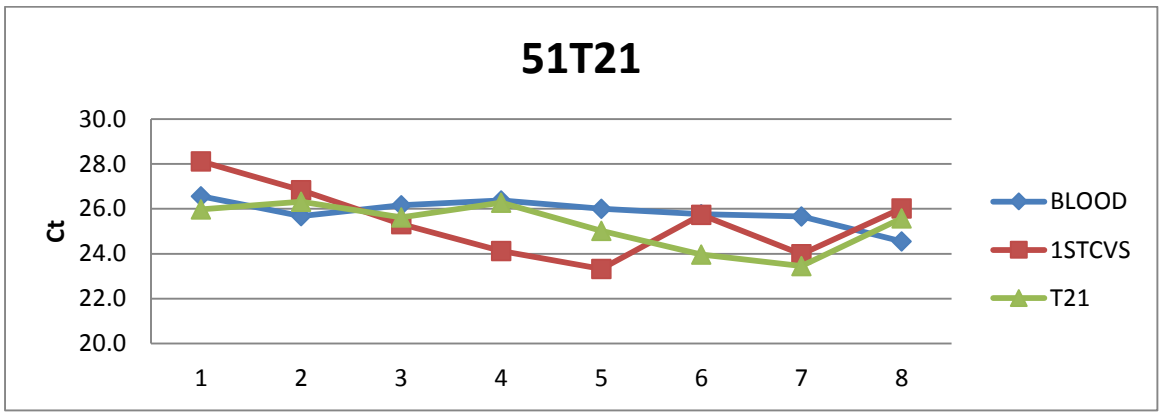


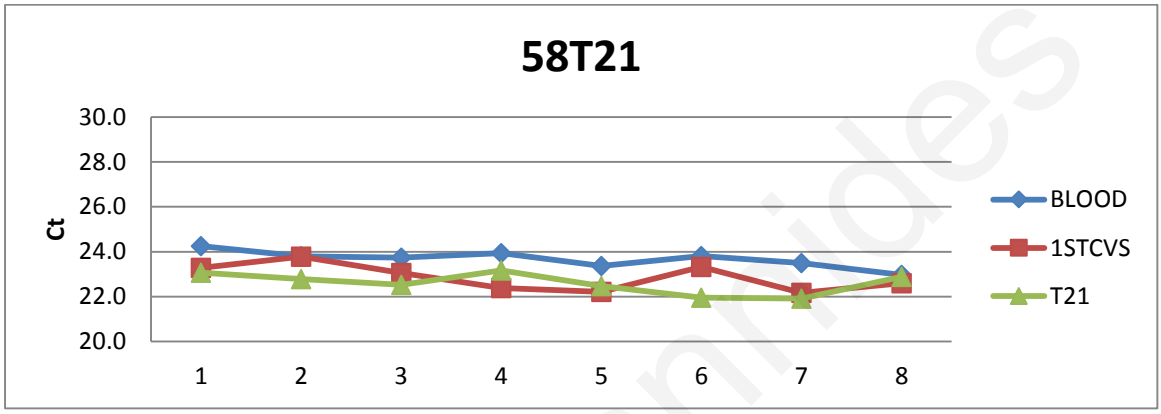


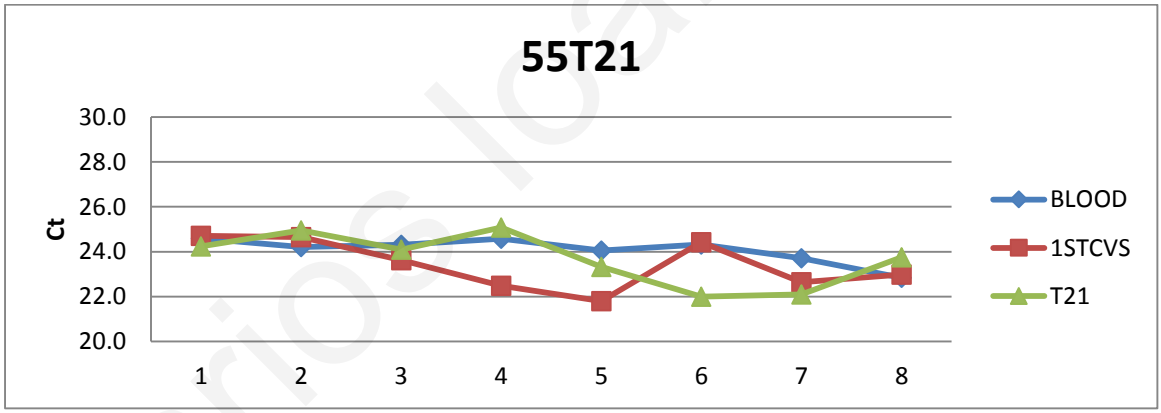


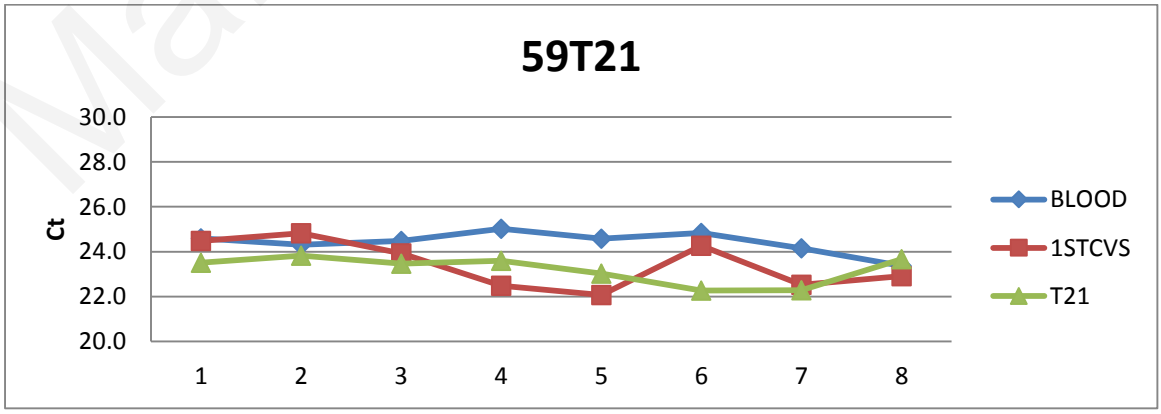


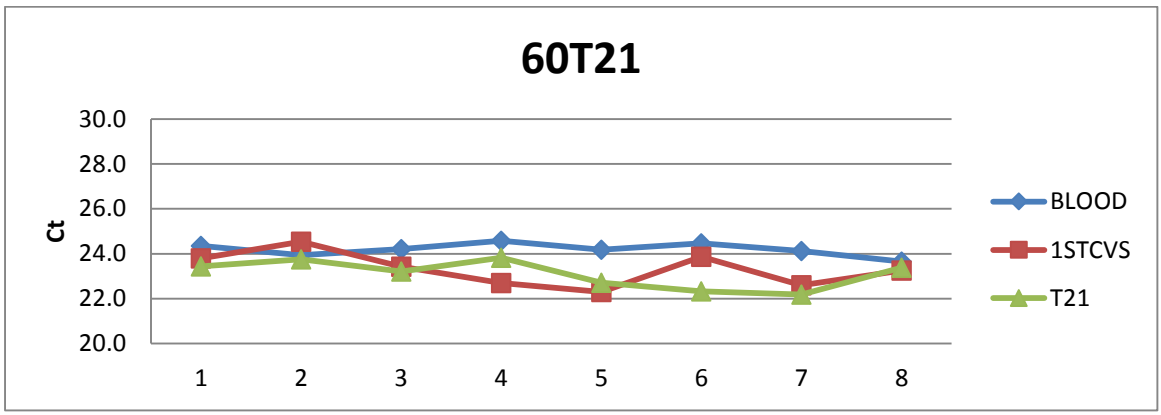


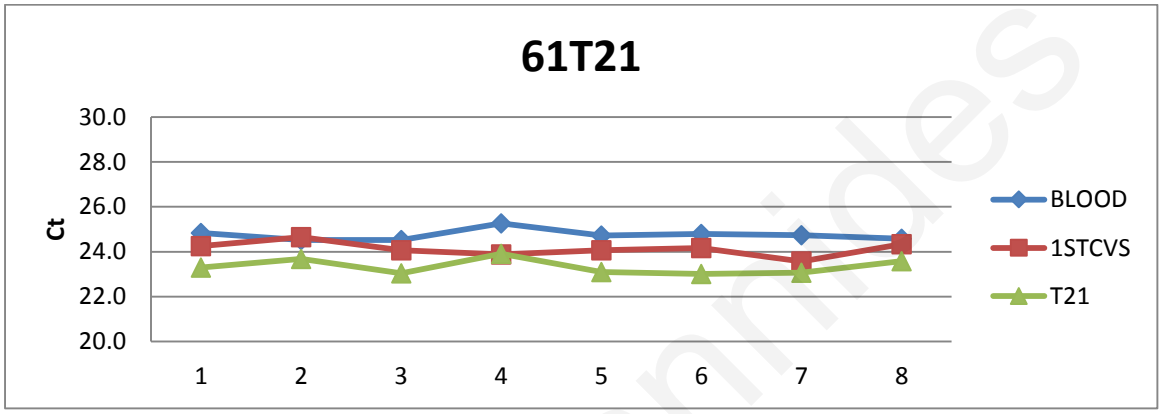


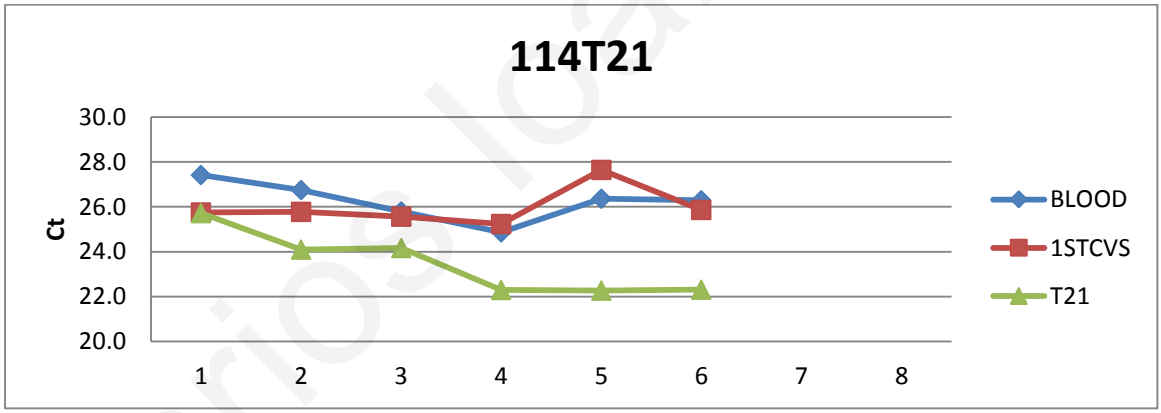


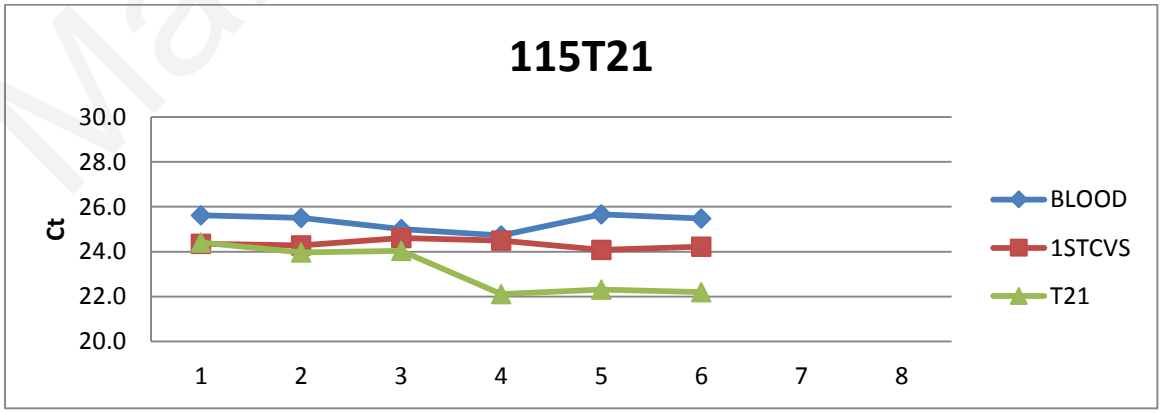


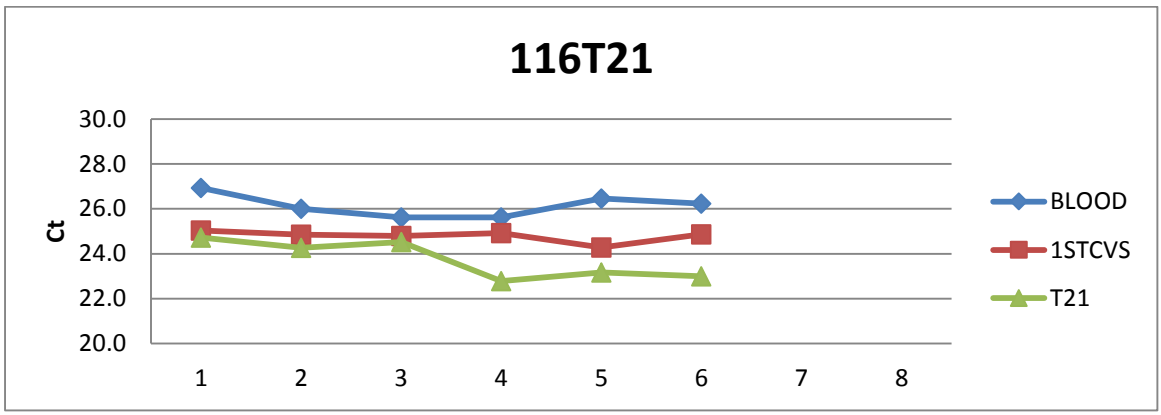


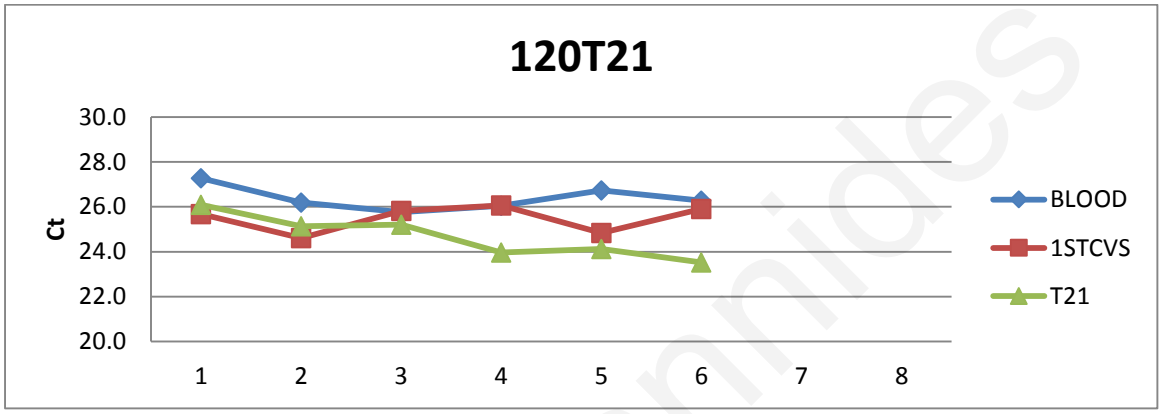


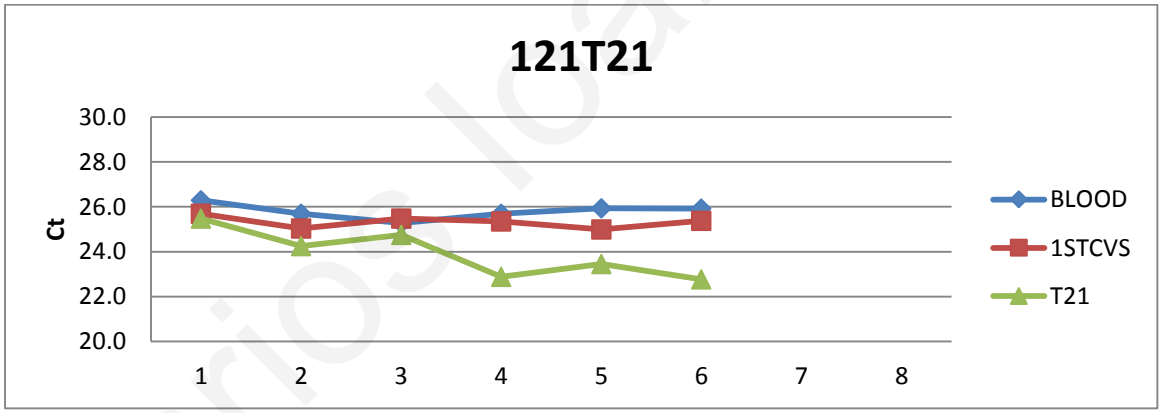


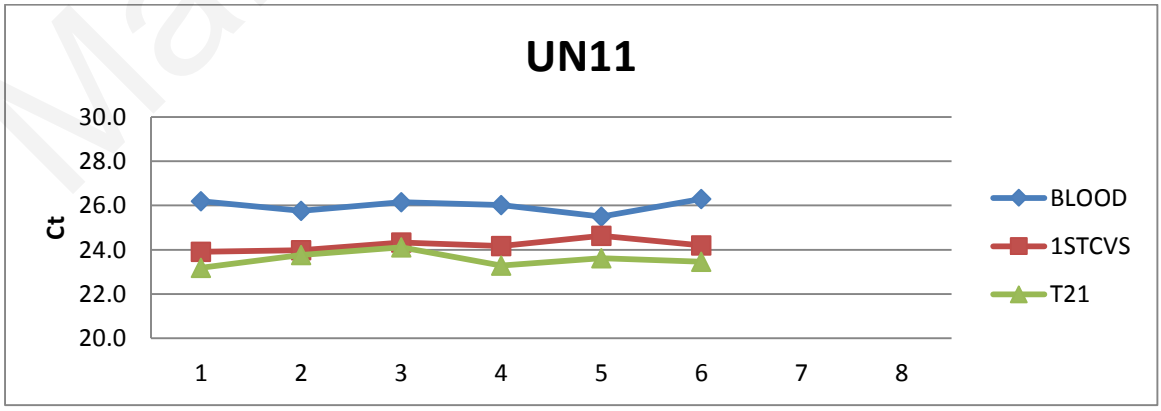


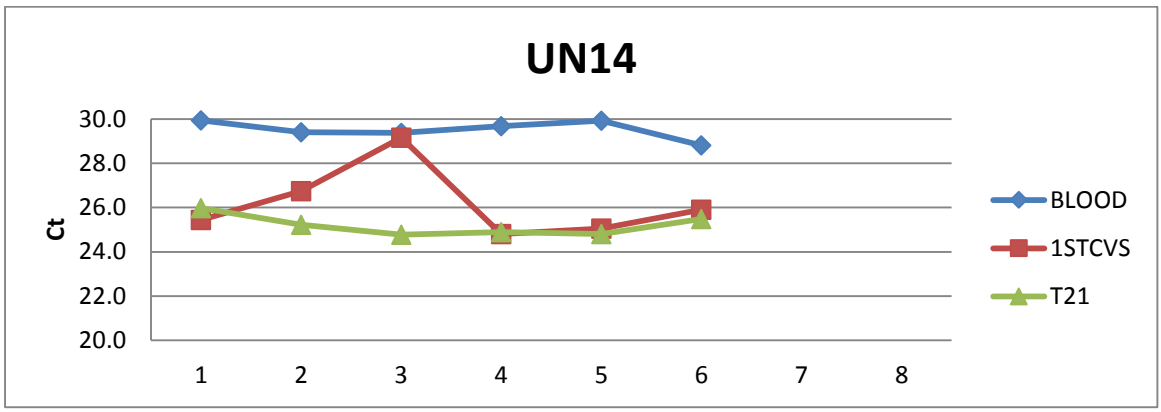


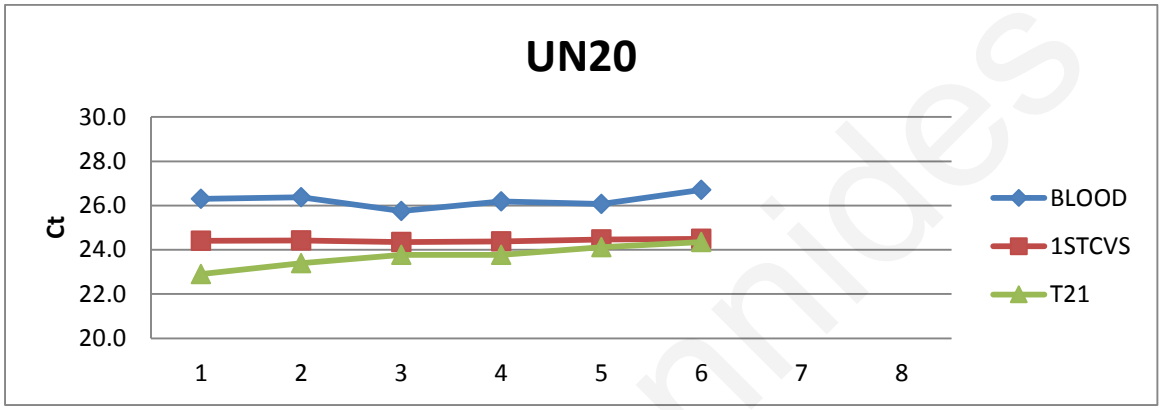


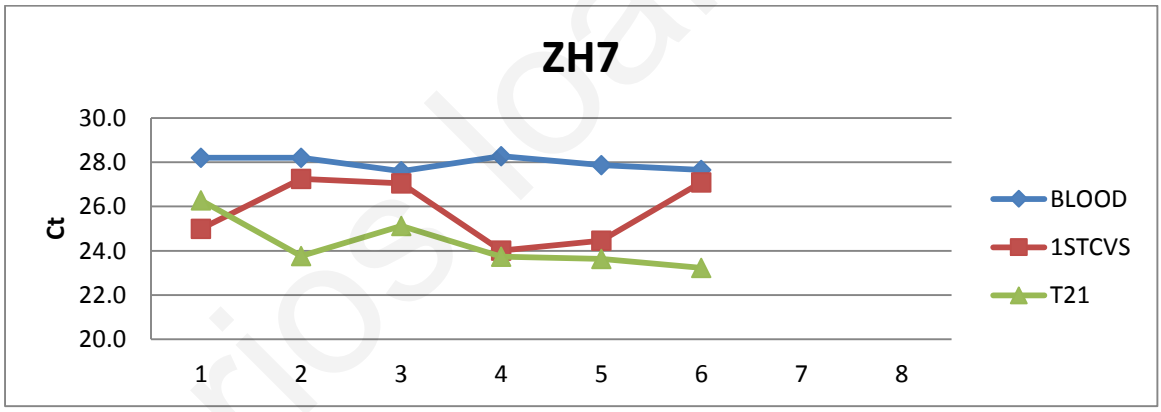


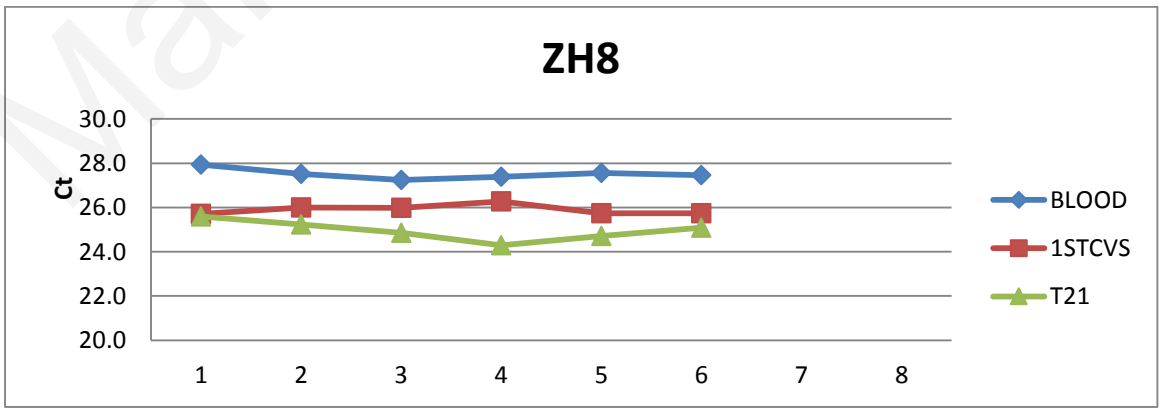


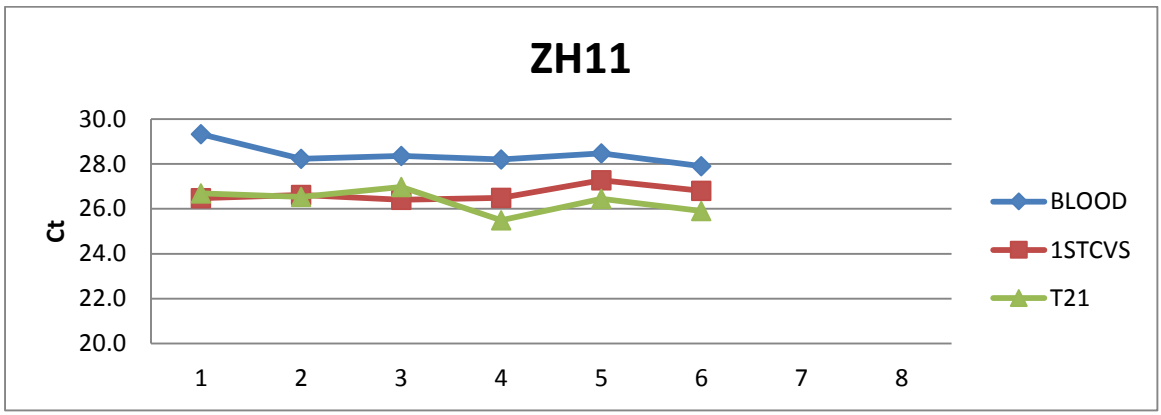


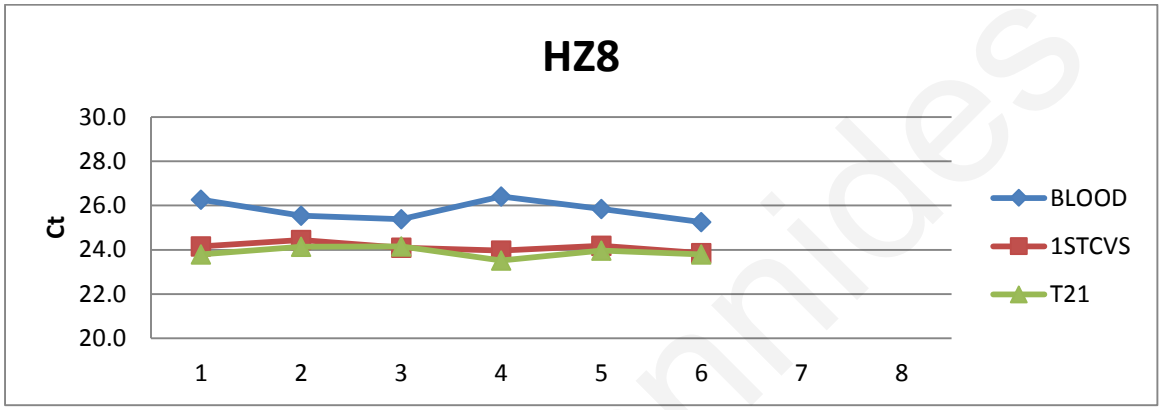




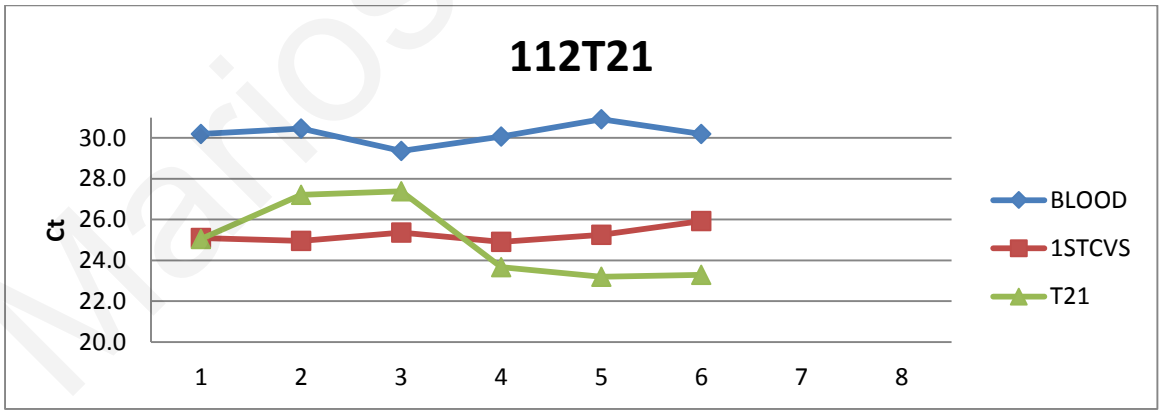


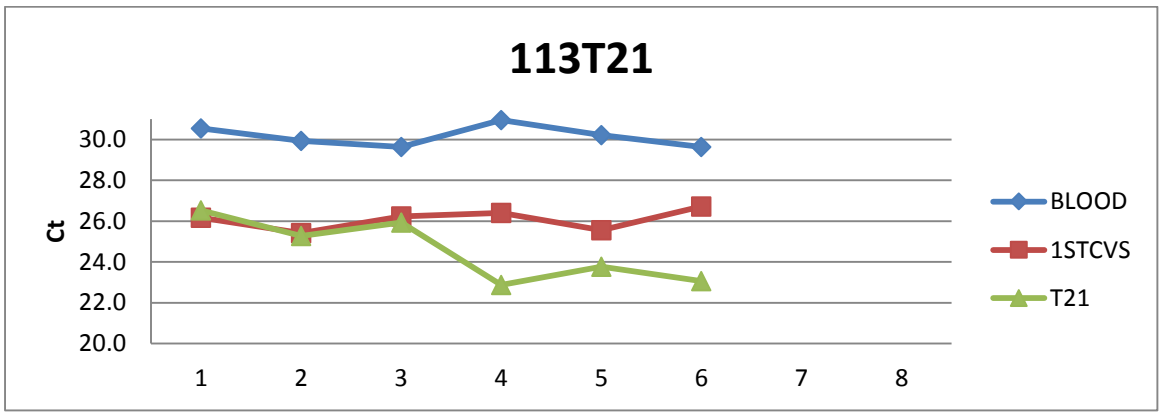


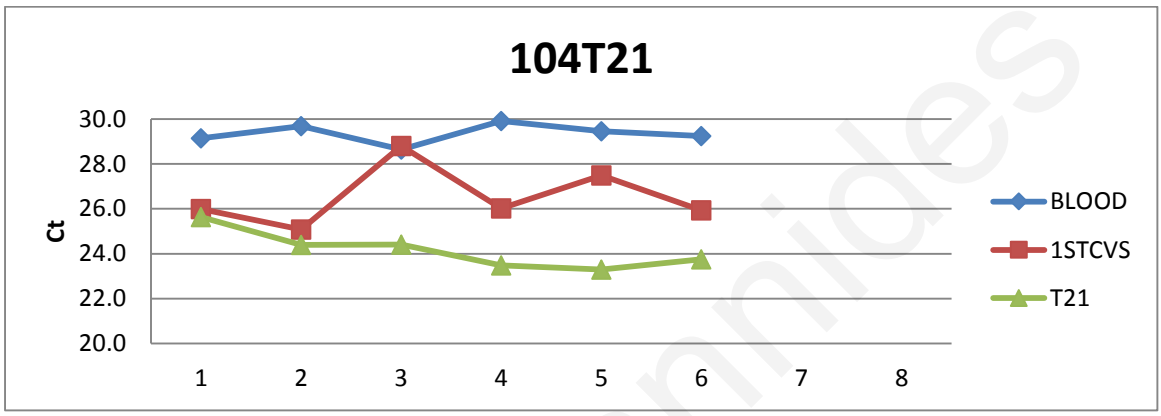


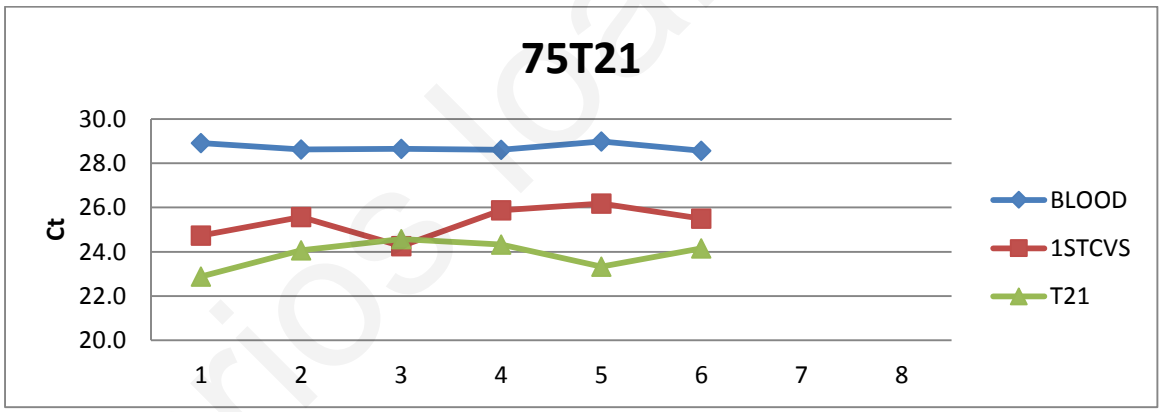


2. "Bad DMRs"

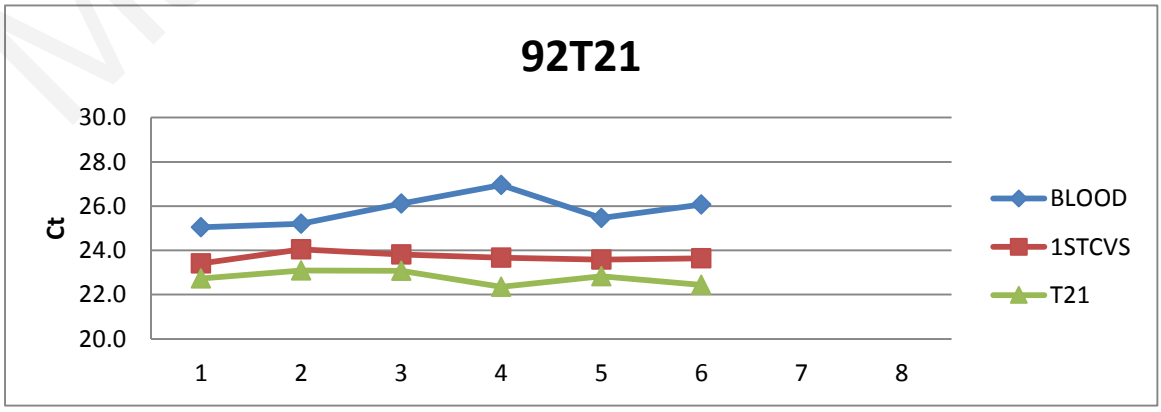


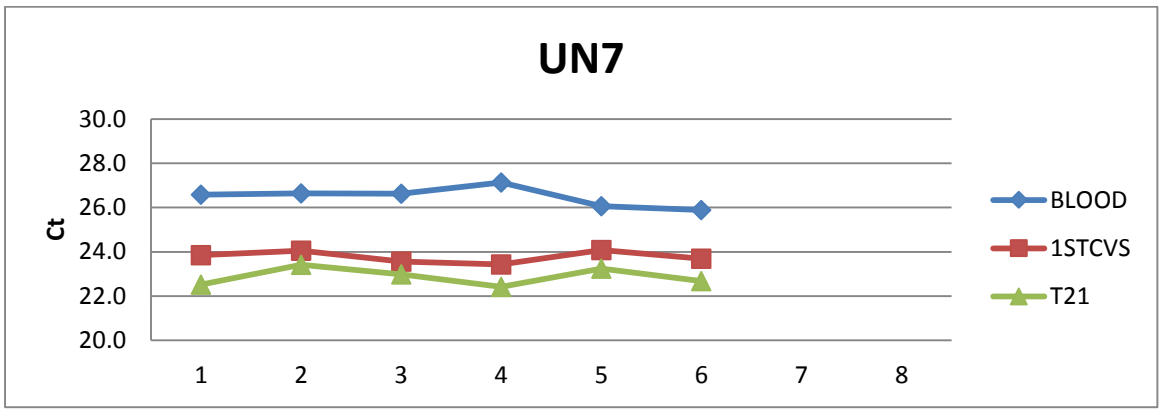


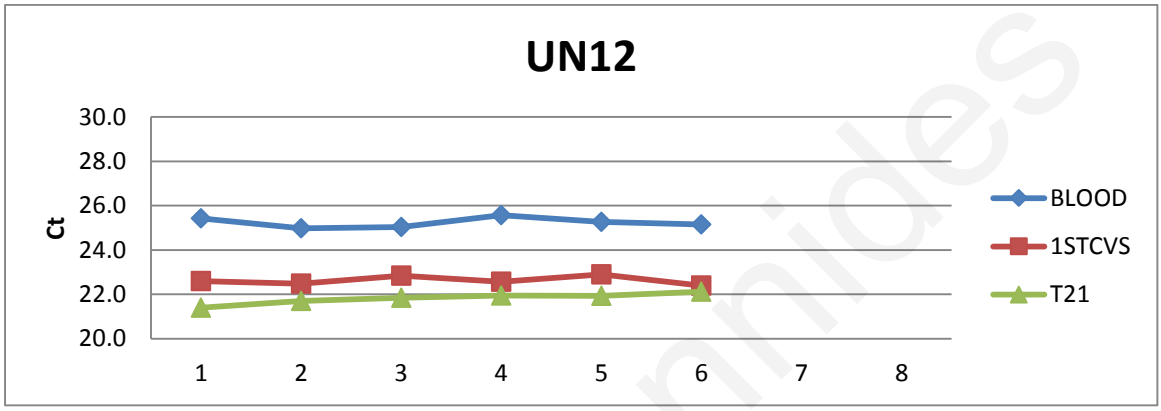


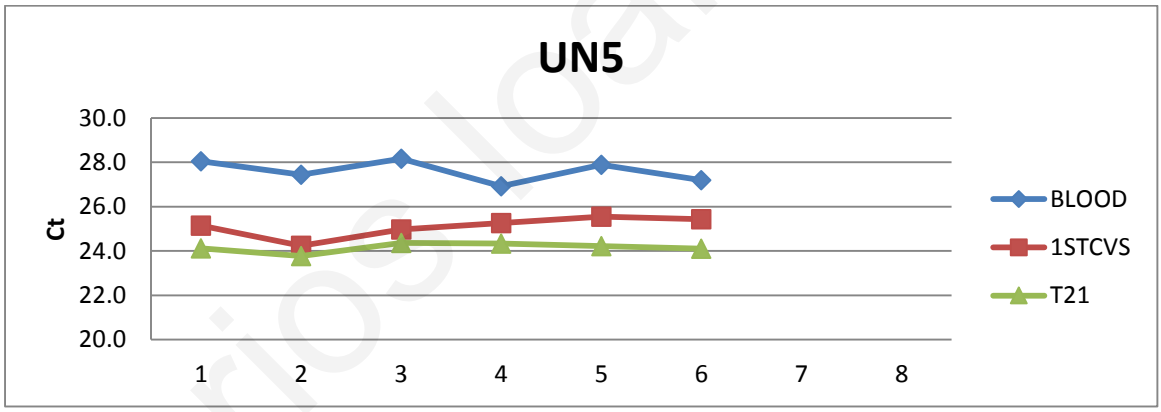


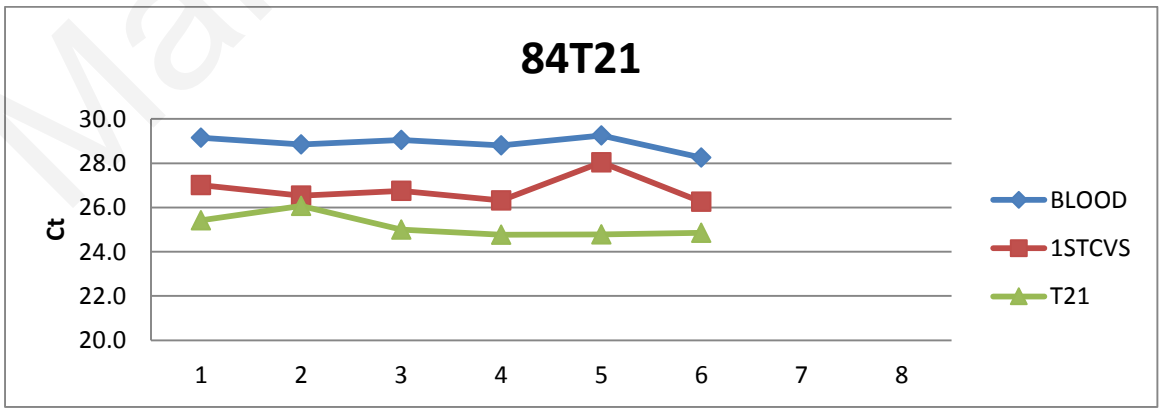
Ct

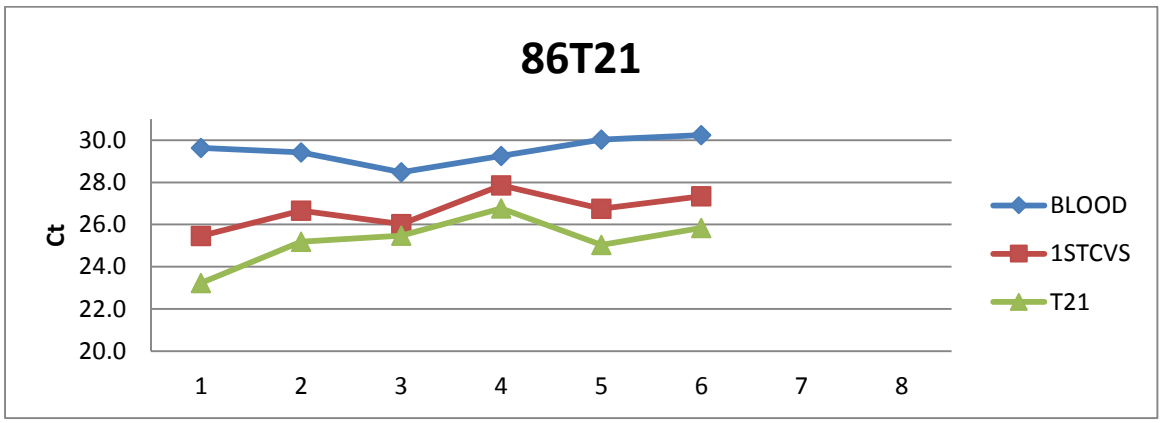


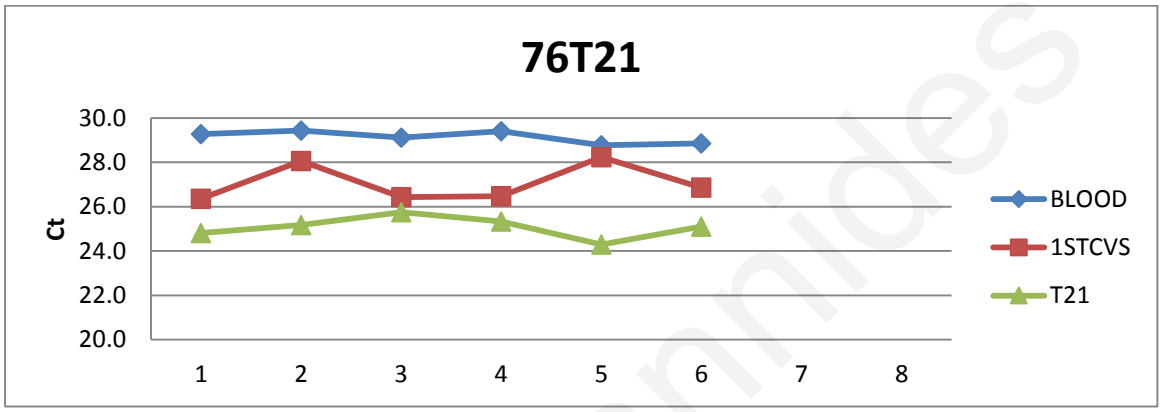


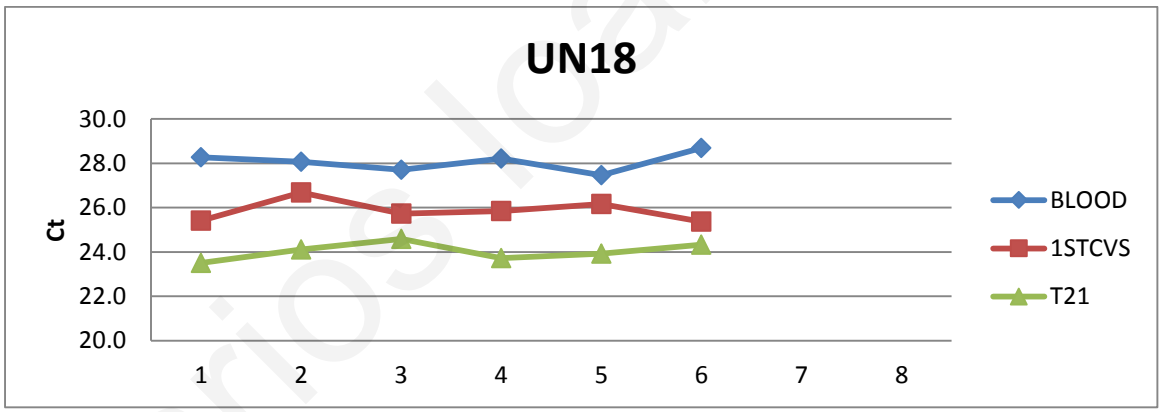


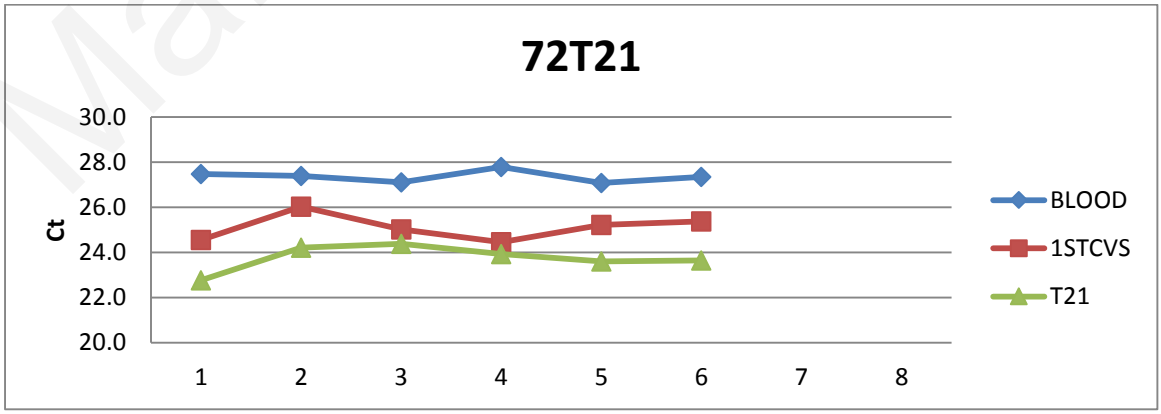


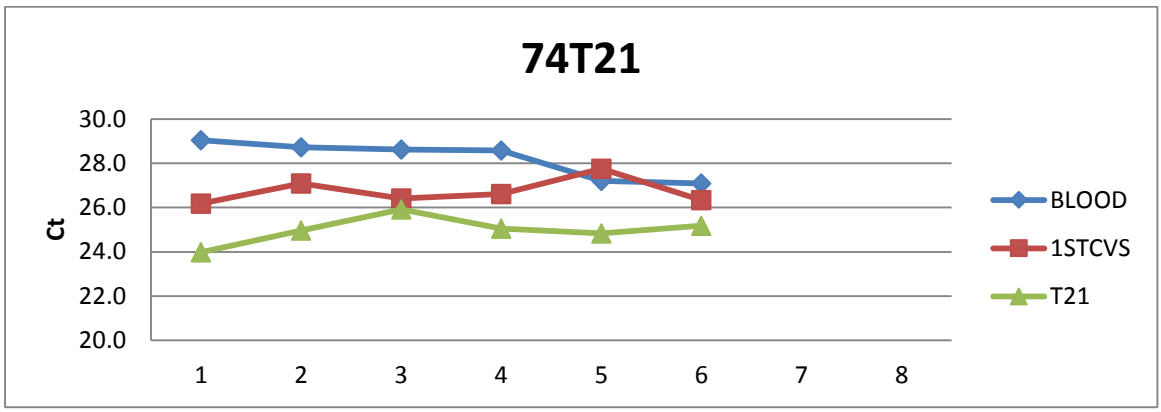


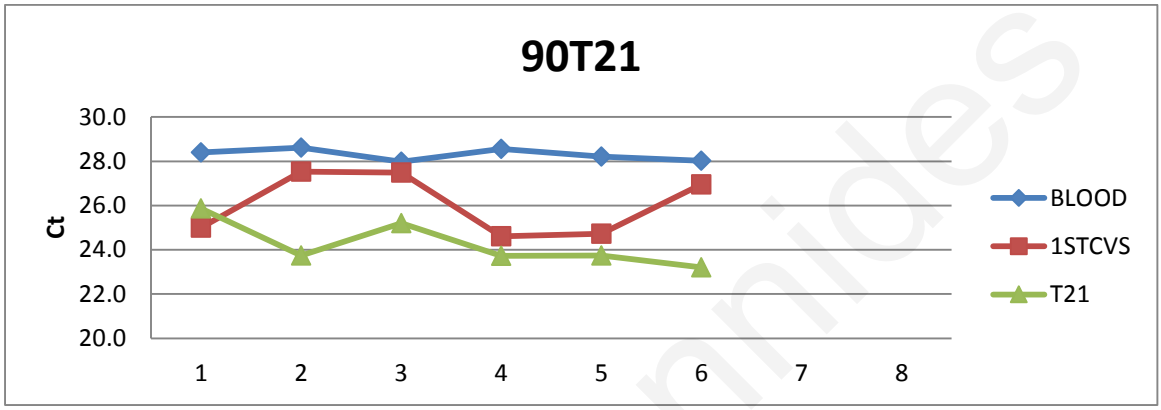


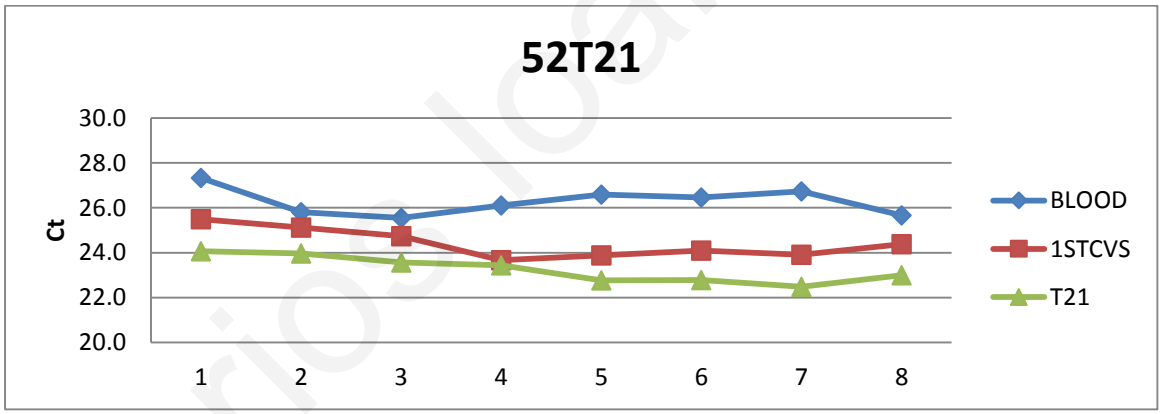


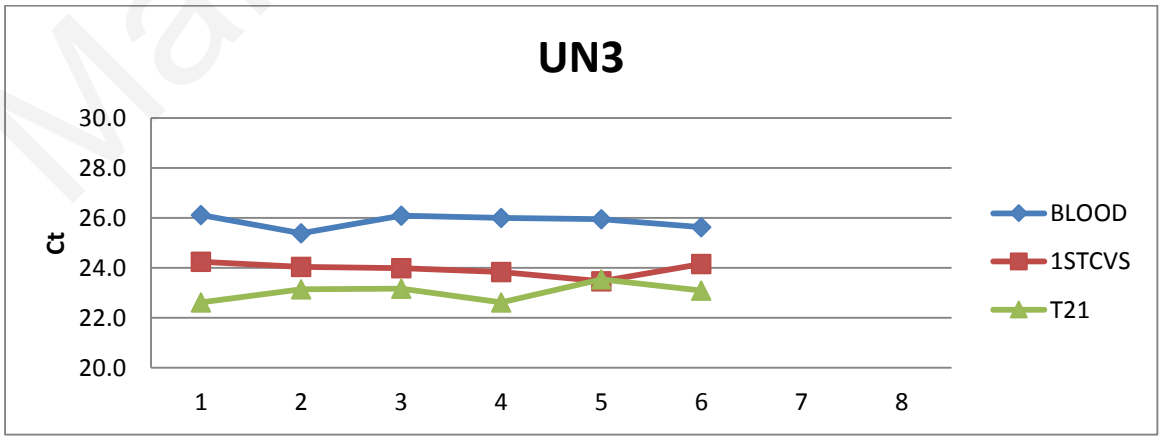


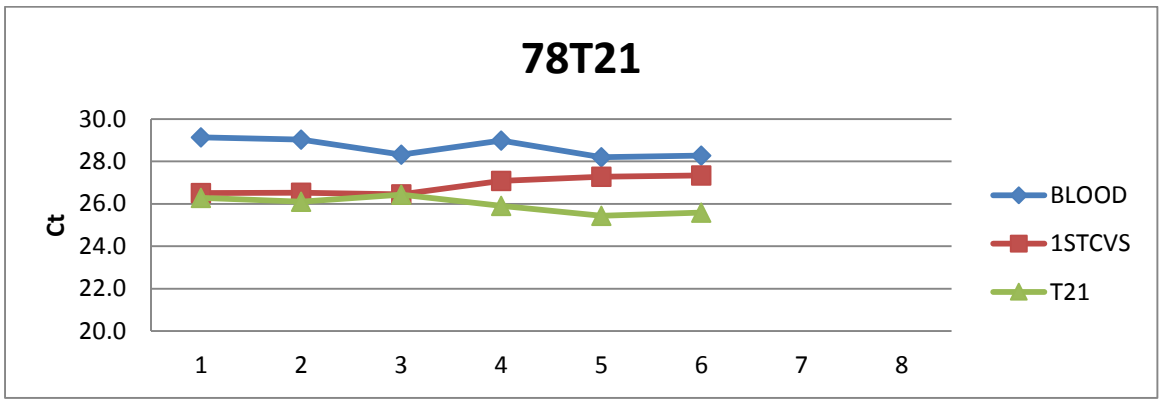


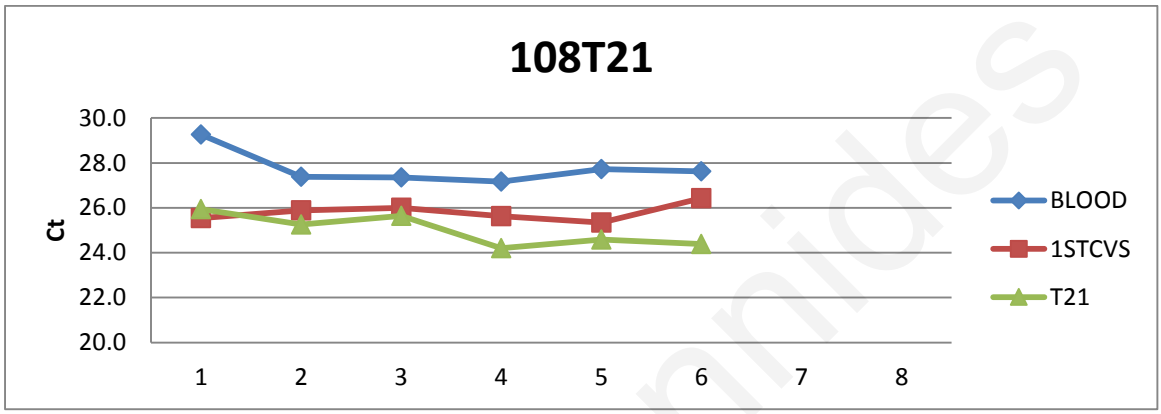


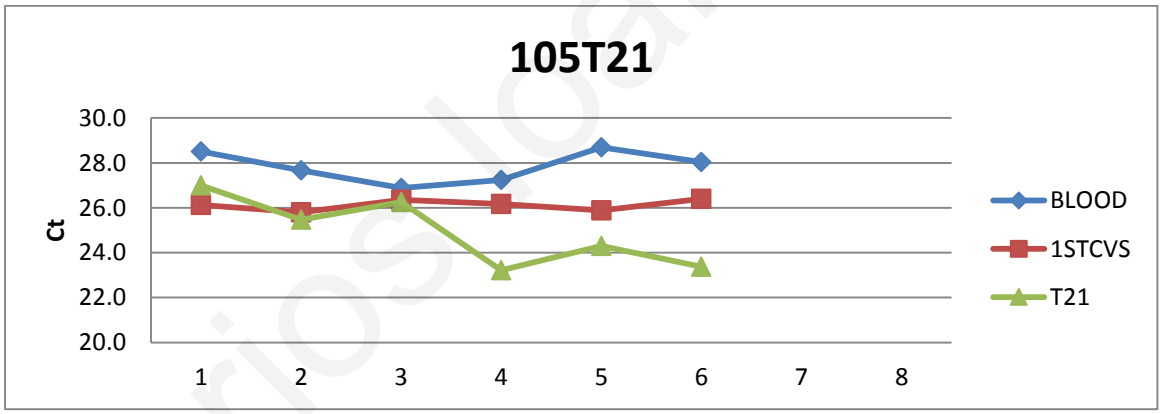


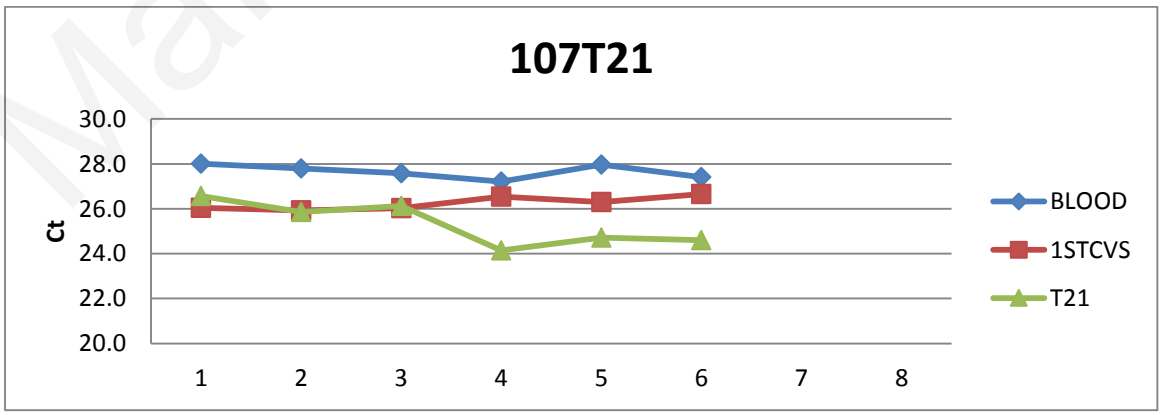


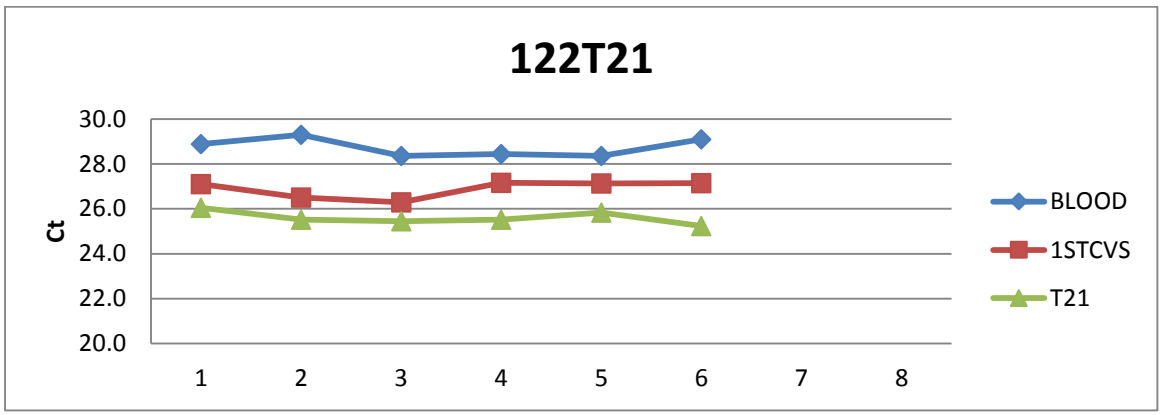


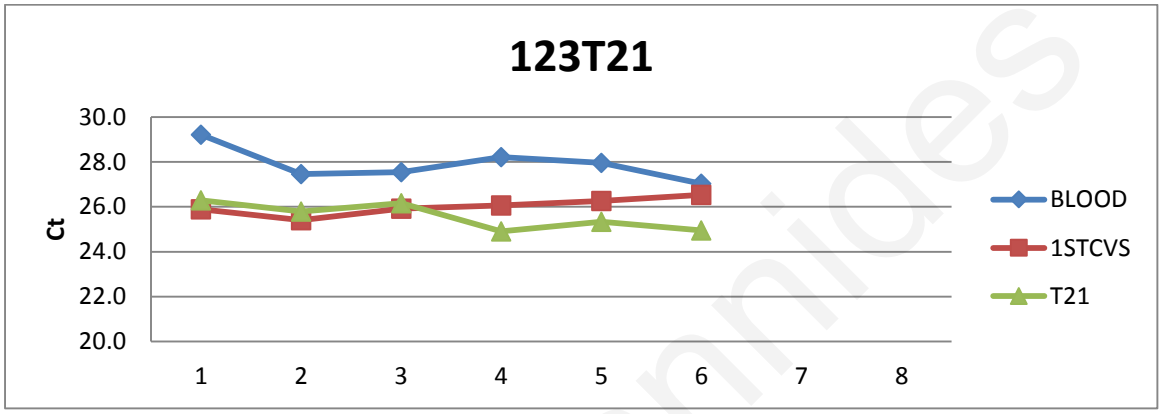


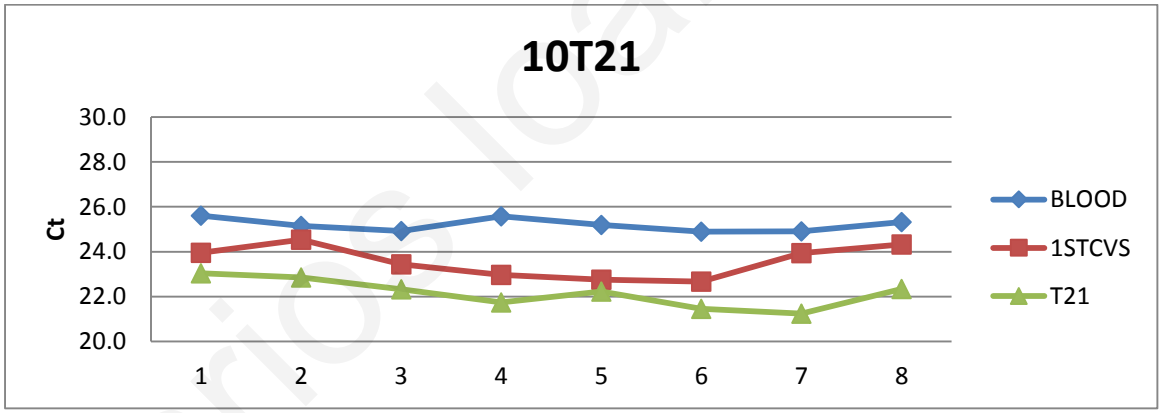


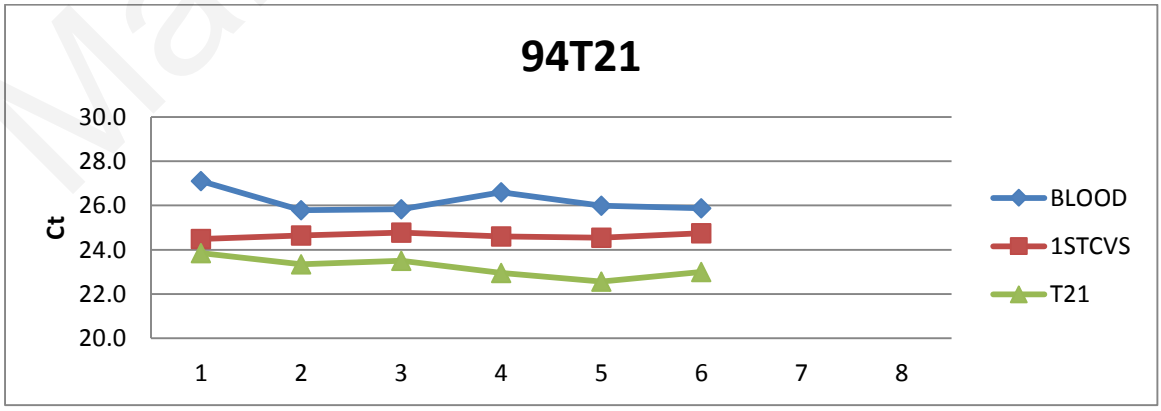


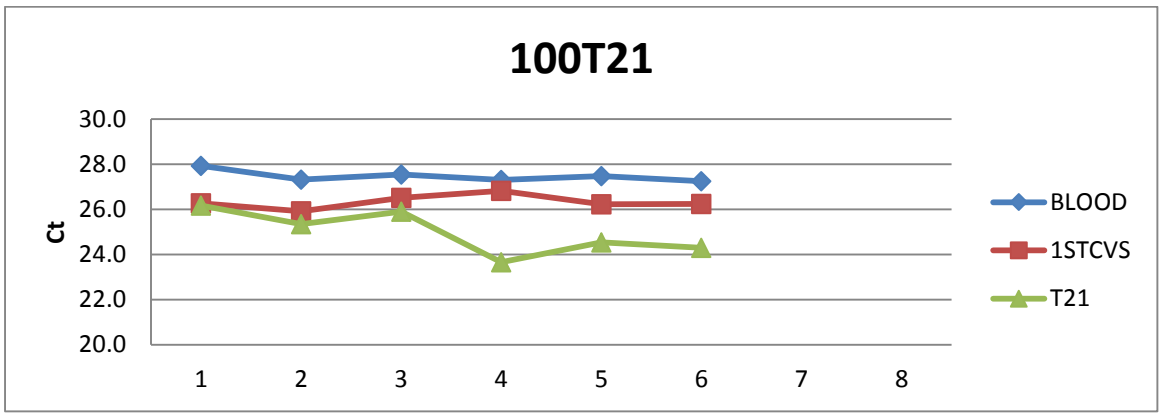


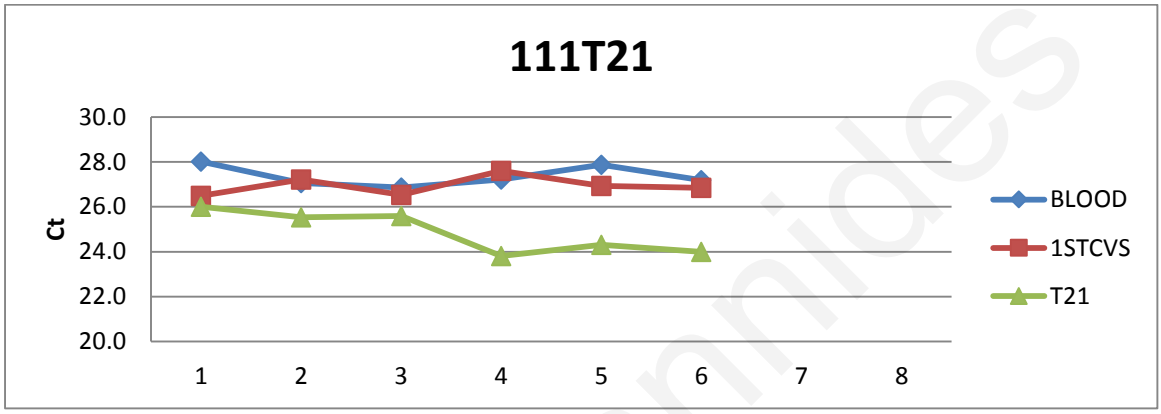


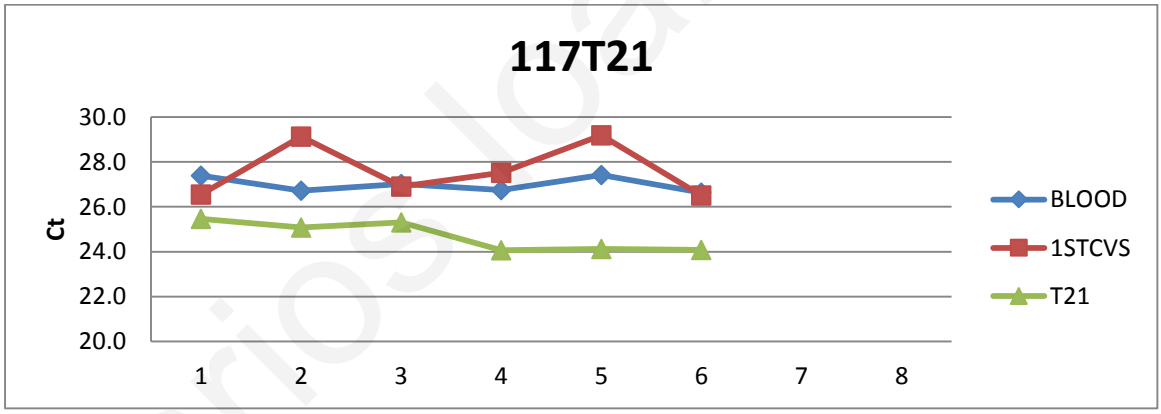




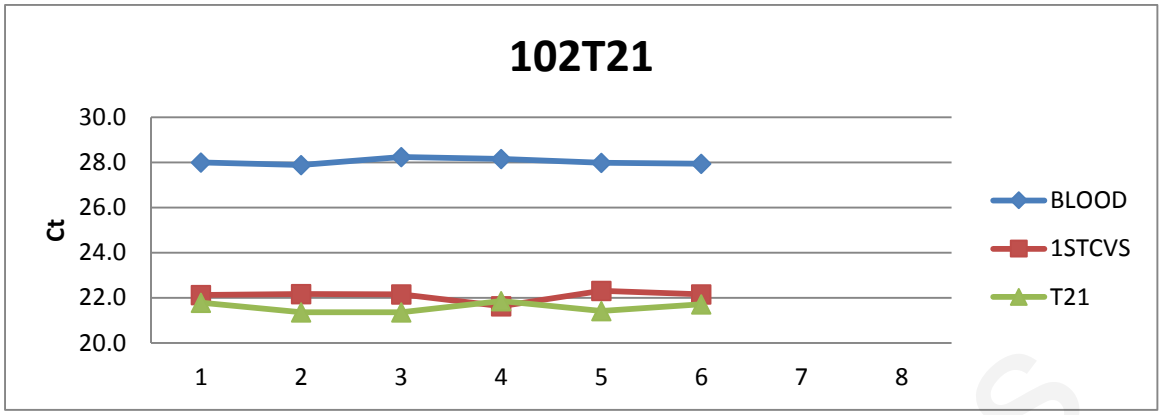


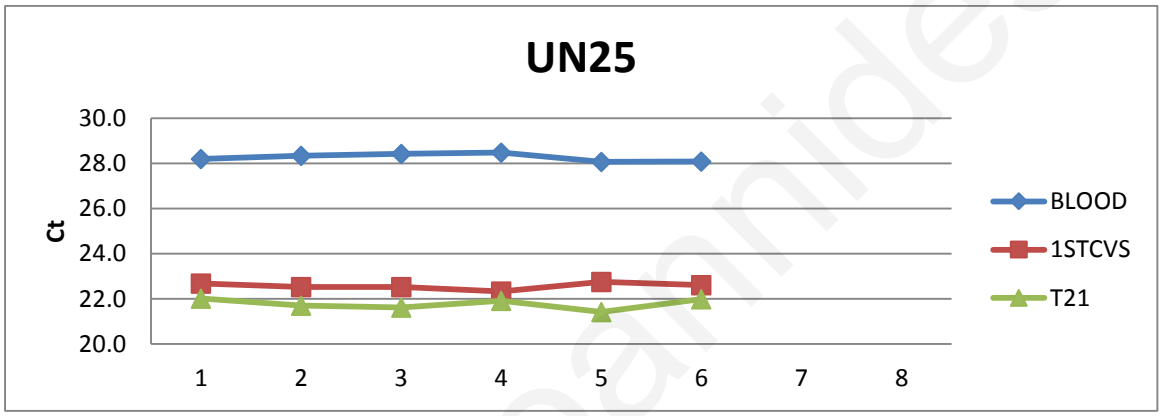


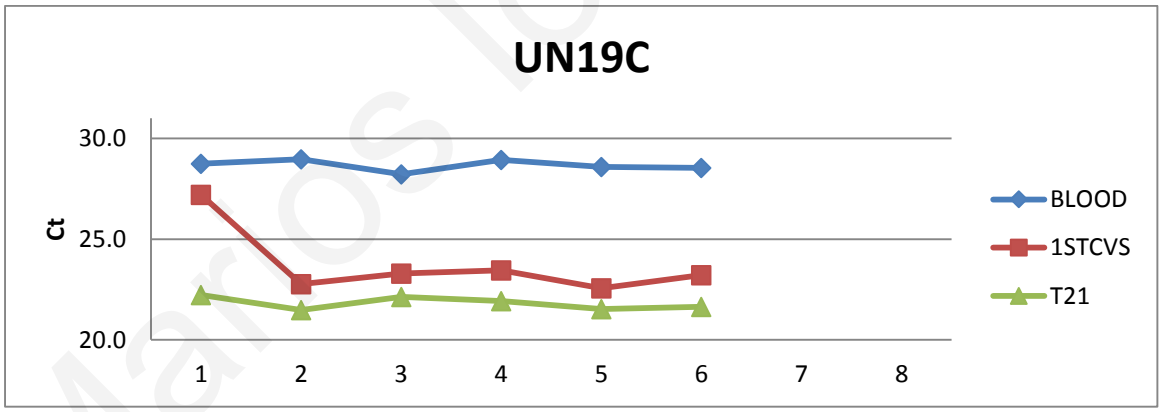


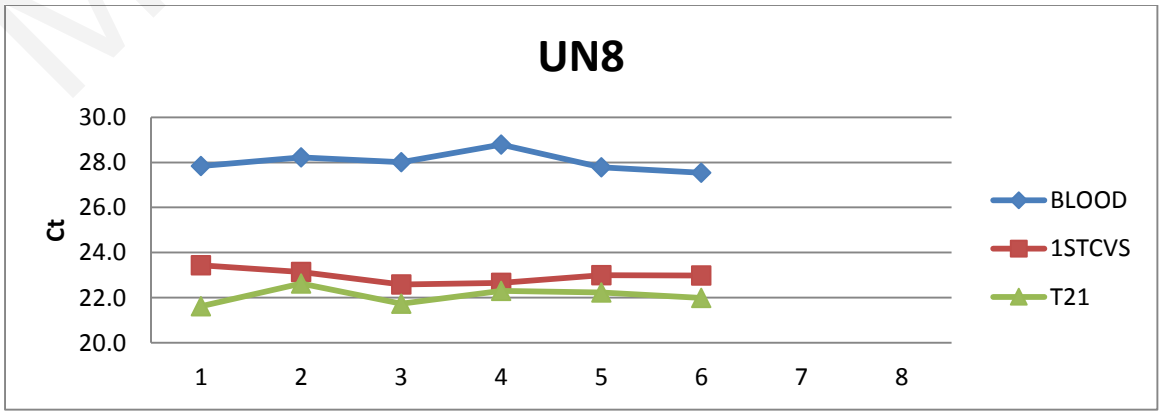


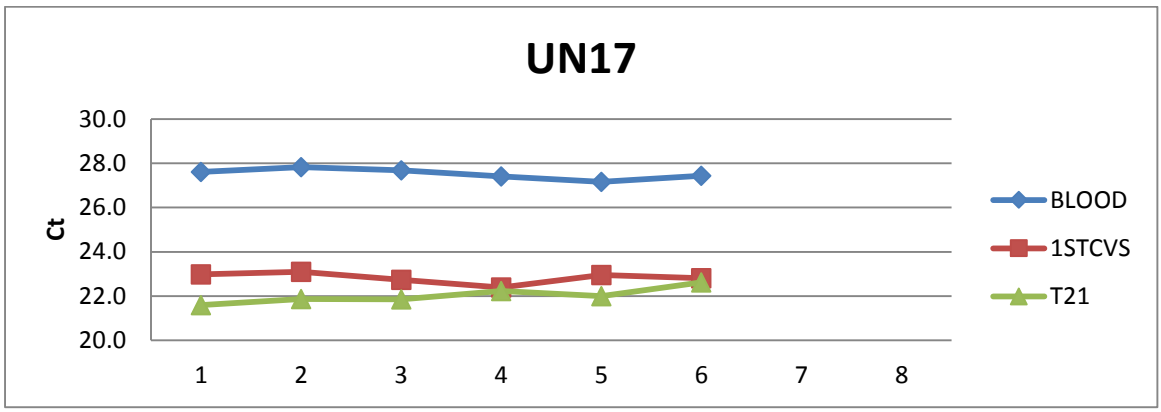
3. “Good DMRs”

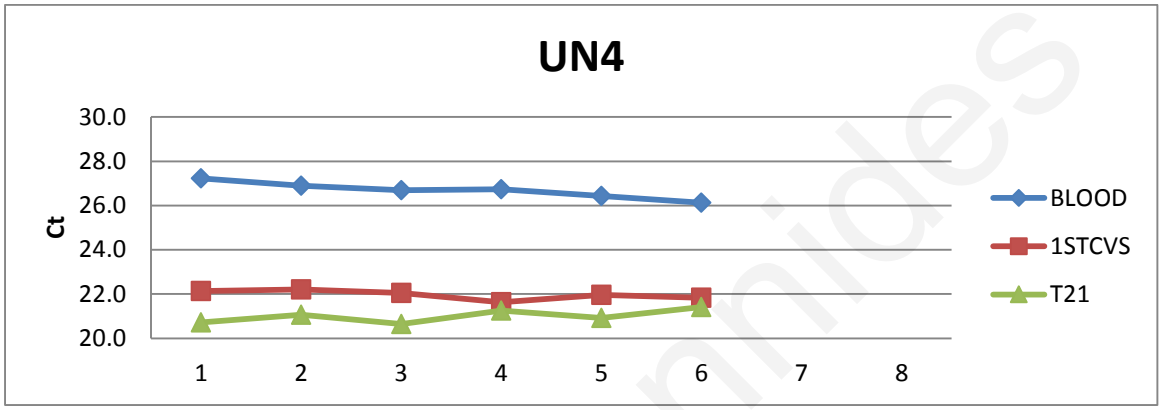


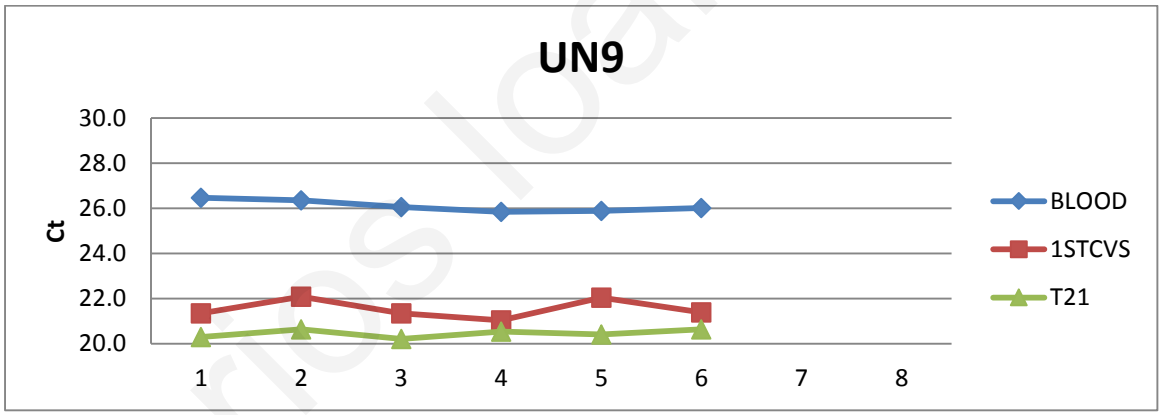


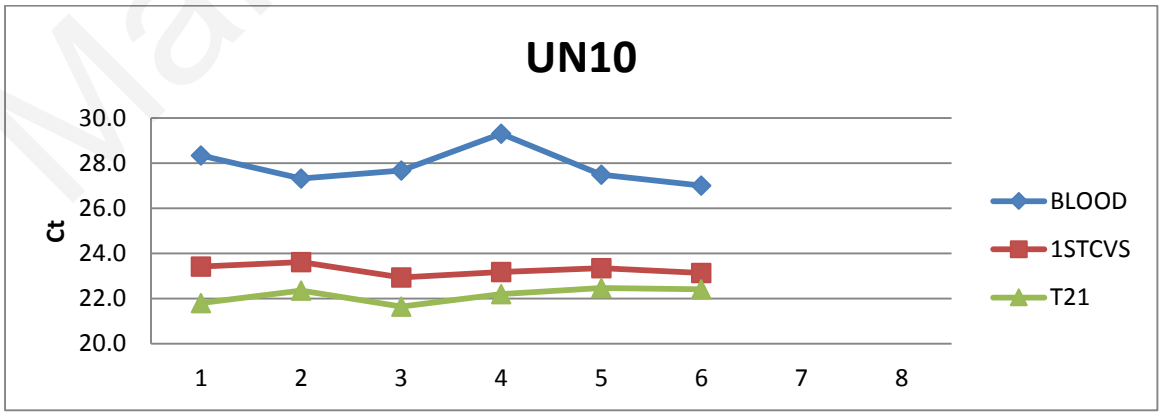


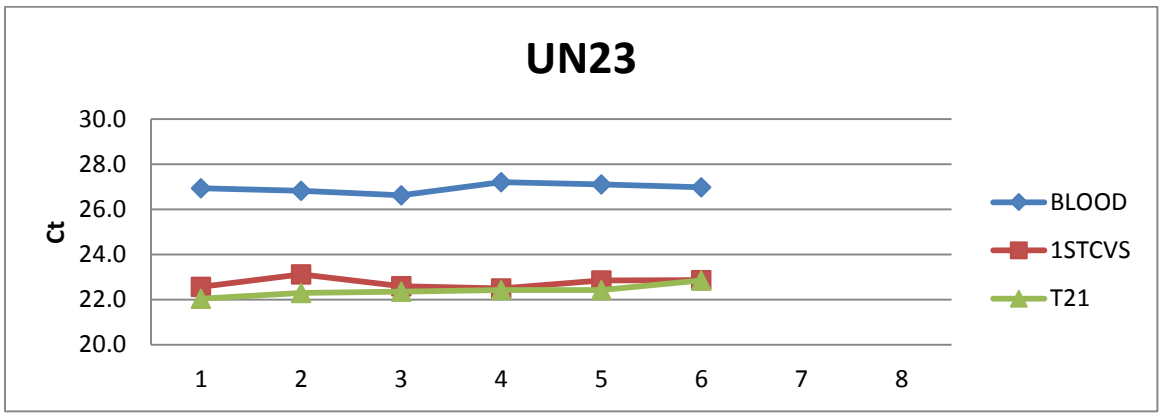


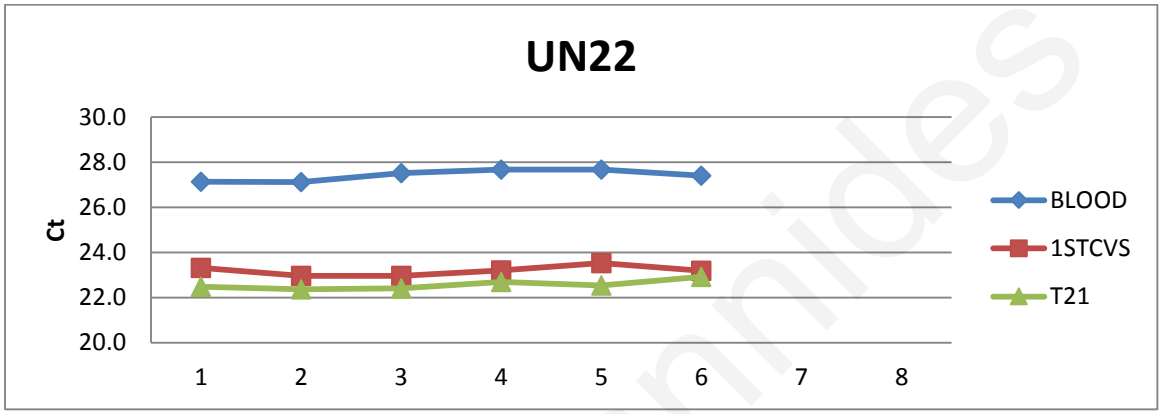


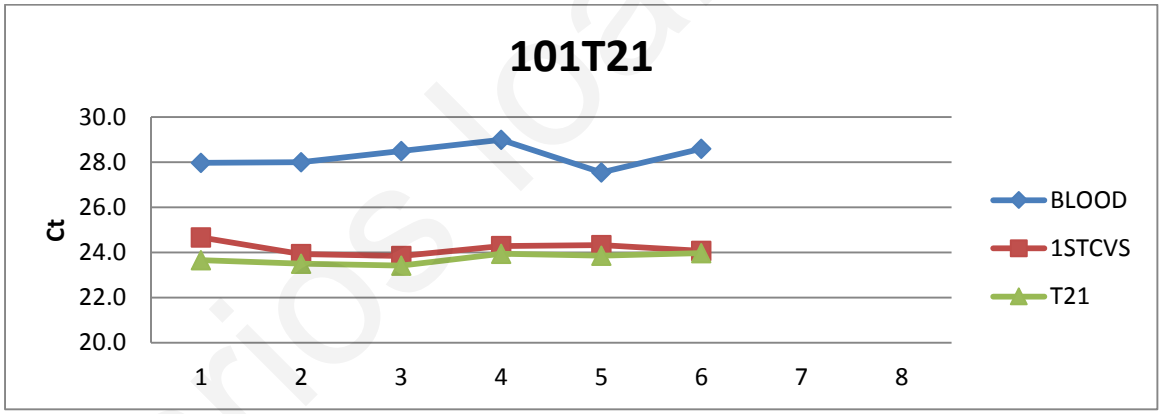


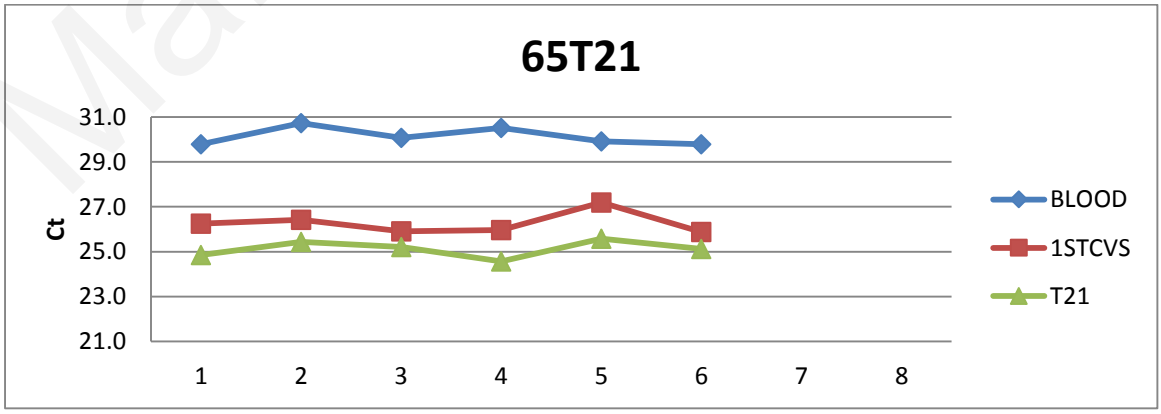


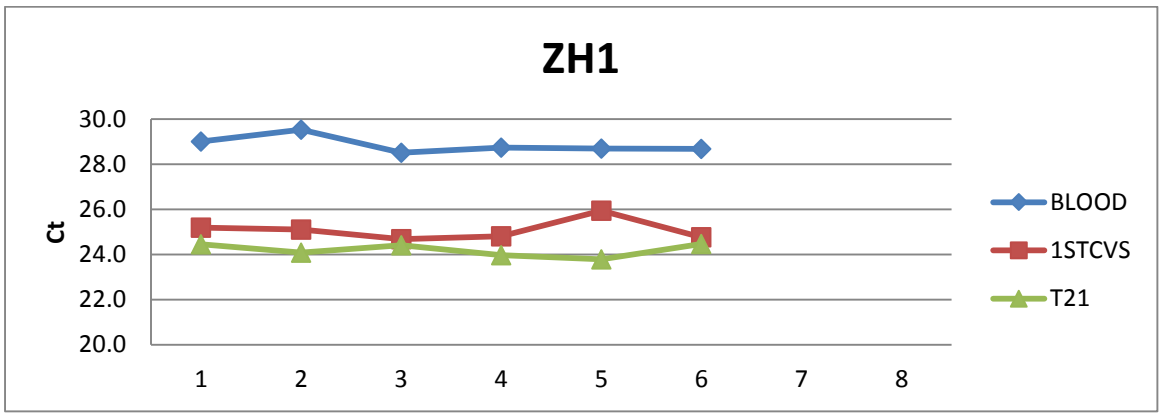


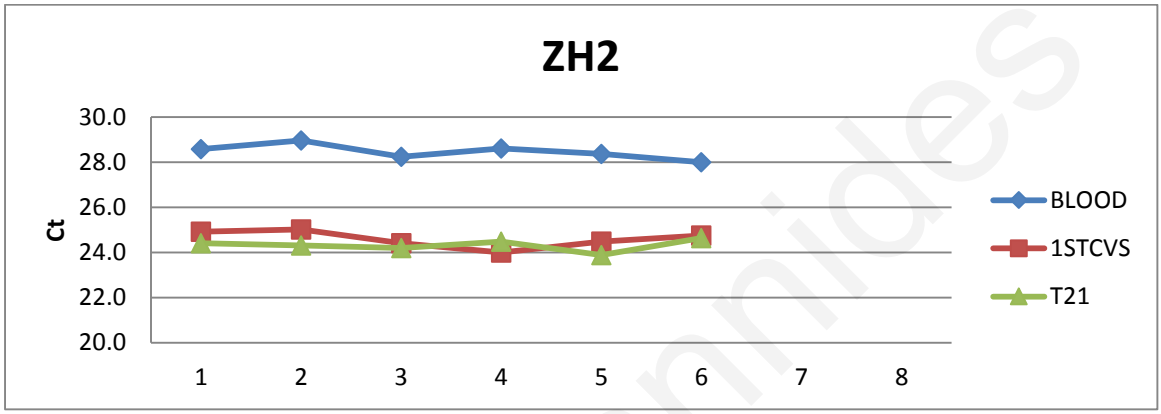


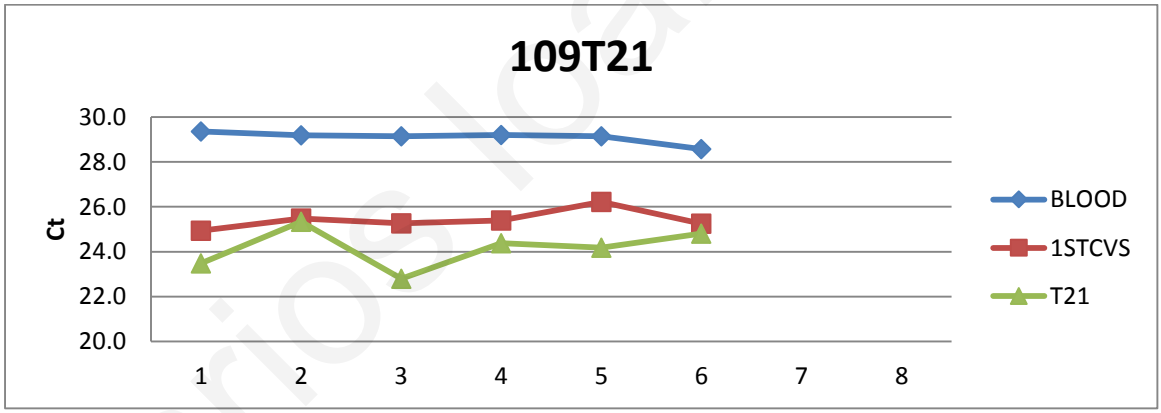


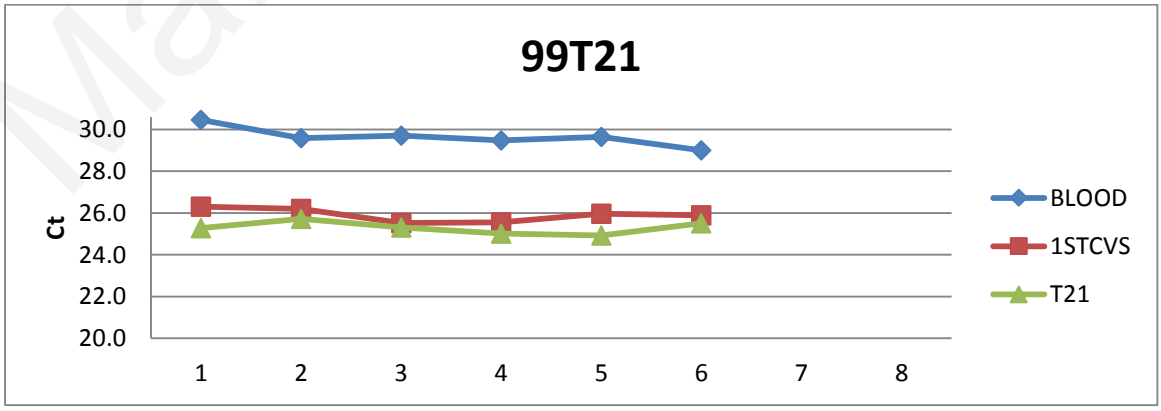


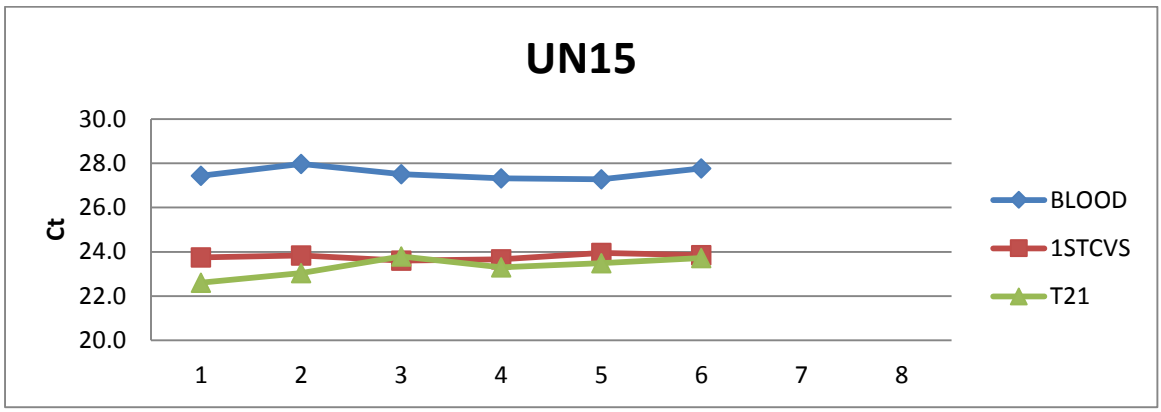


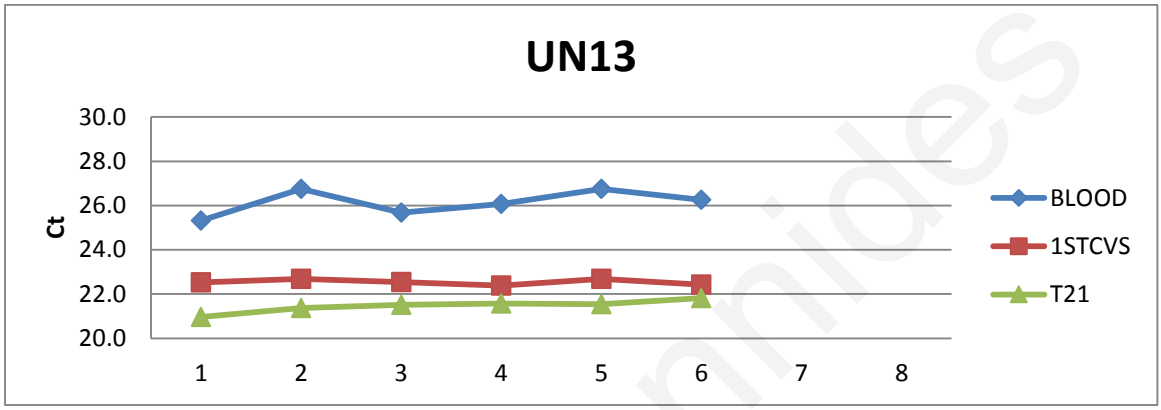


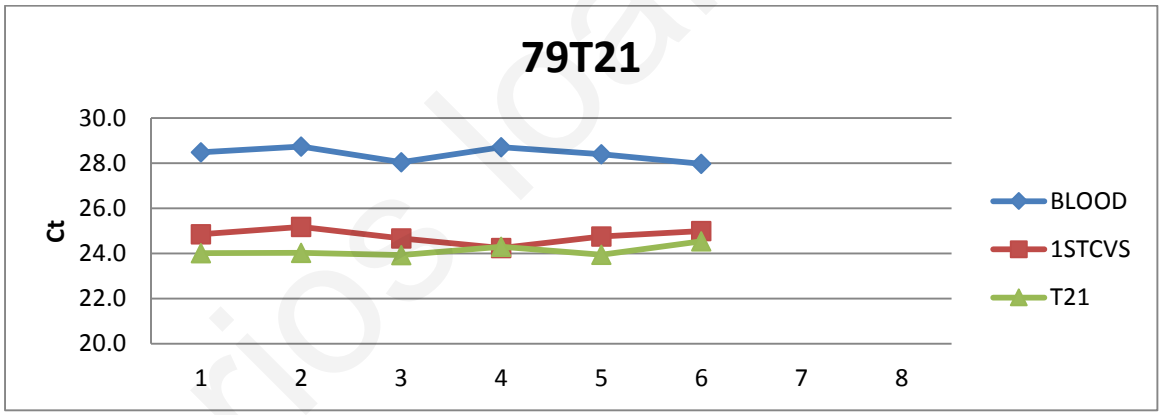


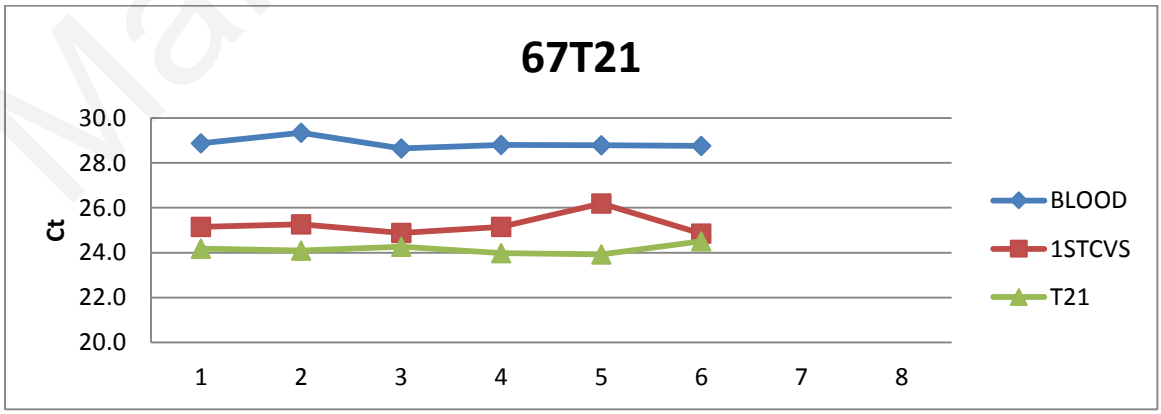


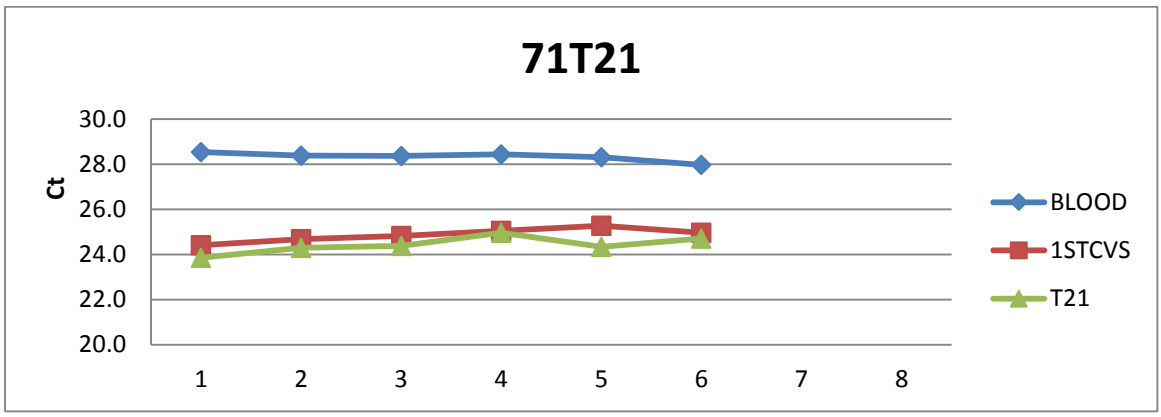


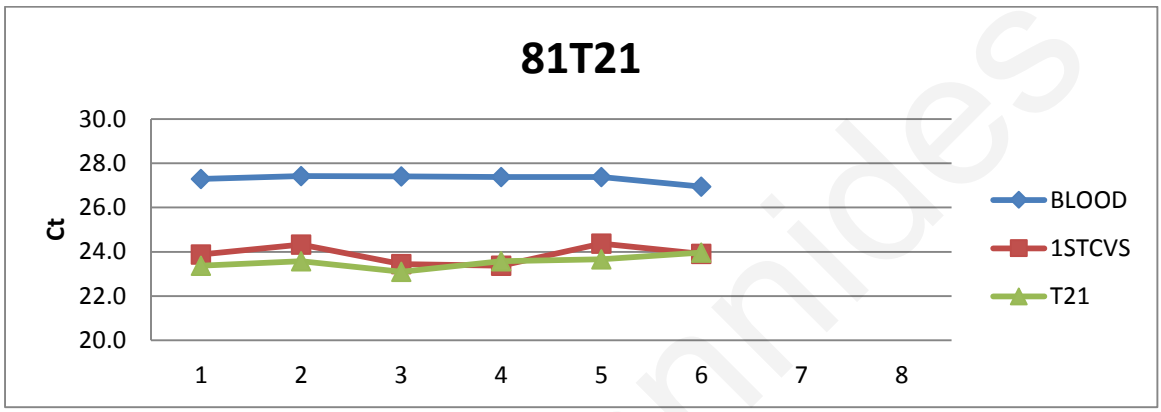


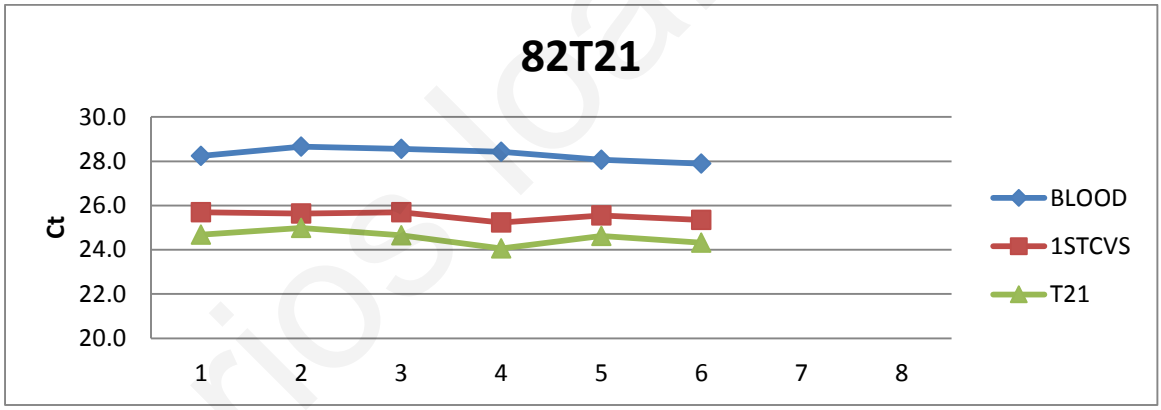


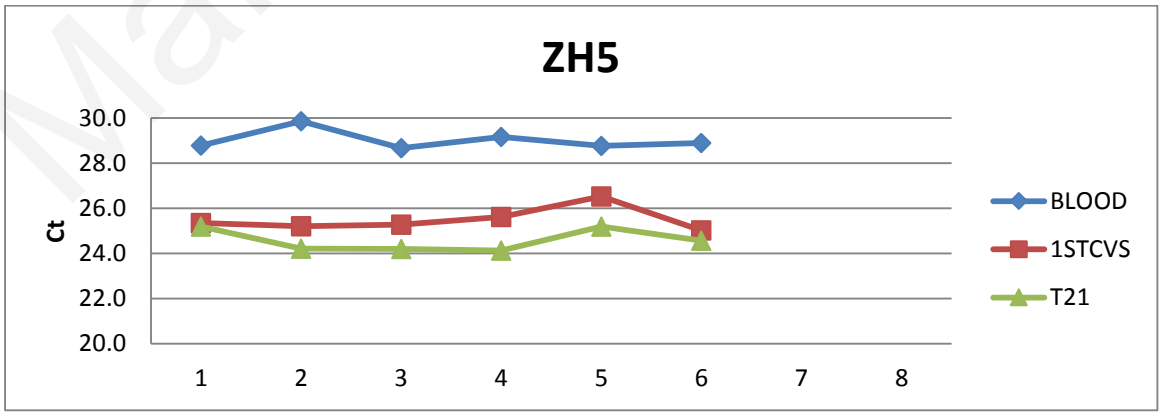


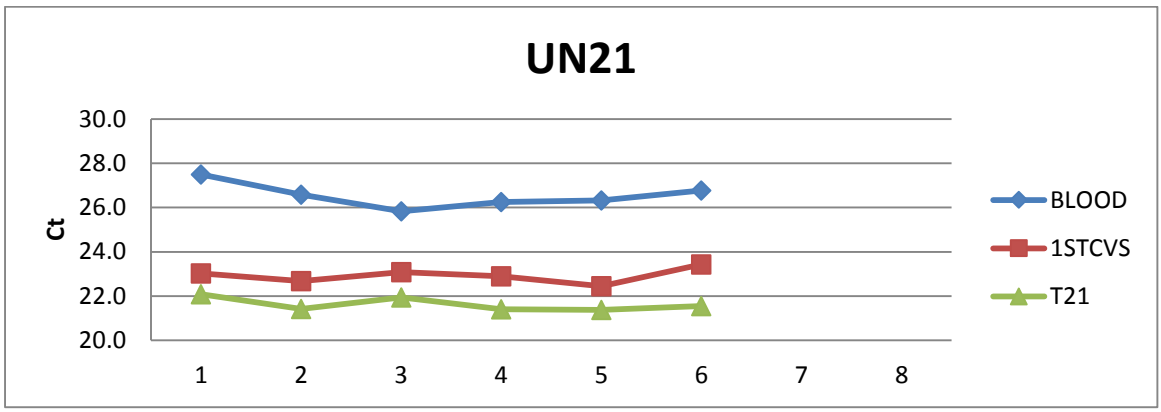


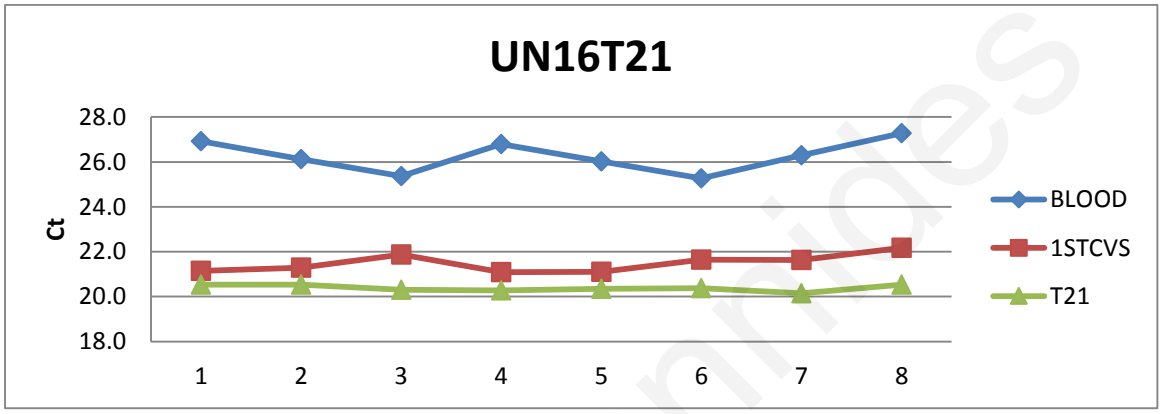


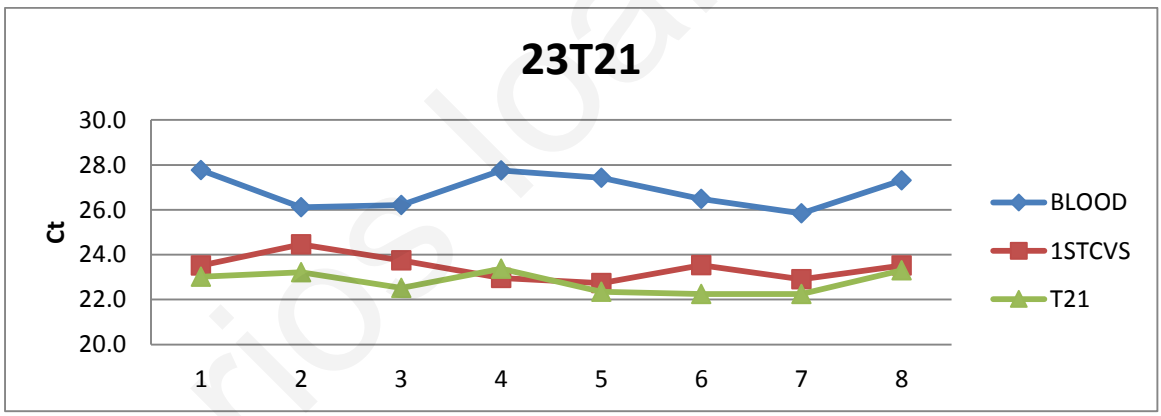


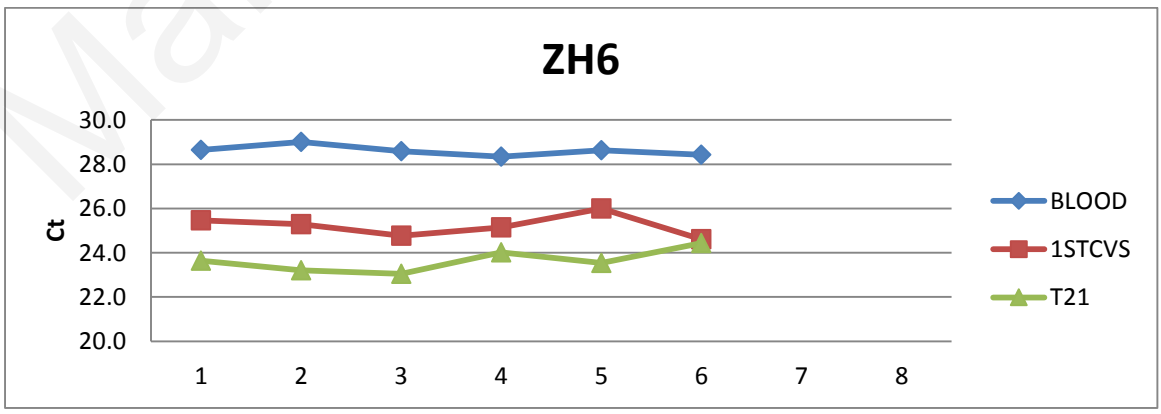


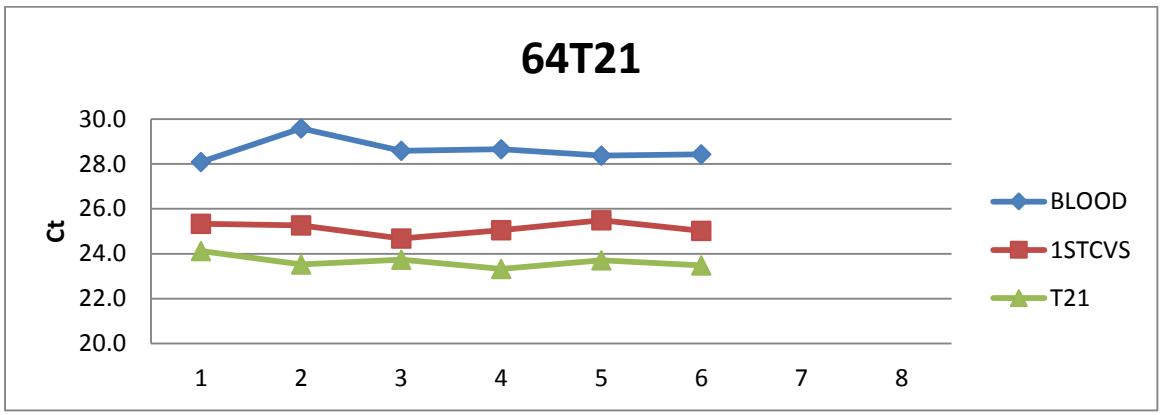


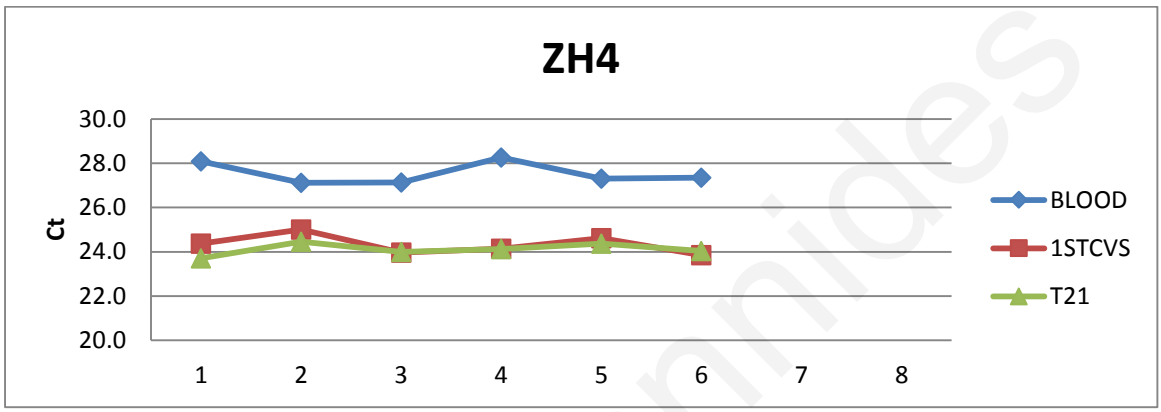


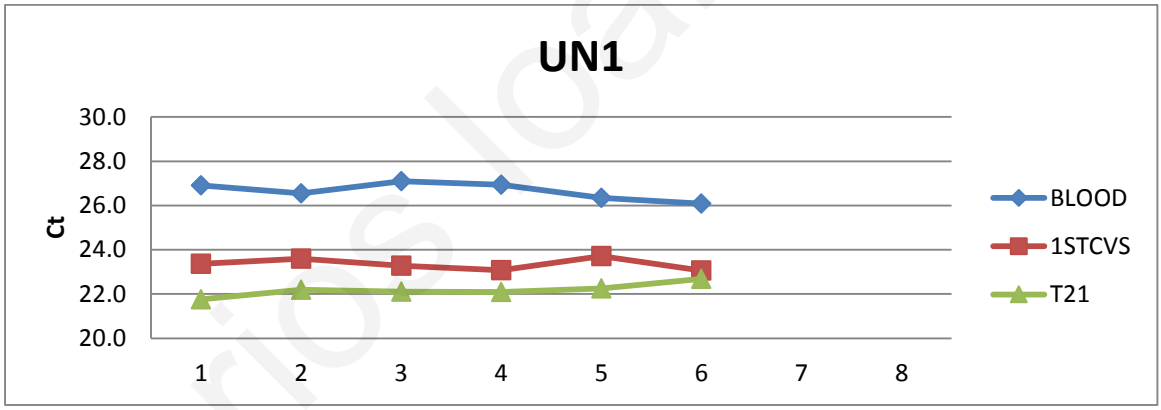


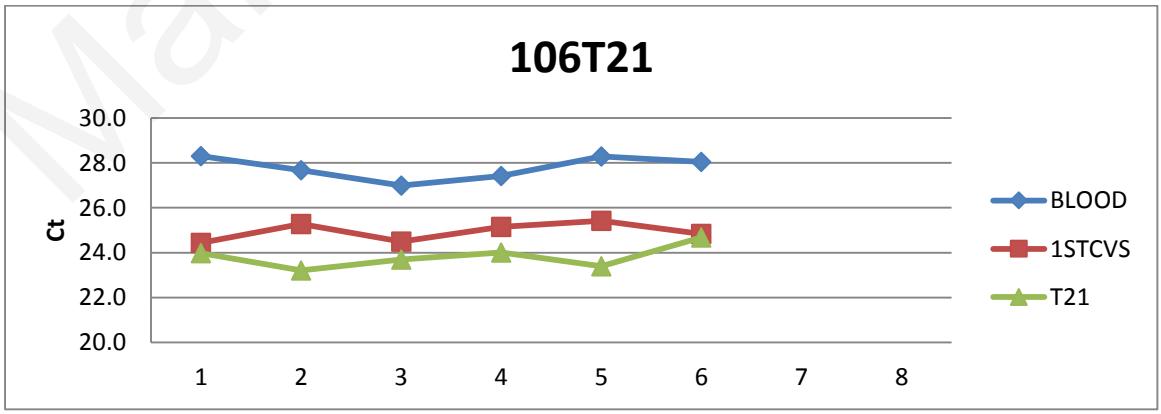


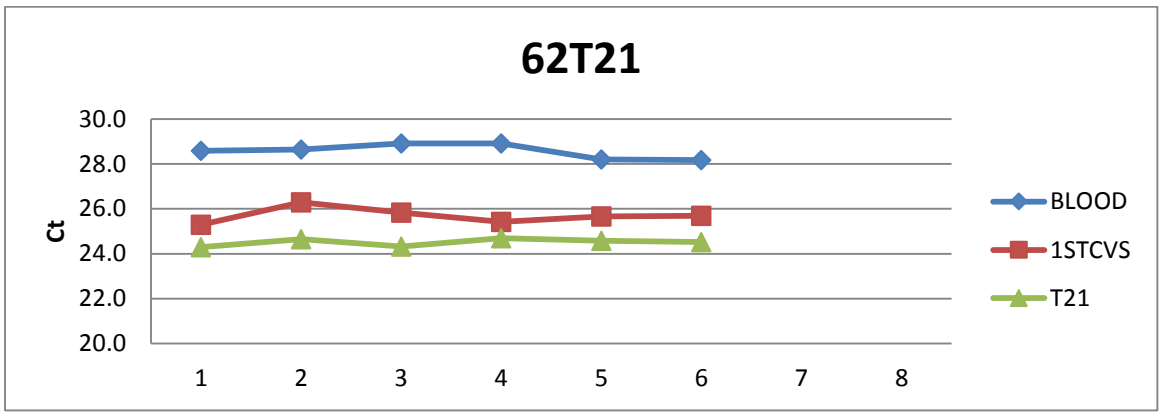


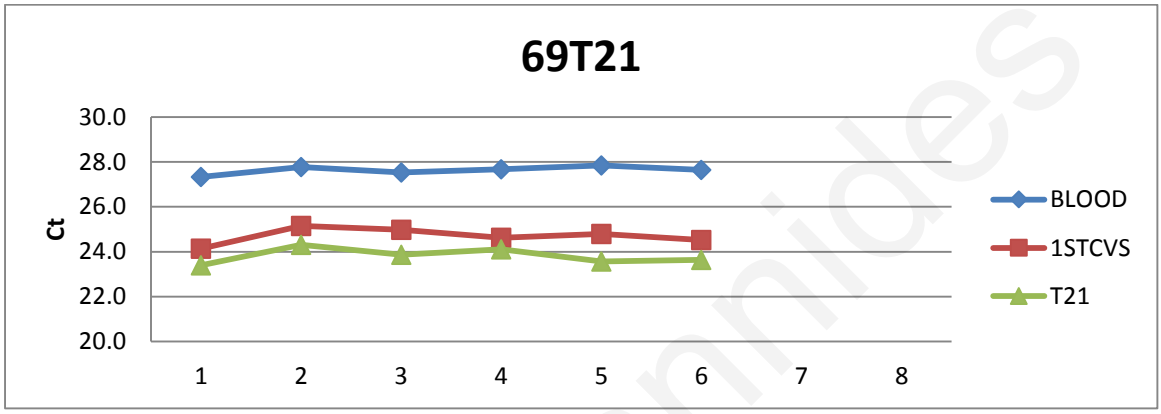


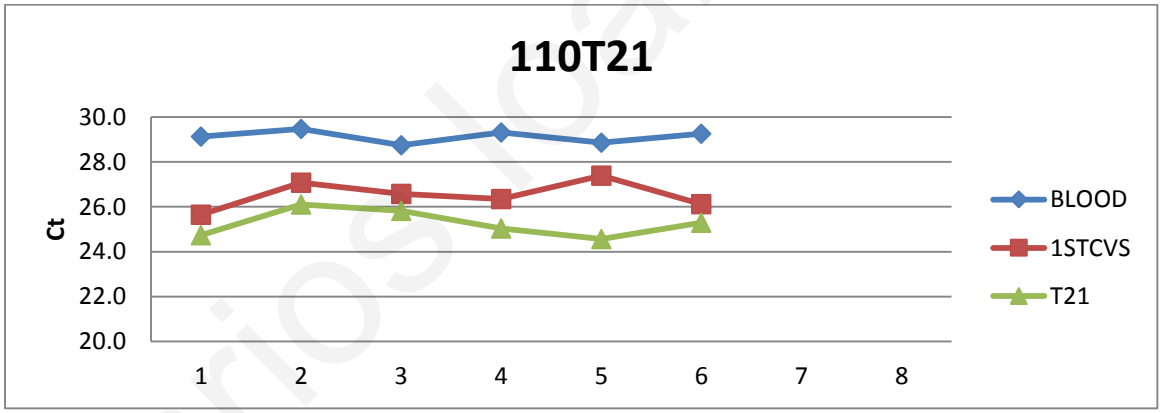


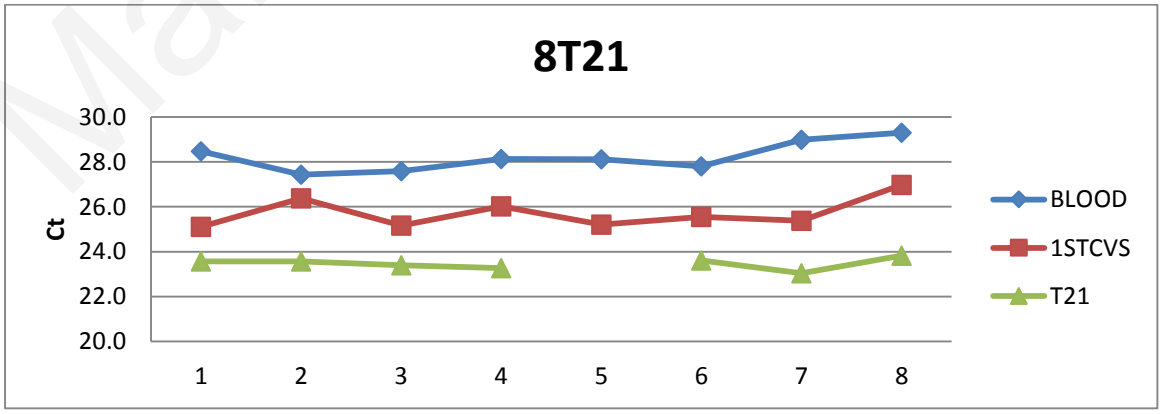


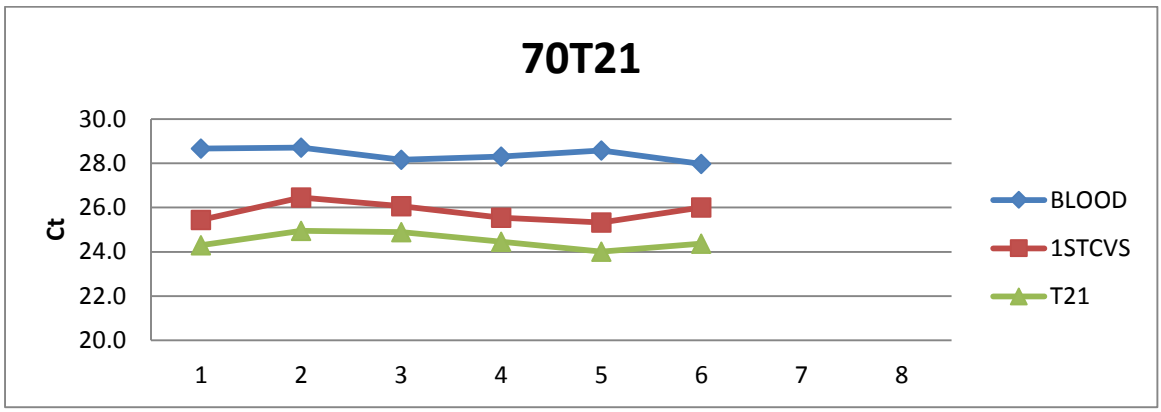


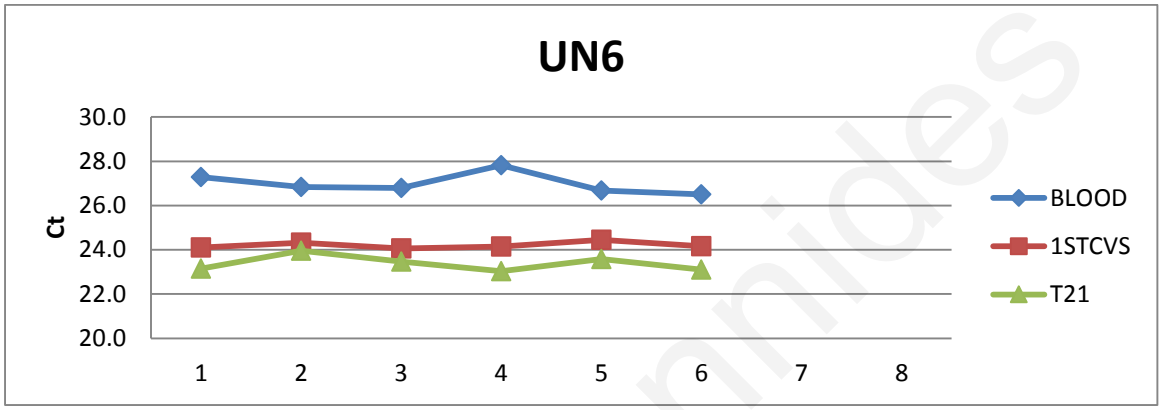


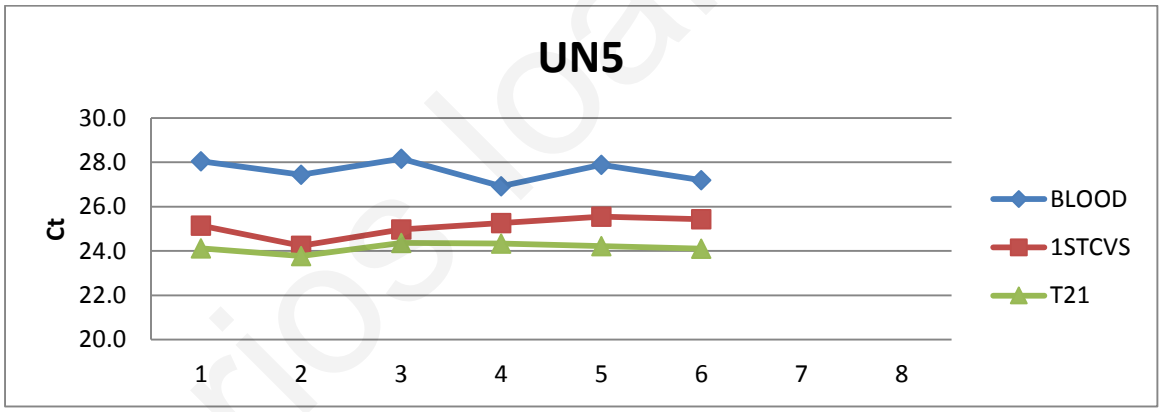


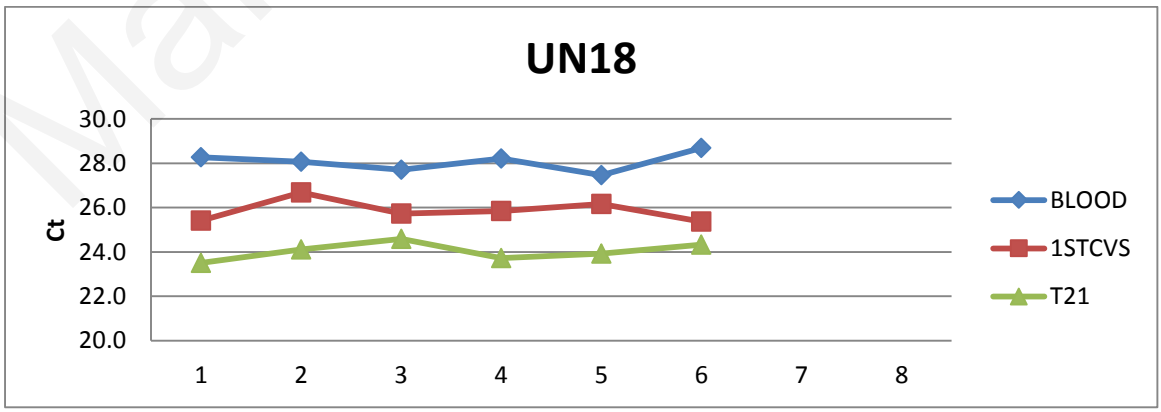


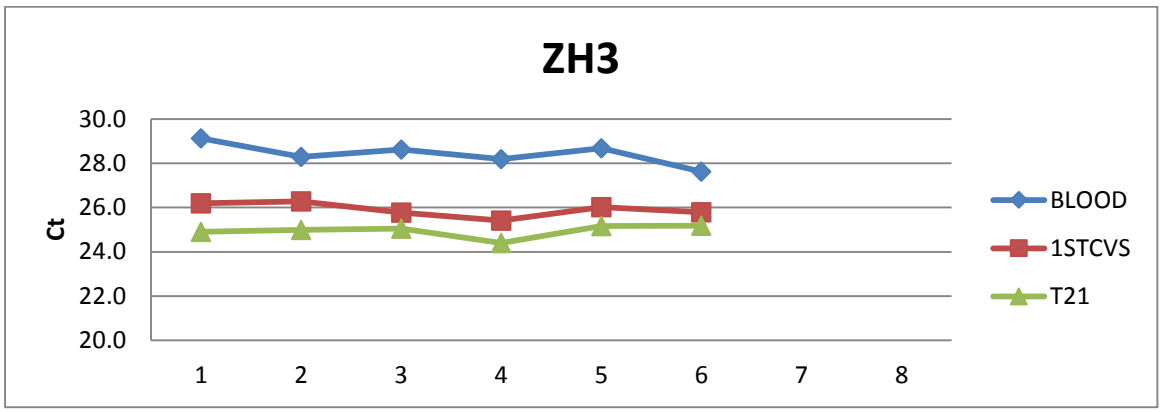


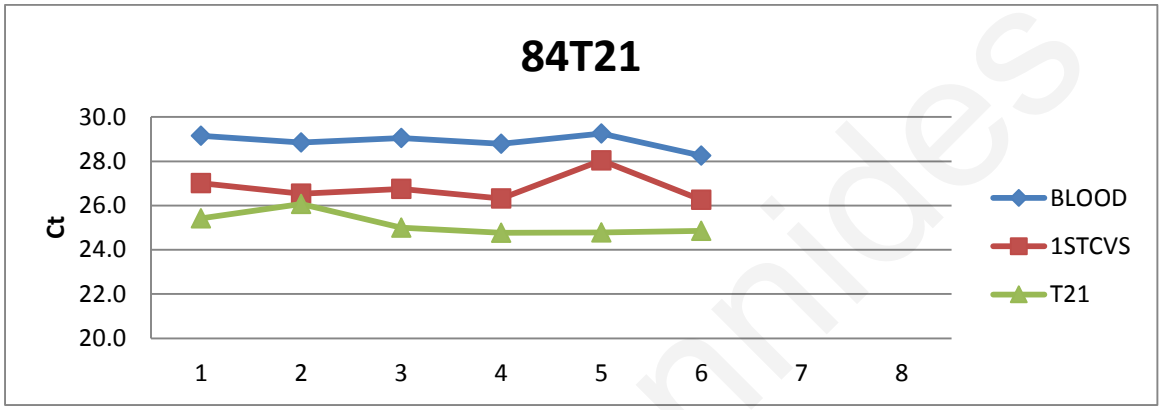




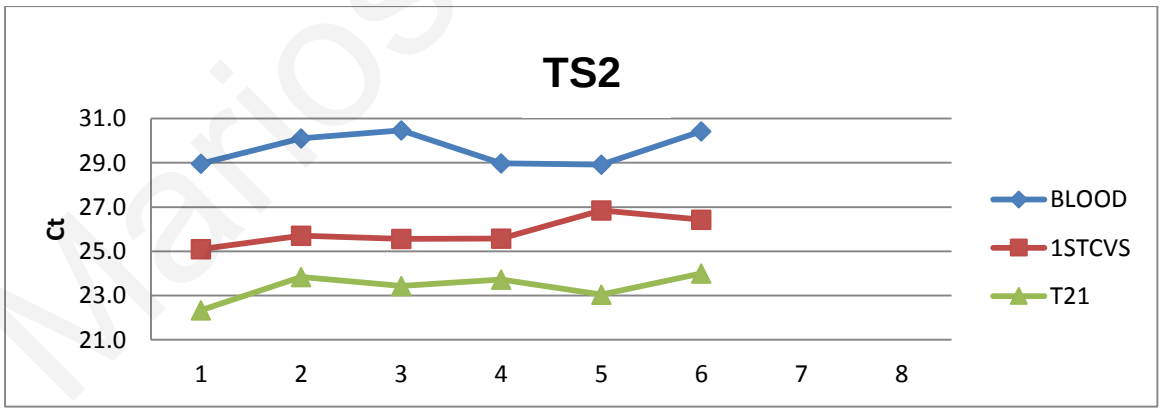


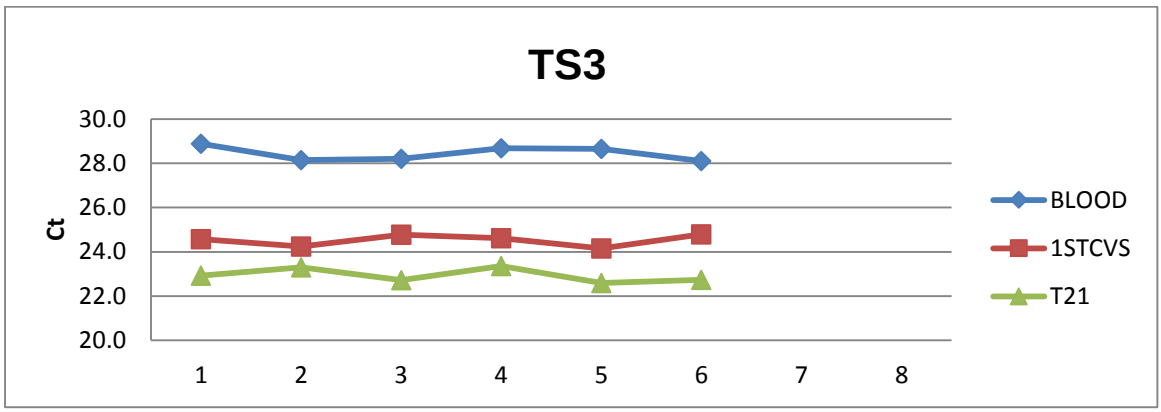


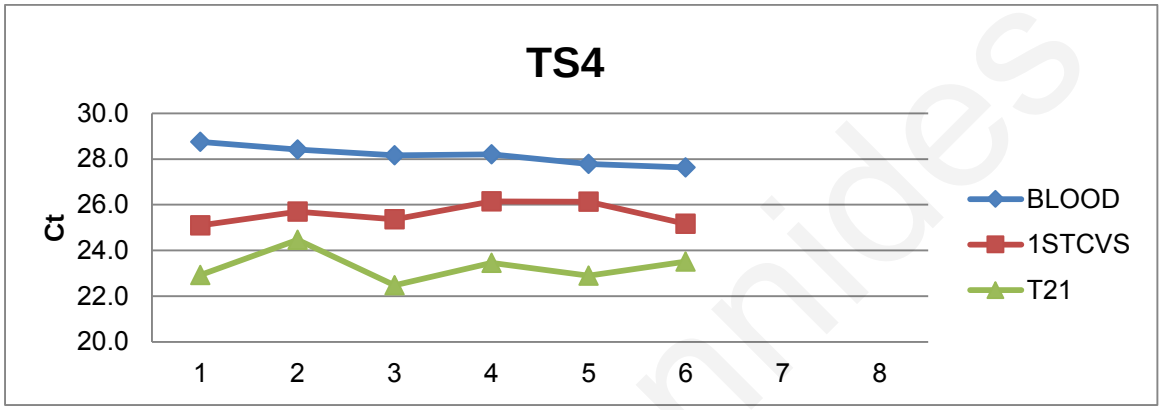


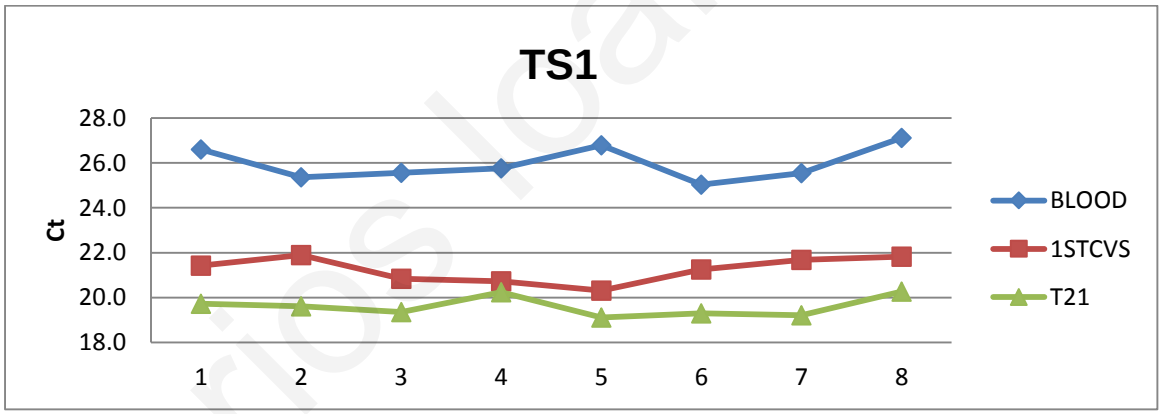


4. "Good DMRs-T21"



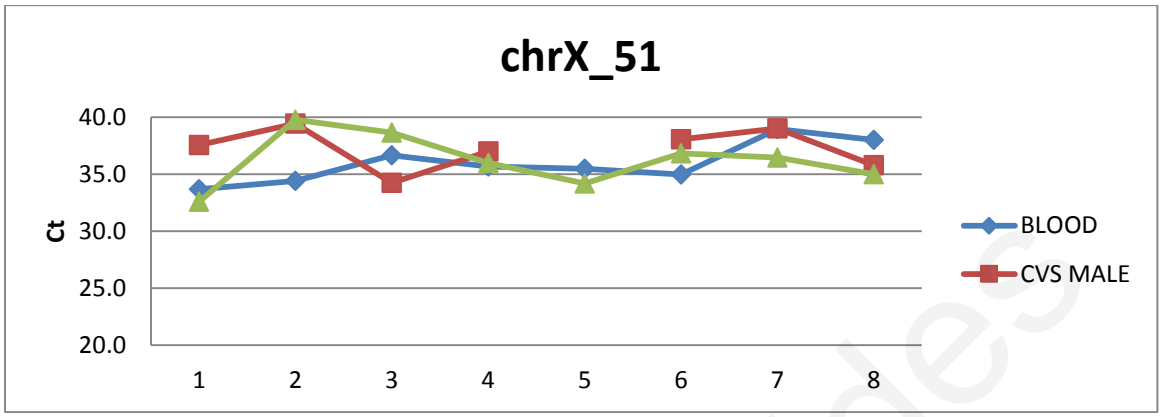


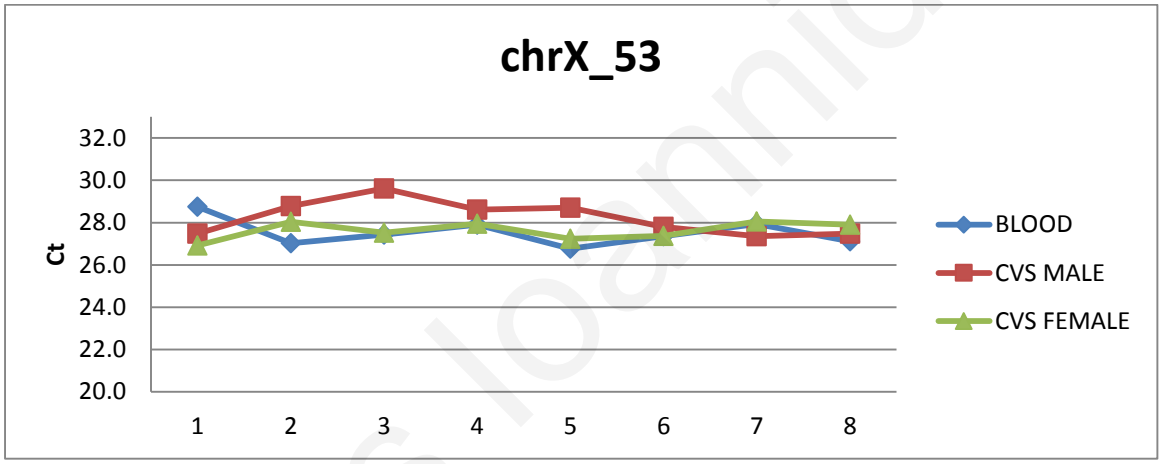


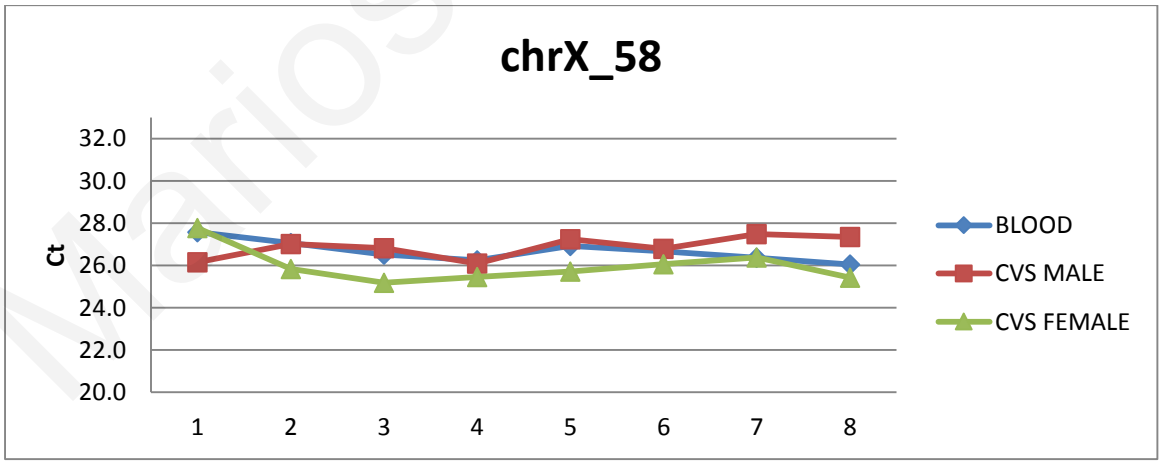


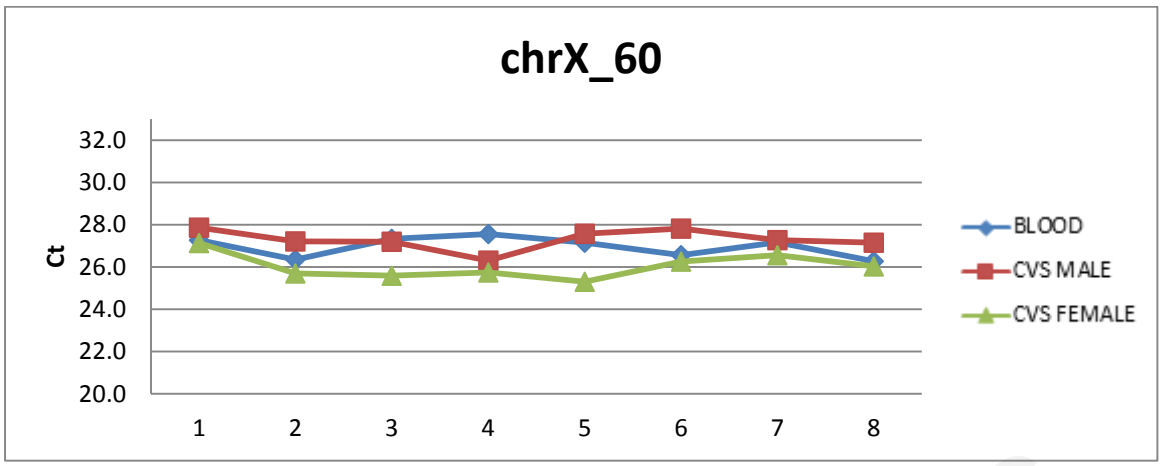
D. Chromosome X

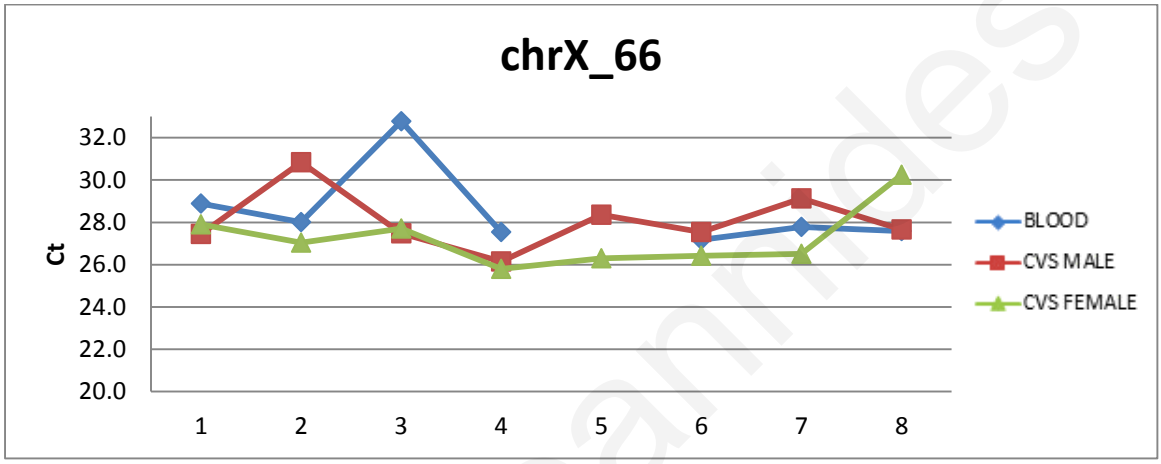
1. "Not DMRs"

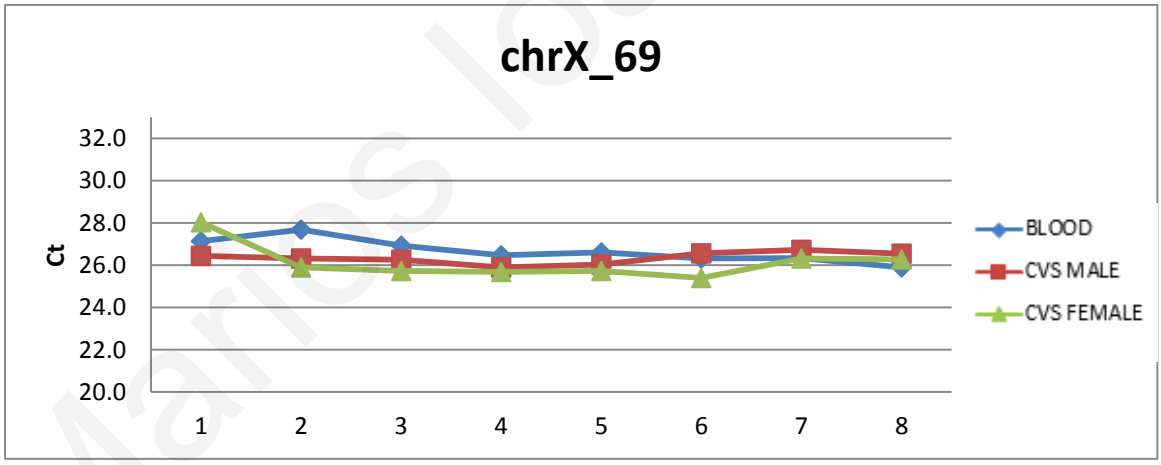


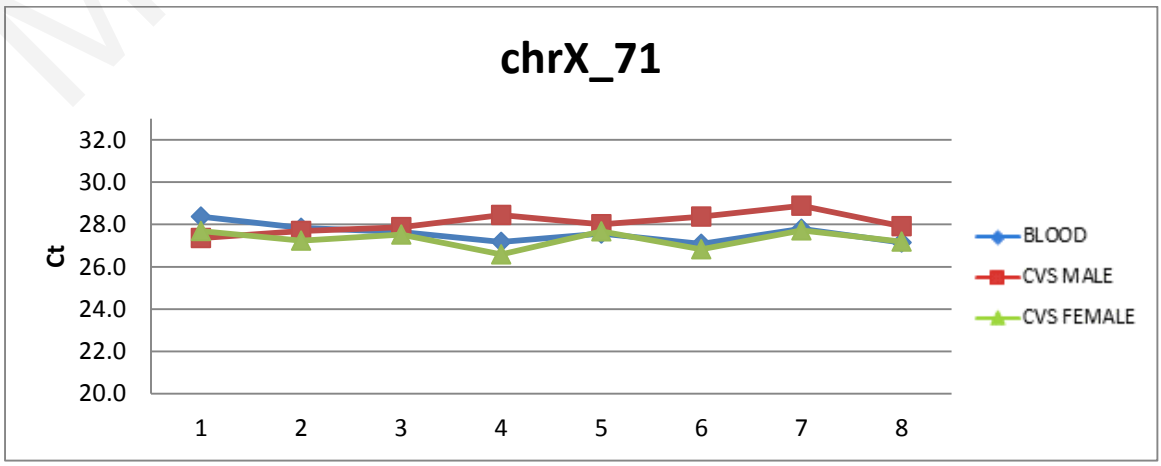


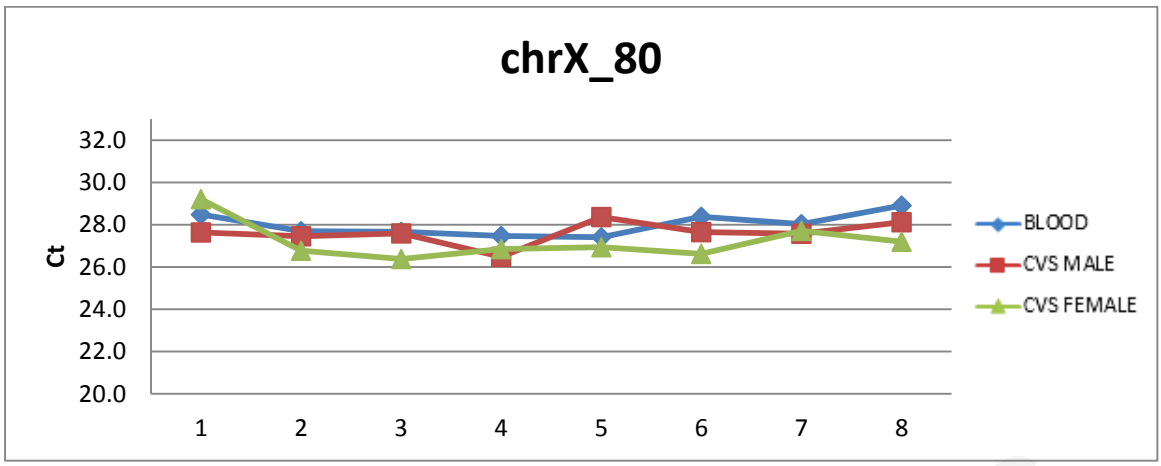


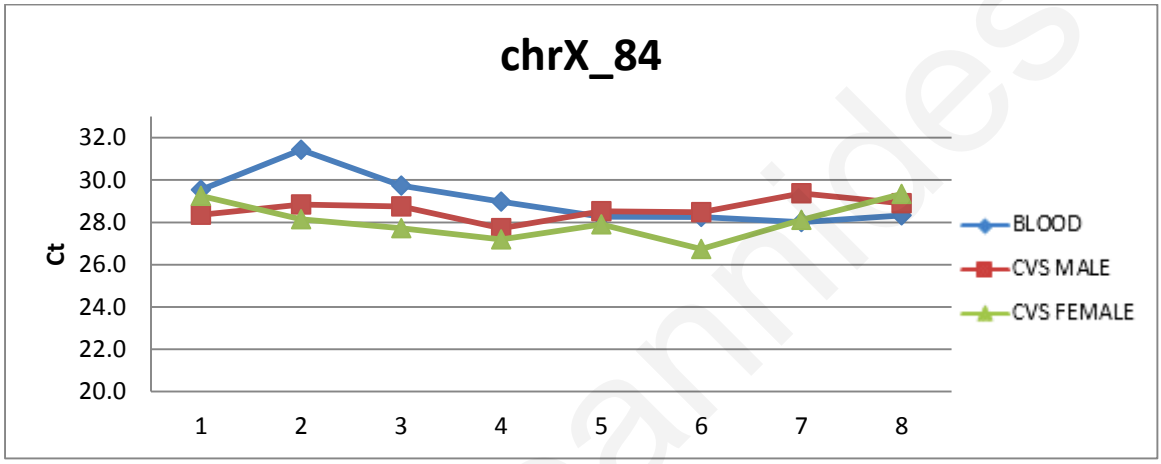


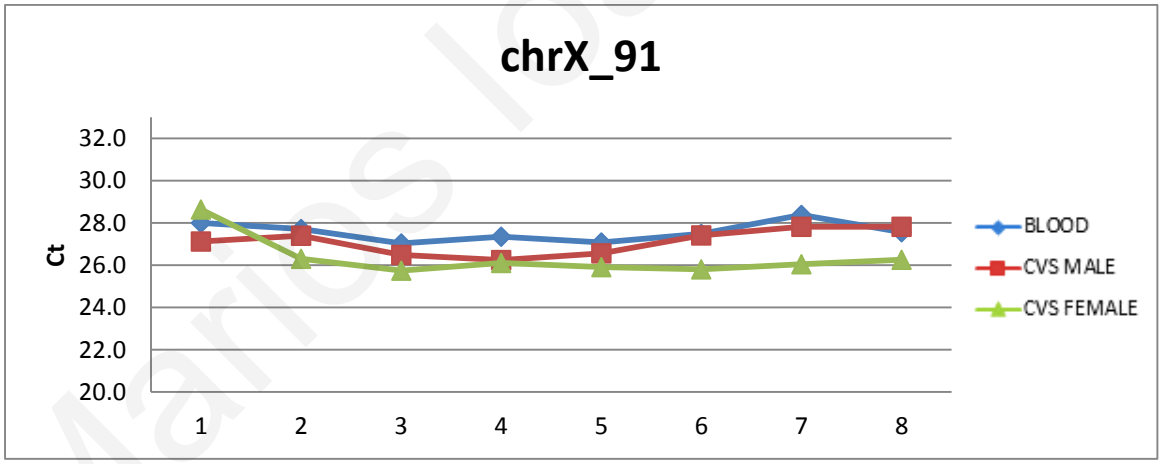


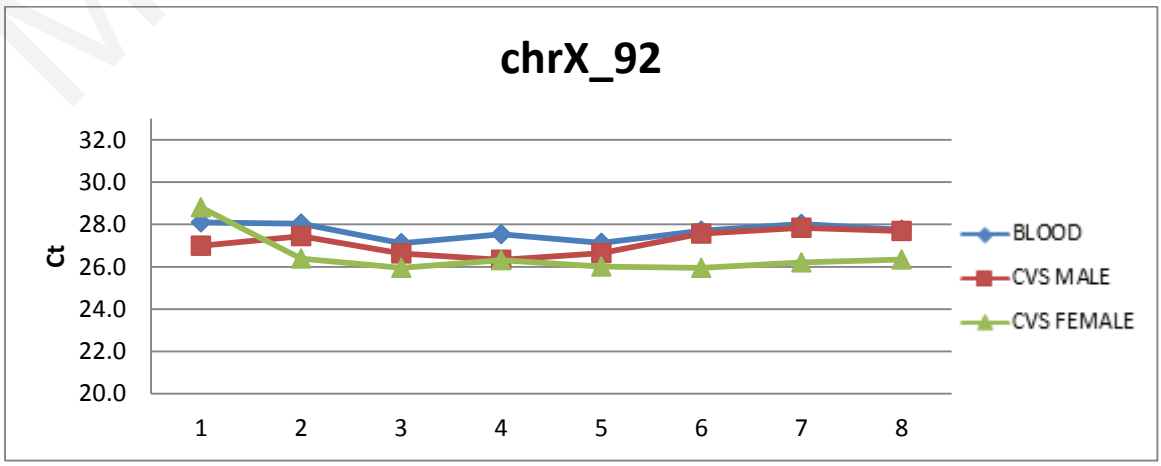


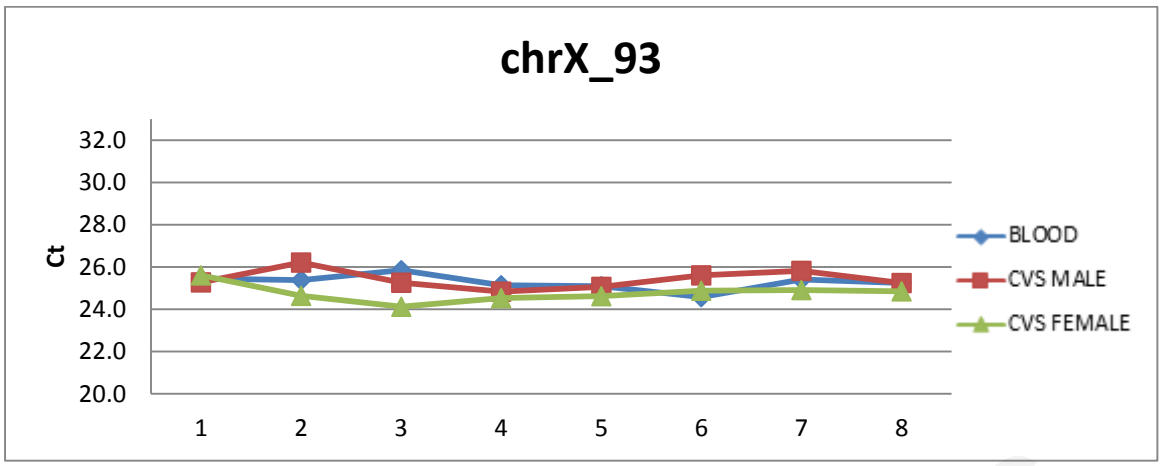


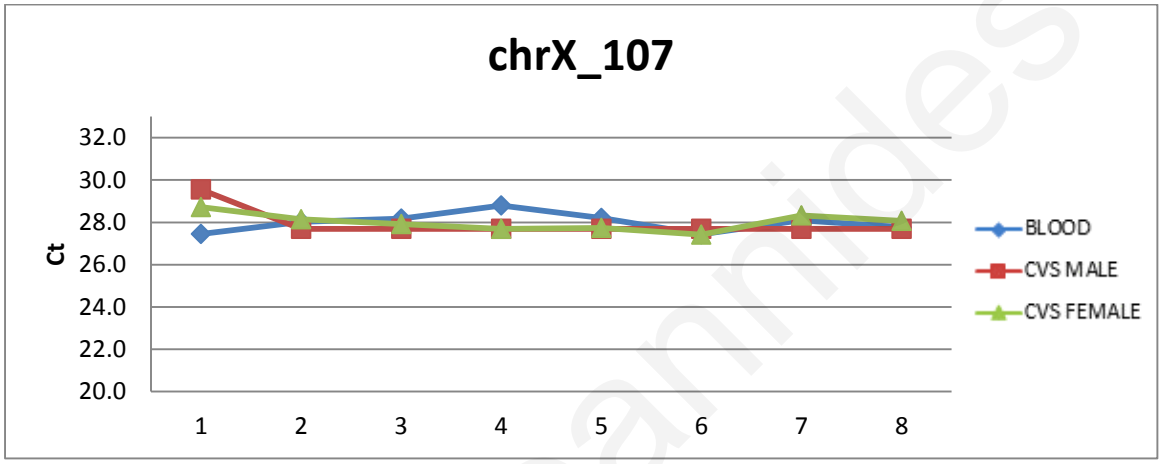


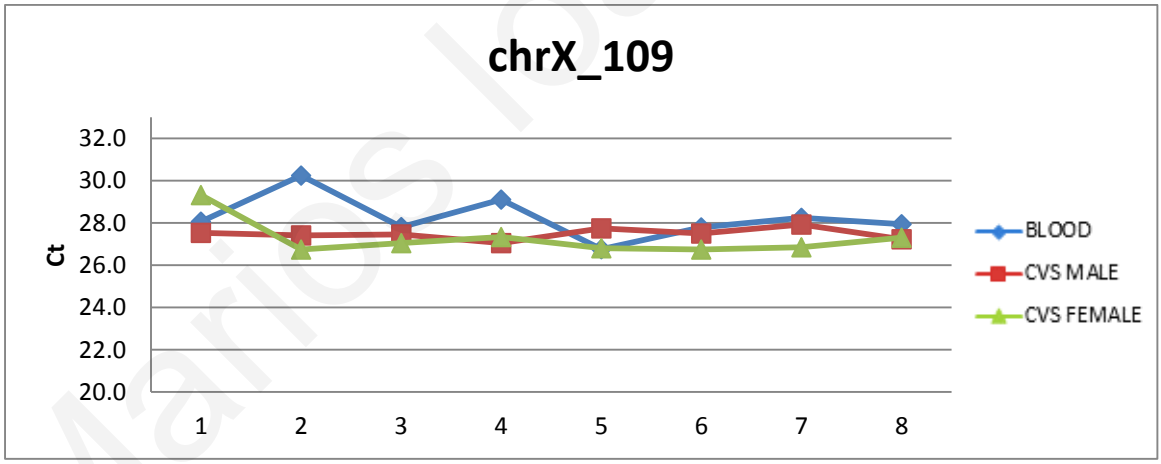


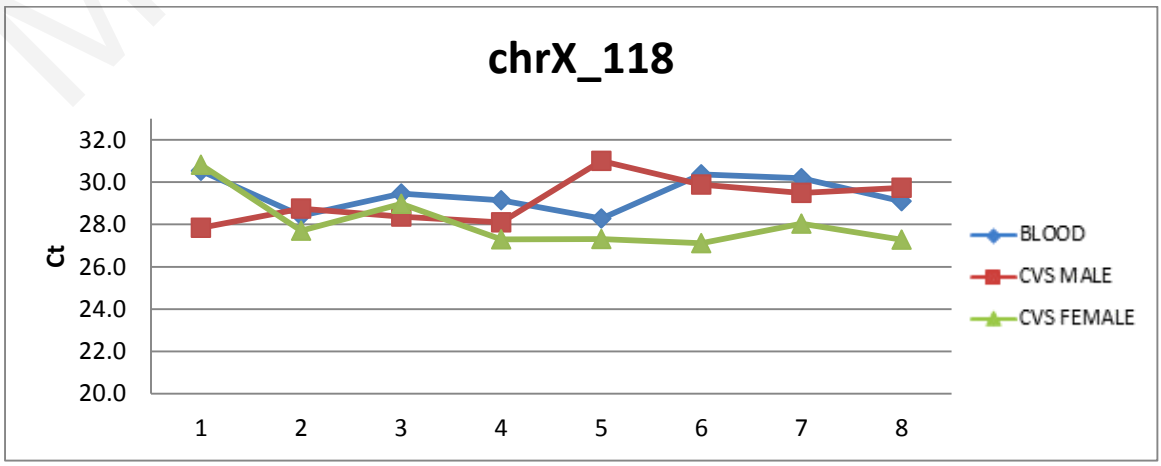


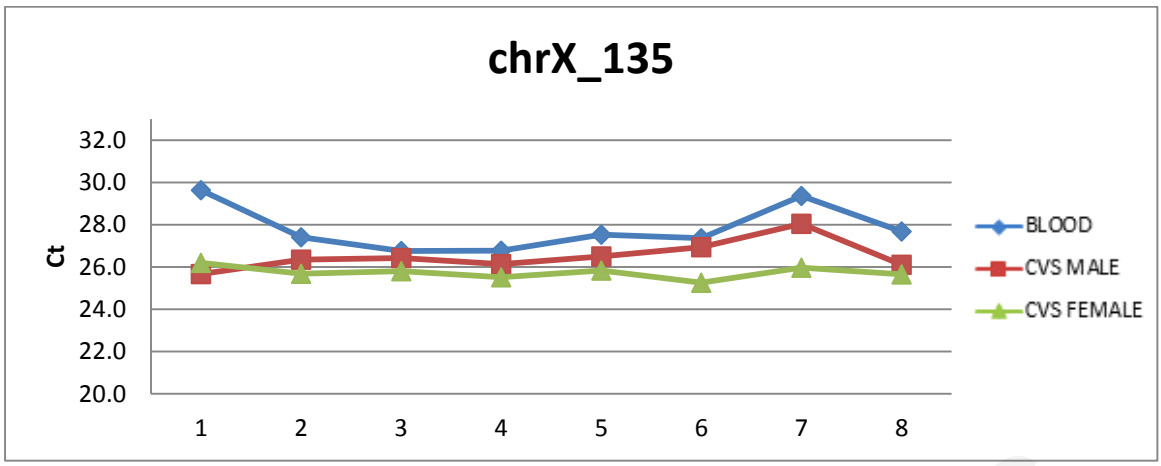


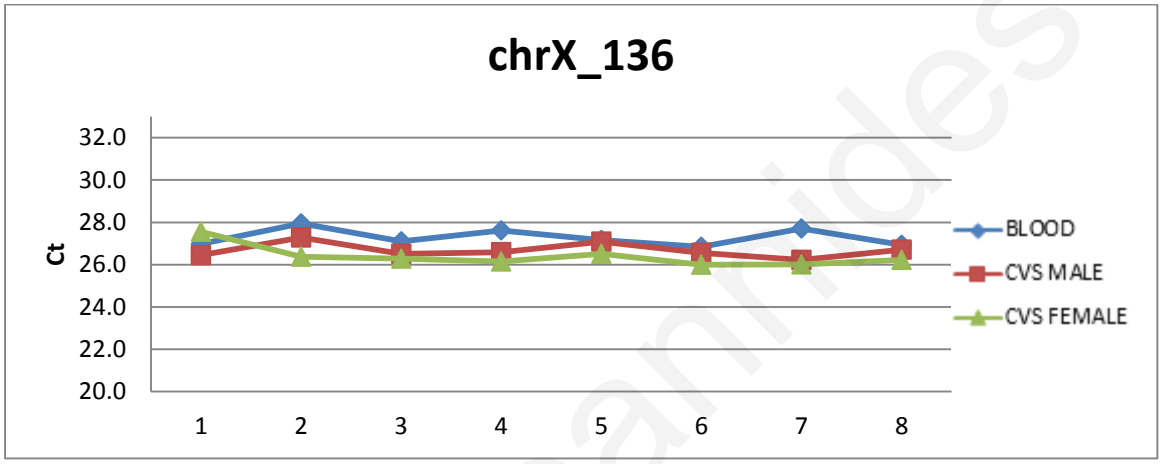


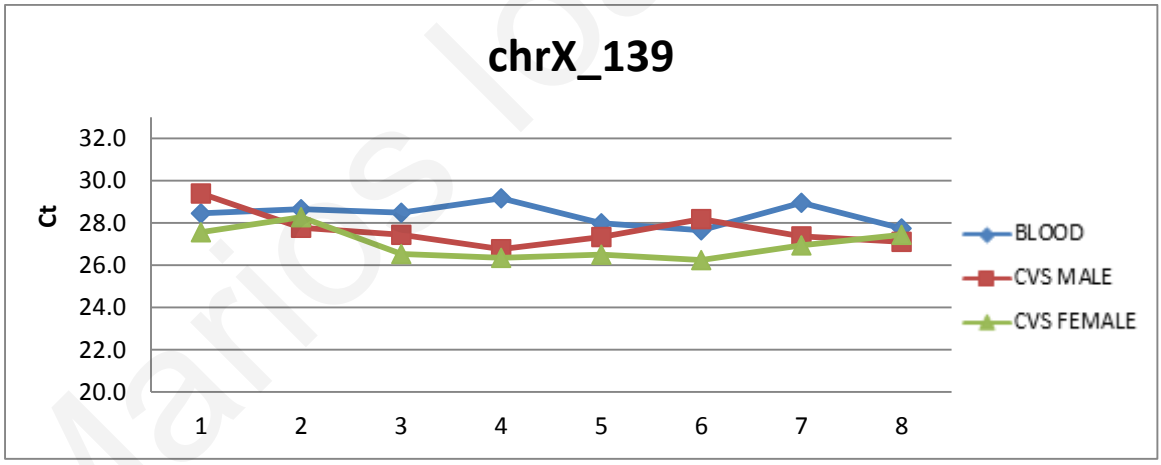


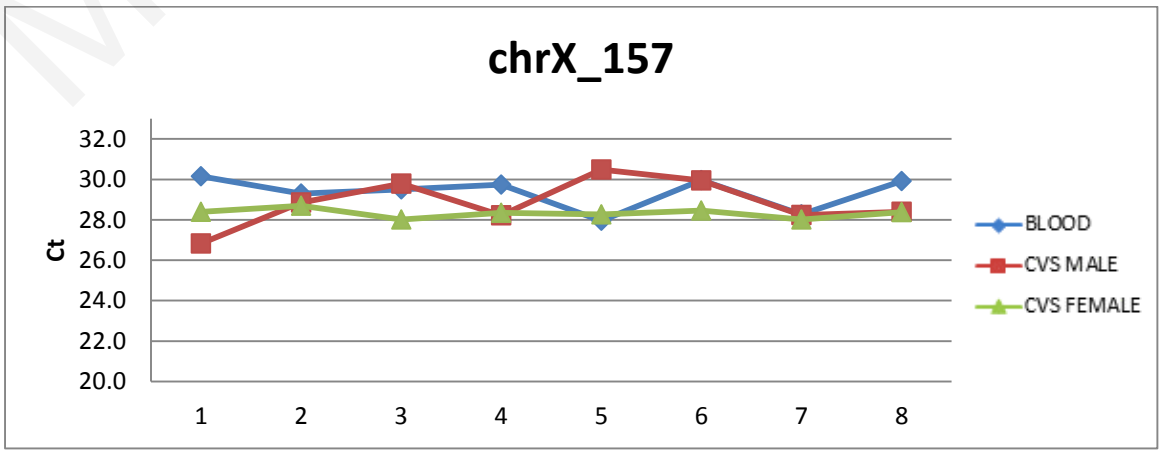


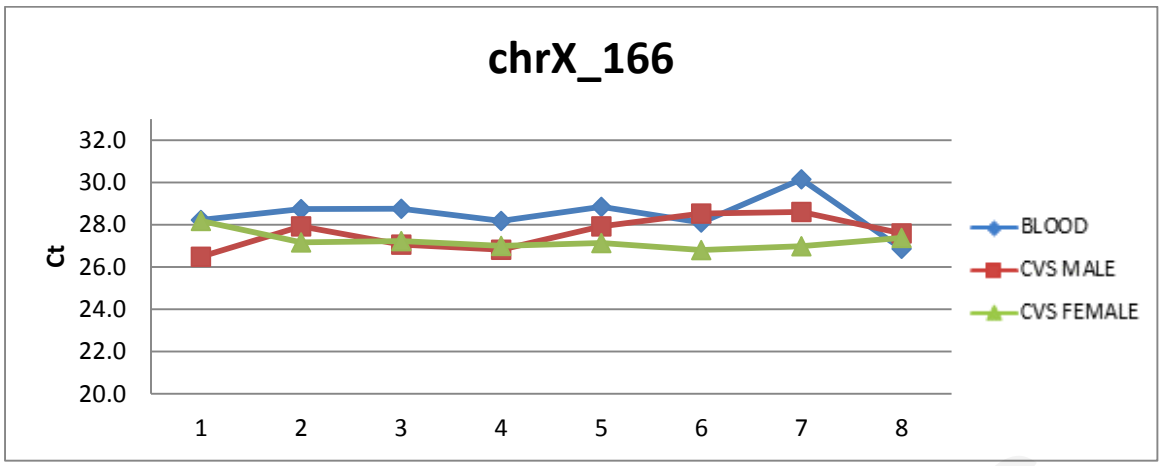


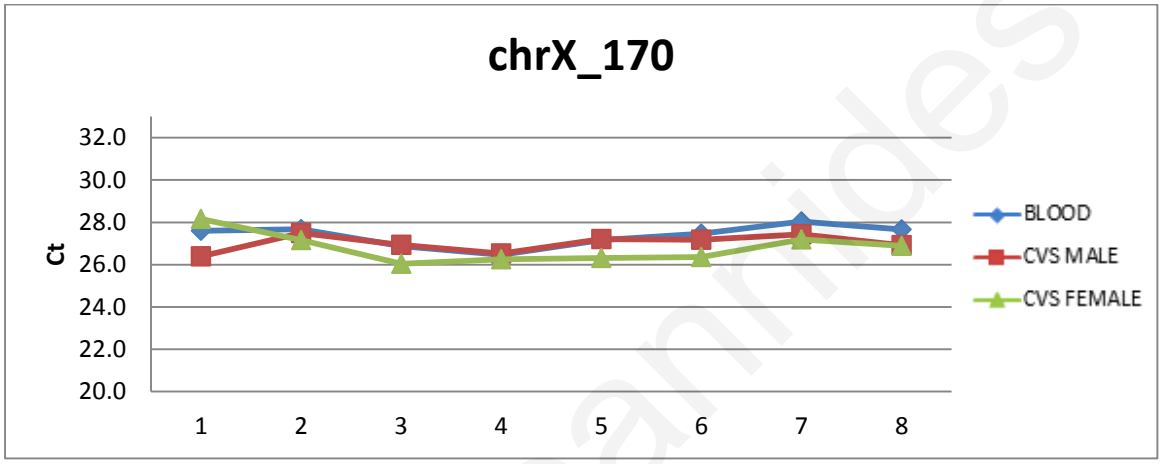


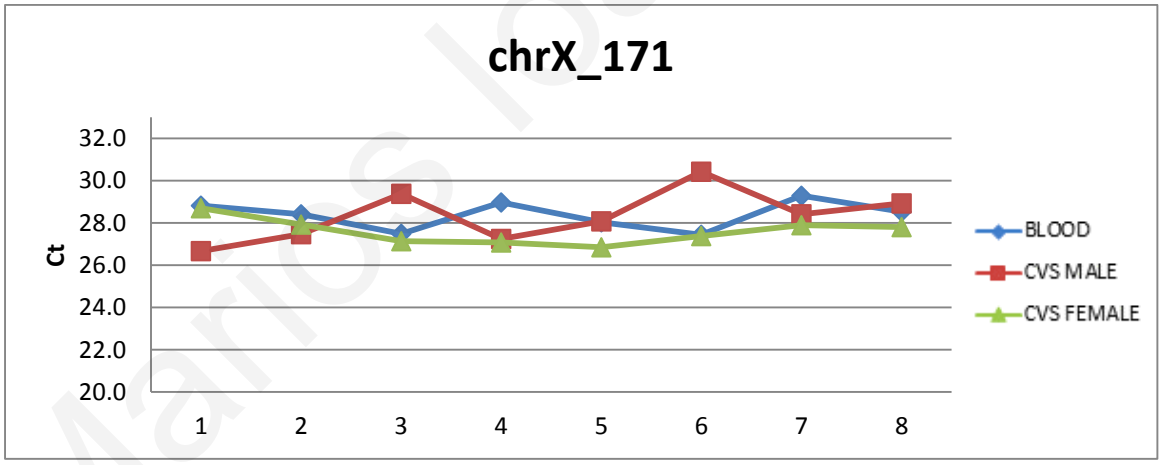


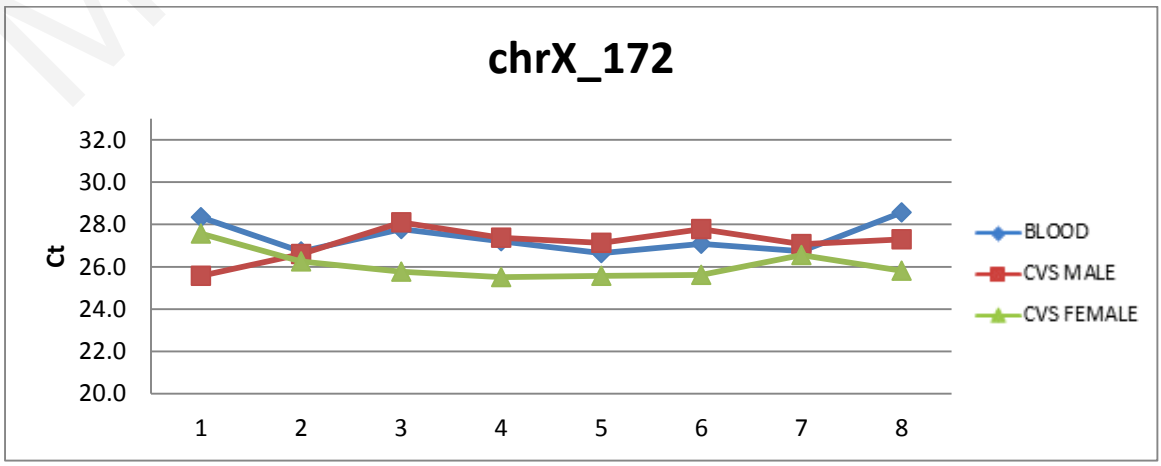


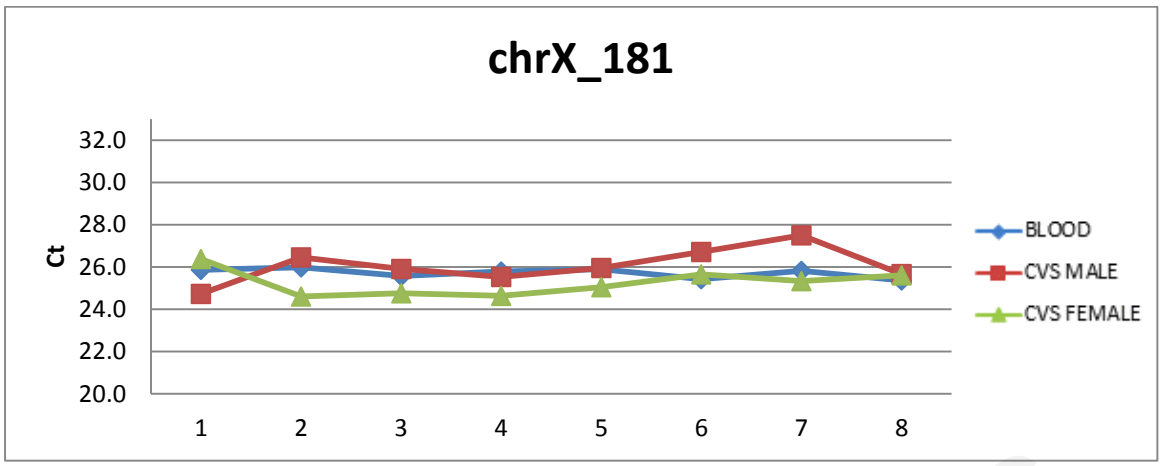


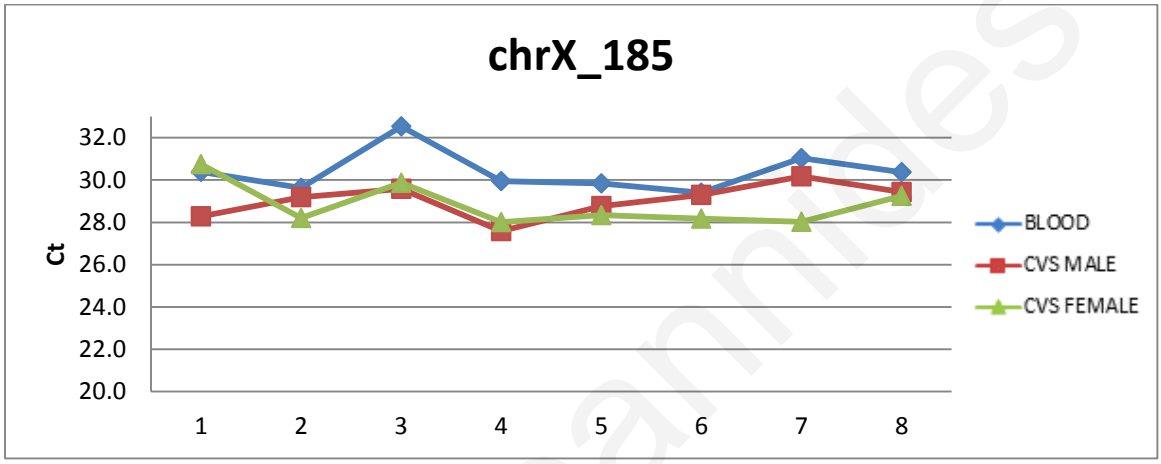


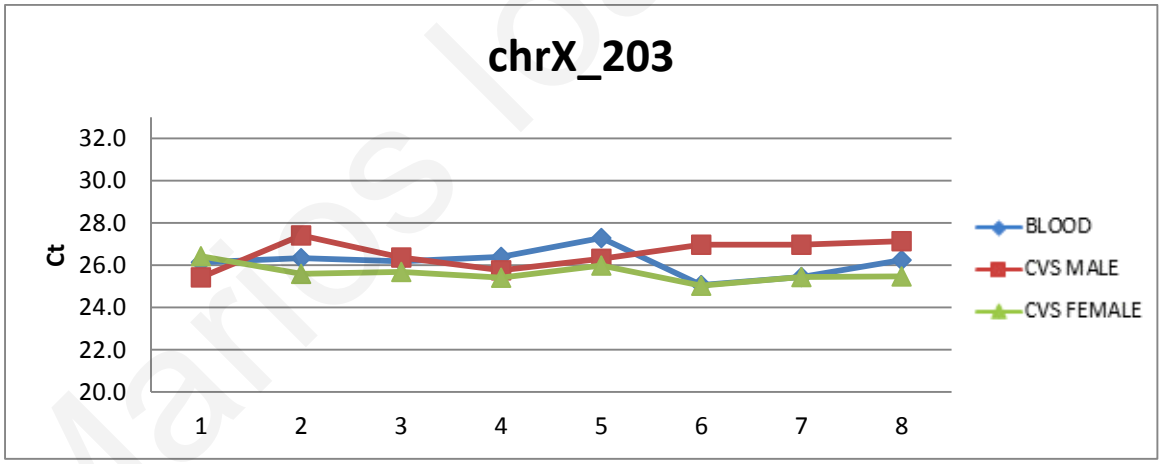


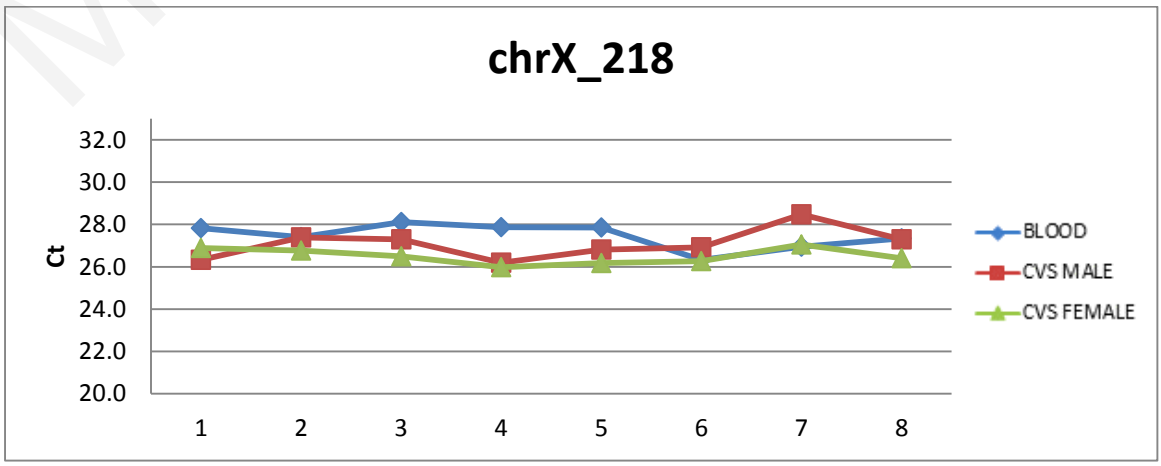


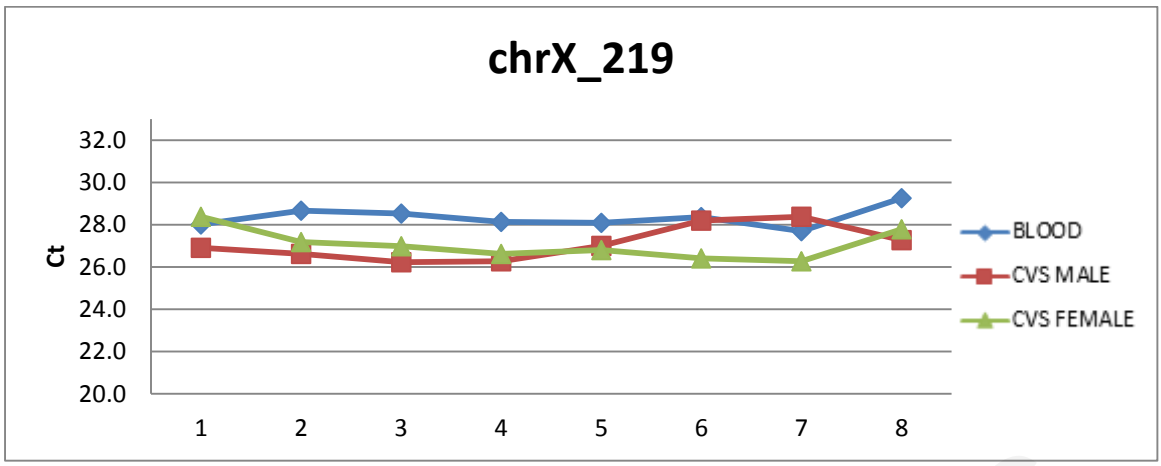


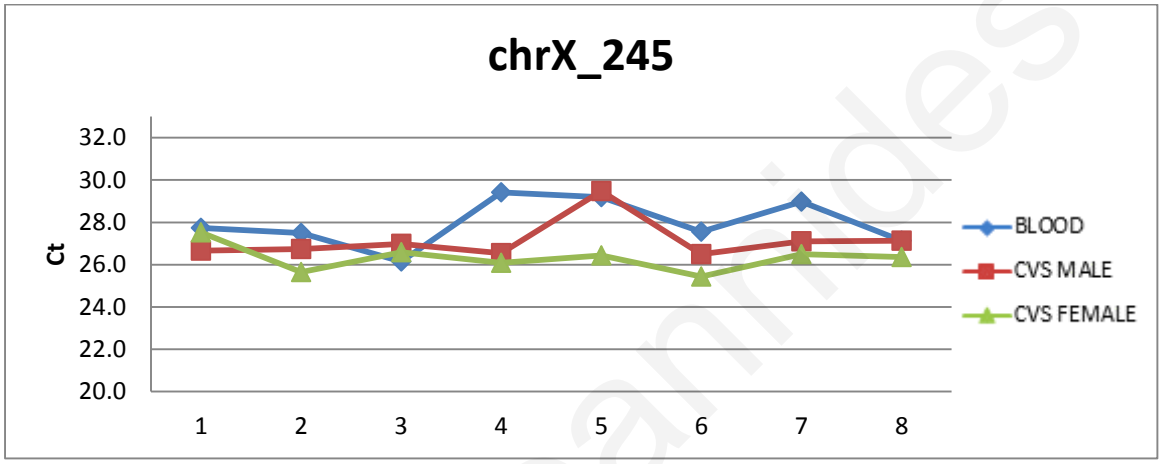


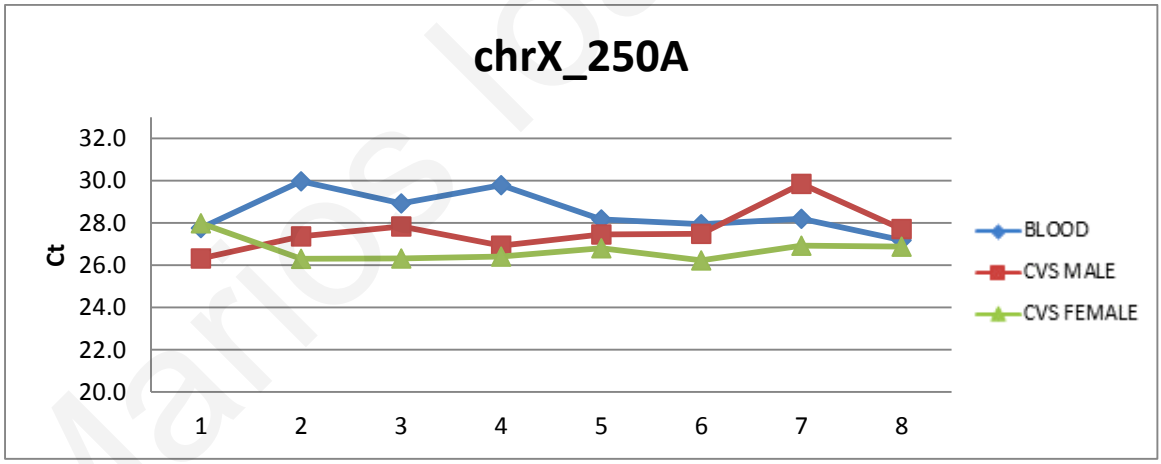


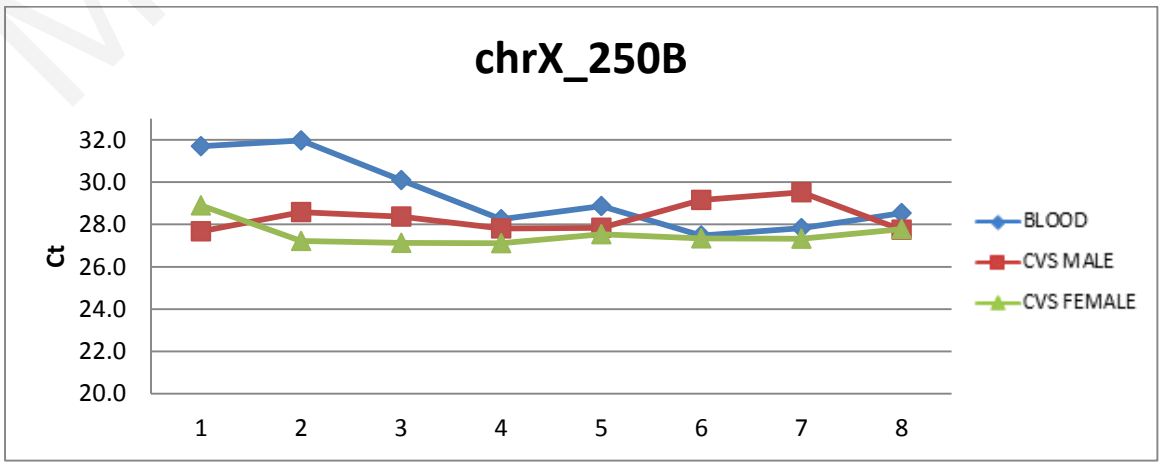


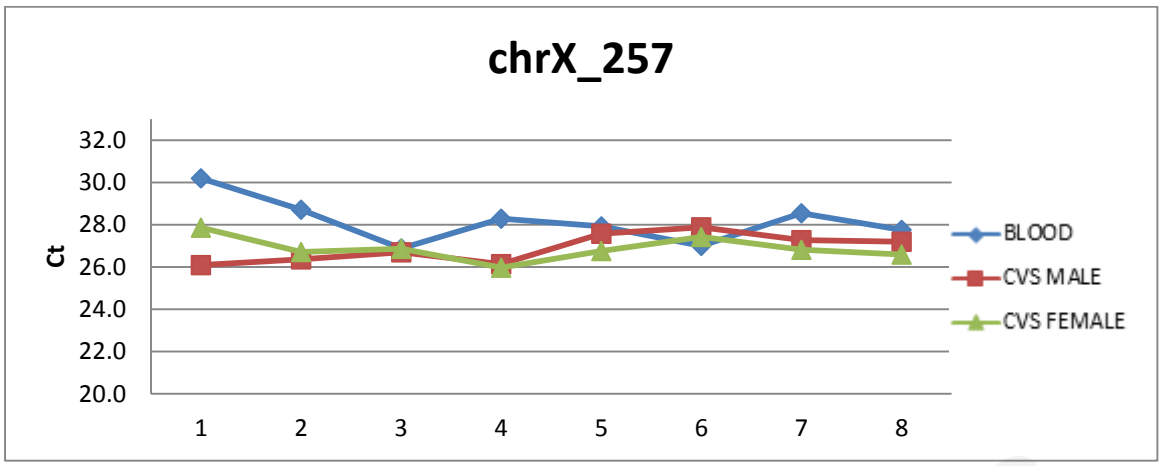


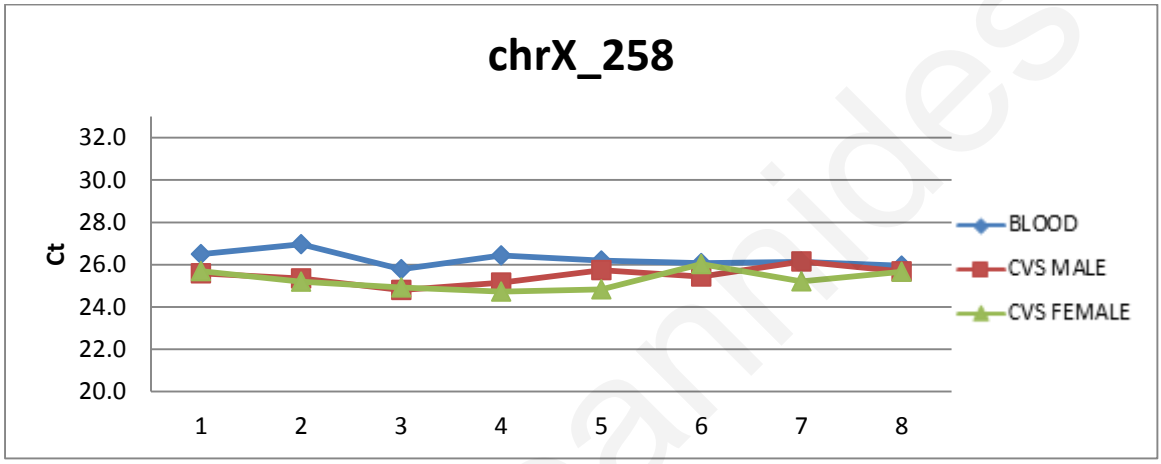


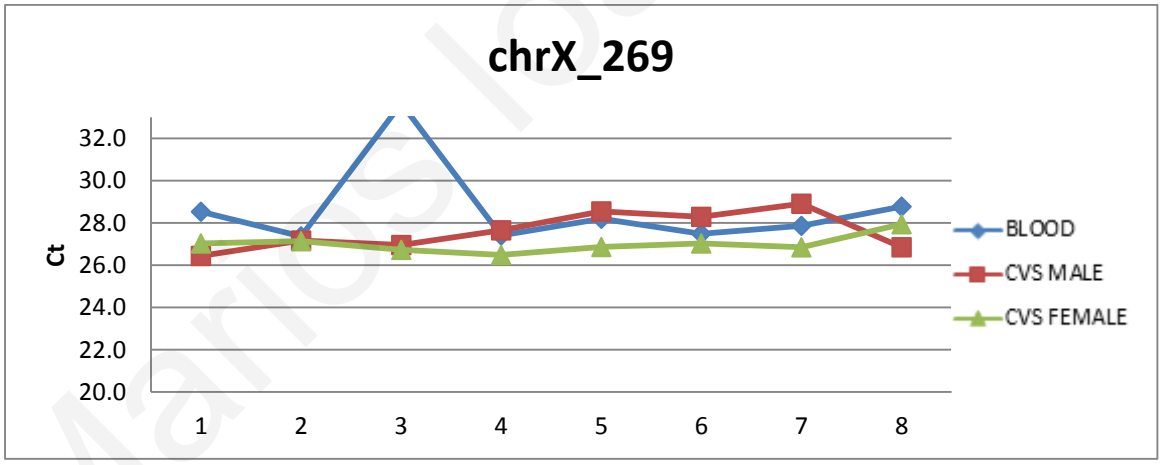


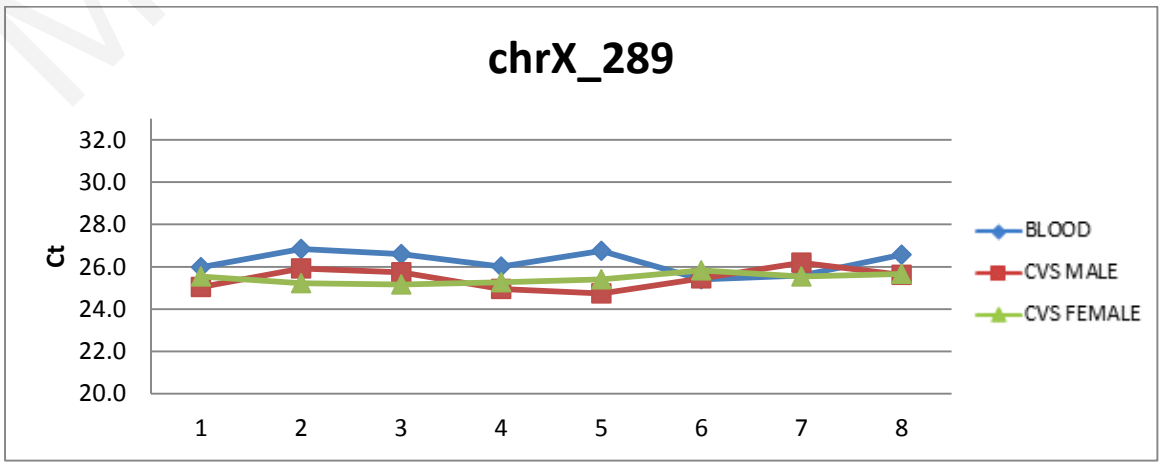


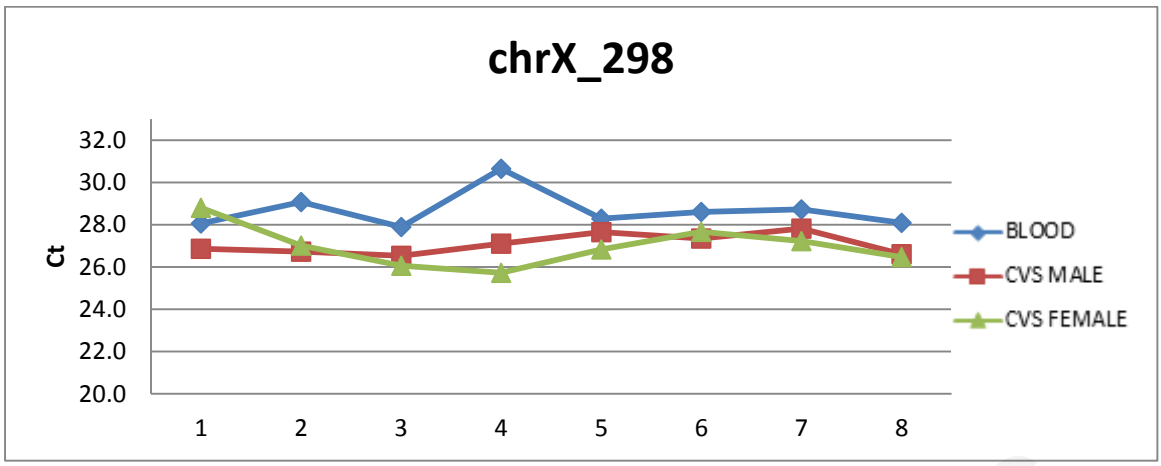


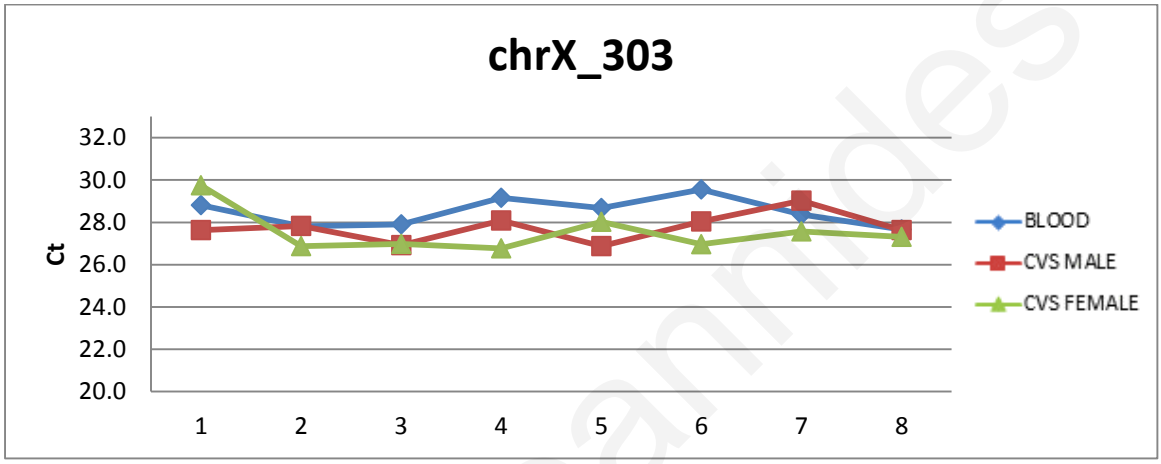


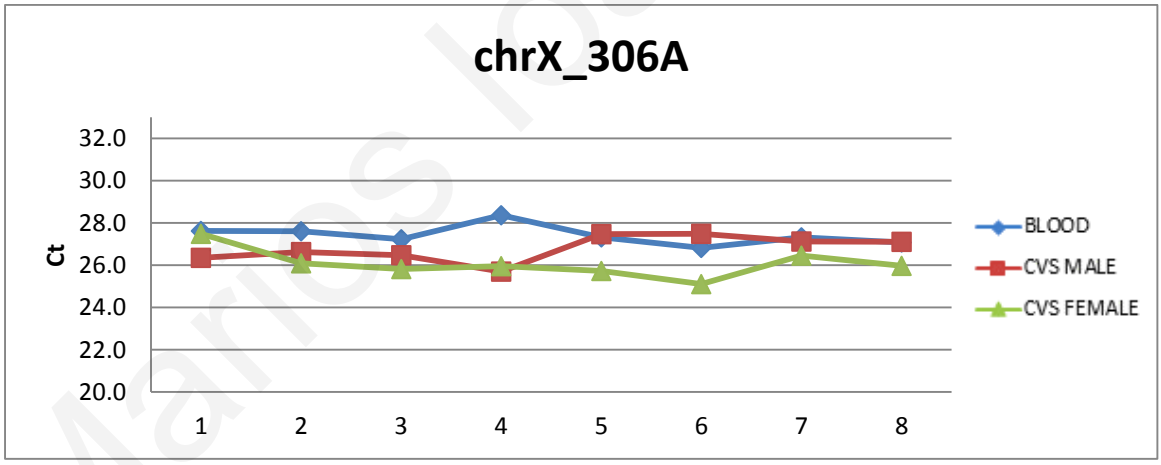


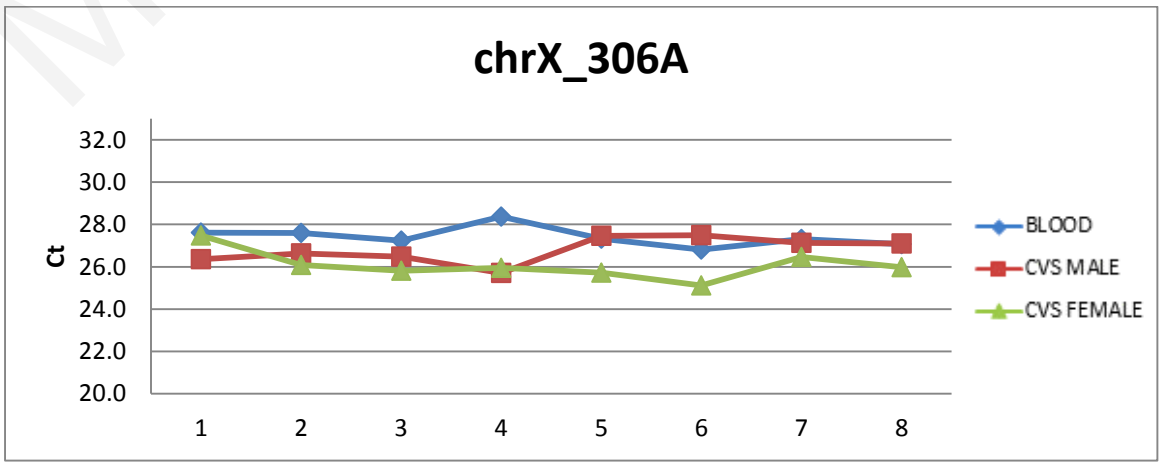


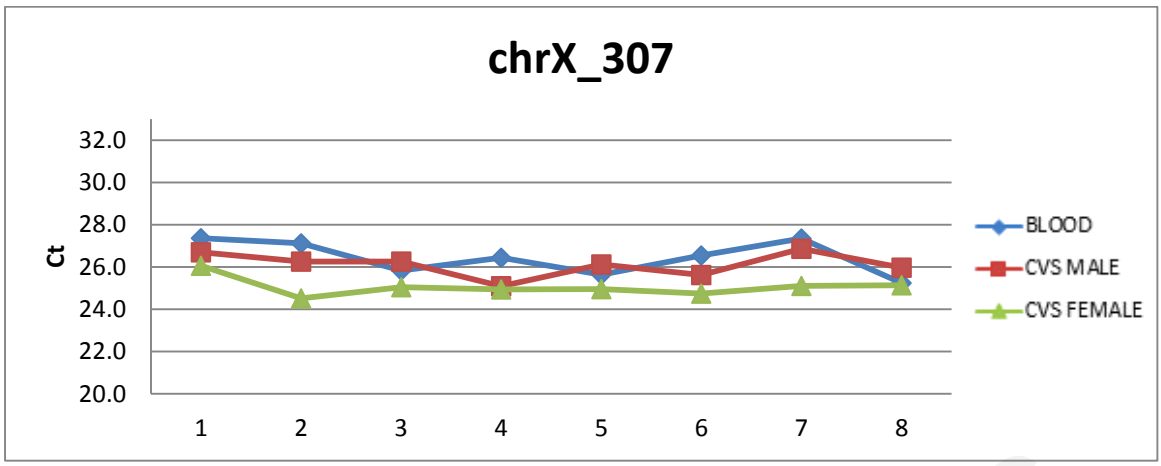


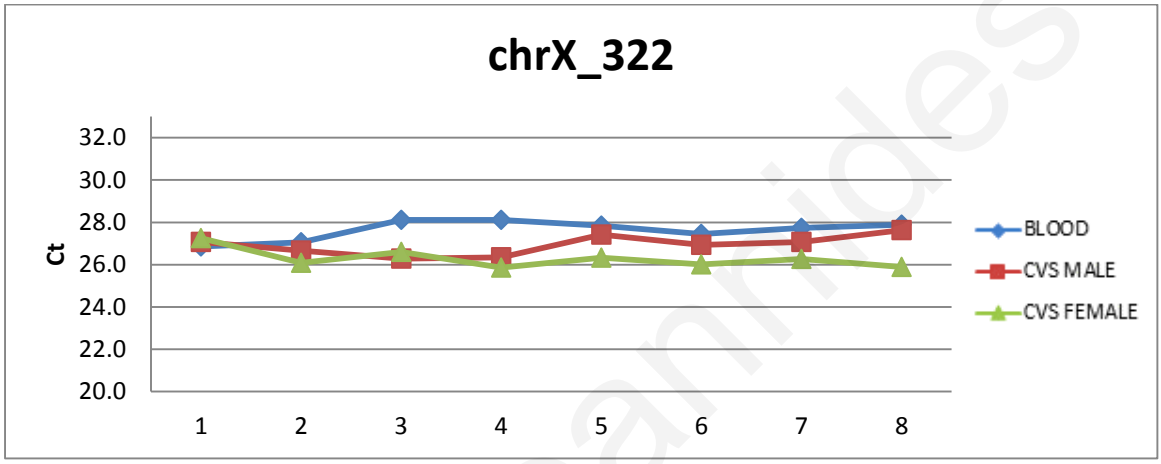


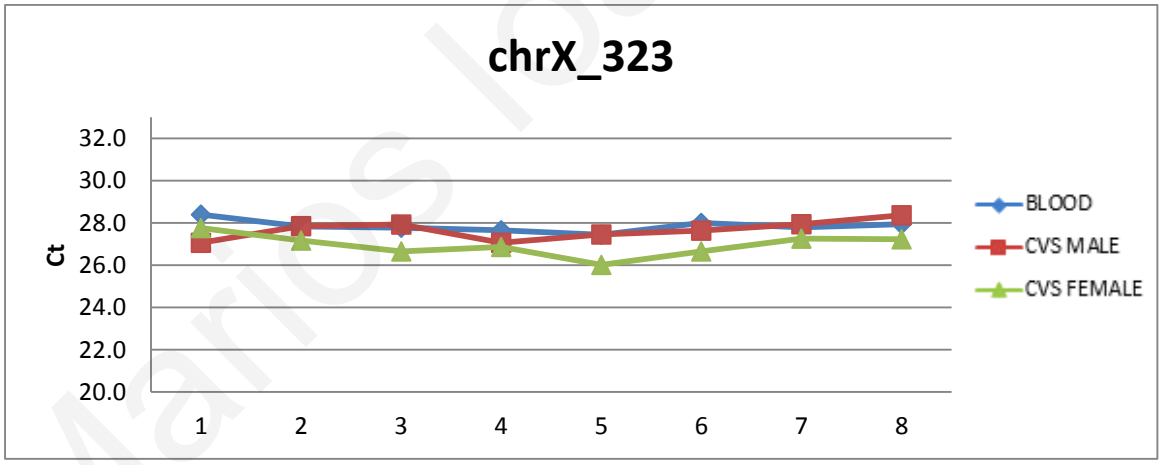


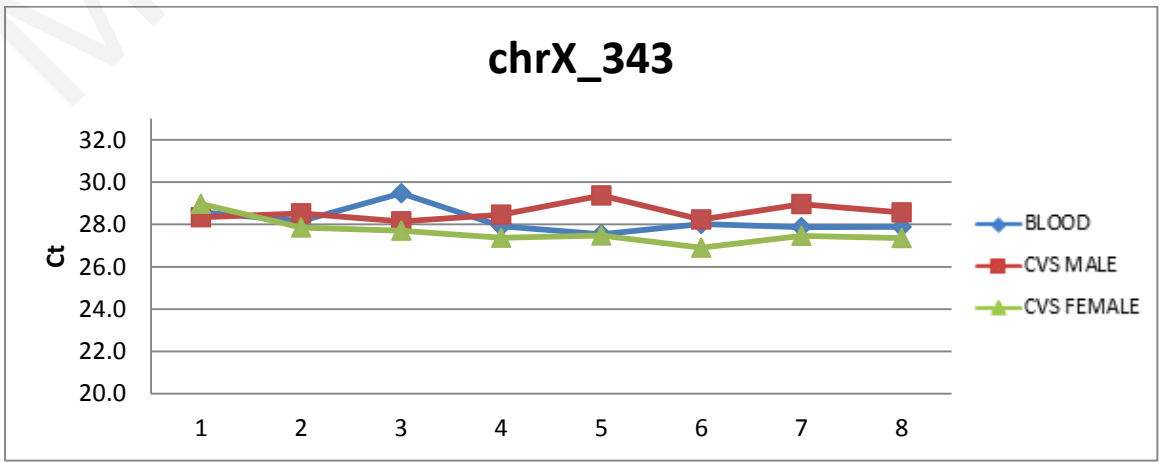


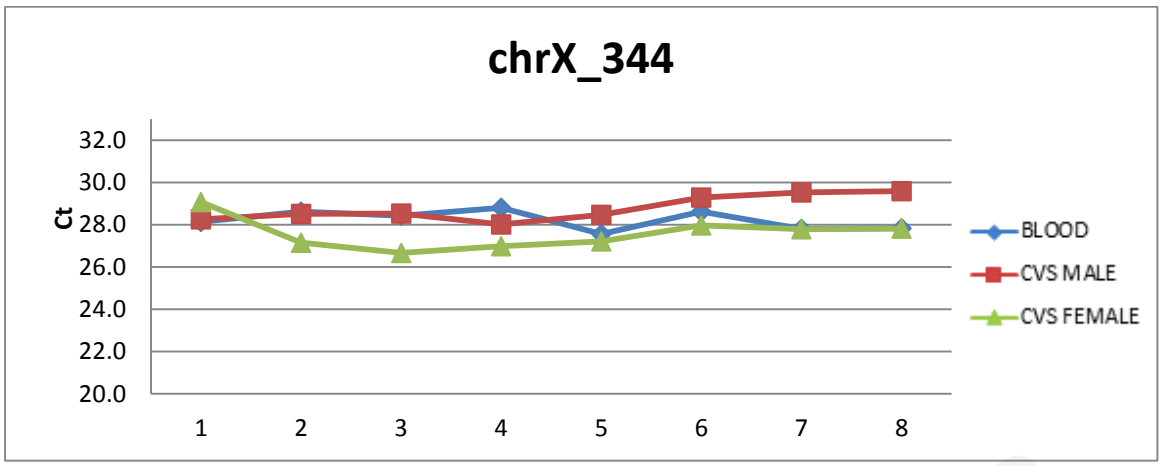


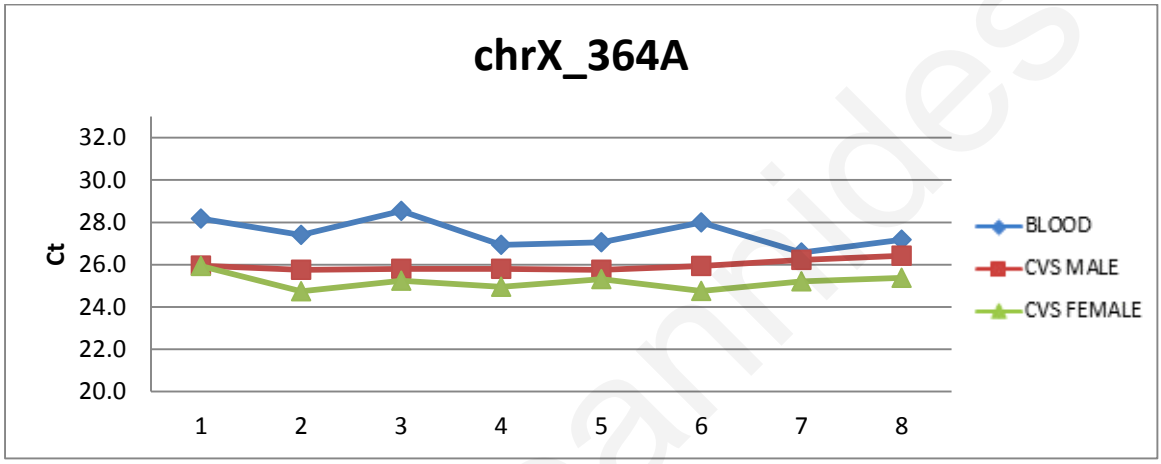


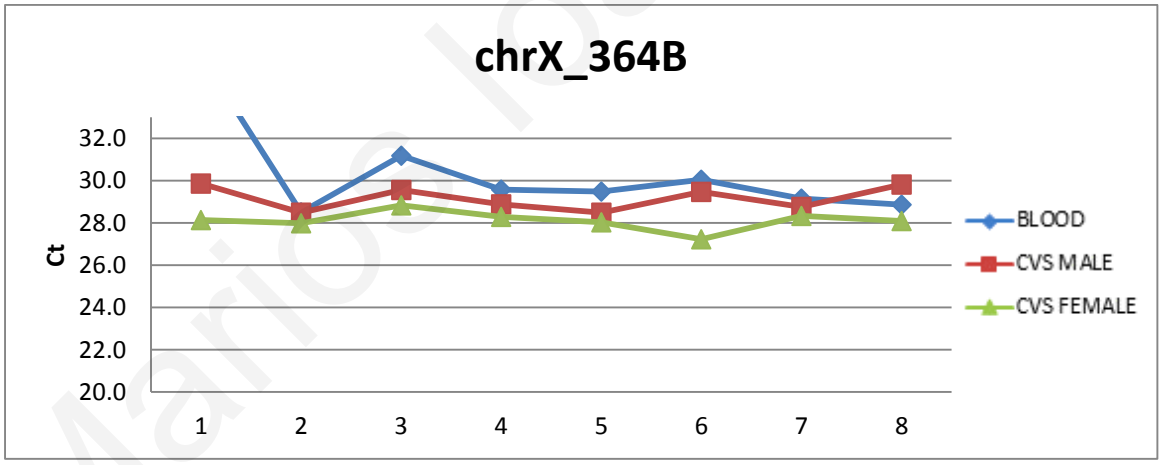


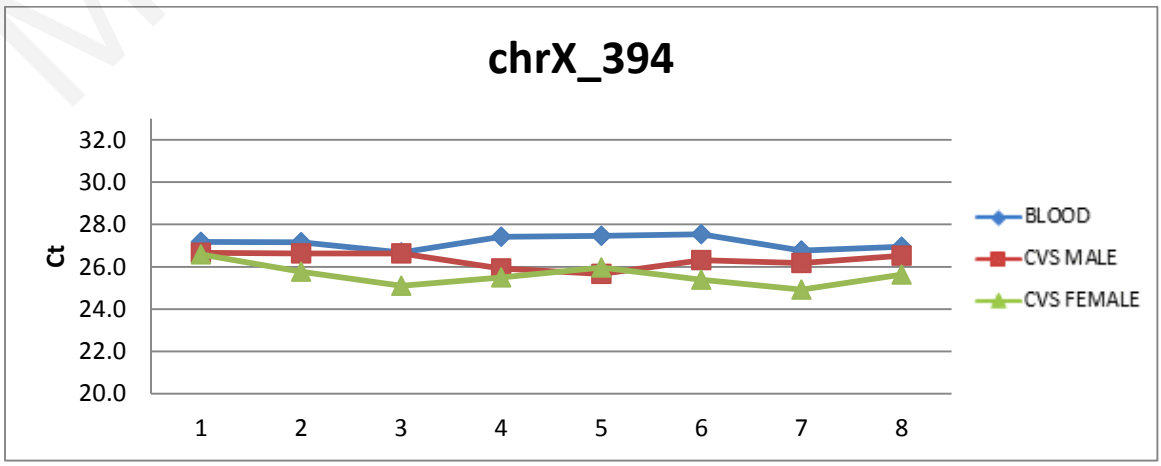


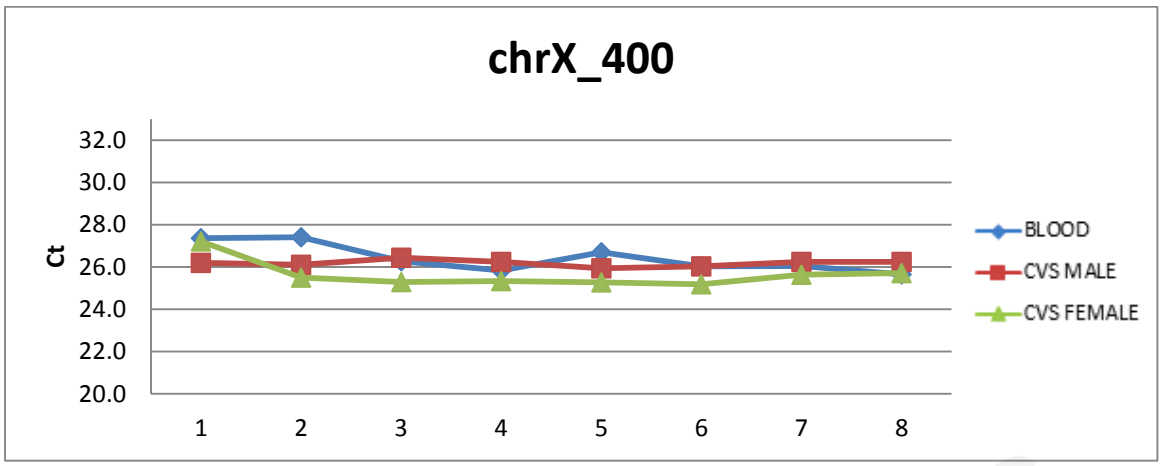


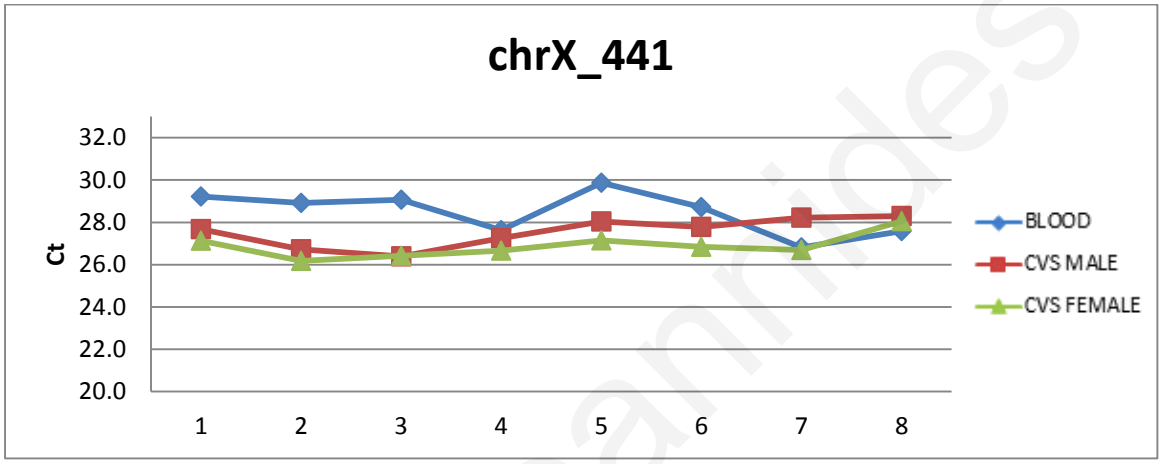


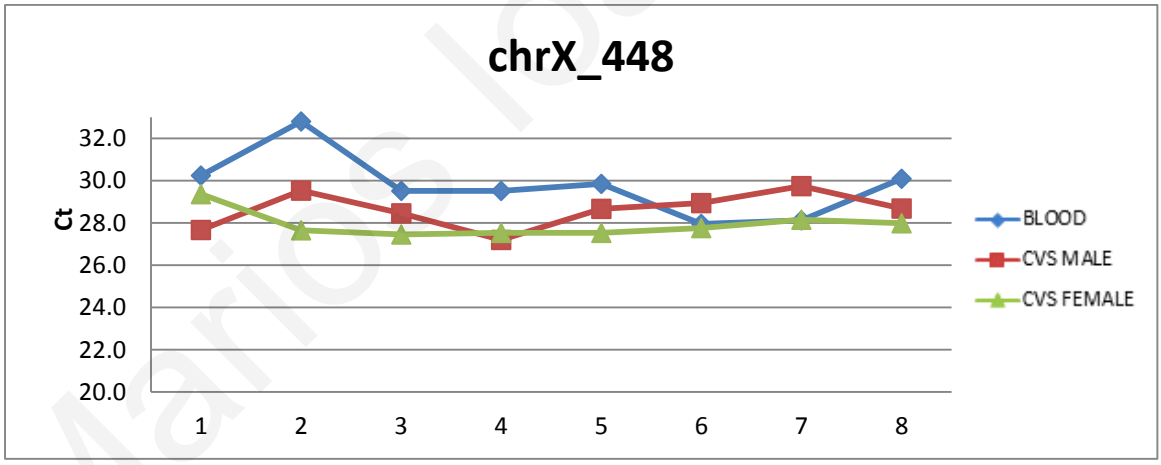


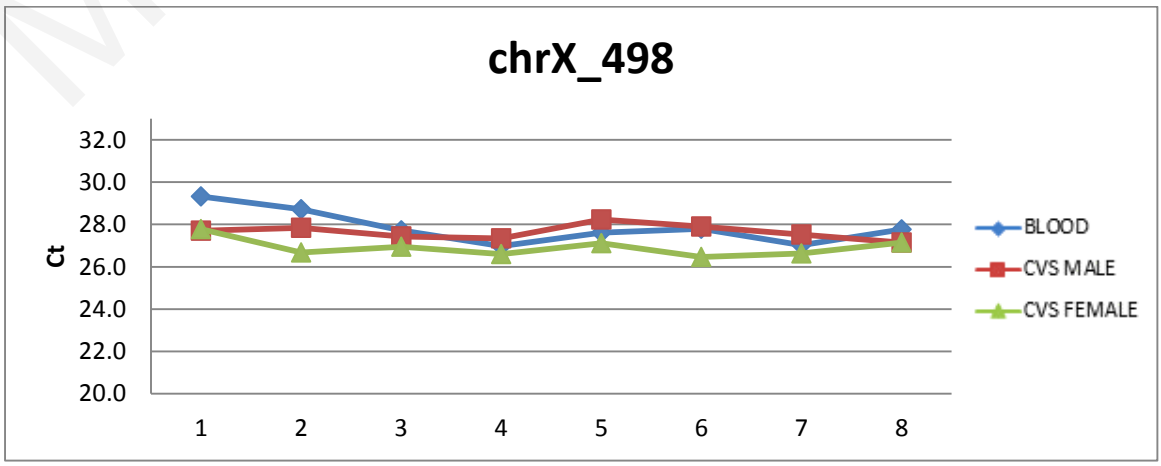


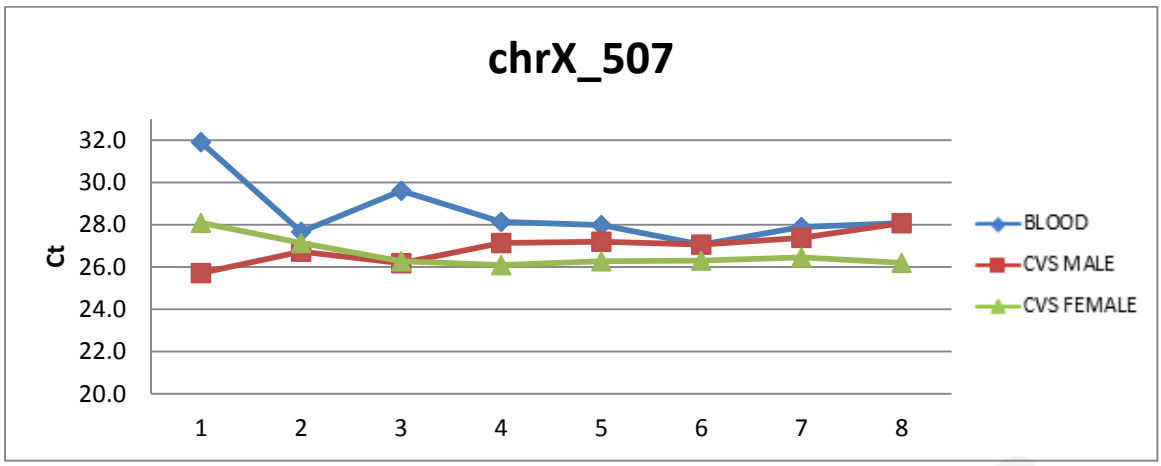


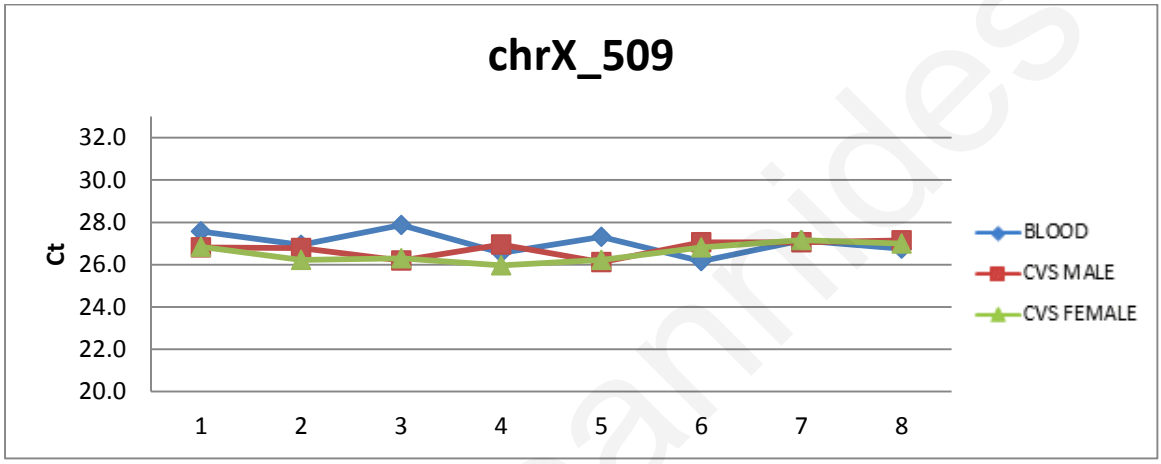


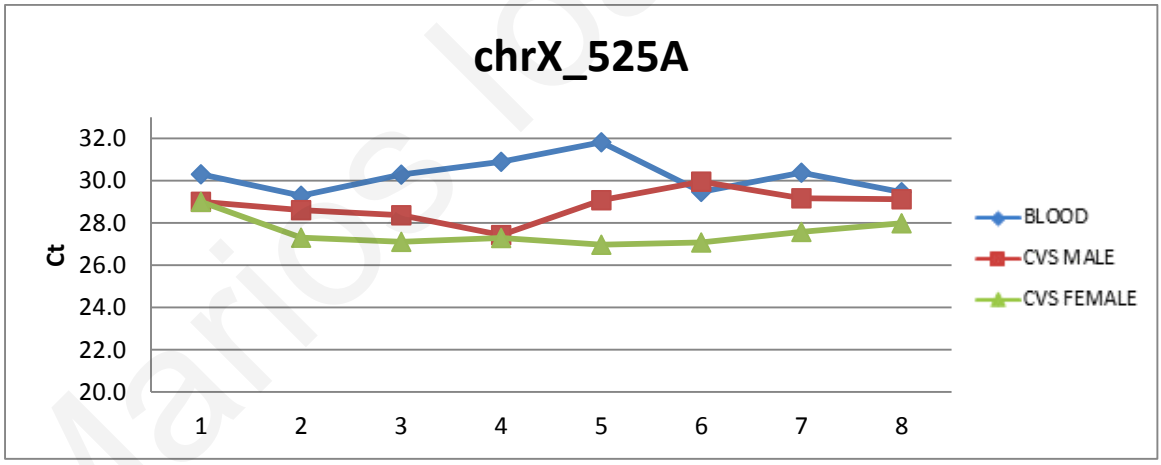


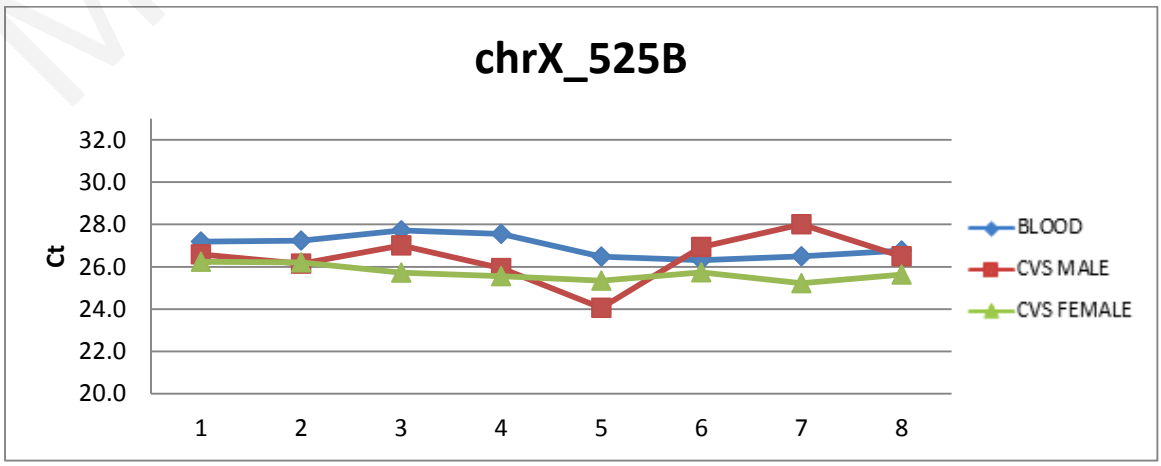


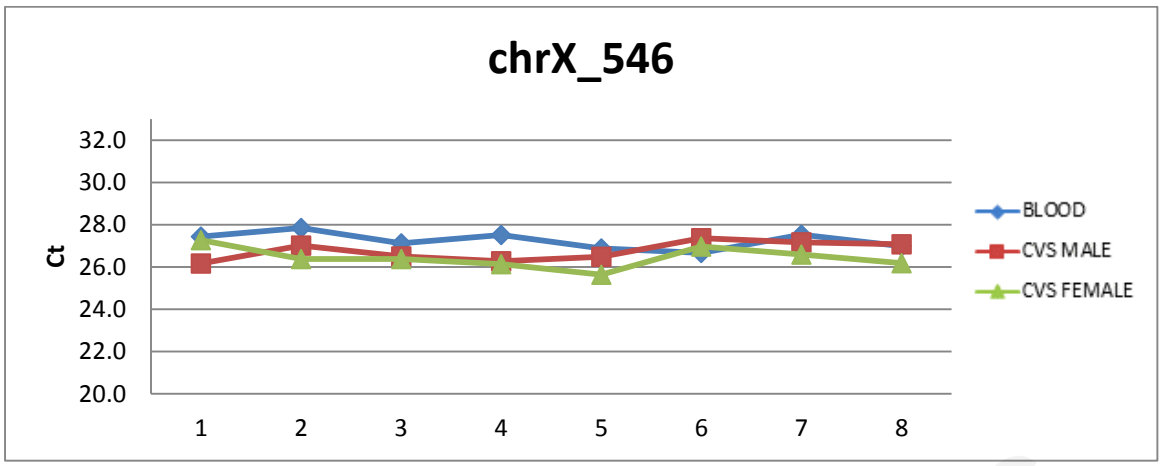


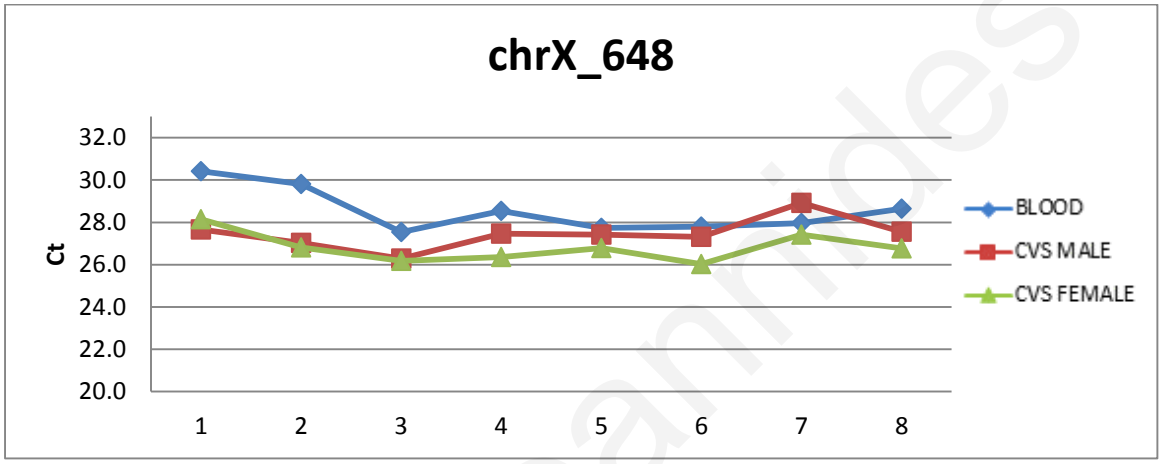


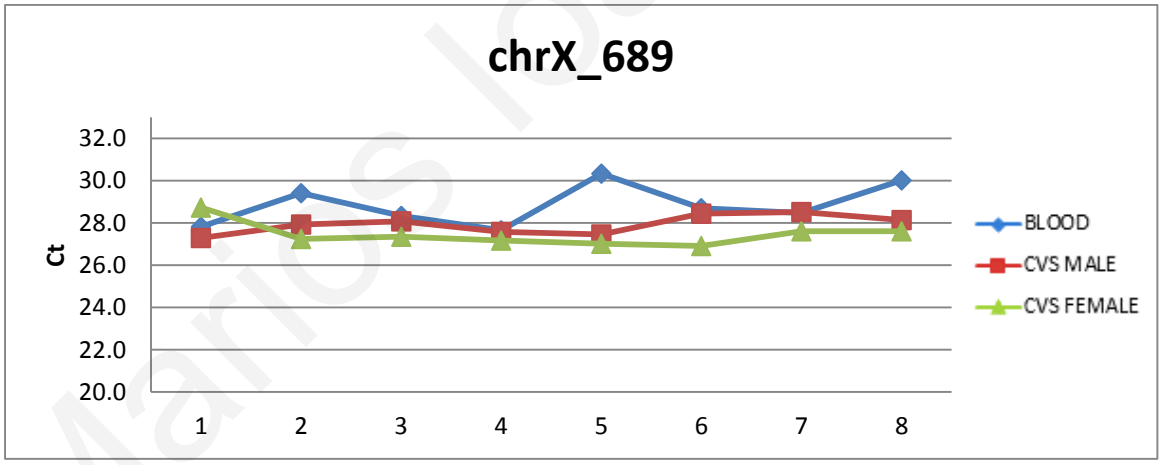


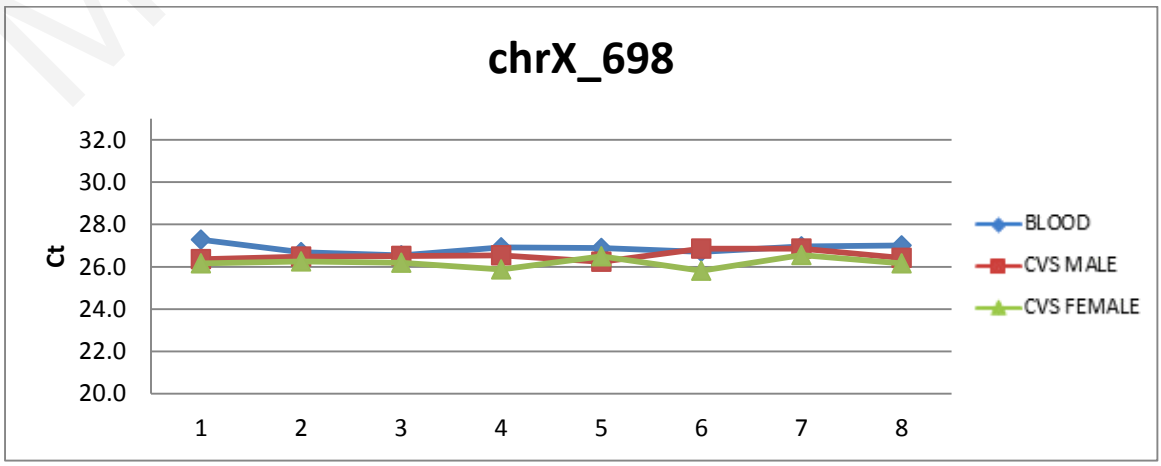


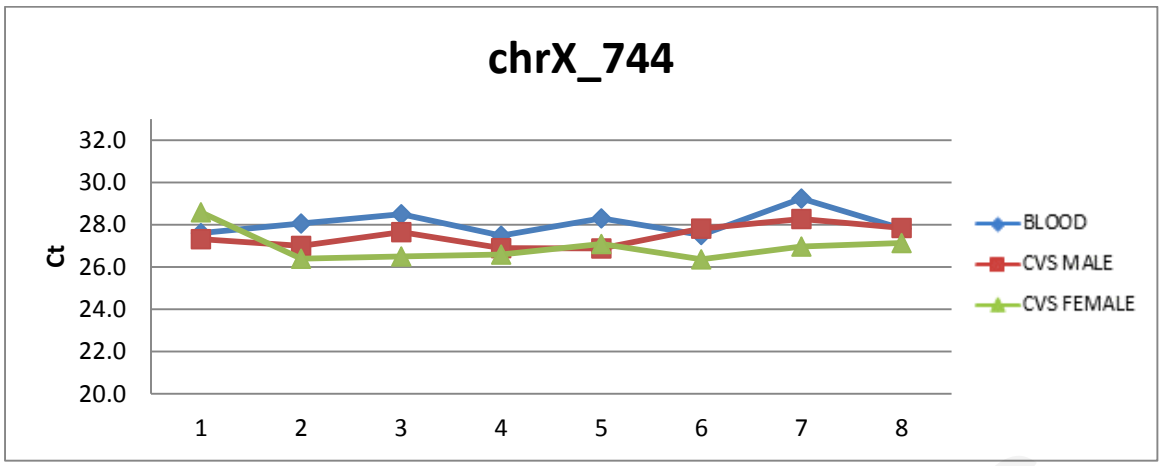


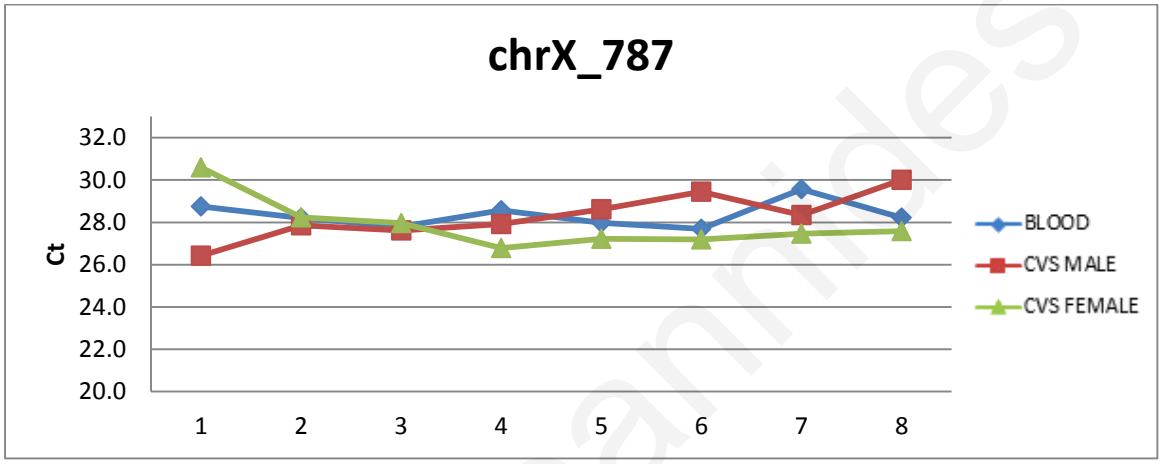


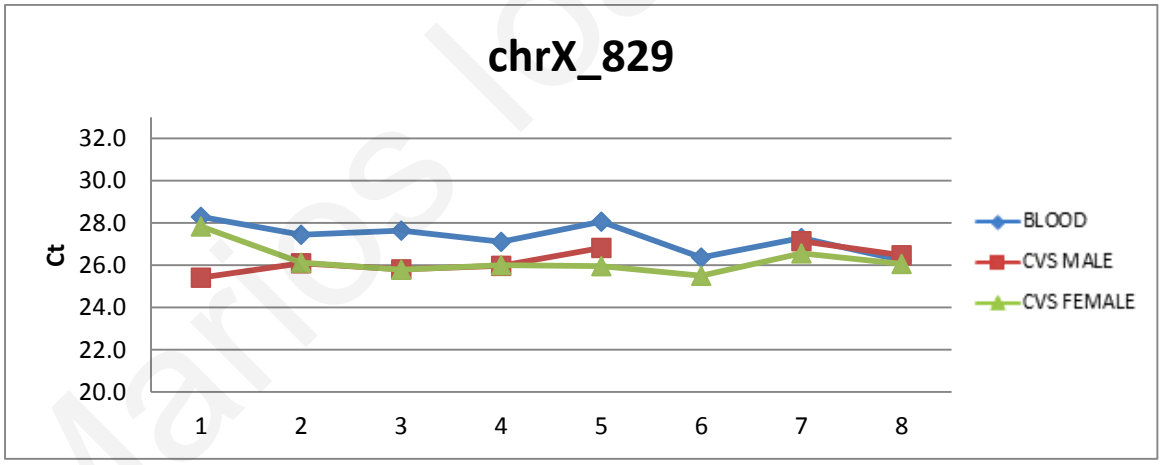


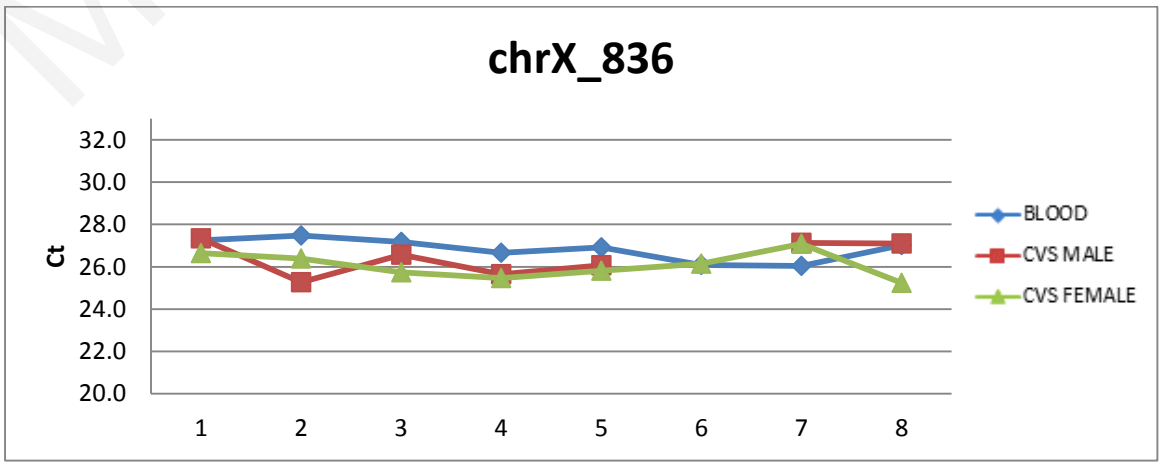


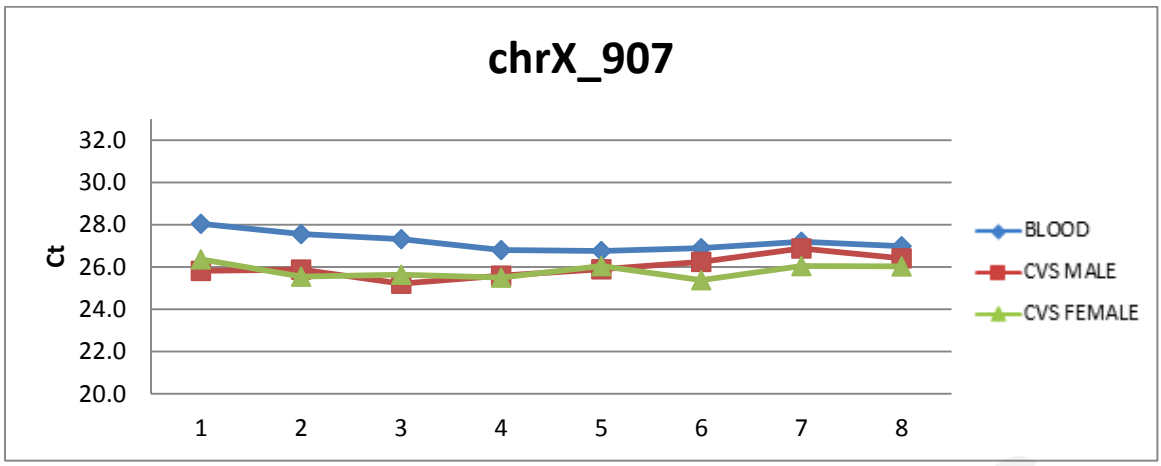


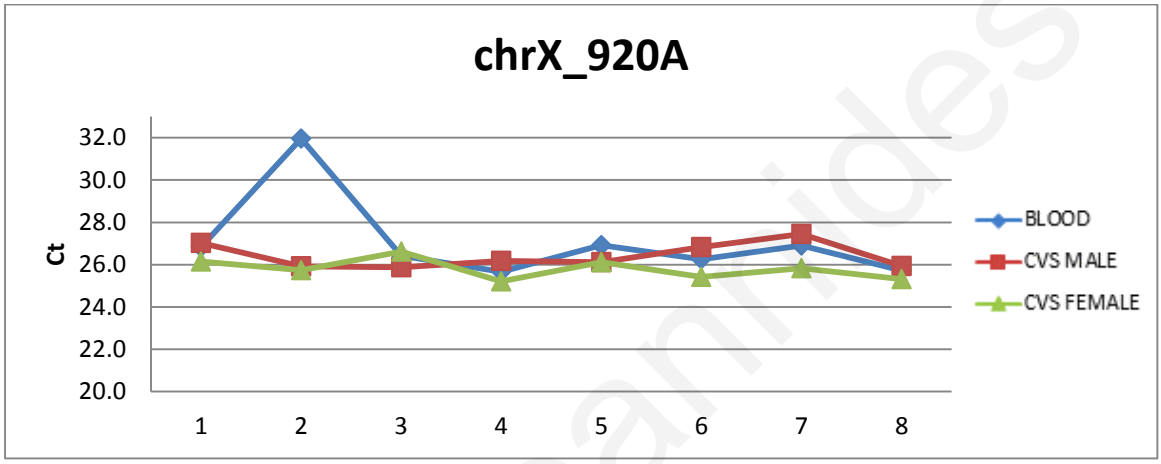


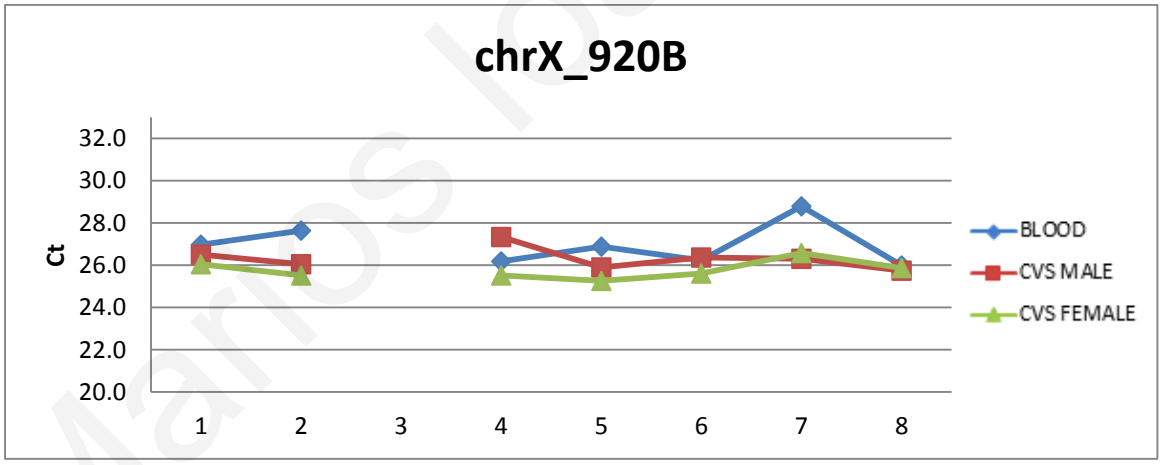


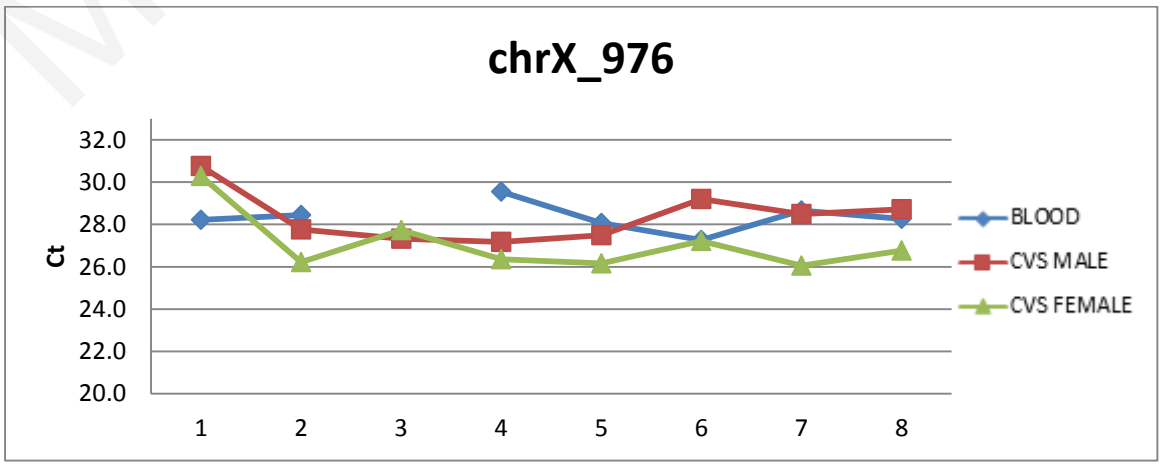


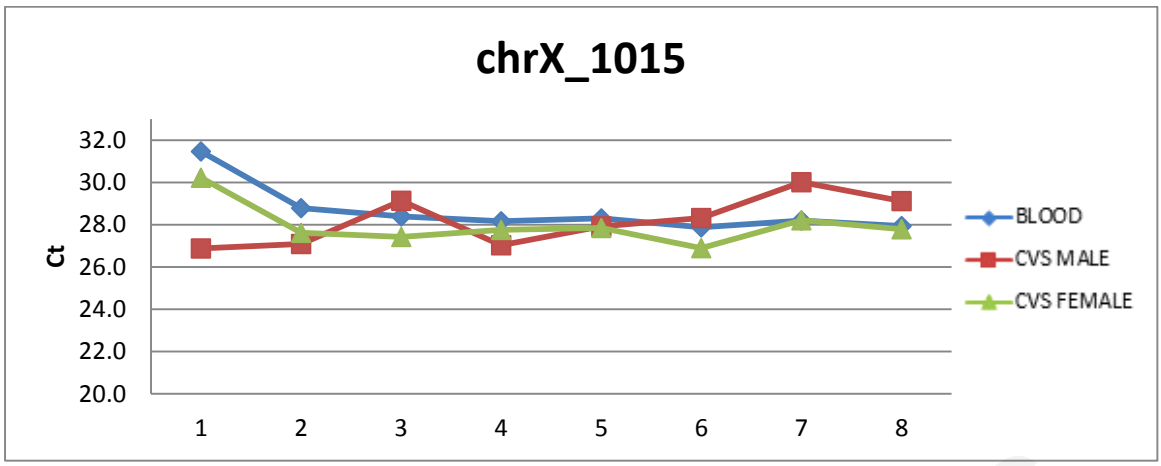


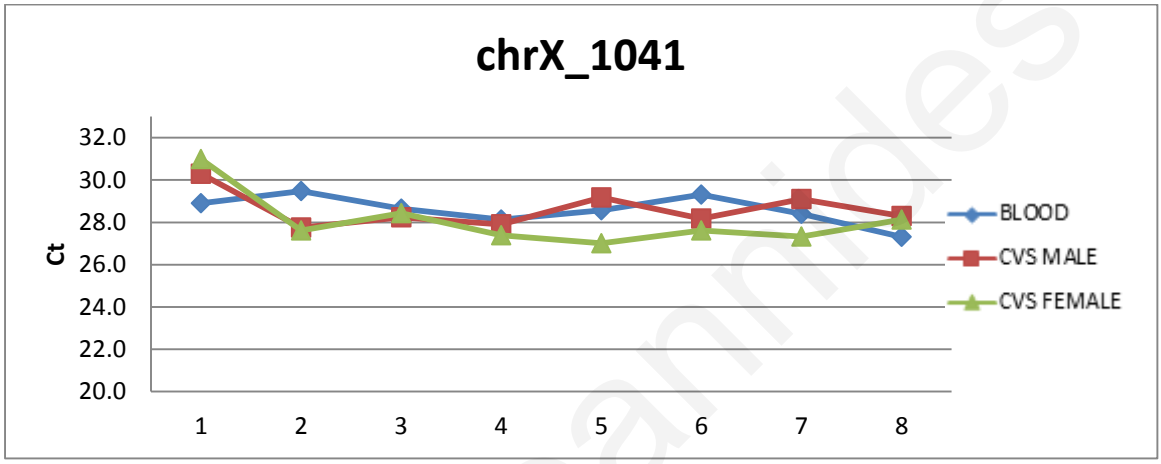


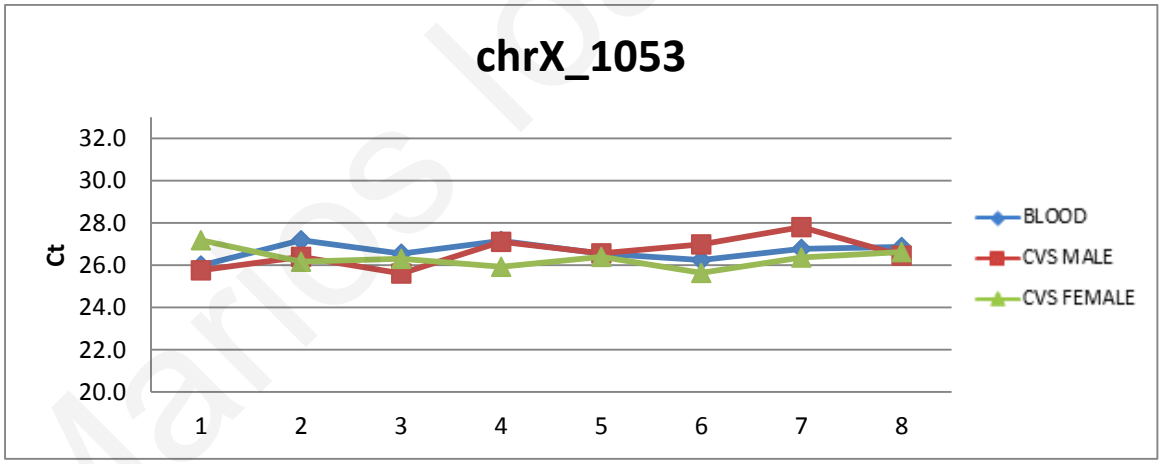


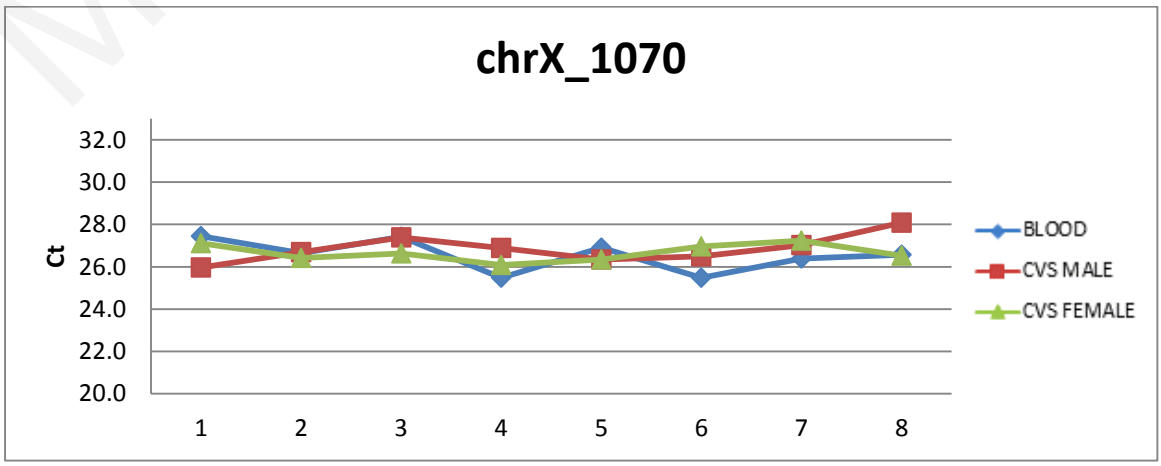


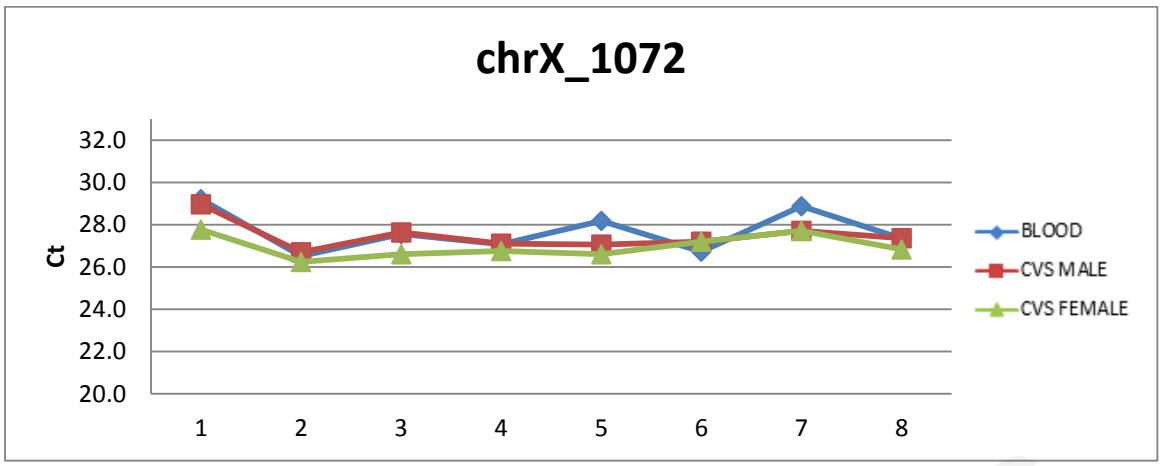


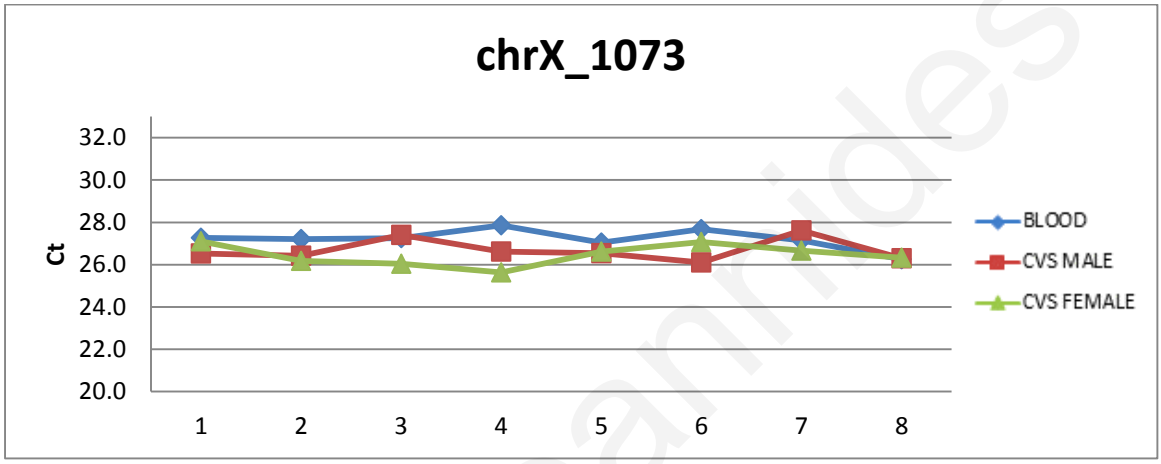


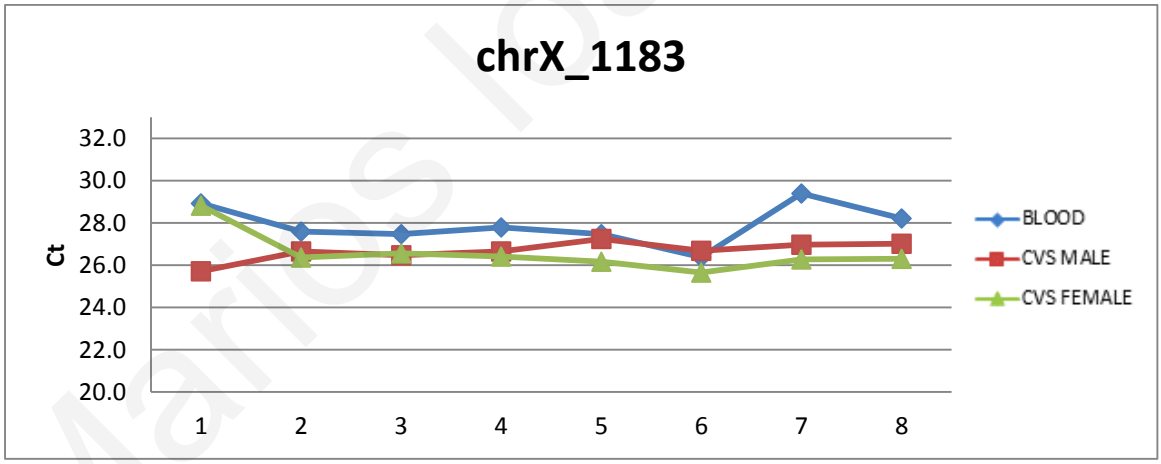


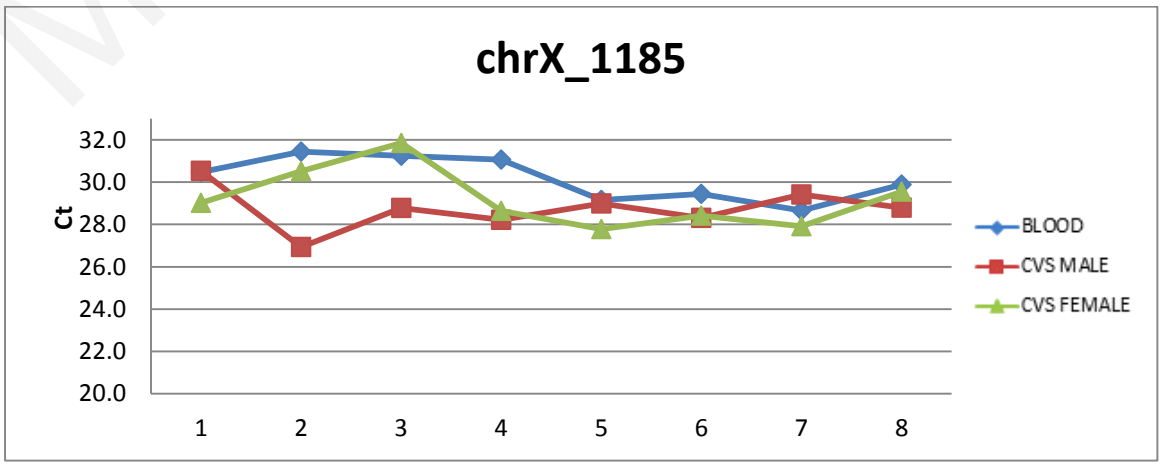


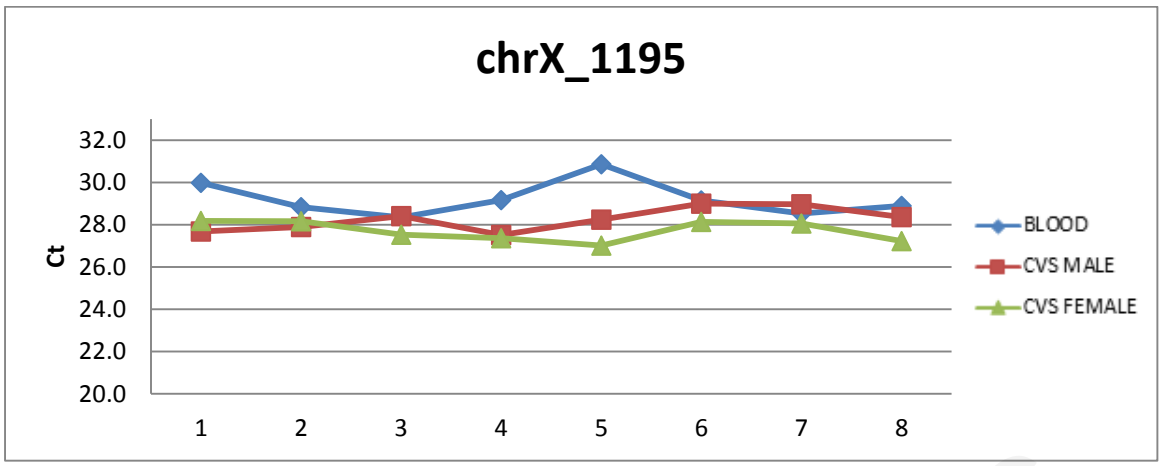


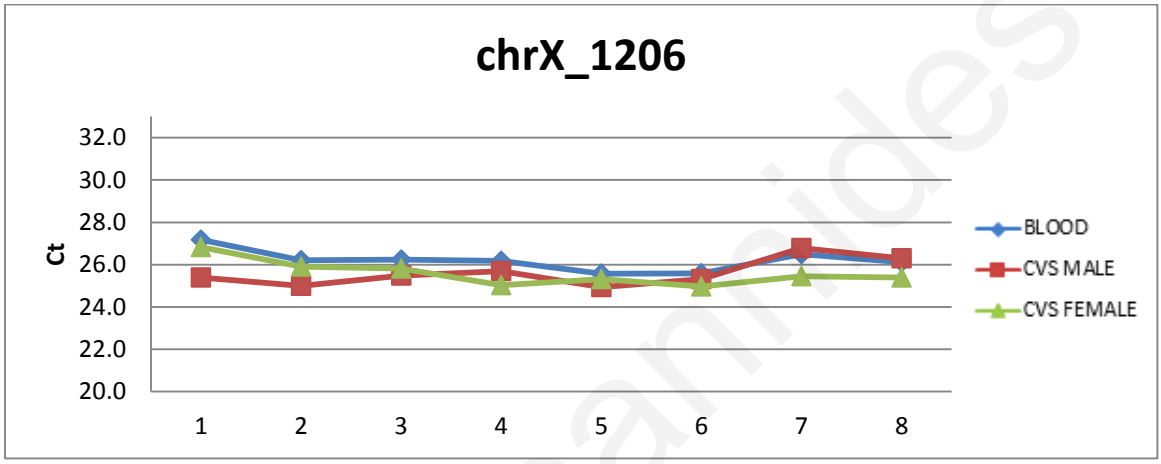


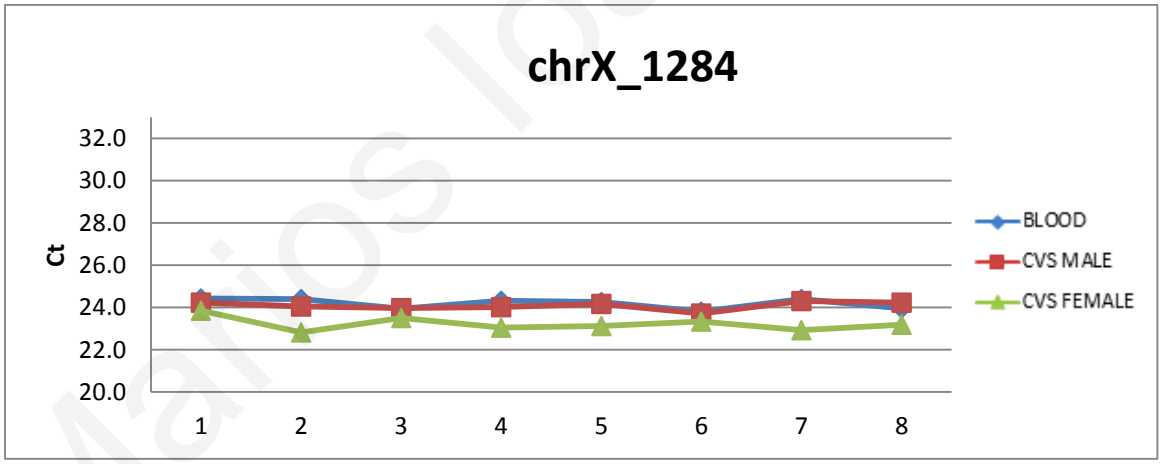


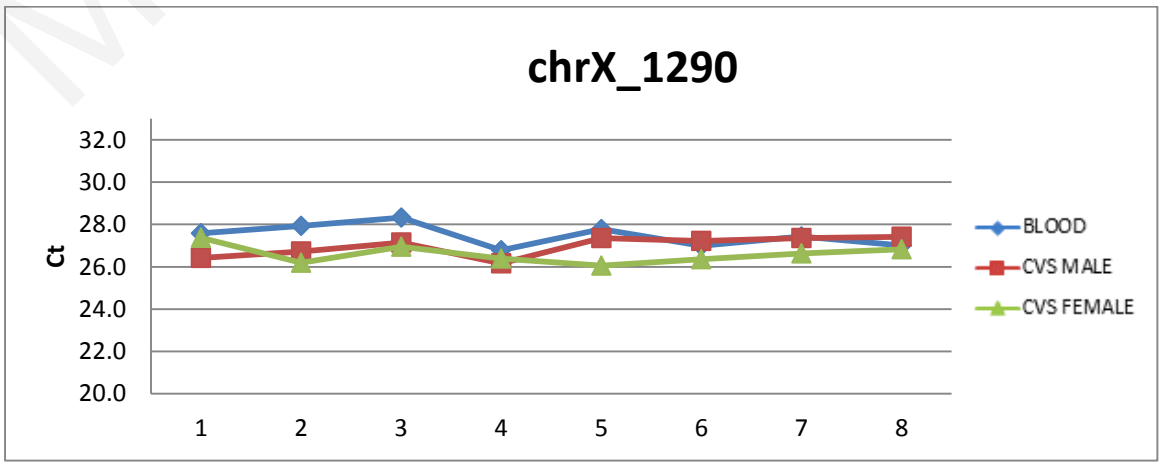


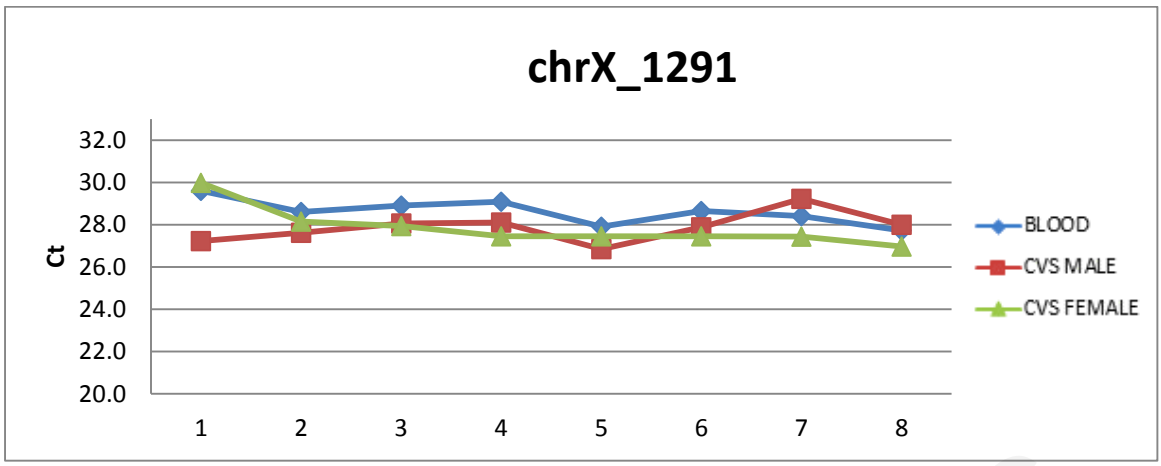


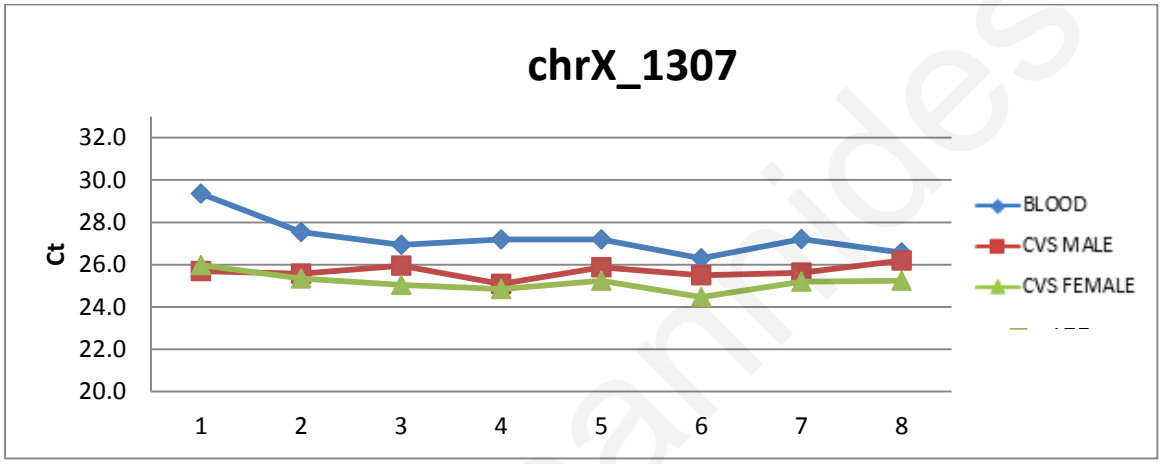




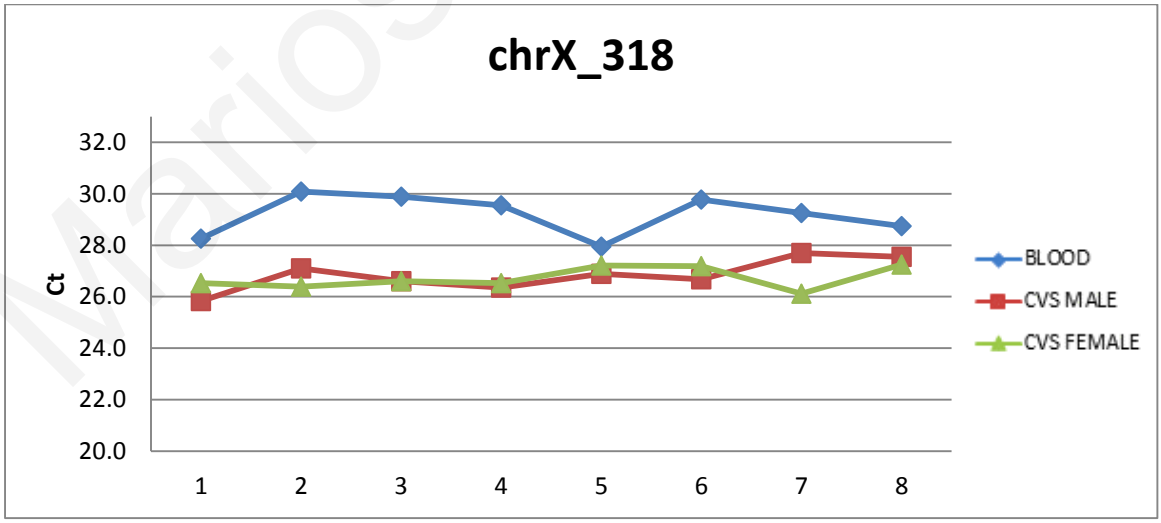


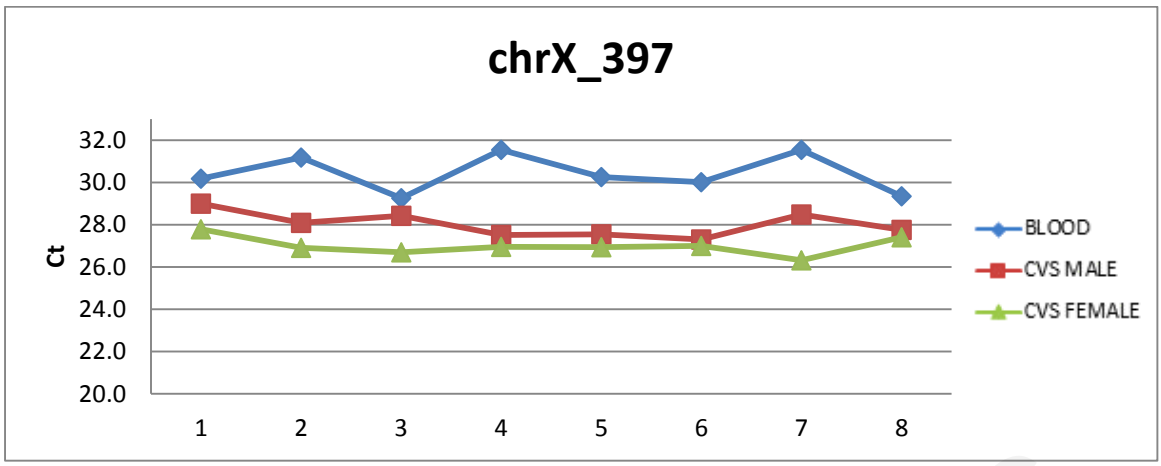






2. "Good DMRs





Appendix IV

Overlapping analysis of control regions and DMRs confirmed using previous aCGH data with MeDIP-seq.

Chromosome	Name	Location	Confirmed	Refined	Comments
chr13	AV2	36045985-36046319	yes		
chr13	AV5	100547305-100548013	Yes	100547080-100547498	
chr13	R2	33484488-33484831	Yes	33484996-33485436	
chr13	*HYPER[100]	21093387-21093465	Yes		
chr18	VA1	518995-519314	yes		
chr18	VA4	2950888-2951347	yes		High WBF
chr18	VA15	60663920-60664240	yes		
chr18	VA17	60958346-60958757	No		
chr18	VA22	74842401-74842914	yes	74842166-74842509	
chr18	AII	56939304-56939625	yes		
chr18	B	45911986-45912277	yes		
chr18	C	60804864-60805624	yes		
chr18	*SERPINB5[64]	61143831-61144295	yes		
chr21	Nn	32504886-32505275	yes		
chr21	On	35570844-35571333	yes	35570590-35571038	
chr21	Fd	43132892-43133147	yes	43132389-43132752	Weak
chr21	*EI [100]	43482297-43482839	yes		
chr21	*EII [100]	43484072-43484332	yes		
chr21	Id	43880608-43880957	yes		
chr21	M1E	46129212-46129426	yes		
chr21	M28	46346587-46346797	yes		
chr21	M18	16409947-16410074	No reads		
chr21	M20	16256542-16256626	Yes?		
chr21	M25	38770994-38771104	No reads		
chr21	M27	43305739-43305939	yes	43305611-43305939	High WBF
chr21	*H[100]	33346794-33347429	yes		
chr21	*C [100]	34398660-34398945	yes		
chr21	*J [100]	38919270-38919995	yes		
chr21	*A [100]	40356788-40358215	yes		
chr21	*D [100]	43315721-43316554	yes		
chr22	*HYPO[100]	31884952-31885055	yes		
chr3	*RASS1F[115]	50375947-50377247	yes		

* Control Regions

Most regions identified and characterized using existing aCGH data were confirmed with MeDIP-seq. Due to the low resolution of the array as compared to the NGS the location of some regions was refined. Highlighted regions indicate the four DMRs that showed overlapping between the enrichment values between CVS and WBF as shown in box plots of Figure 3.2. MeDIP-seq confirmed the results obtained by the qPCR results for these regions

Appendix V

Overlap between confirmed DMRs from MeDIP-Chip with MeDIP-seq.

Chromosome	DMR	Location	Overlap
chr13	V1.3	28000244-28000401	YES
chr13	V2.3	28000968-28001223	YES
chr13	V2.4	28006507-28006702	NO
chr13	V2.5	28492779-28492985	YES
chr13	V2.6	28502766-28502923	YES
chr13	V1.6	28503354-28503805	NO READS
chr13	V1.7	28529246-28529501	NO READS
chr13	V2.10	41237974-41238121	NO READS
chr13	V1.23	45149681-45149936	NO READS
chr13	V2.13	47253569-47253814	NO READS
chr13	V2.15	49079547-49080145	YES
chr13	V2.36	50697634-50697889	NO READS
chr13	V2.16	50706149-50706502	NO READS
chr13	V2.17	52352426-52352681	YES
chr13	V1.29	60026124-60026428	YES
chr13	V1.40	78396779-78397181	NO READS
chr13	V2.39	84453560-84453717	YES
chr13	V1.41	84455373-84455628	NO READS
chr13	V2.20	95356183-95356389	YES
chr13	V2.23	99706202-99706359	YES
chr13	V2.41	99933956-99934260	YES
chr13	V2.27	102572058-102572215	YES
chr13	V2.28	102572352-102572509	NO READS
chr13	V1.48	107143491-107143795	YES
chr13	V1.49	107146281-107146536	YES
chr13	V1.50	107160143-107160477	YES
chr13	V1.52	109550413-109550609	YES
chr13	V2.30	110993171-110993328	YES
chr13	V1.55	110994151-110994406	YES
chr13	V2.33	112331101-112331258	YES
chr18	V2.4	2960419-2960592	NO READS
chr18	V2.12	9244490-9244815	NO READS
chr18	V2.132	29971249-29971574	NO READS
chr18	V2.47	42339235-42339510	NO READS
chr18	V2.51	43785194-43785481	NO READS
chr18	V2.149	54157193-54157480	YES
chr18	V3.1	813609-813787	NO READS
chr18	V3.2	2971153-2971326	YES
chr18	V3.6	5455740-5455951	NO READS
chr18	V3.11	11863670-11863870	YES
chr18	V3.17	20772962-20772121	YES
chr18	V3.20	38970191-38970287	YES
chr18	V3.38	46455444-46455617	YES
chr18	V3.42	32440338-32440487	NO READS
chr18	V3.3	3060530-3060702	YES
chr18	V2.126	22805328-22805463	YES
chr18	V2.57	45911996-45912169	YES
chr18	V2.81	60765646-60765849	YES
chr18	V2.82	60765866-60766077	YES
chr18	V3.39_4	56933527-56933700	YES
chr18	V3.43_2	55862966-55863101	YES
chr21	Un1	32637846-32638067	YES
chr21	Un4	34524265-34524468	YES
chr21	Un5	36080170-36080373	YES
chr21	Un6	36398107-36398274	NO READS
chr21	Un8	36577416-36577619	YES
chr21	Un9	38598143-38598274	YES

Chromosome	DMR	Location	Overlap
chr21	Un10	39869594-39870031	YES
chr21	Un13	43319548-43319931	YES
chr21	Un15	43947842-43948099	YES
chr21	Un16	44166999-44167310	YES
chr21	Un17	47788699-47788992	YES
chr21	Un18	47971306-47971509	YES
chr21	Un19	48087516-48087719	YES
chr21	Un21	44876826-44876957	YES
chr21	Un22	45254802-45255000	YES
chr21	Un23	45509229-45509396	YES
chr21	Un25	45336694-45336915	YES
chr21	8T21	34400677-34400880	NO READS
chr21	TS1	42213717-42214190	YES
chr21	23T21	43146155-43146322	YES
chr21	62T21	18886716-18886883	YES
chr21	64T21	19162137-19162340	YES
chr21	65T21	26935467-26935634	NO READS
chr21	67T21	32699086-32699469	YES
chr21	69T21	34392769-34392936	YES
chr21	70T21	34395003-34395152	YES
chr21	71T21	34393237-34393422	YES
chr21	79T21	35348986-35349153	YES
chr21	81T21	36356646-36356849	YES
chr21	82T21	36421752-36421955	YES
chr21	84T21	38066188-38066355	YES
chr21	TS2	38074043-38074372	YES
chr21	TS3	38077214-38077399	YES
chr21	TS4	38079742-38079944	YES
chr21	99T21	44834618-44834929	YES
chr21	101T21	45254802-45255095	YES
chr21	102T21	45336712-45336889	YES
chr21	106T21	46451064-46451231	YES
chr21	109T21	38083157-38083306	NO READS
chr21	110T21	34400335-34400448	NO READS
chr21	ZH1	32699066-32699487	YES
chr21	ZH2	35348968-35349207	YES
chr21	ZH3	35449473-35449802	YES
chr21	ZH4	36341073-36341523	NO READS
chr21	ZH5	36901300-36901570	YES
chr21	ZH6	38063434-38063799	NO READS

Inspection of the confirmed DMRs showed high degree of overlap with the MeDIP-seq results.

Appendix VI

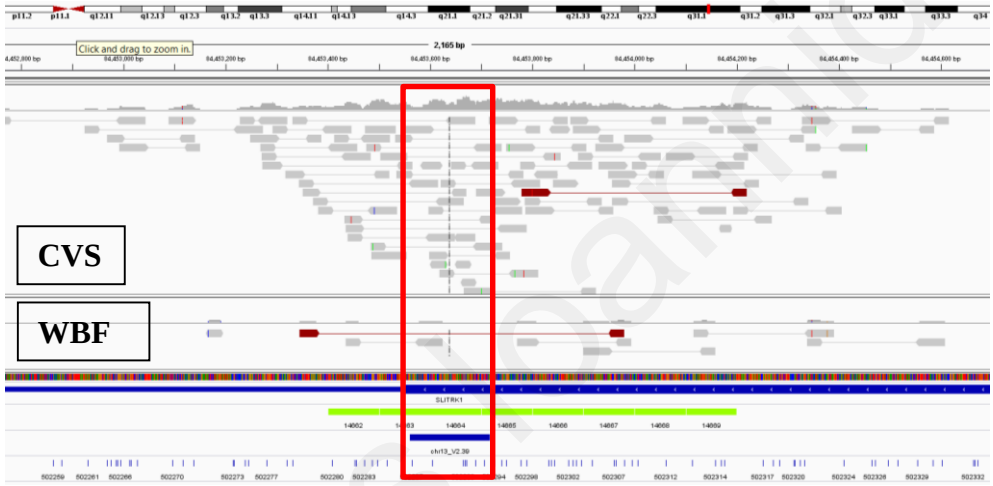
DMRs on chromosomes 13, 18 and 21 that showed overlap between the MEDIPS Software and confirmed regions from MeDIP-Chip (red boxes). Green tracks: Regions detected by MEDIPS software. Blue tracks: Regions confirmed by MeDIP-Chip.

A. Chromosome 13

V1.55

V2.30

V2.27



V2.39

V2.15

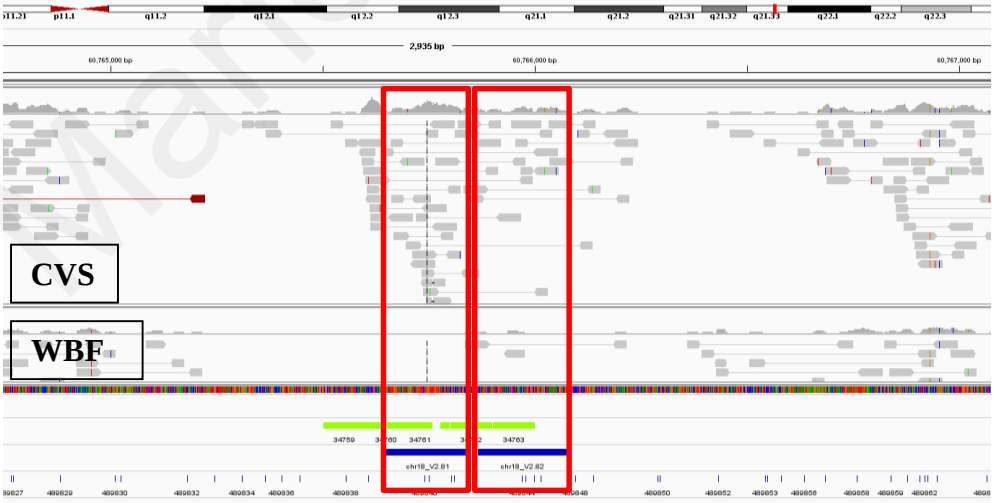
V2.6

V2.5

B. Chromosome 18

V3.43

V3.39



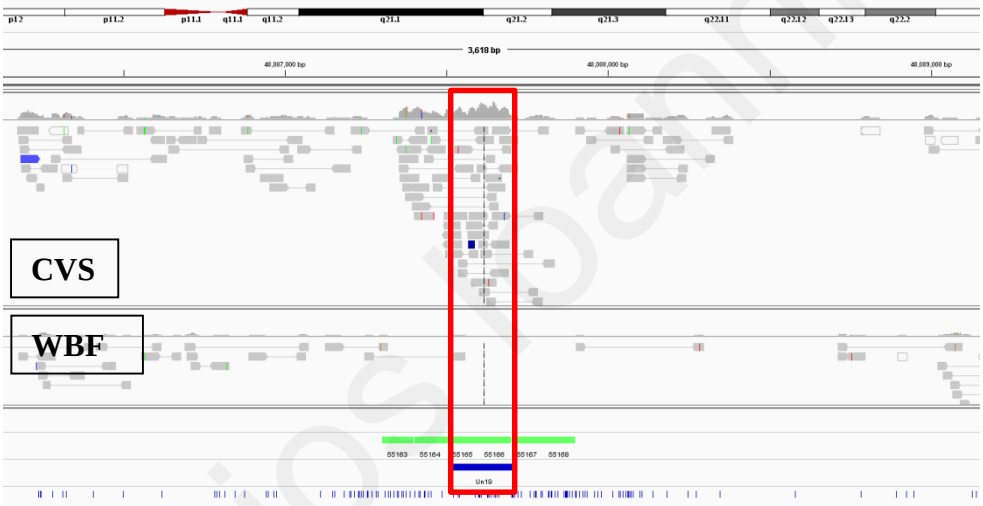
V2.81 AND V2.82

V2.57

V2.126

C. Chromosome 21

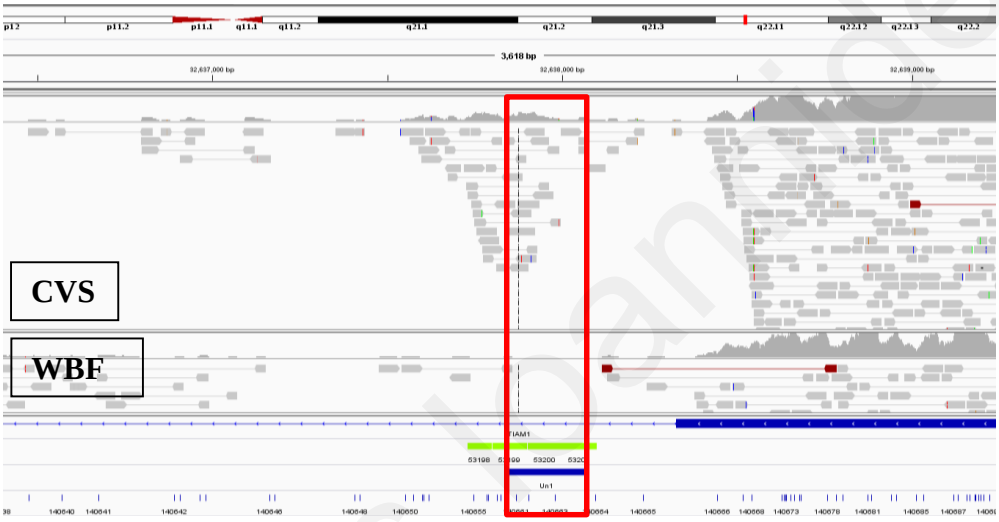
69T21 AND 71T21



UN19

UN10

UN9



UN1

MOLECULAR CYTOGENETICS

RESEARCH Open Access

Inter-individual methylation variability in differentially methylated regions between maternal whole blood and first trimester CVS

Marios Ioannides[1,2], Elisavet A Papageorgiou[1,3], Anna Keravnou[1], Evdokia Tsaliki[3], Christiana Spyrou[3], Michael Hadjidaniel[1], Carolina Sismani[1], George Koumbaris[1,3] and Philippos C Patsalis[1*]

Abstract

Background: DNA methylation is the most studied form of epigenetic regulation, a process by which chromatin composition and transcription factor binding is altered to influence tissue specific gene expression and differentiation. Such tissue specific methylation patterns are investigated as biomarkers for cancer and cell-free fetal DNA using various methodologies.

Results: We have utilized methylation DNA immunoprecipitation (MeDIP) and real-time quantitative PCR to investigate the inter-individual methylation variability of differentially methylated regions (DMRs) on chromosomes 18 and 21. We have characterized 15 newly selected and seven previously validated DMRs in 50, 1st trimester Chorionic villus samplings (CVS) and 50 female non-pregnant peripheral blood (WBF) samples. qPCR results from MeDIP and genomic DNA (Input) assays were used to calculate fold enrichment values for each DMR. For all regions tested, enrichment was higher in CVS than in WBF samples with mean enrichments ranging from 0.22 to 6.4 and 0.017 to 1 respectively. Despite inter-individual variability, mean enrichment values for CVS were significantly different than those for WBF in all DMRs tested ($p < 0.01$). This observation is reinforced by the absence of overlap in CVS and WBF enrichment value distributions for 15 of 22 DMRs.

Conclusions: Our work provides an expansion in the biomarker panel available for non-invasive prenatal diagnosis (NIPD) using the MeDIP-qPCR methology for Down syndrome and can eventually provide the starting point towards the development for assays towards the detection of Edwards syndrome. Furthermore, our data indicate that inter-experimental and inter-individual variation in methylation is apparent, yet the difference in methylation status across tissues is large enough to allow for robust tissue specific methylation identification.

Keywords: Non-invasive prenatal diagnosis, Inter-individual variability, Differentially methylated regions, MeDIP

Background

In vertebrates DNA methylation is a conserved epigenetic modification by which DNA methyltransferases add a methyl group to carbon 5 of cytosine residues present in CpG dinucleotides. This modification is the most studied form of epigenetic regulation and has been strongly associated with chromosomal stability and imprinting control [1]. Furthermore, this epigenetic process also regulates chromatin composition and transcription factor binding to directly influence transcriptional activity [2,3].

DNA methylation occurs primarily in CpG islands (GGIs) and shores both in coding and non-coding regions of the genome, with gene regulatory regions such as promoters and first exons being a frequent methylation target [4]. Due to this integral relationship with gene expression regulation, DNA methylation patterns are very closely associated with developmental processes and differentiation. Consequently, DNA methylation directly modulates phenotype, and distinct methylation patterns have been associated with tissue specificity and a variety of disease states ranging from cancer to neurological disorders [5,6]. These tissue specific differentially methylated regions (tDMRs) are currently under investigation for

* Correspondence: patsalis@cing.ac.cy
[1]The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
Full list of author information is available at the end of the article

their utility as biomarkers for disease progression and prognosis, particularly in the field of cancer research, disease detection and response to treatment [7].

The discovery of cell free fetal DNA (cffDNA) in the maternal circulation has greatly facilitated the development of non-invasive prenatal diagnosis (NIPD) [8]. The direct correlation between phenotype and DNA methylation patterns has allowed the use of DMRs as possible biomarkers in prenatal diagnosis. Several groups have utilized the methylation differences between placenta-derived cffDNA and maternal DNA in order to identify highly specific fetal DMR biomarkers for non-invasive prenatal diagnosis of aneuploidies. Previous studies employed a variety of methods including sodium bisulfite conversion and methylation sensitive restriction digestion, but yielded a relatively small number of fetal specific DMRs including the *SERPINB5*, *RASSF1A* and U-*PDE9A* genes [9-11].

In 2009, Papageorgiou et al. [12] applied methylation DNA immunoprecipitation (MeDIP) coupled with high resolution tiling oligonucleotide array (Chip) analysis to identify DMRs between Chorionic villus sampling (CVS) and female peripheral blood DNA (WBF). They were able to identify thousands of DMRs on chromosomes 13, 18, 21, X, and Y including methylation sensitive restriction sites, CGIs and promoter regions. This MeDIP-Chip approach was the trigger for investigating the utility of MeDIP followed by real-time qPCR (MeDIP-qPCR) for the non-invasive prenatal diagnosis of trisomy 21, yielding 100% sensitivity and specificity [13]. This novel NIPD method was validated by a second study of 175 cases again yielding high sensitivity and specificity [14].

The current study utilizes the MeDIP-qPCR methodology to expand our range of fetal specific DMR biomarkers by selecting and screening 15 additional DMRs on chromosomes 21 and 18. Special emphasis is given on investigating the methylation variability in different samples from these newly selected and previously reported DMRs [12-14] by screening them in a set of 50, 1st trimester CVS and 50 WBF. Overall, this work confirms the distinctively different methylation status of these regions in CVS and WBF.

Results

Using the above criteria we identified a set of 40 candidate DMRs between CVS and WBF from the microarray data [12]. This set was subsequently screened in a cohort of six CVS and six WBF to calculate the enrichment values for each DMR (Additional file 1). Based on this initial screening we were able to select the 15 regions with the highest CVS enrichment for further validation/characterization, using seven previously validated DMRs by Papageorgiou et al. [13] and Tsaliki et al. [14], as a comparison standard.

This DMR validation study was conducted on a set of 50 CVS and 50 WBF samples using the MeDIP–qPCR methodology (Table 1), the efficiency of which was monitored using one hypermethylated (HYPER) and one hypomethylated (HYPO) control regions. The HYPER is a region that showed hypermethylation for both CVS and WBF, while the HYPO is a region that showed hypomethylation for the two tissues [12]. Enrichment values for HYPO were low in WBF and CVS samples while the HYPER control region showed enrichment for CVS and WBF with mean enrichment values of 3.12 and 3.22 respectively, indicating that the MeDIP procedure was highly specific for the methylated regions. Moreover, the previously validated DMRs performed as previously described [12], exhibiting distinctively different enrichment between CVS and WBF.

All tested DMRs showed a significant enrichment ($p < 0.01$) in CVS compared to those of WBF (Table 1). We compared the performance of the 15 newly selected DMRs with the previously validated set and we were able to determine that 11 of 15 DMRs showed enrichment values higher than the lowest of the previously validated DMRs, ranging from 1.9 to 6.4. Additional comparison of the two DMR sets also illustrated that for 11 of these 15 regions the difference of means (mean enrichment CVS – mean enrichment WBF) was again higher than the respective values of the validated DMRs (ranging from 1.6 to 6.4) (Table 1). Our analysis also shows that the enrichment distributions for CVS and WBF have no overlap for these 11 DMRs (Figure 1).

To better investigate tissue specificity (CVS-WBF) in the 15 newly selected DMRs in relation to the previously validated DMRs, we also constructed a heat map and hierarchical clustering of the 50 CVS and 50 WBF samples based on the obtained enrichment values (Figure 2). This analysis shows a clear differentiation between the two tissue types based on the obtained enrichment values. Furthermore, DMR clustering analysis showed that there was no distinct clustering separation between the newly selected and the previously validated DMRs.

Discussion

Our study aimed to validate and characterize a set of differentially methylated regions between CVS and WBF, obtained from MeDIP-Chip data [12]. The methylation characteristics of the 15 candidate DMRs, located on chromosomes 18 and 21, were ascertained in 50 CVS and 50 WBF using the MeDIP-qPCR methodology. To our knowledge this is the first MeDIP based biomarker screening study utilizing such a large sample set. None of the selected DMRs were located on CGIs, but within intergenic or intragenic regions. Such DMR distribution in non-coding intergenic and intragenic sequences is in agreement with data from a large scale investigation of tissue

Table 1 Ranking of DMRs tested according to the difference between mean enrichment values for each DMRs

Marker	Mean WBF	Mean CVS	Mean difference	SD WBF	SD CVS	U pval	Coefficient of variation WBF	Coefficient of variation CVS
EI-4	0.017	6.384	6.367	0.022	2.143	1.53E-17	1.294	0.336
EII-1	0.065	5.319	5.254	0.071	1.937	7.06E-18	1.092	0.364
H2	0.135	4.068	3.933	0.093	1.252	7.06E-18	0.689	0.308
EI-2	0.064	3.894	3.83	0.2	1.338	7.50E-18	3.125	0.344
EI-3	0.116	3.905	3.789	0.312	1.556	1.96E-17	2.690	0.398
B3	0.126	3.86	3.734	0.1	1.268	2.29E-17	0.794	0.328
M27	0.532	4.113	3.581	0.2	1.386	7.06E-18	0.376	0.337
D2	0.317	3.364	3.047	0.179	1.301	7.06E-18	0.565	0.387
M28	0.189	2.777	2.588	0.125	0.919	7.06E-18	0.661	0.331
M1E	0.149	2.636	2.487	0.097	0.742	3.44E-17	0.651	0.281
Id1	0.398	2.682	2.284	0.155	0.932	1.04E-17	0.389	0.348
A5	0.337	2.505	2.168	0.159	0.986	7.06E-18	0.472	0.394
C5	0.18	2.321	2.141	0.106	0.84	7.06E-18	0.589	0.362
C1	0.106	2.229	2.123	0.083	0.635	7.06E-18	0.783	0.285
AII-2	0.065	2.003	1.938	0.084	0.933	1.23E-16	1.292	0.466
On2	0.281	1.993	1.712	0.138	0.552	7.06E-18	0.491	0.277
Nn2	0.245	1.924	1.679	0.107	0.78	7.07E-18	0.437	0.405
J2	0.116	1.707	1.591	0.079	0.519	7.06E-18	0.681	0.304
Fd1	0.135	1.676	1.541	0.101	0.513	7.06E-18	0.748	0.306
M25	1.038	1.822	0.784	0.452	0.655	1.88E-09	0.435	0.359
M20	0.42	0.796	0.376	0.186	0.303	2.37E-10	0.443	0.381
M18	0.097	0.22	0.123	0.07	0.151	6.36E-08	0.722	0.686
HYPER	3.124	3.226	0.102	0.68	0.982	0.951	0.218	0.304
HYPO	0.469	0.508	0.039	0.945	0.948	0.119	2.015	1.866

Despite the statistical significance of all enrichment values ($p < 0.01$), the four markers (M25, M20, M18, Fd1) that showed the lowest difference were not selected as potential DMRs.

specific methylation profiles by Rakyan et al. [15], who reported differential methylation of intergenic and CpG poor promoter regions in addition to CGIs.

Based on our results, all 15 DMRs showed differential enrichment between the two tissues and 11 out of these 15 were strongly and consistently hypermethylated in CVS samples. The ability of these DMRs to distinguish between CVS and WBF was equivalent to that of the seven previously validated DMRs used as performance standards. In fact, the tissue discriminating performance of the DMRs tested here, shows close similarities with the previously validated DMRs as it is illustrated by our heat map distribution and the unsupervised clustering patterns obtained.

DMR enrichment values showed variability among the different samples. This is likely caused by combination of both inter-experimental technical variability and inter-individual methylation variability. The presence of variability in MeDIP based assays has previously been described by Butcher et al. [16]. In addition, the issue of inter-individual DNA methylation variability has been the focus of several studies [17-19]. This high inter-individual variability has been attributed to a variety of factors including environmental conditions, diet, age and psychosocial factors [20-23]. Furthermore, it has been documented that regions with low CpG density, as the DMRs under investigation, show higher inter-individual variability as compared to regions with high CpG density, such as CGIs [24].

Others have also shown that methylation variability can coincide with tissue specific DMRs without obscuring the tissue discriminating properties of those DMRs [17]. It is therefore of no surprise that despite the DNA methylation variability in our study, the newly validated set of 11 DMRs clearly distinguishes between CVS and WBF tissues.

Our work here substantially increases the number of confirmed chromosome 21 fetal specific DMRs, and therefore provides a significant expansion in the biomarker panel available for MeDIP-qPCR-based NIPD of Down syndrome. Such an expansion is predicted to further improve the robustness of the methodology and bolster its diagnostic classification power. It is also very important to

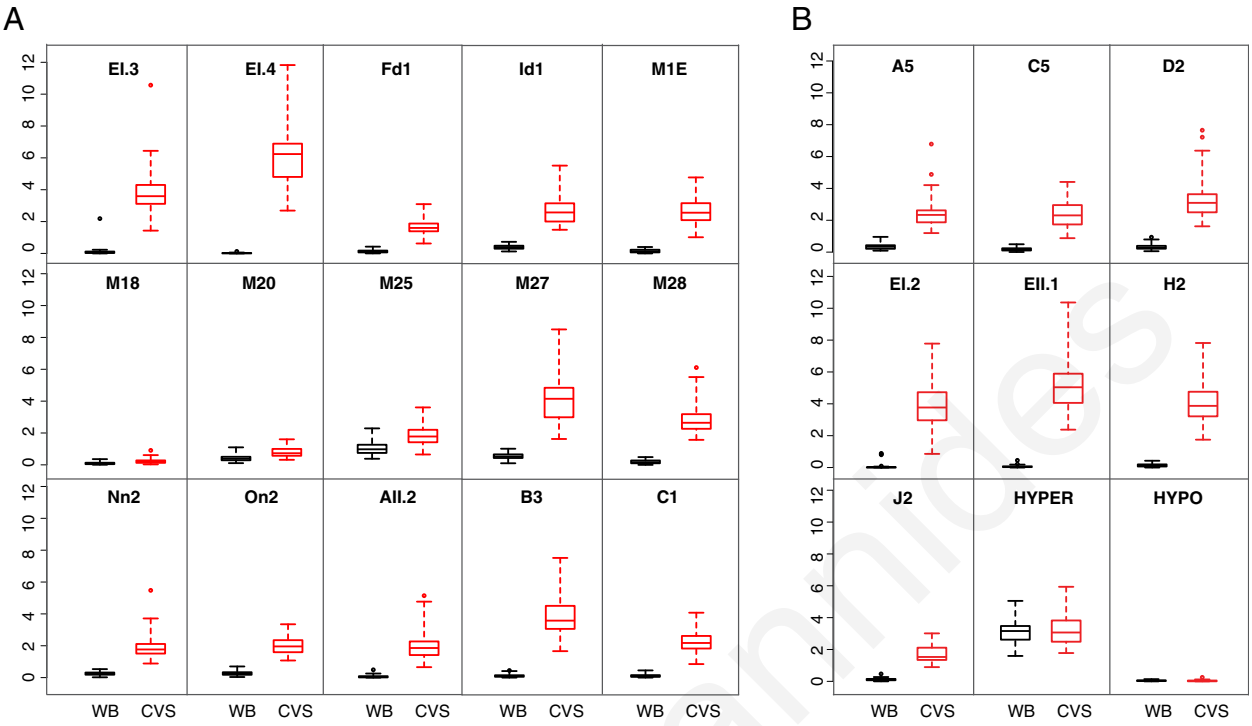


Figure 1 Enrichment profile for all DMRs on 50 WBF and 50 first trimester CVS. Box plots show the distribution of the relative fold enrichment values for WBF (black) and CVS (red) for each DMR. The median value is represented by a horizontal line. The bottom of the box indicates the 25th percentile (lower quartile) and the top the 75th percentile (upper quartile). Whisker lines indicate the outlier boundaries [top: median + 3(75%-25%); bottom: median -3(75%-25%)]. **A**: 11 of 15 newly characterized DMRs show clear separation of the methylation enrichment values between WBF and CVS despite the methylation variability. Fd1, M18, M20, M25 show overlap between the interquartile values, thus they were not characterized as DMRs. **B**: Previously validated DMRs used as a comparison standard show clear separation between WBF and CVS. HYPER: hypermethylated marker for both tissues, HYPO: Hypomethylated marker for both tissues.

note that our current study is the first to validate chromosome 18 fetal specific DMRs in a relatively large sample set. This small panel of chromosome 18 DMRs can potentially provide a very valuable testing platform on which future NIPD assays for Edwards syndrome will be developed.

Conclusions

NIPD has gained a lot of interest the last few years. Utilizing the methylation differences between fetal and maternal DNA, several groups have managed to identify biomarkers using different approaches. This study aimed to characterize and validate fetal specific methylated regions using the MeDIP-qPCR methodology. We were able to show that the selected regions had distinct methylation patterns between fetal and maternal tissue, despite inter-individual and inter-experimental variability. In addition, we have expanded the panel of the existing DMRs on chromosome 21 and have characterized a new set of markers on chromosome 18 which can provide the starting point towards the development for assays towards the detection of Edwards syndrome.

Methods

Human Samples and DNA preparation

WBF samples were obtained anonymously from 50 normal non-pregnant females 20-40 years of age. Fifty, 1st trimester CVS were obtained from the Department of Cytogenetic and Genomics at the Cyprus Institute of Neurology and Genetics (Nicosia, Cyprus). Protocols used for collecting samples for our study were approved by the appropriate Bioethics Committees, and informed consent was obtained from all participants. WBF and CVS samples were used to extract DNA using the QIAamp DNA blood midi kit or the QIAmp DNA mini kit according to the manufacturer's instructions (QIAGEN, Hilden, Germany). All CVS underwent karyotyping and Quantitative-Fluorescent PCR (QF-PCR) analysis in order to confirm their normal status.

Ligation-mediated PCR (LM-PCR) and MeDIP assay

LM-PCR and MeDIP assays were conducted as described previously [12]. Briefly, 2.5 μg of genomic DNA were sonicated using the Bioruptor Twin sonication system (UCD-400, Diagenode, Liege, Belgium) into fragments, 300-1000 bp in size. Fragment size was verified

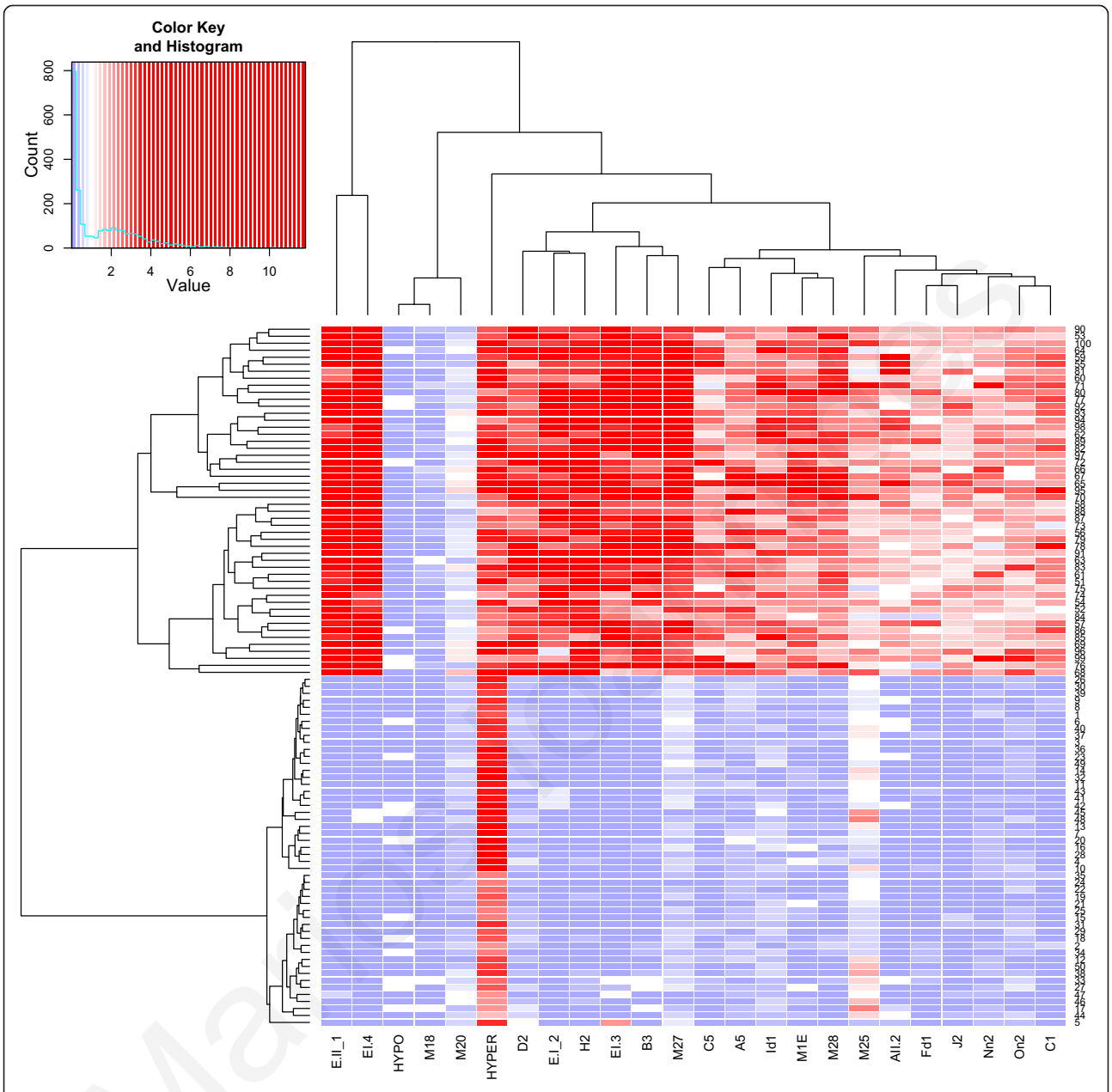


Figure 2 DMRs exhibiting tissue specificity between CVS and WBF. A heat map constructed based on the relative fold enrichment values obtained for the two tissues tested (CVS and WBF) with the MeDIP-qPCR approach shows clustering of the hypermethylated CVS samples (high enrichment-red) and the hypomethylated WBF samples (low enrichment–blue). Furthermore, DMR clustering analysis showed no cluster distinction between the 15 newly selected DMRs and the previously validated DMRs that were used in this study as a comparison standard.

using agarose gel electrophoresis. The fragments were blunt-ended using HPLC water, 1X NEB buffer 2 (New England BioLabs, Ipswich, UK), 10X bovine serum albumin (New England BioLabs) 100 mmol/L dNTP mix (GE Healthcare, Little Chalfont, UK) and T4 DNA polymerase (3 U/μl; New England BioLabs). Fragments were purified using the QiAquick PCR purification kit (Qiagen) and linkers were then ligated onto the blunt ends by overnight incubation at 16°C with T4 DNA ligase (New England) and T4 DNA ligase buffer (New England). Overhangs were subsequently filled in by incubating at 72°C for 10 minutes with 100 mmol/L dNTP mix (GE Healthcare), 1X PCR gold buffer (Roche, Mannheim, Germany), 1.5 mmol/L $MgCl_2$ (Roche) HPLC water and AmpliTaq DNA polymerase (Applied Biosystems, Branchburg, New Jersey, USA). 50 ng of ligated DNA was removed and

kept as input DNA. The remaining ligated DNA (800-1200 ng) was subjected to MeDIP using 3 μg mouse anti-5'methylCytosine (a-5mC) antibody (Eurogentec Saraing, Belgium). Hypermethylated DNA bound to a-mC antibodies was magnetically captured using Dynabeads® M-280 Sheep Anti-Mouse IgG magnetic beads (Life technologies, Carlsbad, California, USA) and subsequently released using proteinase K (Roche). LM-PCR used 12 ng of each input and MeDIP DNA as described earlier [12].

DMR selection

Candidate DMRs on chromosomes 18 and 21 were selected from a set of potential differentially methylated regions previously described [12] according to the following three criteria: a) the region included at least three consecutive microarray probes, b) array results showed consistent DNA hypermethylation in first and third trimester placentas and hypomethylation in WBF samples, c) the region did not include segmental duplications and copy number variable regions based on the Database for Genomic Variants (DGV) [25]. The regions considered for this paper are shown in Table 2.

Real-Time quantitative PCR (qPCR)

Primer design, optimal primer concentration experiments, and efficiency (e) of each qPCR reaction were performed as previously described [12] with the following modifications. Each qPCR reaction was performed on 8 ng of template DNA using SYBR Green PCR mastermix (Eurogentec) in a final reaction volume of 10 μl, using a BIORAD CFX 384 Real time system (BIORAD, Hercules, California). Each MeDIP, or Input template DNA was used to prepare three replicate qPCR reactions that were used for calculating the average Ct value for each template. Primer3 software [26] was used to design the qPCR primer sets that were synthesized by Sigma-Aldrich (Munich, Germany). Primer sets utilized for this body of work are listed in Table 3.

Statistical calculations

MeDIP enrichment values of the CVS and WBF samples were calculated for each region using the following equation:

$\mathbf{Enrichment} = \mathbf{e}^{\mathbf{\Delta Ct}}$ where e corresponds to the efficiency obtained in each real-time PCR reaction $e = 10^{(-1/\text{slope of STD curve})}$) and ΔCt indicates the cycle difference between input DNA and MeDIP DNA [$Ct_{(IN)} - Ct_{(IP)}$].

Table 2 Characteristics of the regions tested

Chromosomal region	Position (hg18)	Location type	Gene involved
Nn	chr21:31426757-31427146	Intragenic	TIAM1
H	**chr21:32268787-32269137**	**Intragenic**	**HUNK**
C	**chr21:33320530-33320815**	**CpG Island**	**OLIG2**
On	chr21:34492714-34493203	Intergenic	
J	**chr21:37841231-37841506**	**Intergenic**	
A	**chr21:39279691-39279971**	**Intergenic**	
Fd	chr21:42005961-42006216	Intragenic	C21orf129
M27	chr21:42178808-42179008	Intragenic	C2CD2
D	**chr21:42189235-42189849**	**LINE-L1**	**C2CD2**
EI	**chr21:42355366-42355908**	**Intergenic**	
EII	**chr21:42357141-42357401**	**Intergenic**	
Id	chr21:42753677-42754026	Intergenic	
M1E	chr21:44953640-44953854	Intragenic	TSPEAR
M28	chr21:45171015-45171225	Intragenic	ITGB2
AI	chr18:55086179-55086755	Intragenic	RAX NM-013435
AII	chr18:55090284-55090605	Intragenic	RAX NM-013435
B	chr18:44165984-44166275	Intergenic	
C	chr18:58955844-58956604	Intragenic	BCL2NM-000633
M18	chr21:15331818-15331945	Intragenic	NR1P1
M20	chr21:15178413-15178497	Intergenic	
M25	chr21:37692864-37692974	Intergenic	DYRK1A

Regions in bold indicate previously validated regions [13,14].

Table 3 Primer sequences on DMRs tested

Primer	Forward	Reverse	Position
CHR21(M27)	ATACGTGTCCTGCCTTCCAC	GCTTTGAGCAGAGAGGGAAA	42178812-42178948
CHR21(M28)	CCCAGAAATTCCATTTGCAG	GAAAGGCTCAACCAACCAAC	45171107-45171192
CHR21(M1E)	TCGCACTGAGGCTTCCTACT	AAGTTGTGGGCTGGGATTTT	44953674-44953772
CHR21(Nn2)	ACCATTGTGGATCACAGCAG	GCTCCGAGGATTAGGGAAAG	31427008-31427139
CHR21(On2)	CTCCTGACCCACTCCCAATA	GGAAACTCAGGGTCAAACGA	34492982-34493090
CHR21(Fd1)	ATGTTGCCTGGGATATGCTT	AACTGGCTGCGTGAGGATA	42006045-42006153
CHR21(EI-3)	GCCTTGGGACAAAAATGACA	TGGGCACAGCCCTAACTAAC	42355352-42355484
CHR21(EI-4)	GGCCAGGTTGTTTCAGATTG	TTCCGGCAGAGTTTATTTGG	42355802-42355908
CHR21(Id1)	ACCGTATCATTTCCCCAGGT	TGACCACATTTCCACCACAG	42753720-42753866
CHR21(A5)	**GCTGGACCAGAAAGTGTTGAG**	**GTGTGCTGCTTTGCAATGTG**	**39279856-39280004**
CHR21(C5)	**CTGTTGCATGAGAGCAGAGG**	**CGTCCCCCTCGCTACTATCT**	**33320735-33320829**
CHR21(D2)	**TGCAGGATATTTGGCAAGGT**	**CTGTGCCGGTAGAAATGGTT**	**42189557-42189683**
CHR21(EI-2)	**TGAATCAGTTCACCGACAGC**	**GAAACAACCTGGCCATTCTC**	**42355712-42355815**
CHR21(EII-1)	**CCGTTATATGGATGCCTTGG**	**AAACTGTTGGGCTGAACTGC**	**42357215-42357341**
CHR21(H2)	**CCACATCCTGGCCATCTACT**	**TTCCACAGACAGCAGAGACG**	**32268843-32268943**
CHR21(M18)	GATGGATGGCCTTTTGGTAA	TATTTGGTTTGCCCCTTCCT	15331818-5331945
CHR21(M20)	CATTAGCGGGTCAGCTAGGA	TGGCAATTACATCTGCCATTA	15178413-5178497
CHR21(M25)	TTGTCTGCCCGTATGGAAGT	ATGGTTGTAGGGCTCATTCA	37692864-37692974
CHR21(J2)	**ATTCTCCACAGGGCAATGAG**	**TTATGTGGCCTTTCCTCCTG**	**37841284-37841411**
CHR18(AII2)	TGTGCCTCTCCCTTGAGACT	AAATTGCAGCCAATGCTTCT	55090427-55090524
CHR18(B3)	TGTGGTTTCAAACATGCACA	CTGAAAAGGCCACTCTGAGG	44166131-44166263
CHR18(C1)	GTGAGAGAGAACGCCAGGAG	TGAGCCAACTCTGGTGTCAG	58956266-58956391
HYPER	CAGGAAAGTGAAGGGAGCTG	CAAAACCCAATGGTCAATCC	19991387-19991465
HYPO	AGGTGCCCAATTCAAGGTA	CTTCCCCACCAGTCTTGAAA	30214952-30215055

Regions in bold indicate previously validated regions [13,14].

The mean enrichment values of each DMR were compared between WBF and CVS samples using the Mann-Whitney U tests [27] and the corresponding p-values were used to decide whether there was significant evidence to claim that the mean enrichments of the two groups were different.

Hierarchical clustering of the DMRs was conducted using an iterative algorithm that joins similar clusters based on the set of dissimilarities of the 100 individuals (calculating the Euclidean distanced between clusters) and re-computing their distances at each stage by the Lance-Williams dissimilarity update formula [28].

Additional file

Additional file 1: Initial screening on six WBF and six CVS for the selection of new DMRs.

Competing interests

GK and EAP are currently employed by, and own shares of NIPD Genetics LTD which is developing a new non-invasive prenatal diagnosis for trisomy 21 based on the MeDIP-qPCR methodology. MI, AK, MH and CS declare that they have no competing interests.

At the time that the study was carried out ET and CS were employed by NIPD Genetics LTD. Currently, they have no affiliations with NIPD Genetics LTD and they declare no competing interests.
EAP and PCP have filed a PCT patent application for the MeDIP real time qPCR based NIPD approach (PCT Patent Application No. PCT/1B2011/000217).

Authors' contributions

MI assisted in the design of the study, collected samples, carried out all the experiments and drafted the manuscript. EAP assisted in the design of the study. AK participated in the Methylation DNA Immunoprecipitation and qPCR assays. ET assisted in the biomarker discovery. CS carried out the statistical analysis. CS assisted in the design of the study. GK and MH assisted in the design of the study and helped to edit the manuscript. PCP conceived and coordinated the study. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Andrea Chrysostomou for her assistance in MeDIP and qPCR assays. This research was partially funded by the 7th Framework Programme as part of the ANGELAB project (http://cordis.europa.eu/projects/rcn/105552_en.html, project number 317635).

Author details

[1]The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.
[2]Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus.
[3]NIPD Genetics Ltd, Nicosia, Cyprus.

Received: 22 September 2014 Accepted: 12 October 2014
Published online: 01 November 2014

References

1. Bird A: **DNA methylation patterns and epigenetic memory.** *Genes Dev* 2002, **16:**6–21.
2. Gopalakrishnan S, Van Emburgh BO, Robertson KD: **DNA methylation in development and human disease.** *Mutat Res* 2008, **647:**30–38.
3. Baylin SB, Jones PA: **A decade of exploring the cancer epigenome—biological and translational implications.** *Nat Rev Cancer* 2011, **11:**726–734.
4. Laird PW: **Principles and challenges of genomewide DNA methylation analysis.** *Nat Rev Cancer* 2010, **11:**191–203.
5. You JS, Jones PA: **Cancer genetics and epigenetics: two sides of the same coin?** *Cancer Cell* 2012, **22:**9–20.
6. Jakovcevski M, Akbarian S: **Epigenetic mechanisms in neurological disease.** *Nat Med* 2012, **18**(8):1194–1204.
7. Rawson JB, Bapat B: **Epigenetic biomarkers in colorectal cancer diagnostics.** *Expert Rev Mol Diagn* 2012, **12:**499–509.
8. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS: **Presence of fetal DNA in maternal plasma and serum.** *Lancet* 1997, **350:**485–487.
9. Della Ragione F, Mastrovito P, Campanile C, Conti A, Papageorgiou EA, Hulten MA, Patsalis PC, Carter NP, D'Esposito M: **Differential DNA methylation as a tool for noninvasive prenatal diagnosis (NIPD) of X chromosome aneuploidies.** *J Mol Diagn* 2010, **12:**797–807.
10. Old RW, Crea F, Puszyk W, Hulten MA: **Candidate epigenetic biomarkers for non-invasive prenatal diagnosis of Down syndrome.** *Reprod Biomed Online* 2007, **15:**227–235.
11. Tong YK, Chiu RW, Akolekar R, Leung TY, Lau TK, Nicolaides KH, Lo YM: **Epigenetic-genetic chromosome dosage approach for fetal trisomy 21 detection using an autosomal genetic reference marker.** *PLoS One* 2010, **5:**e15244.
12. Papageorgiou EA, Fiegler H, Rakyan V, Beck S, Hulten M, Lamnissou K, Carter NP, Patsalis PC: **Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies.** *Am J Pathol* 2009, **174:**1609–1618.
13. Papageorgiou EA, Karagrigoriou A, Tsaliki E, Velissariou V, Carter NP, Patsalis PC: **Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21.** *Nat Med* 2011, **17:**510–513.
14. Tsaliki E, Papageorgiou EA, Spyrou C, Koumbaris G, Kypri E, Kyriakou S, Sotiriou C, Touvana E, Keravnou A, Karagrigoriou A, Lamnissou K, Velissariou V, Patsalis PC: **MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21.** *Prenat Diagn* 2012, **32:**996–1001.
15. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJ, Carter NP, Tavare S, Beck S: **An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs).** *Genome Res* 2008, **18**(9):1518–1529.
16. Butcher LM, Beck S: **AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay.** *Methods* 2010, **52:**223–231.
17. Feinberg AP, Irizarry RA: **Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease.** *Proc Natl Acad Sci U S A* 2010, **1**(Suppl 107):1757–1764.
18. El-Maarri O, Walier M, Behne F, van Uum J, Singer H, Diaz-Lacava A, Nusgen N, Niemann B, Watzka M, Reinsberg J, van der Ven H, Wienker T, Stoffel-Wagner B, Schwaab R, Oldenburg J: **Methylation at global LINE-1 repeats in human blood are affected by gender but not by age or natural hormone cycles.** *PLoS One* 2011, **6:**e16252.
19. Choi SH, Worswick S, Byun HM, Shear T, Soussa JC, Wolff EM, Douer D, Garcia-Manero G, Liang G, Yang AS: **Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer.** *Int J Cancer* 2009, **125:**723–729.
20. Fryer AA, Emes RD, Ismail KM, Haworth KE, Mein C, Carroll WD, Farrell WE: **Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans.** *Epigenetics* 2011, **6:**86–94.
21. Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J: **A longitudinal study of epigenetic variation in twins.** *Epigenetics* 2010, **5:**516–526.
22. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS: **Factors underlying variable DNA methylation in a human community cohort.** *Proc Natl Acad Sci U S A* 2012, **2**(Suppl 109):17253–17260.
23. Schneider E, Pliushch G, El Hajj N, Galetzka D, Puhl A, Schorsch M, Frauenknecht K, Riepert T, Tresch A, Muller AM, Coerdt W, Zechner U, Haaf T: **Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns.** *Nucleic Acids Res* 2010, **38:**3880–3890.
24. Bock C, Walter J, Paulsen M, Lengauer T: **Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping.** *Nucleic Acids Res* 2008, **36:**e55.
25. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW: **The Database of Genomic Variants: a curated collection of structural variation in the human genome.** *Nucleic Acids Re* 2014, **42:**D986–D992.
26. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132:**365–386.
27. Mann HB, Whitney DR: **On a test of whether one of two random variables is stochastically larger than the other.** *The annals of mathematical statistics* 1947, **18**(1):50–60.
28. Lance GN, Williams WT: **A general theory of classificatory sorting strategies II. Clustering systems.** *The computer journal* 1967, **10:**271–277.

doi:10.1186/s13039-014-0073-8
Cite this article as: Ioannides *et al.*: Inter-individual methylation variability in differentially methylated regions between maternal whole blood and first trimester CVS. *Molecular Cytogenetics* 2014 **7**:73.