**THESIS**

# *EPIGENETIC REGULATION OF CANCER-TESTIS GENE EXPRESSION*

**A THESIS SUBMITTED TO**
**THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS**
**AND THE INSTITUTE OF ENGINEERING AND SCIENCE OF**
**BİLKENT UNIVERSITY**
**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR**
**THE DEGREE OF MASTER OF SCIENCE**

**BY**
**AYDAN BULUT**
**AUGUST 2008**

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Dr. Uygar Tazebay

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Dr. Stefan Fuss

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Dr. Ali O. Güre

Approved for the Institute of Engineering and Science

Director of Institute of Engineering and Science
Prof. Mehmet Baray

# ABSTRACT


**EPIGENETIC REGULATION OF CANCER-TESTIS GENE EXPRESSION**

Aydan Bulut
M.S. in Molecular Biology and Genetics
Supervisor: Dr. Ali O. Güre
August 2008, 108 Pages

Cancer-testis (CT) genes are composed of mostly X-linked gene families that are widely expressed in cancer, in a coordinate manner, with ideally no expression in normal tissues except spermatogonia, oogonia and trophoblasts. Exact mechanisms reactivating their expression during carcinogenesis are not known yet. Epigenetic factors emerge as key controllers of CT expression. Selective DNA demethylation of the promoter regions of CT genes has been demonstrated to correlate closely with CT expression in all examples studied so far. Tumor-suppressor genes (TSGs), on the other hand, are known to be frequently down-regulated in cancer by hypermethylation of DNA. In order to elucidate mechanisms that could help explain tumor-specific CT gene up-regulation, we aimed to understand how these seemingly opposite effects could exist in close proximity. For this purpose, we identified eight CT-proximal X-linked putative TSGs (pTSGs) down-regulated in tumors, by screening the SAGE and EST libraries of the Cancer Genome Anatomy Project. By conventional and real-time RT-PCR, we verified that two such genes, ALAS2 and CDR1, were significantly down-regulated in almost all cancers tested, while three others were down-regulated at least in some cancers, and by bisulfite sequencing we demonstrated that the promoter DNA of these pTSGs were hyper-methylated in correlation with their expression levels.

Our search for the presence of insulators between CT and TSG genes did not yield a consensus site. However, we hypothesized that the dynamic organization of CT genes into inverted repeats (identified by the Inverted Repeats Finder program) throughout the X chromosome could be a candidate regulator of CT expression. By using the chromosome conformation capture (3C) assay, we have shown the alteration in higher-order chromosomal structure of the CT gene NY-ESO-1-bearing inverted repeat region in SK-LC-17 as well as in 5-Aza-2'-deoxycytidine-treated HT29 cell lines, correlating with CT as well as ncRNA expression from within the repeat. As it is known that many CT genes are embedded in inverted repeats, our results suggest a general mechanism regarding epigenetic regulation of CT gene expression.

# ÖZET

## KANSER-TESTİS GEN İFADESİNİN EPİGENETİK DENETLENMESİ

Aydan Bulut
Moleküler Biyoloji ve Genetik Yüksek Lisansı
Tez Yöneticisi: Dr. Ali O. Güre
Ağustos 2008, 108 Sayfa

Kanser-testis (KT) genleri, çoğunlukla X kromozomu üzerinde yer alan, normal dokular içerisinde ideal olarak sadece spermatogonia, oogonia ve trofoblastta ve fakat birçok kanser türünde de, koordine olarak, ifade edilen gen ailelerinden oluşurlar. Hangi mekanizmaların kanserde tekrar çalışmalarına yol açtığı büyük ilgi odağı olduğu halde halen tam olarak bilinmemektedir. Epigenetik etmenler bu bağlamda ana denetleyiciler olarak ortaya çıkmaktadır. Promotor bölgesi seçici DNA demetilasyonunun KT ifadesiyle sıkıca bağlantılı olduğu çalışılan tüm örneklerde gösterilmiştir. Diğer yandan, tümör baskılayıcı genlerin (TBG) ise DNA'nın aşırı-metillenmesi sonucu kanserlerde ifade azalmasına uğradıkları bilinmektedir. Kanser-testis genlerinin tümör-özgün ifade artışına yol açan mekanizmaları izah edebilmek amacıyla, öncelikle bu iki ters işleyişin birbirine yakın olarak nasıl var olabildiğini anlamaya çalıştık. Bu amaçla Kanser Genomu Anatomi Projesi'nin SAGE ve EST kütüphanelerini tarayarak sekiz adet KT-komşusu X-bağlantılı varsayımsal tümör baskılayıcı gen belirledik. Geleneksel ve gerçek zamanlı PZR ile bu tarz iki genin, ALAS2 ve CDR1'in, ifadelerinin test edilen hemen bütün kanserlerde önemli ölçüde azaldığını ve başka üç genin ifadelerinin ise en az birkaç kanser türünde azaldığını tespit ettik. Bisülfit dizileme tekniğini kullanarak, bu varsayımsal TBGlerin promotor DNAlarının aşırı-metillenmesinin, genlerin ifade düzeyleriyle orantılı olduğunu da gösterdik.

KT ve TBG genleri arasında yalıtkan elemanların olup olmadığını araştırdık, bu çalışmanın sonucunda konsensus bir yalıtım bölgesi olmadığını belirledik. Ne var ki, KT genlerinin X kromozomu boyunca, tersine tekrarlı DNA bölgeleri içinde yer almasını; kromozom yapısının KT ifadesi üzerinde denetleyici olabileceği olasılığı olarak önerdik. Biyokimyasal 3C yöntemini uygulayarak, NY-ESO-1 içeren tekrarlı DNA bölgesinin kromozom yapısının SK-LC-17 ve 5-Aza-2'-deoksisitidin ile muamele edilmiş HT29 hücre hatlarında değiştiğini gösterdik. Bu değişim aynı zamanda protein kodlamayan-RNA ve KT ifade seviyelerinin değişimiyle bağdaşıktır. Birçok KT geninin tersine tekrarlı DNA bölgeleri içerisinde yer aldığı bilindiğinden dolayı, sonuçlarımız genel bir KT epigenetik denetleme mekanizmasını ileri sürmektedir.

*ana, baba ve sevgiliye…*

# ACKNOWLEDGEMENTS

It’s my responsibility to thank my supervisor Dr. Ali O. Güre for his always-accurate guidance, his patience and very good temper. He has been a good model for me. I would like to thank my group-mates also; Rasim, Duygu, Şükrü, Esen and Derya for their help and friendship. It was my pleasure and honor to be part of the Bilkent MBG family.

My family and my dear darling always supported and encouraged me. I can never compensate for; nevertheless my deepest thanks are to them.

I was supported by TÜBİTAK-BİDEB scholarship # 2228 throughout my M.S. study. This work was also supported by grants from TÜBİTAK and European Commission to Dr. Ali O. Güre.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABBREVIATIONS

| | |
|---|---|
| AZA | 5-Aza-2'-deoxycytidine |
| bp | Base Pair |
| CT | Cancer-Testis |
| $C_t$ | Threshold Cycle |
| DNMT | DNA Methyltransferase |
| E | Efficiency |
| EST | Expressed Sequence Tags |
| HAT | Histone Acetyltraferase |
| HDAC | Histone Deacetylase |
| IRF | Inverted Repeats Finder |
| m | Number of Technical Replicates |
| *m* | Slope |
| NF | Normalization Factor |
| Pc | Polycomb |
| SAGE | Serial Analysis of Gene Expression |
| SD | Standard Deviation |
| SE | Standard Error |
| Trx | Trithorax |
| TSG | Tumor-Suppressor Gene |

# 1 INTRODUCTION

## 1.1 Cancer-Testis (CT) Genes and Antigens

Cancer-testis antigens constitute a group of mostly X-linked testicular antigens, the familial and immunogenic properties of which were realized by mid-90's, and new members were identified rapidly since then. CTs are immunogenic and have vaccination potential because they are widely expressed in cancers with no expression in normal tissues except testis; and cytotoxic T cells can recognize the CT antigens. The initial discovery of the first CT-antigens MAGEA1, BAGE and GAGE1 between 1991-95 was through T cell epitope cloning, thus they were found as cytotoxic T lymphocyte-recognized antigens [1-3]. SSX2 and NY-ESO-1 were discovered after 1995 by another technique, SEREX (serological expression cloning), developed by Pfreundschuh et al [4-7]. From then on, the application of SEREX technique led to identification of many additional CT-antigens, the number of which was further increased by high-throughput transcript analyses [8, 9]. Today 82 CT gene families with 263 genes can be found in literature [10]; although only a subgroup have testis-restricted expression and are immunogenic. An up-to-date list of CT genes is given in **Appendix A**.

Individual CT antigens proposed to be used as tumor biomarkers are MAGE A1/A3 for hepatocellular carcinoma and lung adenocarcinoma, BORIS for breast cancer, NY-ESO-1 for lung adenocarcinoma, etc [11-13]. However, CT antigens are more likely to be candidate biomarkers when assessed in combination since their expression is thought to be triggered by the same mechanism [11]. In addition, some CT antigens bear the potential to be used for early diagnosis of cancer, demonstrated by few studies [13, 14].

### 1.1.1 Structure

Cancer-testis genes can be divided into two groups according to chromosomal location. Most CTs are on the X chromosome and are referred to as CT-X genes [15]. Members of CT-X group exist mostly as multi-copy gene families, some also containing splice variants. They reside in tandem or inverted repeats and consist approximately 10% of the genes on the X chromosome. Duplications of the unstable repeat regions are thought to be the primary mechanism increasing CT copy number on chromosome X [10].

CT genes located on autosomal chromosomes are mostly single-copy genes that are not associated with segmental duplications (yet).

Regarding the similarity of the CT genes, different families have slight homology with each other (e.g. NY-ESO-1 and LAGE1, CAGE-HAGE and XAGE-GAGE-PAGE homologies) while members of the same family generally have sequence similarity of >90% [16].

The best-studied of all CT genes is the MAGE superfamily. Individual families within the MAGE superfamily are related to each other, such that they all have the MHD (MAGE homology domain) that encodes a ~200 residue stretch of conserved protein domain [17, 18]. The rest of MAGE proteins are nearly completely different in each family. According to sequence analysis, the MAGE superfamily involves 3 acidic groups, MAGE A, B and C; and a basic group, MAGE D [19].

None of other CTs but CAGE gene bearing a DEAD box exhibit structural domains, in correlation with their limited functionality and evolutionary history [20]. These aspects are reviewed in sections 1.1.2 and 1.1.3.

### 1.1.2 Function

Many of the CT antigens do not have characterized biological functions. Early-identified major CT families fall into this group with the major exception being the MAGE superfamily. Some MAGE orthologs have defined functions. Mouse MAGE

B4 plays role in germ cell development; mouse necdin binds to E2F1 and negatively regulates G1 to S progression [19]. In a similar fashion, Drosophila necdin controls neural precursor proliferation in postembryonic neurogenesis [21]. Similarly MAGE A4 binds to the Gankyrin or p28 protein and inhibits the adhesion-independent growth of Gankyrin-overexpressed cells [19]. Recently MAGE-A3/6 was identified as a novel target of fibroblast growth factor 2-IIIb (FGFR2-IIIb) signaling in thyroid cancer cells, such that FGF7/FGFR2-IIIb activation resulted in H3 methylation and deacetylation of the MAGE-A3/6 promoter, to down-regulate gene expression [22].

Within those CTs that are known to have a function, a subgroup consists of meiosis-related proteins, like SCP1 (CT8) which is a synaptonemal complex protein involved in chromosome reduction in meiosis, and OY-TES-1, which encodes the proacrosin binding protein sp32 precursor [16]. Both SCP1 and OY-TES-1 are non-X CT antigens.

Within other functional CT genes are BORIS and CAGE. BORIS (brother of the regulator of imprinted sites) is the paralog of the abundant transcription factor CTCF, and seems to play role in CT-regulation according to the studies showing BORIS and CTCF binding to NY-ESO-1 and MAGE A1 promoters, resulting in derepression of both antigens [23-25]. CAGE, on the other hand, is a DEAD box helicase protein, expression of which is sensitive to the cell cycle. CAGE was thus suggested to play a proliferative role [20]. SSX antigen is still a candidate to be placed among functional CTs with a possible role in stem-cell migration and cancer cell metastasis, since it overlaps in cytoplasm with matrix metalloproteinase 2 (MMP2) and its down-regulation impairs cell migration with a reduction in MMP2 levels [26].

An interesting finding that may shed light to the functionality of CTs was published in 2006, declaring the first direct physical interaction between two CT antigens, the homology domain of MAGE C1 and NY-ESO-1 [27]. Although it is not yet known whether this interaction is associated with the function of either of the two antigens, or rather an artifact resulting from putatively same CT-regulatory

mechanism (like a non-functional weak interaction resulting from the abundance of CT antigens within the cell in cases of coordinate hypomethylation); yet it would be beneficial to examine the possible interactions between other CT antigens.

### 1.1.3 Conservation

CT genes are almost exclusively specific to primates, with few exceptions. Few members of the MAGE superfamily are found in mouse genome, the mostly conversed one being necdin because of its functionality. Indeed, MAGE genes have expanded in both rodent and primate lineages, but independently [17]. In general, human CT genes have orthologs in primates, especially in great apes. CT genes in the chimpanzee genome are highly similar to human orthologs and generally located on the same chromosomes [10]. Members of the GAGE family are tandemly arranged on the X chromosome only in human, chimpanzee and macaque genomes. Phylogenetic analyses reveal that the GAGE family began to duplicate after the split of human and chimpanzee [28].

As a recent estimate of the future evolution of CT genes, the divergence rate between human and chimpanzee CT orthologs was analyzed, which shows that CT-X genes are evolving faster than non-X CT genes [10].

The absence of CT genes in genomes of model organisms is a challenge for CT research, since generation of knock-out strains and other manipulations are limited. Only the necdin knock-out was generated in mouse [19]. As a result, CT research is based on patient samples and germ cells.

### 1.1.4 Expression (and Acting Epigenetic Mechanisms)

Most cancer-testis genes are expressed in none of the normal adult tissues but testis, and then, only at the spermatogonial stage. However, recent research suggest they may be among genes escaping MSCI (meiotic sex chromosome inactivation) thus being expressed in spermatids also [29]. Based on their expression profiles in a panel of normal tissues, CT genes are classified as testis-

restricted (1), tissue-restricted (expressed in ≤2 of 13 non-gametogenic tissues tested) (2), differentially expressed (expressed in 3-6 non-gametogenic tissues, among 13 tested) (3) and ubiquitously expressed (4) [30]. The initially-discovered, major CT gene families fall in the testis- and tissue-restricted groups. Some genes referred to as 'CT' are thus controversially classified if the testis-specific expression rule is strictly applied. CT genes are also expressed, in a coordinate manner, in a wide-variety of cancers including lung cancer (especially non small-cell lung cancer), hepatocellular carcinoma, cancers of prostate, ovary, esophagus, hematological malignancies, multiple myeloma, head and neck squamous carcinoma, lymphoma, multiple myeloma, colorectal and gastric cancers and Hodgkin's sarcoma; and are generally correlated with poor prognosis [11, 16]. What reactivates their expression during carcinogenesis is of great interest, however the exact mechanisms are not known yet. Genetic mutations are not responsible for this task, since these genes are expressed in germline with no known mutation. Although some CT family members are frequently involved in translocations, like SSX2 within t(X:18) region [4, 31], this seems to be the exception rather than the rule. Epigenetic factors therefore emerge as key controllers of CT expression. DNA demethylation associated with CT gene expression is the most frequently reported mechanism shown for many cancer types [12, 32]. Expectedly, treatment with the DNMT inhibitor, 5-AZA-2'-deoxycytidine, induces CT gene expression [33]. 'Genome-wide hypomethylation' is a phenomenon believed to associate or possibly cause CT gene expression; however promoter-dependent selective demethylation (restricted to the 5' region) was shown to be the case for MAGE-A1 [34]. Histone modifications, namely (de-)acetylation, are known to affect CT expression, shown both *in vivo* and pharmaceutically by trichostatin A (HDAC inhibitor) treatment [32]. Acetylation-induced transcription was shown to be required for DNA demethylation of CTs, at least for the GAGE family; and since CT genes are *coordinately expressed*, this may be part of the real mechanism resulting in gene expression [35].

Since CT-X genes are organized in *repeated regions* on the X chromosome, we hypothesized that at some point before or after histone acetylation the higher-order chromosomal structure (and may be the nuclear localization) should alter leading to formation of permissive structures for CT gene expression. We further hypothesized that this could resulting in, or happen as a consequence of the generation of non-coding transcripts. Transcription factors known to have a role in CT expression, like Sp1, CTCF, BORIS, and others, should also have a role within this scenery before initiation of transcription. Adequate understanding of the CT-activation mechanisms is required for therapeutic purposes, since inhibition of CT-derepression might help us proceed in our combat with cancer and other complex diseases.

## 1.2 Epigenetic Regulation of Transcription

Whatever exceeds beyond the borders of genetics falls in the fields of epigenetics and RNA biology. Investigation of epigenetic statuses and mechanisms assisted us understand how cell fates are different, how the zygote is not just a product of the egg and sperm, why identical twins raised together are not really identical? General knowledge regarding epigenetic mechanisms, namely, DNA methylation, histone modifications, and chromatin remodeling, are summarize in the following sections; how these mechanisms are altered in cancer, germ, and stem cells is also summarized.

### 1.2.1 Methylation/Demethylation of DNA

In mammals, methylation of DNA appears almost exclusively in the form of cytosine methylation (on the 5' end) of CpG dinucleotides. Other types of DNA methylation include CNG and CNN residues. Although not studied as widely as CpG methylation, recent research suggests that non-CpG methylation might play important roles in the mammalian epigenome as well [36]. DNA methylation exerts its effect by suppressing transcription and is an indispensible process for mammalian development, and the post-developmental maintenance of the

healthy state. Genomic imprinting, X chromosome inactivation, gene- and tissue-specific expression are some basic phenomena that directly include CpG methylation. Bearing in mind that 50 - 70% of all human genes contain CpG islands might clarify the abundant role of DNA methytransferases. Methylation statuses are very dynamic, yet heritable upon cell division enabling long-term memory of transcriptional levels [36, 37].

Three enzymes act as CpG methyltransferases in mammals: Dnmt1, Dnmt3a and Dnmt3b [38, 39]. Dnmt1 is the maintenance methyltransferase which acts on hemi-methylated DNA during replication and is essential for X inactivation and imprinting [40]. Dnmt3a and Dnmt3b are *de novo* methyltransferases playing both collaborative and unique roles mostly in development [41]. Dnmts also act as co-repressors in protein complexes.

How reversal of DNA methylation is achieved is a controversial issue. In the classic 'replication-coupled demethylation' point of view, DNA methylation is reduced progressively upon each replication [42]. Inhibition of Dnmt1 activity by nucleoprotein complexes or histone modifications associated with transcription (e.g. acetylation) is required [42]. A more novel aspect is the existence of active DNA demethylases, either removing the methyl group from the cytosine residue, or removing the 5'-methylcytosine itself by glycosylase activity which requires endonuclease activity afterwards for repairing the missing base [42]. The former pathway is mediated by a protein that belongs to the family of MBD (methyl CpG-binding domain) proteins *in vitro* but it is not yet isolated *in vivo* [42].

Direct analysis of DNA methylation is possible through sodium bisulfite treatment followed by sequencing or restriction digestion (COBRA assay: COmbined Bisulfite Restriction Analysis) [43]. Alternatively, methylation-sensitive restriction endonucleases may be employed and afterwards the product is amplified by PCR reaction. One drawback of restriction-mediated assays is their limitedness to restriction sites. Genome-wide maps of methylation states can be generated through microarray hybridization or high-throughput sequencing of bisulfite treated DNA samples. Lastly, treating cells with 5-AZA-2'-deoxycytidine, a cytosine

analog which cannot be methylated, and investigation of transcript levels would provide an indirect analysis of DNA methylation [44].

### 1.2.2 Histone Modifications

Human DNA of approximately 2m length is large enough to exceed the borders of a cell nucleus, which is the major reason why higher-order chromatin structures exist. Nucleosome is the basic structural unit that consists of four core histones – H2A, H2B, H3 and H4 – around which 147 bp DNA is wrapped. Linker histone H1 is employed between two nucleosomes and brings them closer to generate a more compact structure named the '30nm fiber'. Further compactness is enabled by wrapping this fiber further utilizing protein scaffolds, finally generating chromosomes. The core histones play additional roles other than being a scaffold for wrapping DNA: the N-terminal and C-terminal tails bear more than 60 residues that can be covalently modified. Known histone modifications include lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, lysine ubiquitination, lysine sumoylation and proline isomerization [45]. Methylation is more complex than other modifications since there are mono-, di- and trimethyl states for lysines and mono- and dimethyl (asymmetric and symmetric) for arginine [45]. The enormous number of combinations that the histone modifications might exist in, dectates a 'histone code' that regulates accessibility of DNA in first hand [46].

The putative 'histone code' implies specific and heritable combination of histone modifications for specific functional roles. This has 3 consequences; firstly, organization of DNA into global chromatin environments - euchromatin and heterochromatin - that generates the banding pattern of chromosomes. The borders of chromosome bands have shown to be separated by boundary elements flanked by methylated H3K4 and H3K9 [47]. Heterochromatic states of the inactive X chromosome and pericentric heterochromatin are maintained by recruitment of specific proteins to methylated H3K27 and H3K9, respectively. Second consequence is the short-term regulation of biological processes like transcription, replication and repair through recruitment of chromatin remodeling complexes.

Although which combinations of modifications directly imply activation or repression is not still clear, H3K4me1, 2 and 3, H3K27me1, H3K9me1, H3K79me1, H4K20me1 and all acetylations are generally associated with transcription whereas H3K9me3, K3K27me3 and H3K79me3 are generally connected to transcriptional repression [47, 48]. A third effect is the long-term memory of transcriptional state, which plays putative roles in development and differentiation [46].

Almost each of the histone modifications are exerted by different enzymes. In general, histone acetyltransferases (HATs) and deacetylases (HDACs) carry out (de-)acetylation; histone methyltransferases (HMTs) - specifically lysine and arginine methyltransferases- and lysine demethylases carry out (de-)methylation; and serine/threonine kinases carry out phosphorylation. No arginine demethylases have been identified to date [45]. An up-do-date list of histone-modifying enzymes is given in **Appendix B**.

Apart from the core histones and the linker histone H1, there are other histone proteins named 'histone variants', which are employed in more specific processes. Such histone-variant genes have introns and are often polyadenylated unlike major histone RNAs. H2A and H3 variants are the widely studied ones. H2A.X plays role in maintaining genome integrity, it acts in double-strand break DNA repair, apoptosis, [V(D)J] recombination and replication [49]. H2A.Z was shown to be indispensible for survival however its exact role is yet to be determined. There are controversial data suggesting a role for H2A.Z in both transcriptional activation and repression, but it is certainly enriched in boundary elements between euchromatic and heterochromatic regions [50]. Another H2A variant, macroH2A, is thought to be involved in transcriptional repression and replaces H2A on the inactive X chromosome. The histone H3 variant H3.3 varies from H3 by only 4 amino acids, however it carries out remarkably different functions. It is enriched in some tissues during development. It is associated with actively transcribed regions of the genome, replacing H3 during transcriptional elongation, and constitutively expressed during the cell cycle accordingly. CENP-A, another H3 variant, is also essential for survival and localizes to centromeric heterochromatin [49]. Apart

from these, there are also testis-specific histone variants that may even play role in cancer-testis gene expression. Three models - replication-coupled histone deposition, transcription-coupled histone deposition and histone exchange (by certain factors) - have been proposed to explain how variant histones incorporate into the chromatin, though none has been proven [50].

Histone modifications are usually analyzed by chromatin immunoprecipitation (ChIP) using antibodies specifically recognizing modified residues. The generation of histone maps of human, mouse and yeast have been successfully performed by high throughput assays based on hybridizing ChIP samples on arrays (ChIP on chip) or by massively parallel signature sequencing of ChIP samples (ChIP-Seq) [47, 48]. Alternatively, histone modifications might be analyzed using mass spectrometry.

### 1.2.3 Polycomb/Trithorax Group Proteins and Chromatin Remodeling

Transcriptional regulation through modifications of DNA and histones discussed above seem to be the final steps in the regulation of a gene. To understand which mechanisms precede histone and DNA modifications, one needs to take into account polycomb group (PcG) and trithorax group (trxG) protein complexes, and other chromatin remodeling complexes. For the reasons explained below, these complexes are candidate major players determining the active/suppressed states of transcription, higher order chromosomal architecture, and even non-coding RNA transcripts.

PcG and trxG proteins were firstly discovered in *D.melanogaster* as activators or repressors, respectively, of Homeobox (Hox) genes, which define cell identity along the anteroposterior axis. In this context, they are required to maintain the state of expression, not to initiate or cease. Most of the involved proteins have defined human orthologs, that regulate Hox genes as well as many others. PcG and trxG proteins play role in some basic phenomena like genomic imprinting, X inactivation, pluripotency and cell proliferation [51].

Polycomb repressor complexes (PRC) are divided into two groups based on their physical association in different multiprotein complexes. Class II PRCs are directly

involved in repression of transcription while class I PRCs are methyltransferase complexes that maintain the repressed state [52]. PRC2 is a class I PRC that has four core components: E(z), Esc, Su(z)12 and Nurf-55 (these are the Drosophila proteins). Enhancer of Zeste (E(z)) is a the SET1- domain containing histone methyltransferase that trimethylates H3K27 and H1K26, the former of which is recognized by the chromodomain of Polycomb (Pc), which Is a core component of PRC1 complex (class II inhibitory complex). PRC1 has four core components in stoichiometric amounts: Pc, Polyhomeotic (Ph), Posterior sex combs (Psc) and and dRING. A recently identified third PcG complex is PhoRC, which includes sequence-specific DNA binding protein Pho. The mechanism of PRC action at target sites is rather sophisticated and not fully understood yet. PcG proteins are recruited to target genes' polycomb response elements (PREs) via sequence-specific DNA-binding factors and/or by specific histone modifications (namely, H3K27me3). PREs are often several kilobases long and contain recognition sites for some protein motifs; however, they are proposed to have a bifunctional nature since trxG proteins also bind to trithorax response elements (TREs) overlapping PREs [51]. PcG recruitment might even involve components of the RNAi machinery, since recent data showed that while siRNA-mediated silencing of a gene promoter, AGO1 recruited EZH2 (human homolog of *Drosophila* E(z)) [53]. Most probably, binding of PcG and trxG complexes at target sequences is mediated by combinatorial signals from DNA and histone motifs, and requires other DNA-binding factors.

Once recruited, PRC complexes might suppress transcription via direct or indirect mechanisms. Direct inhibition of the transcriptional machinery is a choice, and supporting this notion, TAFIIs are stoichiometric components of PRC1. Generation of repressive chromatin environments by heterochromatinization is the indirect method, which is supported by *in vitro* data showing that PRC1 complex bound to chromatin does not allow remodeling by Swi/Snf complex as it does not allow transcription by RNA polymerase II [54, 55]. Long-range interactions of DNA and PcG complexes are putative further steps of regulation, either in the form of looping or PcG polymerization. These interactions might be required because of

large distances between PREs and promoters of target genes. *In vitro*, recombinant PcG complexes of Drosophila and mammals bound to an immobilized chromatin template can silence a second free chromatin *in trans* [56]. On the other hand, many PcG proteins can oligomerize or self-associate, which may be an important functional property for silencing [52].

Trithorax activatory complexes are divided into five groups: The ATP-dependent chromatin remodelers SWI/SNF and NURF (a member of the ISWI family), and SET-domain factors TAC1, Ash1 and MLL1-3 [51]. TrxG-mediated activation is less wellcharacterized than PcG-mediated silencing. Like PcG recruitment, trxG recruitment involves DNA and histone motifs (TREs and H3K4me3) and non-coding RNA. According to a recent study, non-coding transcripts from the *bxd* regulatory region of the *Drosophila Ubx* gene recruit the Ash1 protein to this region, inducing activation [57]. Current models of trxG-mediated activation suggest that the SET-domain factors induce H3K4 trimethylation which is recognized by ATP-dependent chromatin remodelers [51]. Remodeling of chromatin either by exchange of nucleosomal subunits, by sliding the nucleosomes, or by looping out of nucleosomal DNA provides permissive environment for transcription [58].

Apart from PcG and trxG proteins, there are other chromatin-modifier complexes that regulate transcription. Among these are the ATP-dependent chromatin remodelers similar to SNF: NuRD/Mi-2/CHD and INO80 families [58]. Such as the bromodomain of SWI/SNF proteins recognizing acetylated histone tails, NuRD/Mi-2/CHD family members have unique tandem chromodomains that recognize methylated histone tails. INO80 family members, on the other hand, are characterized by split ATPase domains. All ATP-dependent chromatin remodelers are highly conserved from yeast to human, indicating their essential functions in variety of cellular processes. They have both shared and unique roles. All are capable of both inducing and repressing transcription, depending on the local chromatin context. Unique processes involve roles of ISWI family members in chromatin assembly after DNA replication, SWI/SNF and INO80 family members in double strand-break repair, and NuRD/Mi-2/CHD members in maintenance of

pluripotency in ES cells [58]. Deregulation of these essential complexes therefore contributes to uncontrolled cell growth and complex diseases.

### 1.3 Epigenetics of Germ- and Stem-Cells

Stem cells and germ cells bear completely different epigenetic regulation of the genome when compared with somatic cells, which also are significantly different from each other. This is reasonable, since differentiation genes are switched off in these cells; gene expression should be adjusted to maintain the dedifferentiated state. How these choices are made and the mechanisms involved are of great interest both because the differentiation phenomenon is not completely understood yet and because of the therapeutic potential of stem-cell research. Recent knowledge regarding stem- and germ-cell epigenetics is therefore presented here.

There are two main classes of stem cells: embryonic and adult (or somatic) stem cells. Embryonic stem cells (ESCs) are pluripotent ('plural potency'), meaning that they can give rise to the three different cell lineages; ecto-, meso- and endo-derm, but not to extra-embryonic tissues. ESCs are established from the totipotent blastomeres within the inner cell mass of blastocyst. The zygote and early blastomeres are totipotent ('total potency'), they can give rise to the whole organism. The difference between totipotent and pluripotent states is vastly an issue of epigenetics. Zygotic division is termed cleavage and regulated by maternal factors that exist within the egg that contribute to zygotic transcription also. Among the maternal factors are pluripotency factors like Sox2 and Oct3/4 and many chromatin-modifying enzymes like PcG proteins, histone-modifiers and chromatin remodelers that help reactivate the quiescent genome. The zygote is thus intrinsically regulated and does not have self-renewal capacity. ESCs, on the other hand, operate their own expression, are responsive to external signals and are capable of self-renewal. Pluripotency only exists transiently in a developing embryo, until the gastrulation stage. The exception is the germ-cell lineage that expresses pluripotency factors after gastrulation [59, 60].

How pluripotency is maintained is a highly debatable issue; however transcription of pluripotency factors were shown to be a pre-requisite. ChIP-Chip experiments show that in human ES cells, 3 major factors, OCT3/4, SOX2 and NANOG, together bind to the promoter regions of 2260 genes, 1303 of which are actively transcribed and 959 of which are inactive genes, creating a complex gene network that can be regulated by these factors [59]. However, the criteria for choice of some genes instead of others and the resulting epigenetic modifications that lead to maintenance of pluripotent state have not been adequately delineated. Bivalent chromatin marks, meaning the chromosomal regions that carry both repressive and permissive histone modifications in adult tissues, are considered signs of pluripotency-related genes however this hypothesis has not been proved yet [47, 61].

Adult stem cells emerge when ESCs somehow differentiate and lose pluripotency so that they now can give rise to cells of a specific lineage. Different types of adult stem cells, mainly mesenchymal and hematopoietic, have self-renewal capacity but are restricted in differentiation, thus they are multi-potent [61]. Adult stem cells are found in minute amounts in bone marrow and stem-cell niches specific to some organs. Umbilical cord blood is a rich source of hematopoietic stem cell isolation [62].

Germ cells are the only cells that express pluripotency factors after gastrulation, as mentioned above. Since germ cells , when fused, are expected to give rise to a totipotent zygote at the end of their developmental program, their specialization includes repression of somatic differentiation followed by extensive erasure of the epigenetic modifications - like DNA and H3K9 methylation - in order to reprogram the whole genome. As a result, pluripotency factors like Sox2, Nanog, Stella and Nanos3 are reactivated. Although germ cells cannot give rise to diverse cell types, pluripotent cells can be obtained from them *in vitro [60]*.

All cells within the body of an individual are differentiated under one specific lineage, except from germ cells and adult stem cells. Epigenetic mechanisms rule repression of inappropriate developmental programs temporally and spatially,

while maintaining heritability of current phenotypic states. The maintenance of cell fate is highly fulfilled by PcG and TrxG proteins. Strikingly, the nucleus of a differentiated somatic cell that has lost cellular potency can be reprogrammed by the maternal factors in the cytoplasm of an unfertilized egg, but also by ES and EG cells [59].

Expression of cancer-testis genes in germ cells, especially in testis, results from the differentially regulated epigenome of these cells. Germ cells and dedifferentiated cancer cells, or 'cancer stem cells', absolutely share some of the epigenetic mechanisms and pluripotency factors, that make them both express CT genes; however these factors are yet to be fully discovered and characterized.

### 1.4 Epigenetics of Cancer

Cancer is a complex disease that needs to combine many aberrant operations within the cell to emerge. A cancer cell acquires more and more capabilities as it accumulates genetic mutations and misregulation in non-mutated genes. Genetic and epigenetic aberrations cooperate within the pre-cancerous cell to make it acquire six capabilities as 'hallmarks' of cancer, meaning general to most cancers, as listed by Hanahan and Weinberg: Self-sufficiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitness replicative potential, sustained angiogenesis, tissue invasion and metastasis [63]. The order of these events varies among different cancer types [63]. But the main questions are to understand the reason why and how a healthy cell quits the established intracellular and extracellular regulatory circuits and turns to uncontrolled cell growth. Cancer research yielded sound improvements in answering these questions, which are concerned by thousands of people all over the world. Tumor suppressor genes and oncogenes were discovered as a result of cancer research, as genes inactivated and promoted in cancer, namely cancer genes [64, 65]. They sure have roles in prognosis of cancer because there is an established idea that each cancer gene brings an advantage to the cancer cell in terms of gaining the capabilities mentioned above [64, 65]. The cellular pathways abused by cancer cells are well-

characterized, however the causes or effects of these events on the whole genome are still not exactly known.

Although Hanahan and Weinberg did not mention about epigenetics as an important acting mechanism in carcinogenesis, we today know that epigenetic aberrations are as destructive as genetic ones in the route to uncontrolled cell growth [66]. In fact, regarding these two phenomena as two different incidents would be a big mistake since they are closely linked to each other. For example, candidate genetic alterations affecting genes involved in the great epigenetic machinery, like DNMTs, histone acetylases and deacetylases, and the SWI/SNF chromatin remodeling complex were proposed to be causes of epigenetic aberrations that are in turn exerted by these proteins within cancer cells [67]. The epigenetic aberrations shown to contribute tumorigenesis include changes in DNA methylation, histone modifications, chromatin remodeling, and regulation of small RNAs [67-72]. Involvement of chromosomal higher-order structure and non-coding transcripts are newly being studied [73-75]. In other words, every aspect of epigenetic systems is affected from the carcinogenesis process and they also exert broad range of changes to many sites of the genome. Another scenario of cancer progression attributes a much more major role to epigenetic factors. In this scenario, polyclonal epigenetic disruption of cancer stem cells leads to accumulation of later genetic and epigenetic aberrations which are required for tumor progression [76].

Among the above mentioned epigenetic alterations in cancer, aberrant DNA methylation is the most detailedly studied of all; thus will be the main focus of this section. Both aberrant hypo-methylation and hyper-methylation take place, examplified by misregulation of cancer-testis genes and tumor suppressor genes, respectively [68, 77] . Since hyper-methylation of TSGs have more direct affect on cellular proliferation as boosting cell growth [78], many researchers are trying to understand which TSGs are aberrantly methylated and at what stage of carcinogenesis these events take place. The first TSG promoter shown to be hyper-methylated in a human cancer belongs to the Retinoblastoma gene, and was discovered in 1989 [79]. However, hypermethylation of CpG islands being a

common mechanism of TSG inactivation in cancer was widely accepted when $p16^{INK4a}$ was also shown to be inactivated by hypermethylation in 1995 [80-82]. After this discovery, many researchers focused on searching new tumor suppressor genes that are epigenetically inactivated and succeeded in finding BRCA1, RASSF1, estrogen, androgen, progesterone and retinoic acid receptors in this class as striking examples [83-87]. On the other side, many key TSGs are proved to be normally methylated, like BRCA2, PTEN and $p19^{INK4d}$ [88-91]. Today, 100-400 CpG islands are estimated to be hyper-methylated in a given tumor [92] and many key cellular pathways suffer from inactivation of the TSGs, like the p53 network, cell cycle, DNA repair, apoptosis, hormone and vitamin response, etc. [93].

CpG hypermethylation of TSGs are rather specific to different types of tumors, both in sporadic tumors and in familial-inherited cancers, where it acts as a 'second hit' [94]. The 'methylotype' of a given tumor is the term used for the complete profile of CpG island methylation throughout the genome. The methylotype data for various tumors, although limited to known CpG islands, indicate that tumors of gastrointestinal origin (esophagus, colon, stomach) are significantly more methylated than others like ovarian tumors [92]. Exposure to external carcinogens within this tract is the most possible answer [67]. The methylotype data points methylation hot spots on human chromosomes 3p, 1p35 and 11p15 where multiple genes are hyper-methylated, thus the region behaves like a large loss of heterozygosity locus [67, 95-97].

CpG island hypermethylation is common to many types of cancer and seems to have a deep impact in cancer progression. In order to proceed in answering the key questions regarding hyper-methylated TSGs and all other components affected from aberrant DNA methylation (miRNAs, DNA repeats, oncogenes, CT genes etc.), the necessity to plot a human epigenome map arouse. The Human Epigenome Project aims to produce the tissue-specific DNA methylation profiles of the human genome and is thus under construction [98]. The Human Epigenome Pilot Project is completed and first data of the main project came from human chromosomes 6, 20 and 22 [98]. Once finished, the results will profoundly

enhance our understanding of both the targets of epigenetic machinery and the operation criteria of this machinery as well.

# 2 MATERIALS & METHODS

*Cell Lines and Culturing*

HCC cell lines HCT15 and Colo205 were grown in RPMI media (HyClone #SH30027.01); HCT116 and LoVo were grown in DMEM (HyClone #SH30307) for bisulfite sequencing analysis. All media were supplemented with 10 % FBS (HyClone #SH30160.03), 1% penicllin/streptomycin (HyClone #SV30010), 1 % L-glutamine (HyClone #SH3004.01) and 1 % non-essential amino acids (HyClone #SH30238.1). Media were changed every three days; cells were split while exponentially growing. DNA from above cell lines plus Ls174T and Lim1215 was used for PCR detection.

*DNA and RNA extraction*

DNA from appropriate cell-lines for bisulfite sequencing experiments was extracted with UltraCleanTissue DNA Isolation Kit (MoBio #12334). Peripheral blood lymphocyte DNA was extracted from whole blood with hypotonic swelling method. RNA was extracted from all cells by Tri-reagent (MRC #TR 118). DNase I treatment (DNA-Free Kit, Ambion #AM1906) was performed after extracting the RNA.

## 2.1 Expression Analyses of Putative TSGs

*Selection of Putative Tumor Supressor Genes on X chromosome*

X-chromosome genes that are downregulated, with respect to their normal counterparts, in any human cancerous tissue except from embryonic and germline origin were identified by analyzing SAGE Digital Gene Expression Displayer and cDNA Digital Gene Expression Displayer databases of the Cancer Genome

Anatomy Project based on SAGE and EST libraries [99]. Among existing EST libraries, 269 libraries of cancerous tissues versus 339 libraries of healthy tissues were screened. Among existing SAGE libraries, 78 libraries of cancerous tissues versus 182 libraries of healthy tissues were screened. Significance filter was adjusted to $p<0.05$. 59 genes were obtained based on the above criteria. Extracted data were checked by using Monochromatic SAGE/cDNA Virtual Northern. Among the 59 genes, 8 genes that have a neighboring cancer-testis (CT) gene were chosen and expression analyses were made to verify the database data. These 8 genes were located at least 30kb and at most 560kb from a CT gene.

*Conventional and Real-Time PCR*

Primer Ts were calculated according to the formula: $Tm = 69.3^{o}C + 0.41$ (%GC) - 535/n, where n is primer length. Conventional PCR was carried out under the conditions of $94^{o}C$ for 10 min followed by 35 cycles of $94^{o}C$ for 60s, appropriate melting temperature for 60s, $72^{o}C$ for 60s, with a final extension at $72^{o}C$ for 10 min, in Techgene thermal cycler (Techne), and run on 1.5% agarose gel at 90V for 45 min. DyNAzyme II HS DNA polymerase (Finnzymes #F-504), dNTP mix (Finnzymes #F-560), forward and reverse primers at final concentrations of 0.03 U/µl, 250 µM, 500 nM, respectively, were used. GAPDH was used as positive control. For real-time PCR, efficiency curves of candidate genes were plotted and 1:5 dilution was chosen. 4 housekeeping genes, GAPDH, 18S rRNA, TFCP2 [100] and GOLGA1 [101], were tested as candidates to normalize the real-time data. Analyses of these genes by GeNorm program [102] lead to choice of TFCP2 and GOLGA1 for normalization. Real-time PCR experiments were carried out in Bio-Rad iCycler under the same setup explained above followed by melting curve analyses. Sybr Green dye (DyNAmo HS SYBR Green qPCR Kit, Finnzymes #F-410L) was used for measurement of DNA amount. Genes were normalized according to delta-delta Ct method.

**2.2 Methylation Analyses of Putative TSGs**

*Detection of CpG Islands*

A region including 3000 bases upstream of transcription start sites were analyzed for presence of CpG islands by using CpG Island Searcher [103] and CpGPlot [104] programs according to the following criteria: Observed/expected ratio > 0.60, percent C + percent G > 50.00, length > 200

*Bisulfite Sequencing*

400 ng DNA from HCT15, HCT116, Ls174T, LoVo, Colo205 and Lim1215 cells was sodium bisulfite treated with EZ DNA Methylation-Gold Kit (Zymo Research #D5005) according to manufacturer's instructions. 2 µl of 1:10 diluted bisulfite-modified DNA was amplified by conventional PCR under conditions of 94$^{o}$C for 10 min followed by 35 cycles of 94$^{o}$C for 60s, appropriate annealing temperature for 60s, 72$^{o}$C for 60s, with a final extension at 72$^{o}$C for 10 min, in Techgene thermal cycler (Techne), run on 1.5% agarose gel at 90V for 45 min, gel extracted (QIAquick Gel Extraction Kit, Qiagen #28704) and sequenced (Iontek). Cytosines within CpG dinucleotides were accepted to be methylated if C/T (or G/A) ratio is more than 50 % among the cell population. Primers used for analyzing sodium bisulfite modified DNA are listed in **Table 3**.

**2.3 Analyses of Higher-Order Chromosomal Structure**

*Detection of Inverted Repeats*

Sequences that correspond to regions encompassing NY-ESO-1/LAGE1, CT45 and MAGE-A family of genes were analyzed by using the Inverted Repeats Finder Program [105] Public Database project. Sequence complementarities were verified by MegAlign (DNAStar).

*Chromosome Conformation Capture (3C)*

HCC1937, FOCUS, Colo205, HT29, SK-LC-17 cell lines were grown to 80% confluency in RPMI media (PAA #E15-039) supplemented with 10 % FBS (PAA #A15-103), 1% penicllin/streptomycin (Gibco #15140) and 1 % L-glutamine (Gibco #25030). Cells were trypsinized (Hyclone #SH30042), centrifuged and resuspended in 45 ml of fresh RPMI media. For performing the 3C experiment with tissues,

normal thyroid and gall bladder were first dissected, passed through a 70 µm cell-strainer (BD Falcon #352350), centrifuged and resuspended in 45 ml of fresh RPMI media. In presence of 1% formaldehyde, cells were crosslinked for 10 min at RT; quenched, resuspended in ice-cold lysing solution supplemented with 100 µl complete protease inhibitor cocktail (Sigma #P8340) and homogenized in a bounce homogenizer with pestle B. Following wash steps, nuclei from $5x10^6$ cells were digested with 30U of appropriate restriction endonuclease, overnight with conventional restriction endonucleases (Fermentas BglII #ER0081,EcoRI #ER0271, GsuI #ER0462, XceI #ER1471) and for 2 hours with Fast-Digest enzymes (Fermentas BglII #FD0074, EcoRI #FD0274, XceI #FD1474). $1x10^6$ restriction-digested nuclei were used for ligation for 2 hours at $16^oC$ with 30U of T4 DNA ligase (Promega #M1804, Fermentas #EL0012). Cross-links were reversed overnight at $65^oC$ by Proteinase K (100 µg/ml). DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in appropriate amount of TE buffer. DNA concentrations were measured at Nano-Drop UV/Vis spectrophotometer (Thermo). Chromatin in the absence of formaldehyde and T4 DNA ligase were used as not cross-linked and unligated negative controls and processed similarly in the following steps.

Final chromatin preparations were analyzed by conventional PCR with the primers listed in **Table 4** under the setup: $94^oC$ for 10 min followed by 35 cycles of $94^oC$ for 30s, appropriate annealing temperature for 30s, $72^oC$ for 30s with a final extension at $72^oC$ for 10 min, in Perkin-Elmer 9700 thermal cycler. DyNAzyme II HS DNA polymerase (Finnzymes #F-504), dNTP mix (Finnzymes #F-560), forward and reverse primers were used at final concentrations of 0.03 U/µl, 250 µM, 500 nM, respectively. Samples were run on 1.5 % agarose-gel at 90V for 45 min. Detected bands were gel extracted and sequenced for verification.

Analyzed interactions were normalized for bias in 3C efficiency to a control interaction at the γ-actin region [106]; and for PCR-bias of primer pairs to interactions of similarly processed BAC (bacterial artificial chromosome) clones RP11-103M23 (for NY-ESO-1 region), RP11-111C16 and RP11-299H1 (for CT45 region).

*5-Aza-2'-deoxycytidine Treatment*

HT29 and Colo205 cells were treated for 2 days with 5-Aza-2'-deoxycytidine (Sigma #A3656) at 1 µM final concentration in RPMI media. The cells were then washed with PBS and used for 3C analysis at subsequent steps. RNA from AZA-treated and -untreated cells was also isolated and expression levels were analyzed by RT-PCR (Finnzymes DyNAmo cDNA Synthesis Kit #F-470).

**Table 1 Genetic information regarding selected putative TSGs**

| Gene Name | Accession # | Chromosomal Region | Gene Length (bp) | mRNA length (bp) |
|---|---|---|---|---|
| Aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anemia) (ALAS2) | NT_011630.14 | Xp11.21 | 21924 | 1941 (NM_000032)<br>1830 (NM_001037967)<br>1937 (NM_001037968)<br>1745 (NM_001037969)† |
| Cerebellar degeneration-related protein 1 (CDR1) | NT_011786.15 | Xq27.1-q27.2 | 1299 | 1299 (NM_004065.2) |
| Gamma-aminobutyric acid (GABA) A receptor, alpha 3 (GABRA3) | NT_011726.13 | Xq28 | 283304 | 2785 (NM_000808) |
| Hypothetical protein LOC286411 | NT_011786.15 | Xq27.1 | 20529 | 1607 (XM_932619.3) |
| Ras-related GTP binding B (RRAGB) | NT_011630.14 | Xp11.21 | 41029 | 2217 (NM_016656)<br>2133 (NM_006064)† |
| Solute carrier family 9 (Na+/H+ exchanger), member 6 (SLC9A6) | NT_011786.15 | Xq26.3 | 61843 | 4742 (NM_001042537)<br>4646 (NM_006359)† |
| SLIT and NTRK-like family, member 2 (SLITRK2) | NT_011681.15 | Xq27.3 | 8011 | 5022 (NM_032539) |
| Zinc finger, CCHC domain containing 12 (ZCCHC12) | NT_011786.15 | Xq24 | 3145 | 2212 (NM_173798.2) |

† the isoforms used when indicating amplified regions for gene expression analyses

**Table 2 Primers used for expression analyses**

| Gene Name | Sequence of Primer | Region amplified and amplicon length (bp) | $T_m$ ($^oC$) |
|---|---|---|---|
| ALAS2 | F: 5`- AGGGTGCGAGATTTACTCAGAC<br>R: 5`- ATTCTAGAGCTCCAGAGAGCAC | 678 - 1156 (479) | 65<br>65 |
| CDR1 | F:5`- TCGGAAGCTATGGATTTGAGGG<br>R:5`- TCCAGCCAATATATGTCTTCCTGAAG | 400 - 802 (403) | 65<br>66 |
| GABRA3 | F:5`- GCTAGTAAGATCTGGACACCG<br>R:5`- TTCCACGGATTTGTTCTTTCCGAG | 666 - 935 (270) | 65<br>66 |
| LOC286411 | F:5`- GGTGAAAGACACTCATCCAGATG<br>R:5`- TCGGTGTGCCTTTGGATATGTC | 135 - 516 (382) | 66<br>65 |
| RRAGB | F:5`- TCTGGTAAGACCAGCATGAGG<br>R:5`- GGAAGATGTTGTCCCGTTGG | 733 - 928 (196) | 65<br>65 |
| SLC9A6 | F:5`- CTTGGGTCTATCCTAGCATACG<br>R:5`- TCGCTGTGACATCAAAGGTGTG | 561 - 897 (337) | 65<br>65 |
| SLITRK2 | F:5`- CCCACTTCCTGAAGATGCTAAAC<br>R:5`- CACCTTCGGTGTCAACTCAG | 1333 - 1585 (253) | 66<br>65 |
| ZCCHC12 | F:5`- CTTGAATTCTGTGCAGCTGATTGC<br>R:5`- ACACGTGCAATGATGCTAGCC | 204 - 496 (293) | 66<br>65 |

**Table 3 Primers used for bisulfite sequencing experiments**

| Gene Name | Sequence of Primer | Region amplified and amplicon length (bp) | $T_m$ ($^oC$) | # of CpGs |
|---|---|---|---|---|
| ALAS2 | F: 5`- GTAGTAAGTTTTGGGTGGGAAGTTG<br>R: 5`- CTCCCCTTTTACTAACAACCTCTTC | (+4865) to (+5188) of TSS (324)<br>[(-99) to (+225) of 3rd exon] | 66<br>66 | 3 |
| CDR1 | F: 5`- ATTGGGTAGTTGTTGGAAGATATGAAG<br>R: 5`- CTTCCATCAAATTAATATCTTCCAACCTAC | (-9) to (+334) of TSS (343) | 65<br>65 | 8 |
| ZCCHC12 | F: 5`- AGTTTAGGGAGAAGGGAYGG<br>R: 5`- CCRCCCAATACTAATAAATCTTCCC | (-263) to (+249) of TSS (512) | 65<br>66 | 41 |

**Table 4 Primers used for the 3C assay**

| Primer number | Sequence of Primer | Tm ($^{o}$C) | Amplicon length (bp) |
|---|---|---|---|
| 1 | 5'GGGGAAGCCTAACACACTGG | 67 | 276 |
| 9 | 5'TTGCCCAGGCTGGTCTCAC | 67 | |
| 2 | 5'AGGAAGGAAGACACAGAGCTG | 65 | 192 |
| 10 | 5'CTCCGAAAGTGCTGGGATTATAG | 66 | |
| 3 | 5'TGGACTCGTGGCCAGGTTTG | 67 | 281 |
| 11 | 5'GGTCTCACCAGGGACAACTG | 67 | |
| 4 | 5'TCCTCTAACCGCGTGTGACG | 67 | 174 |
| 12 | 5'GGATATTGTTAGTTGTCACAACAGGAC | 66 | |
| 5 | 5'AGAGAGCAAGTCAGGACTGTG | 65 | 336 |
| 7 | 5'GACAAAATTCGGTCAGCACAAGTAAATAG | 66 | |
| 6 | 5'CACCCACATCTGATCTTGAAGTG | 66 | 173 |
| 8 | 5'TGCACATTCTCCAACAGGCATC | 65 | |

**Table 5 Primers used for analyses of non-coding transcripts**

| Primer name | Sequence of Primer | Tm ($^{o}$C) | Amplicon length (bp) |
|---|---|---|---|
| A_F | 5`- AGCACAGCGTGCAGGTGGAC | 69 | 209 |
| A_R | 5`- GAGATCTTCCAGCTGCATTCC | 65 | |
| B_F | 5`- CACTGGCCCCAATTAGGAAGAAC | 67 | 275 |
| B_R | 5`- GAAGGCCTCATATCCCAATTCTAGC | 68 | |
| C_F | 5`- CAAGGAAGTTTTGTGGAGTAAGGAAGG | 68 | 234 |
| C_R | 5`- CTAGGCTTTCTTCAGTCCCCAAAC | 68 | |
| D_F | 5`- GACAGGGTCACATGCACTTTAC | 65 | 387 |
| D_R | 5`- CCCGACTTGATCATTACATCGTG | 66 | |
| E_F | 5' - CAGGGCTGAATGGATGCTGCAGA | 70 | 337 |
| E_R | 5' - GCGCCTCTGCCCTGAGGGAGG | 70 | |
| F_F | 5`- TGCATACCCTTCCAGCTGTAGG | 67 | 365 |
| F_R | 5`- GGAGAAACCTTGGACAATACCCG | 67 | |
| G_F | 5`- GTTAAATTAGAGCGCATTCATATTGCG | 65 | 176 |
| G_R | 5`- CTCACCCACTGCAAACATTCAATG | 66 | |

**Table 6 Primers of normalization genes used for all assays**

| Primer name | Sequence of Primer | Tm ($^{o}$C) | Amplicon length (bp) |
|---|---|---|---|
| GAPDH_F | 5`- GACCACAGTCCATGCCATCACT | 67 | 454 |
| GAPDH_R | 5`- TCCACCACCCTGTTGCTGTAG | 67 | |
| GOLGA1_F | 5`- AGATACGGAAGTTAGAGGCCAG | 65 | 200 |
| GOLGA1_R | 5`- GTCCTTTCTGGCTAATGCCAAAG | 66 | |
| TFCP2_F | 5`- AAGAAGAGTCGAGTTTGCCTCC | 65 | 200 |
| TFCP2_R | 5`- CTTCACCAATTTGCCATTAATTTCTGG | 65 | |
| 18S rRNA_F | 5`- CGTGCATTTATCAGATCAAAACCAACC | 66 | 135 |
| 18S rRNA_R | 5`- ATGGTAGGCACGGCGACTAC | 67 | |
| γ-actin_F | 5' - GCTGTTCCAGGCTCTGTTCC | 67 | 337 |
| γ-actin_R | 5' - GCGCCTCTGCCCTGAGGGAGG | 67 | |

# 3 RESULTS

## 3.1 Identification of X-linked CT-Proximal Putative Tumor-Suppressor Genes

CT genes are known to be induced in many cancer types by hypomethylation of promoter or the whole gene. 5-Aza-2'-deoxycytidine induction of CTs verifies activation by DNA hypomethylation also. On the other hand, region-specific hypermethylation is known to affect many genes, especially tumor suppressors, which contributes to tumorigenesis/metastasis through their silencing. Although the hypo- and hyper-methylation phenomena are widely referred, exact mechanisms of how these controversial events occur are not known yet. The only knowledge in hand is about the enzymes exerting these effects in the final steps, namely DNMTs and putative de-methylases. In order to approach this mechanistic problem in a more detailed manner, we have identified candidate regions on the X chromosome which involve both hypomethylated and hypermethylated sites in close proximity. Since CT genes are already known to be hypomethylated in cancer, we selected CT-proximal genes that are hypermethylated - and thus down-regulated- in cancer, which we refer as putative tumor suppressor genes (pTSGs). X-linked genes that are down-regulated, with respect to their counterparts in normal tissues of same origin, in any human cancerous tissue except from embryonic and germline origin were identified by analyzing SAGE Digital Gene Expression Displayer and cDNA Digital Gene Expression Displayer databases of Cancer Genome Anatomy Project based on SAGE and EST libraries [99]. 59 genes were obtained based on the above criteria (available in **Appendix C**). Extracted data were checked by using Monochromatic SAGE/cDNA Virtual Northern. Among the 59 genes, 8 genes that have a neighboring cancer-testis (CT) gene were chosen and expression analyses were made to verify the *in silico* data, as presented in

sections 3.2 and 3.3. The 8 chosen putative TSGs are tabulated in **Table 7** together with the neighboring CT genes and chromosomal locations.

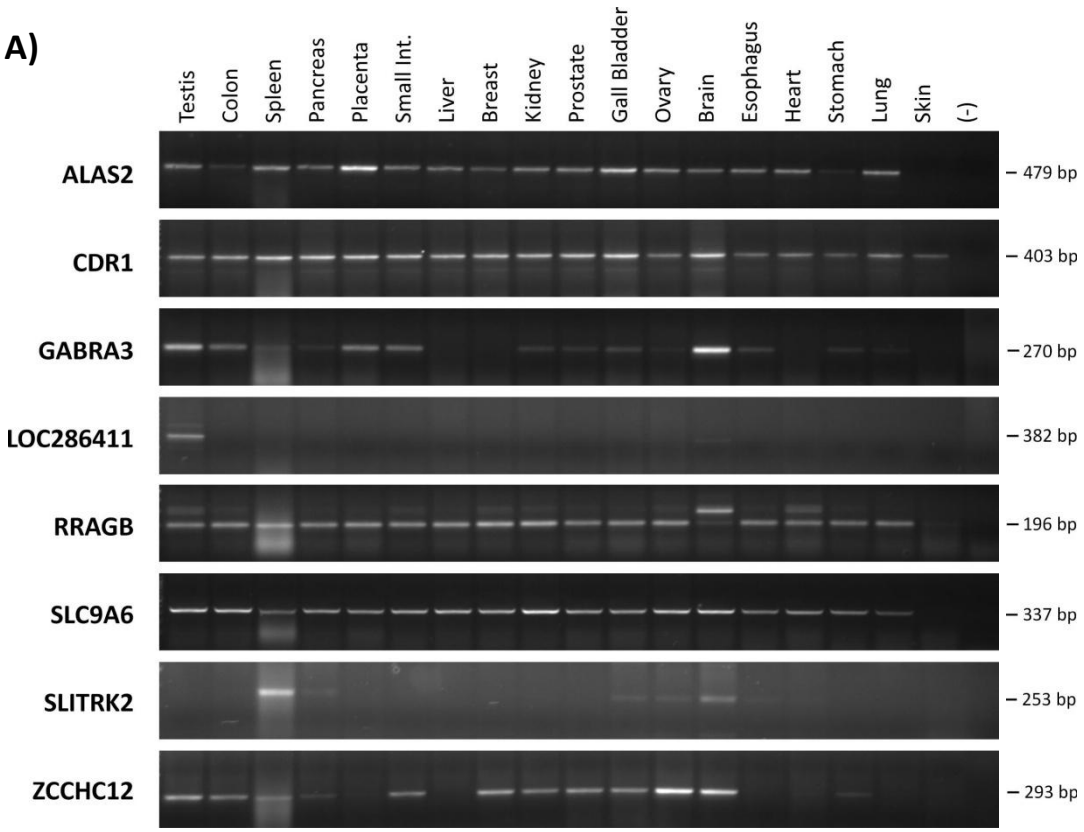
**Table 7 X-linked putative TSGs and the proximal CT genes**

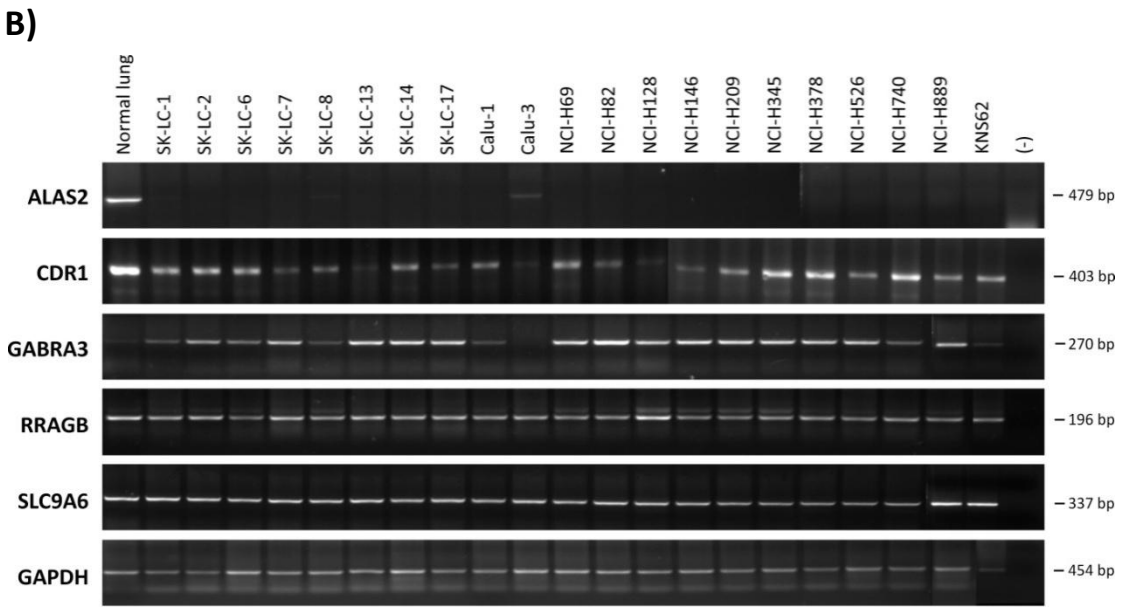
| pTSG | Neighboring CT gene(s) | Approximate distance in between (kb) | Chromosomal location |
|---|---|---|---|
| ALAS2 | PAGE2B, PAGE2 | 44, 58 | Xp11.21 |
| CDR1 | SPANXB1 | 220 | Xq27.1 |
| GABRA3 | MAGE-A10, -A6 | 30, 250 | Xq28 |
| LOC286411 | SPANXB1 | 290 | Xq27.1 |
| RRAGB | XAGE4 | 60 | Xp11.21 |
| SLC9A6 | SAGE1, CT45 cluster | 50, 80 | Xq26-26.3 |
| SLITRK2 | SPANXN1 | 560 | Xq27-27.3 |
| ZCCHC12 | IL13RA1 | 30 | Xq24 |

### 3.2 Validation of Virtual Data by Conventional RT-PCR

Firstly, mRNA expression of the 8 genes was tested using conventional RT-PCR in a panel of normal tissues. All genes but LOC286411 and SLITRK2 exhibit some expression in normal tissues, indicating that their products are not strictly tissue-specific, although GABRA3 and ZCCHC12 seem to show some tissue specificity (Figure 1A). LOC286411 is likely to encode a testis-specific product, thus is probably a false-positive of the CGAP database. Those genes that have mRNA expression in normal colon and lung tissues were tested in lung and colon cancer panels. 2 genes, ALAS2 and CDR1 showed dramatic down-regulation in several lung and colon cancer cell lines. Since CDR1 is an intronless gene, negative RT PCRs were carried out in DNAse I treated and non-treated samples (**Appendix D**). GABRA3, SLC9A6 and ZCCHC12 genes showed down-regulation only in some cell lines; whereas RRAGB did not exhibit significant down-regulation in neither lung nor colon cancer cell lines, however, this could occur due to the low-sensitivity of

the assay (**Figure 1B,C**). As a result of the RT-PCR assay, 6 out of the 8 genes, namely ALAS2, CDR1, GABRA3, RRAGB, SLC9A6 and ZCCHC12 were selected to be putative TSGs and assessed by further experiments.



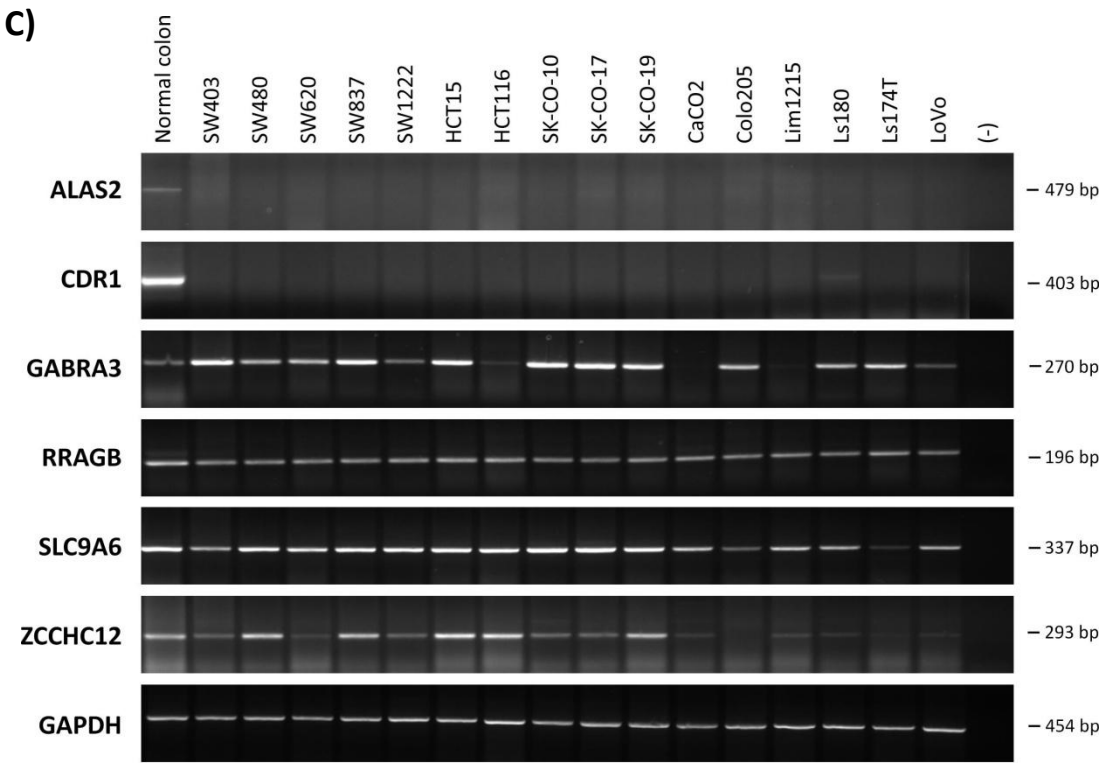


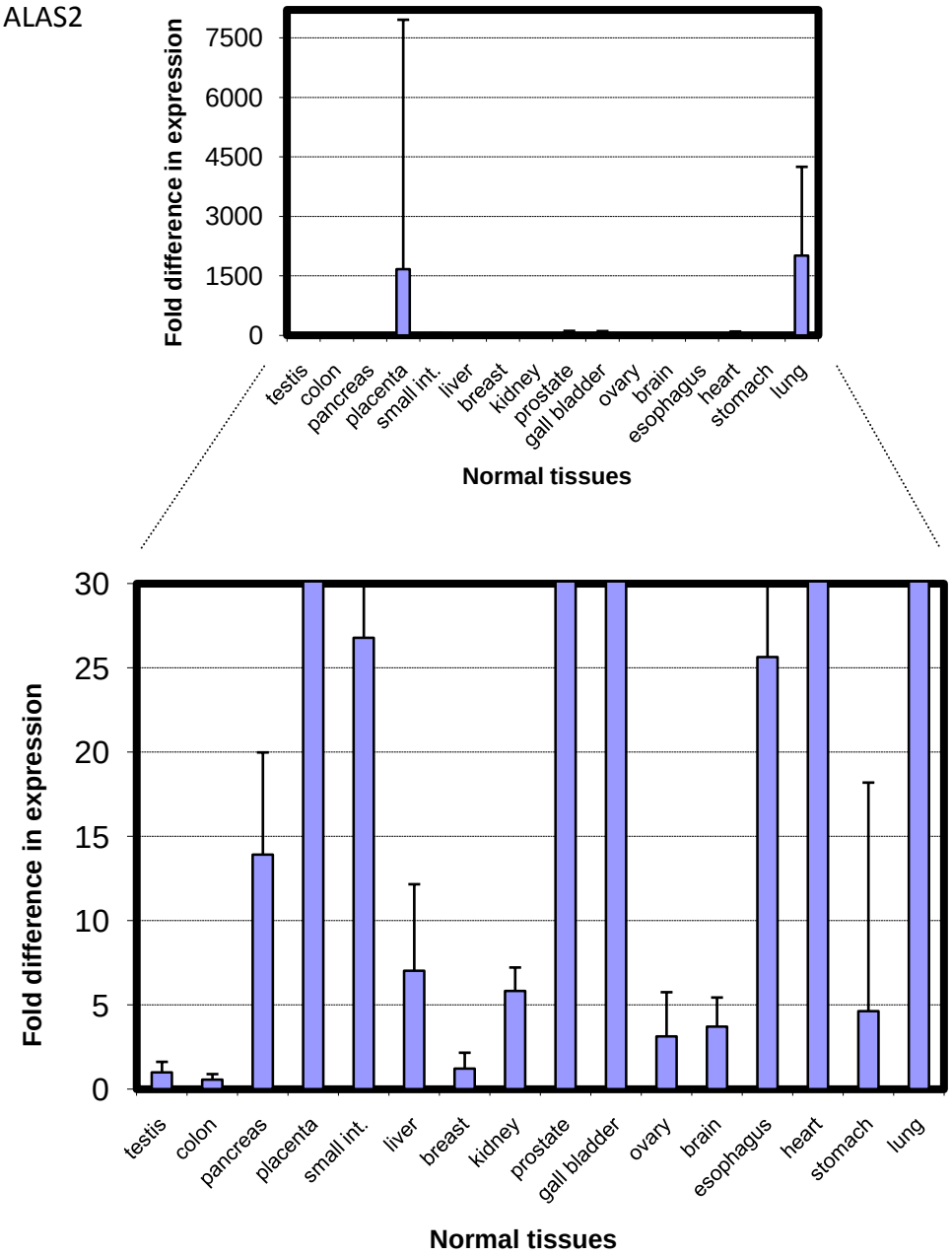

**Figure 1 mRNA expression levels of 8 putative tumor suppressor genes on X chromosome in normal tissue (A), lung cancer (B) and colon cancer (C) panels, by conventional RT-PCR. Two genes, ALAS2 and CDR1 are significantly down-regulated in most lung and colon cancer cell lines, while others showed variable down-regulation.**

### 3.3 Validation of Virtual Data by Real-Time RT-PCR

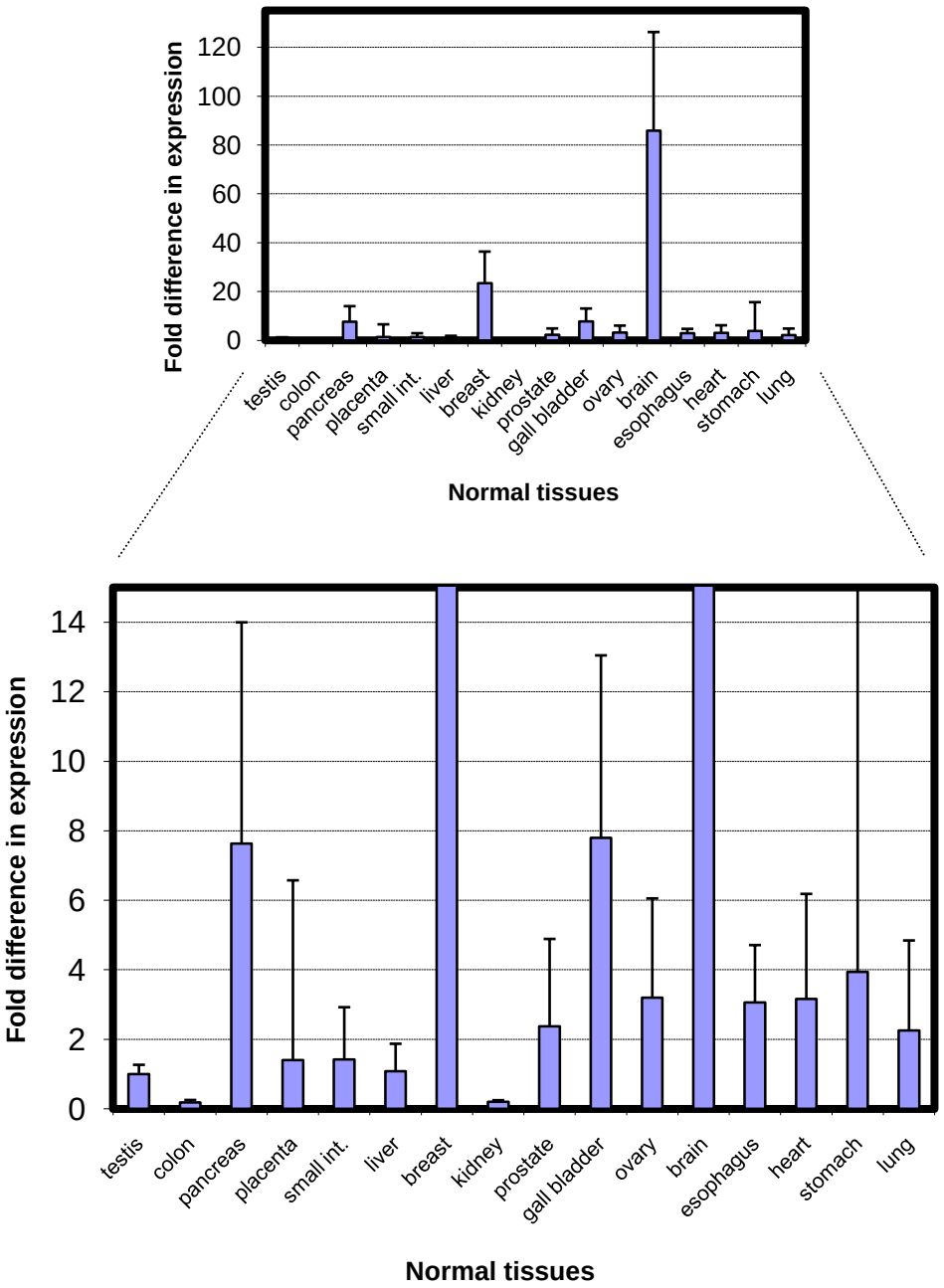
In order to quantify the mRNA expression levels, we aimed to analyze the transcripts of 6 genes whose expression were subject to change, ALAS2, CDR1, GABRA3, RRAGB, SLC9A6 and ZCCHC12, by real-time RT-PCR. Firstly, we had to identify 'housekeeping' genes with mRNA expressions more stable than GAPDH, to be used for normalization of expression in our normal tissue, lung cancer and colon cancer panels. 4 genes - GAPDH, 18S rRNA, GOLGA1 [101] and TFCP2 [100] - were analyzed as candidates for normalization by the GeNorm software and 2 of these genes – TFCP2 and GOLGA1 - were shown to be the best choices (**Appendix E**). The background information and basics of GeNorm software are presented in **Appendix E**. After identification of genes for normalization, we analyzed the 5

selected genes by real-time PCR. Data were normalized according to the delta-delta $C_t$ method. All normal tissues were normalized relative to mRNA expression in testis tissue. Lung and colon cancer cell lines were normalized relative to their normal counterparts. **Figure 2** shows that, these individual genes were expressed at similar levels in different healthy tissues with some exceptions: ALAS2 was highly expressed in placenta and lung; CDR1 was highly expressed in brain and breast; and GABRA3 was highly expressed in brain.

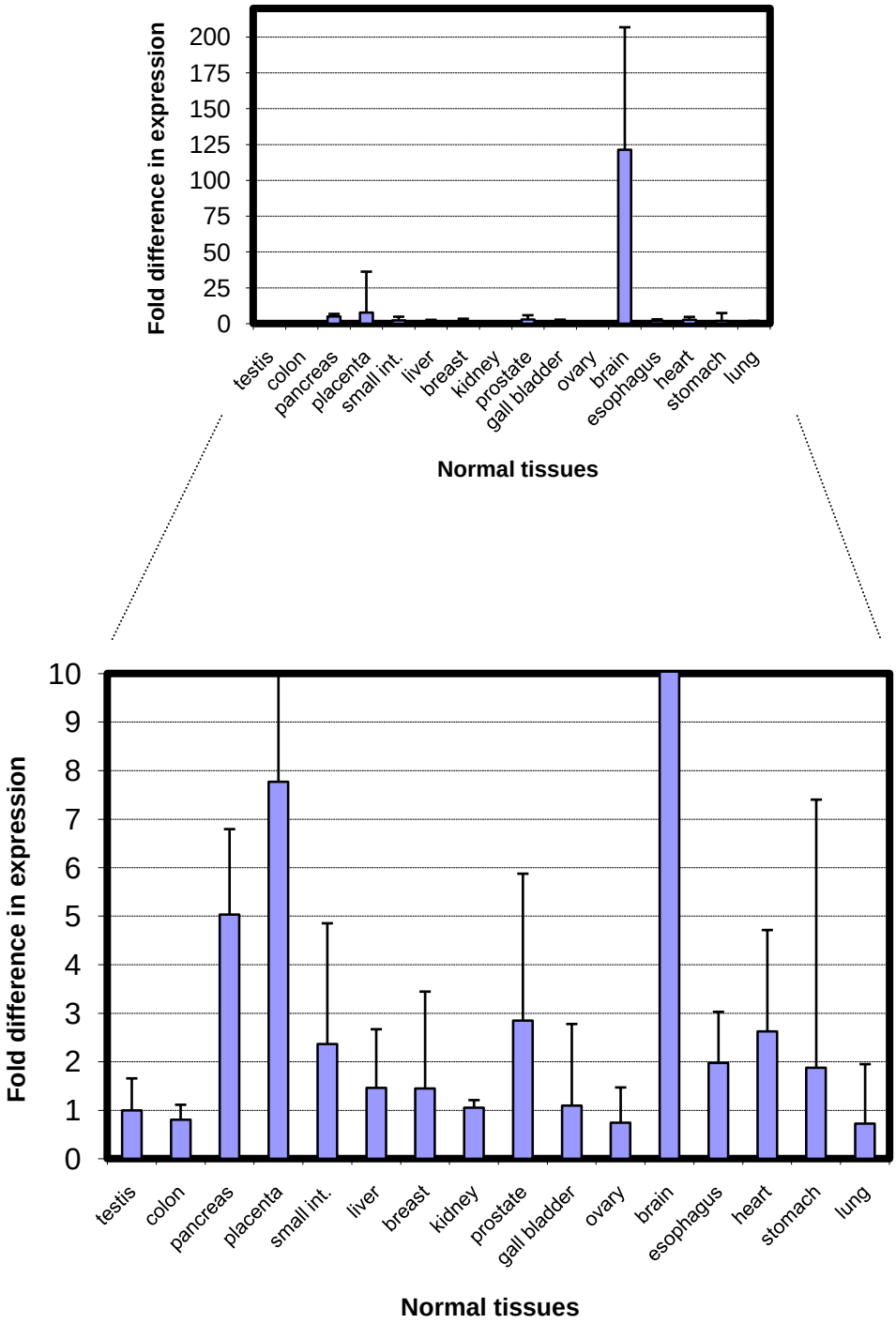
**A)** ALAS2



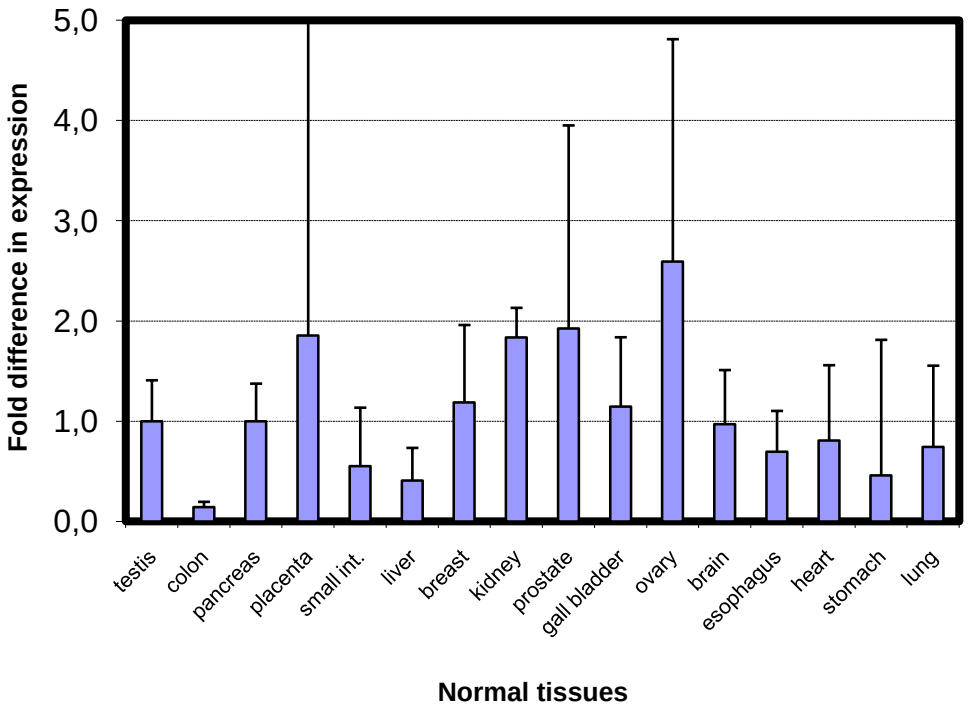
**B)** CDR1


Fold difference in expression
120
100
80
60
40
20
0
testis
colon
pancreas
placenta
small int.
liver
breast
kidney
prostate
gall bladder
ovary
brain
esophagus
heart
stomach
lung
Normal tissues
Fold difference in expression
14
12
10
8
6
4
2
0
testis
colon
pancreas
placenta
small int.
liver
breast
kidney
prostate
gall bladder
ovary
brain
esophagus
heart
stomach
lung
Normal tissues

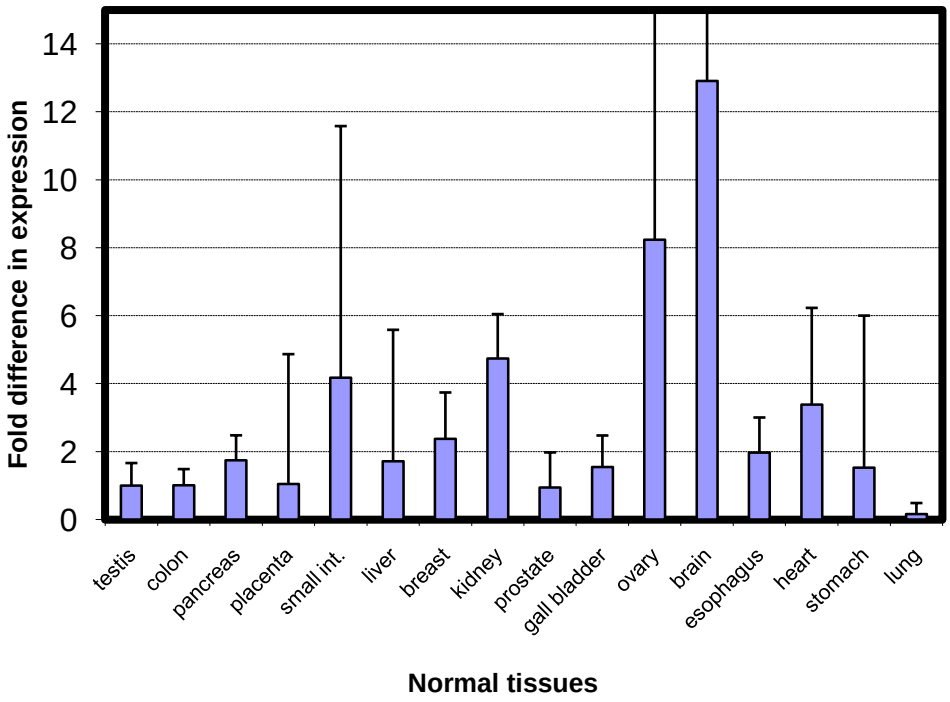
**C)** GABRA3


Fold difference in expression
200
175
150
125
100
75
50
25
0
testis
colon
pancreas
placenta
small int.
liver
breast
kidney
prostate
gall bladder
ovary
brain
esophagus
heart
stomach
lung
Normal tissues
Fold difference in expression
10
9
8
7
6
5
4
3
2
1
0
testis
colon
pancreas
placenta
small int.
liver
breast
kidney
prostate
gall bladder
ovary
brain
esophagus
heart
stomach
lung
Normal tissues

**D)** RRAGB




**E)** SLC9A6



**F)** ZCCHC12

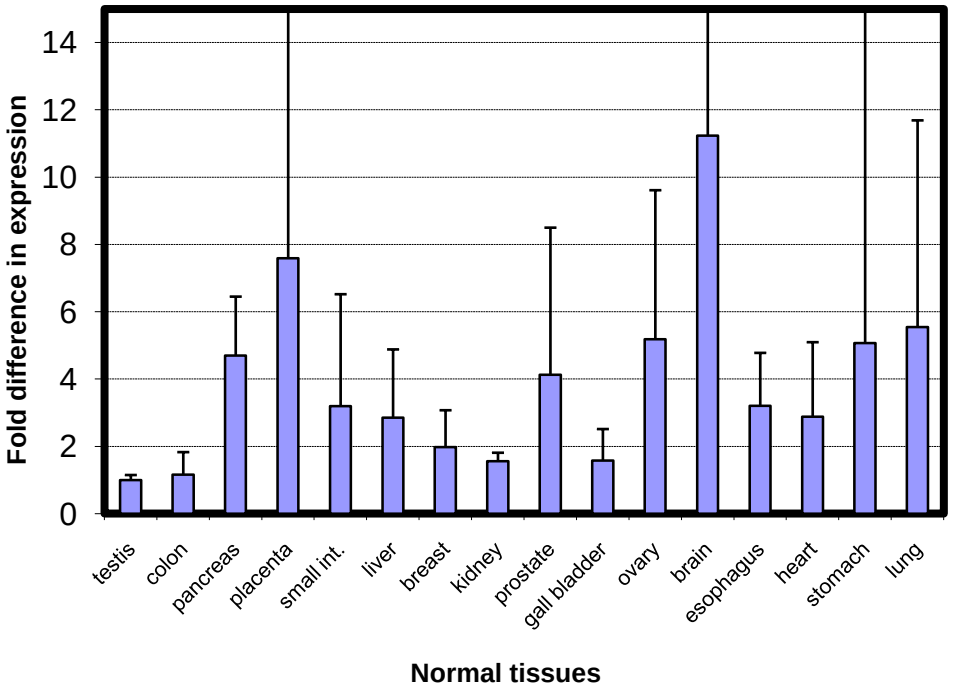
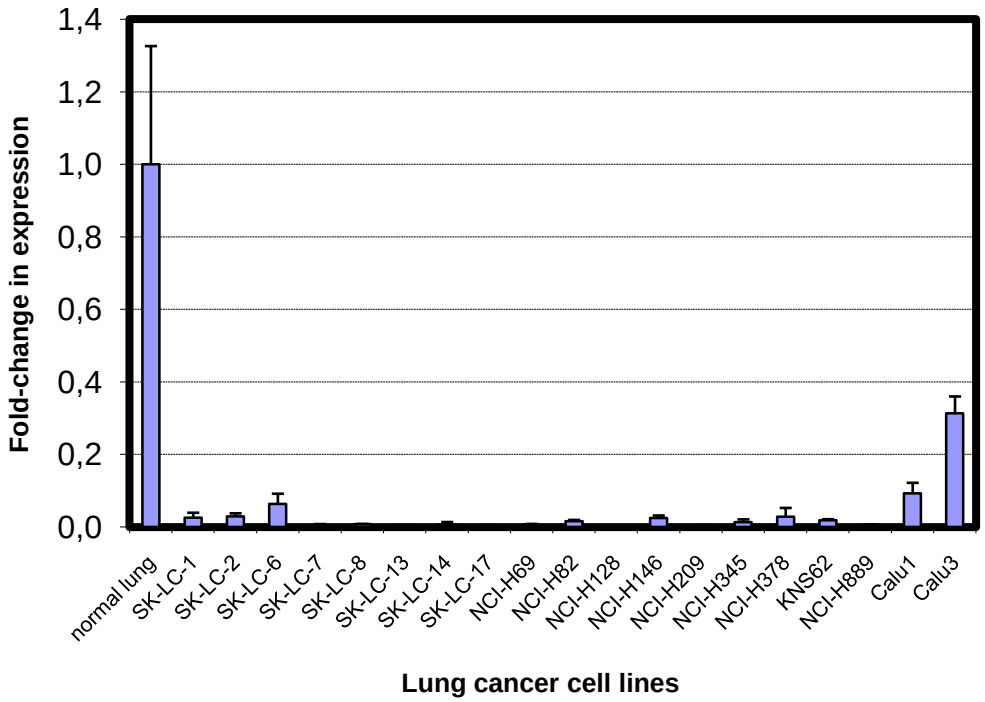


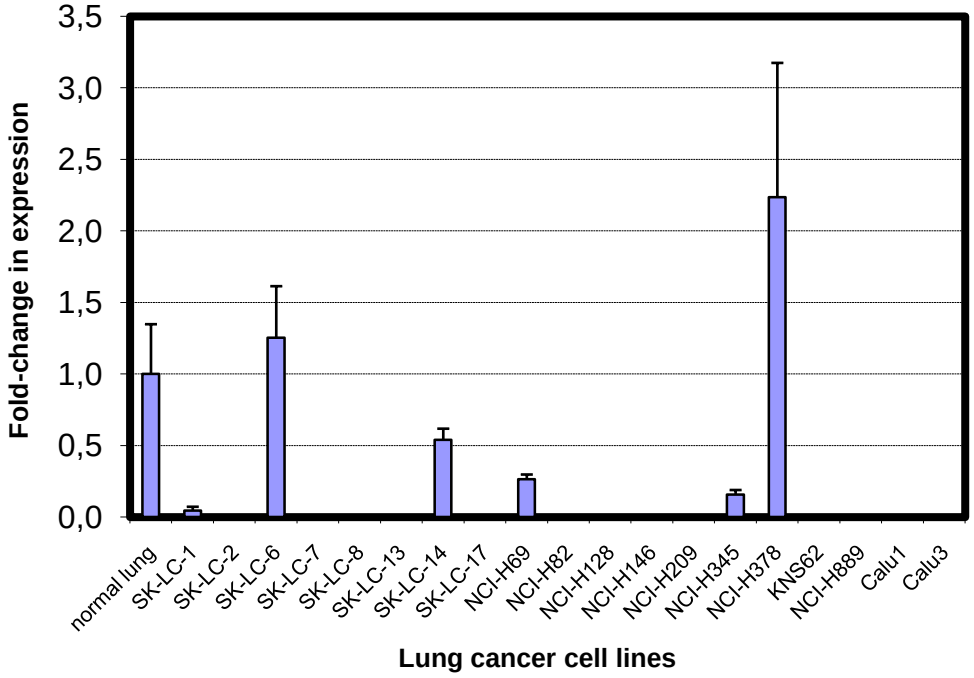

**Figure 2 Quantified expression levels of 6 pTSGs in normal tissues. (A-F) Graphs show mRNA expression levels on a large scale (small) and a zoomed-in scale to show details (big), where applicable. Graphs show overall mRNA expression levels in other cases.**

Secondly, mRNA levels of ALAS2, CDR1, RRAGB and SLC9A6 were analyzed in a panel of lung cancers, relative to healthy lung tissue. Strikingly, ALAS2 and CDR1 were significantly down-regulated in most cell lines (**Figure 3A-B**). RRAGB and SLC9A6 were only slightly down-regulated in some cell-lines (**Figure 3C-D**).

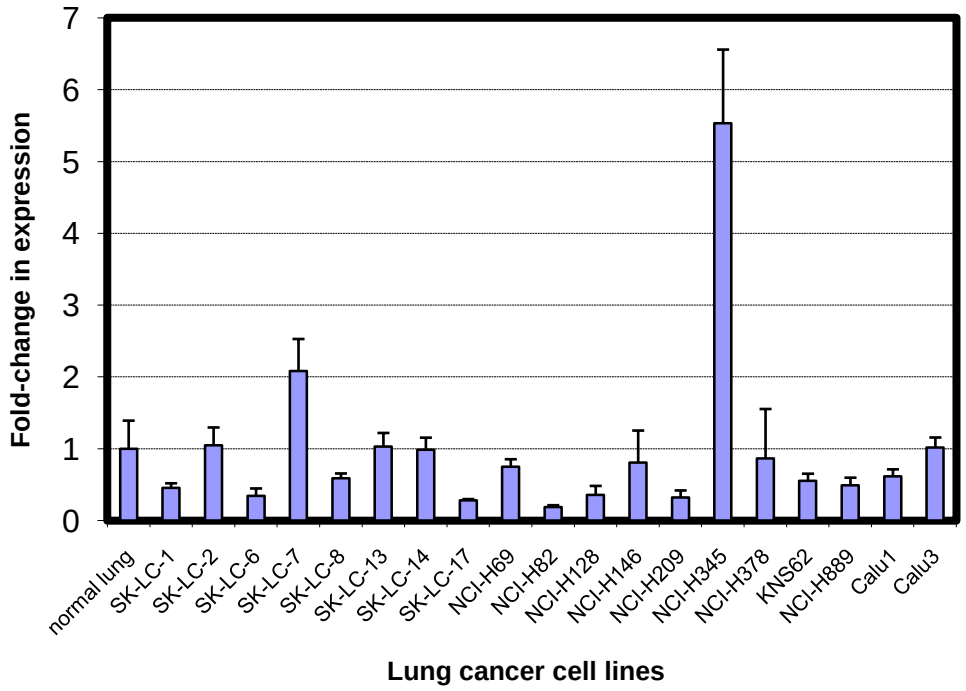
**A)** ALAS2




**B)** CDR1



**C)** RRAGB




**D)** SLC9A6

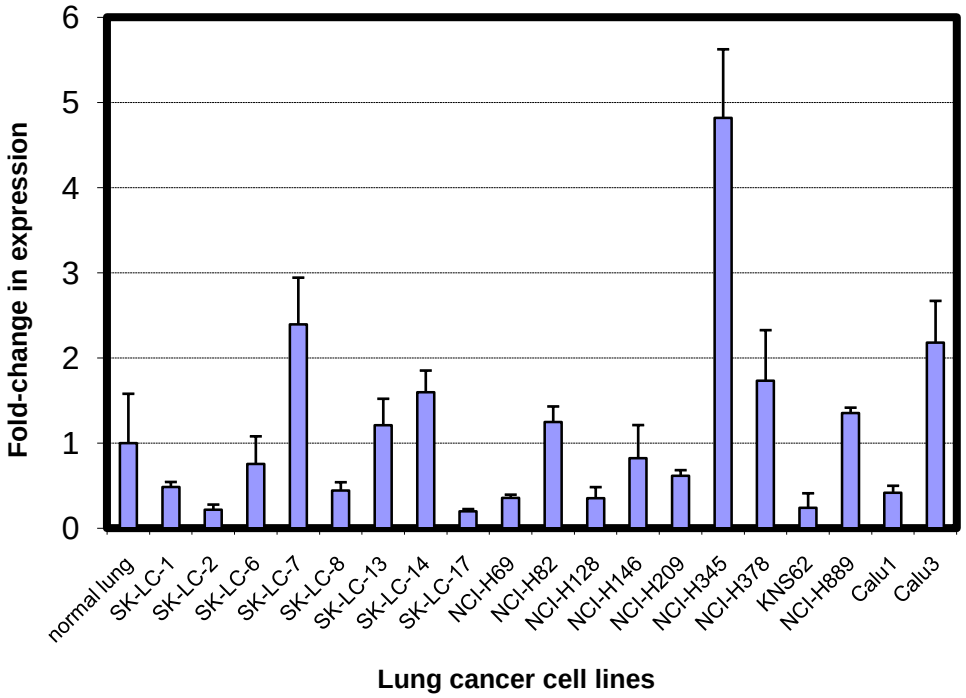
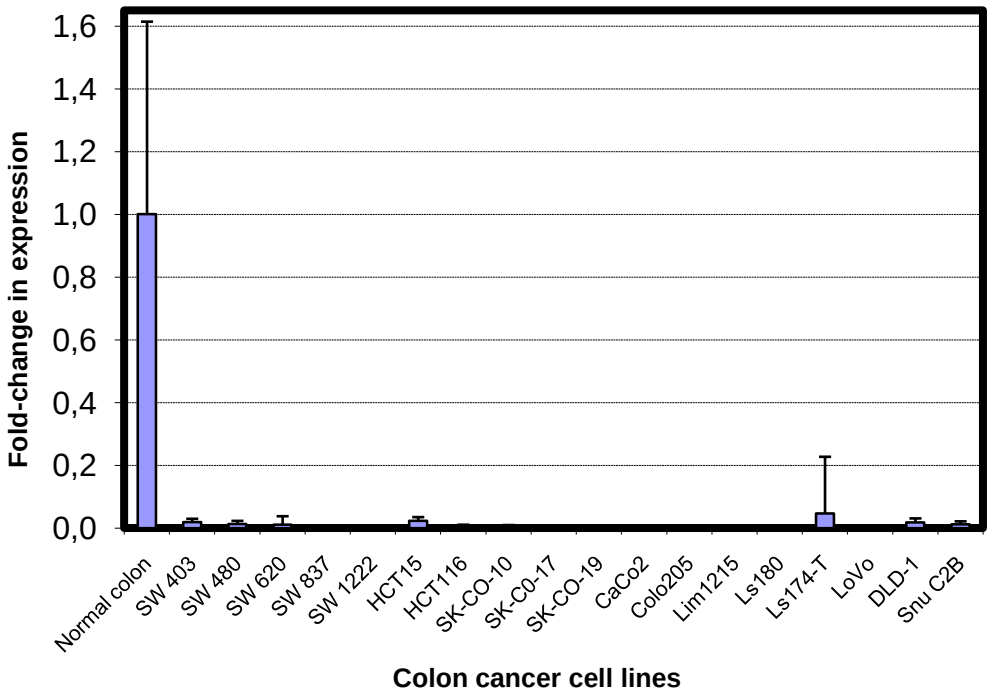


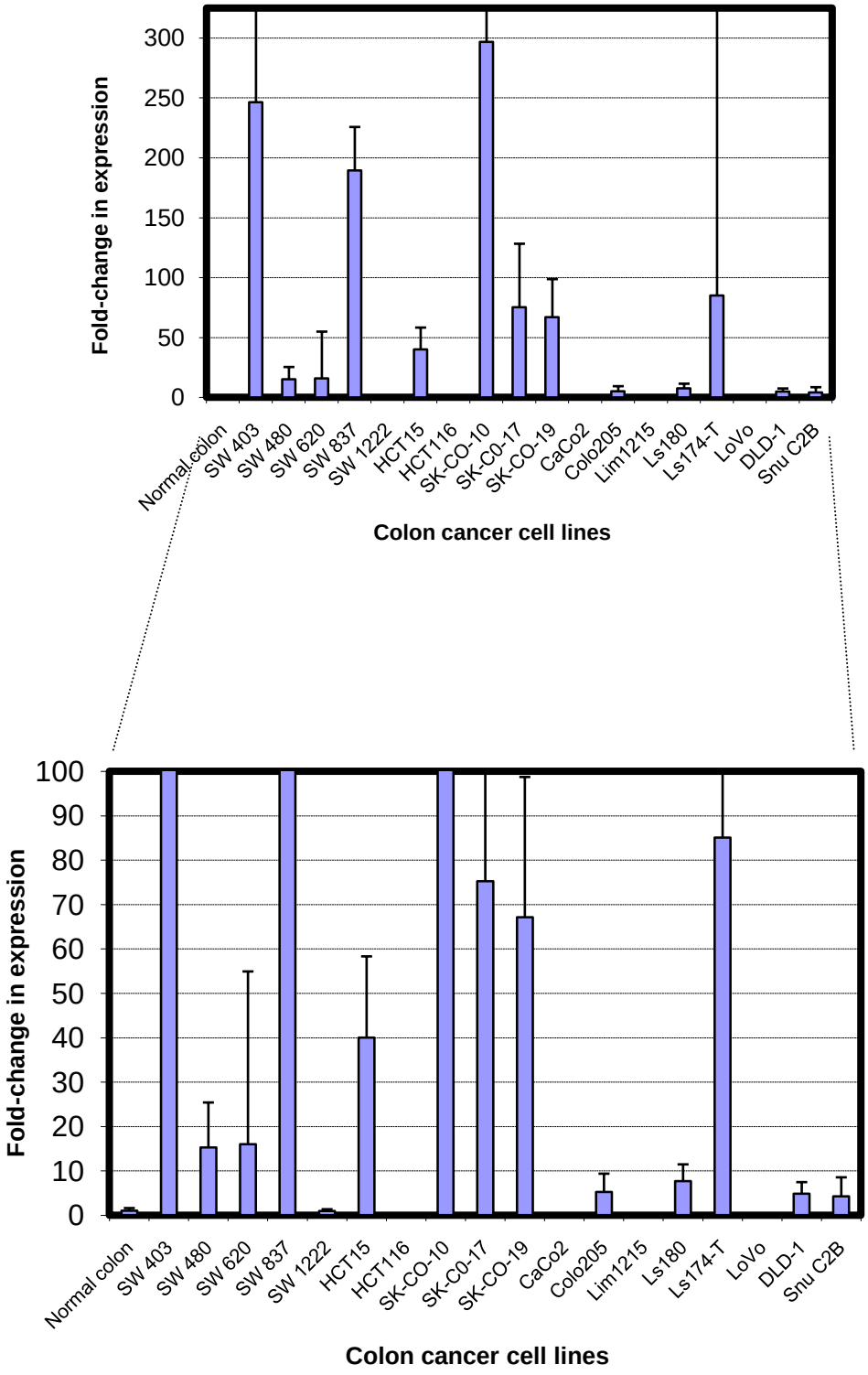

**Figure 3 Quantified expression levels of 4 pTSGs in a panel of lung cancer cell lines. ALAS2 and CDR1 were significantly down-regulated in most cell lines.**

Investigation of expression of CDR1, GABRA3, RRAGB, SLC9A6 and ZCCHC12 in colon cancer cell lines, relative to healthy colon tissue, yielded similar results. CDR1 was significantly down-regulated in all cell-lines (**Figure 4A**). RRAGB, SLC9A6 and ZCCHC12 were down-regulated only in some cell lines (**Figure 4C-E**). GABRA3 was mostly up-regulated (**Figure 4B**). Unexpectedly, dramatic down-regulation of ALAS2 that was detected by conventional PCR could not be verified by the real-time PCR assay because of high amount of primer dimers that emerge in the absence of ALAS2 transcripts. Performance of probe-based real-time PCR is necessary for this purpose.

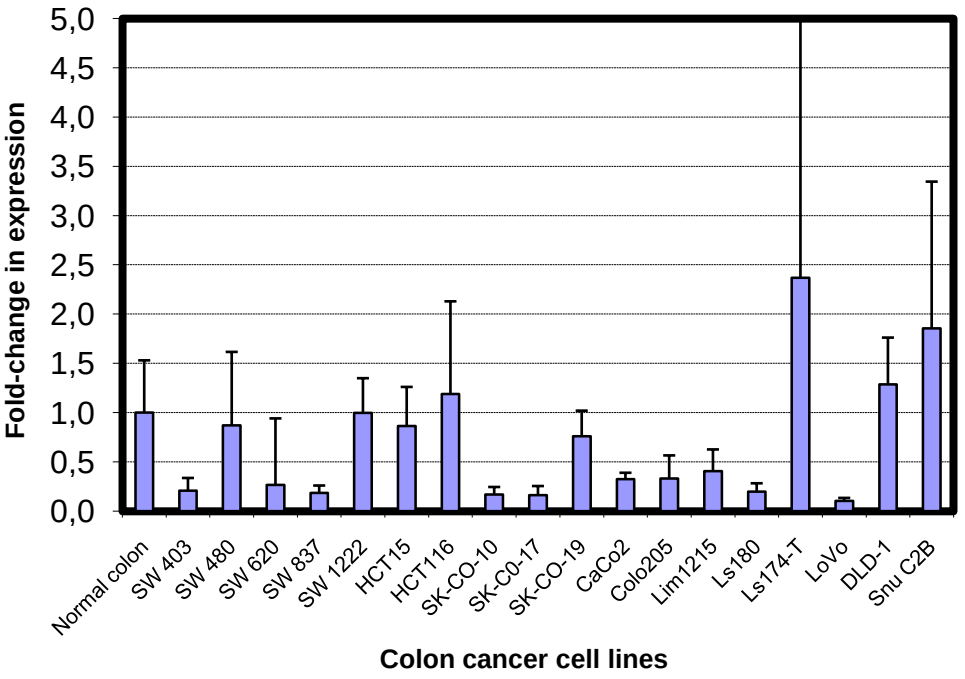
**A)** CDR1



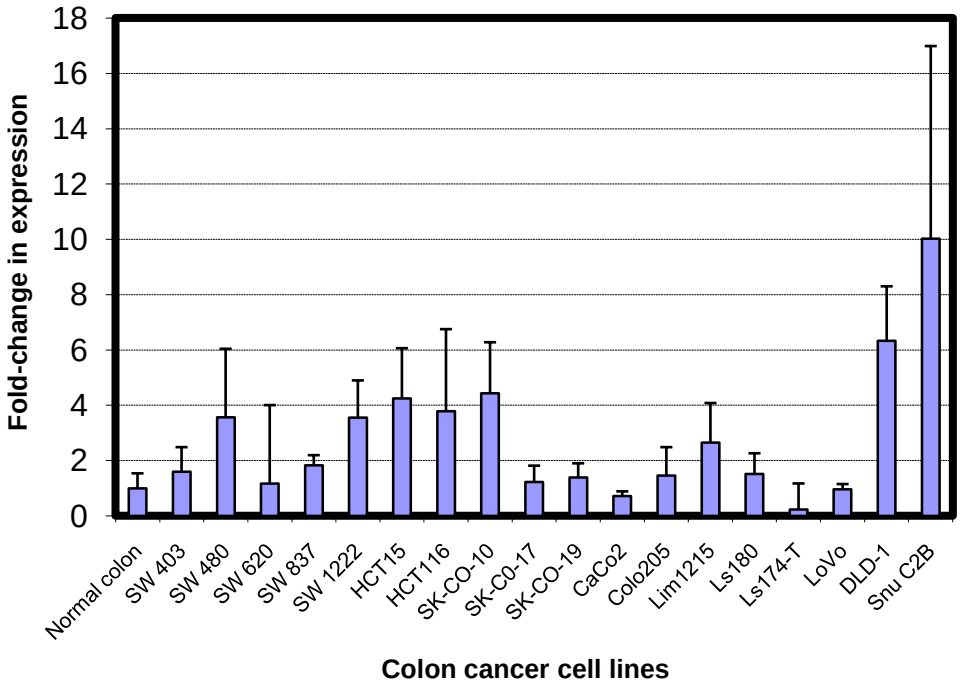
**B)** GABRA3


Fold-change in expression
300
250
200
150
100
50
0
Normal colon
SW 403
SW 480
SW 620
SW 837
SW 1222
HCT15
HCT116
SK-CO-10
SK-C0-17
SK-CO-19
CaCo2
Colo205
Lim1215
Ls180
Ls174-T
LoVo
DLD-1
Snu C2B
Colon cancer cell lines
Fold-change in expression
100
90
80
70
60
50
40
30
20
10
0
Normal colon
SW 403
SW 480
SW 620
SW 837
SW 1222
HCT15
HCT116
SK-CO-10
SK-C0-17
SK-CO-19
CaCo2
Colo205
Lim1215
Ls180
Ls174-T
LoVo
DLD-1
Snu C2B
Colon cancer cell lines

**C)** RRAGB




**D)** SLC9A6



**E)** ZCCHC12

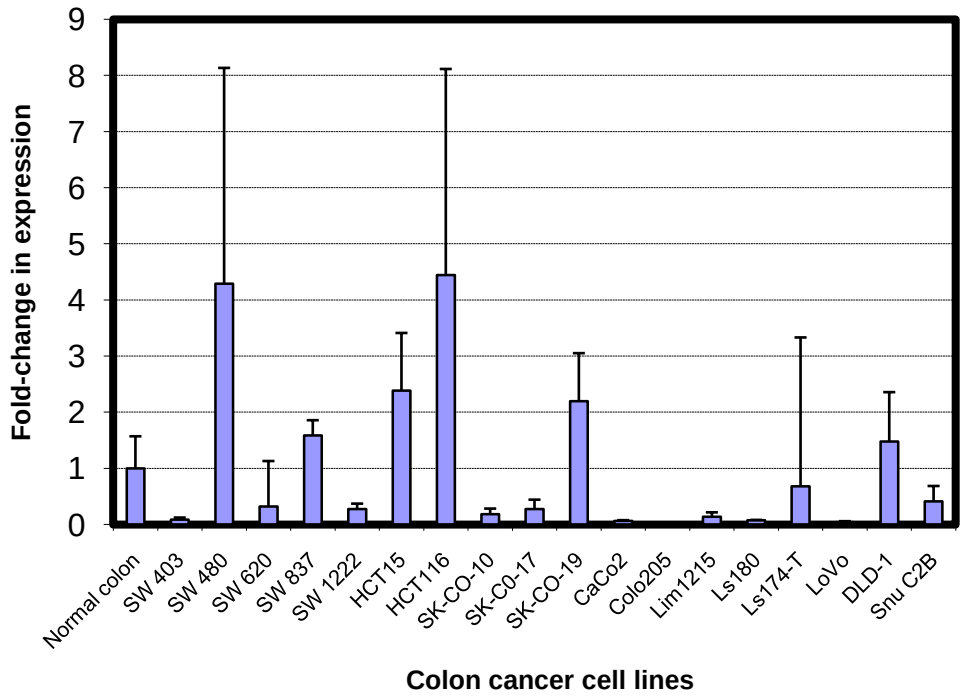



**Figure 4 Quantified expression levels of the 6 pTSGs in a panel of colon cancer cell lines. Graphs show mRNA expression levels on a large scale (small) and a more zoomed-in scale to show details (big), where applicable. CDR1 was significantly down-regulated in all cell lines.**

As a result of the mRNA expression experiments, we chose ALAS2 and CDR1 as the first priority putative TSGs, down-regulated in most colon and lung cancer samples. The other 4 genes might also act as pTSGs under specific conditions. Consistent down-regulation of GABRA3, SLC9A6 and ZCCHC12 in the same cancer cell lines, shown by conventional and real-time RT-PCR experiments, suggests these three genes as second priority tumor suppressors, although their down-regulation is valid to a limited extent. Detailed calculations of the real-time PCR assay are presented in Appendix E.

### 3.4 Analysis of Methylation Statuses of Putative TSGs

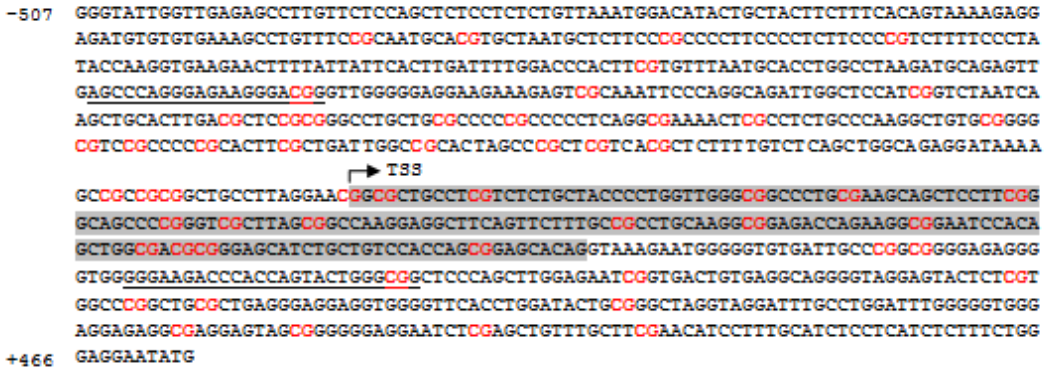
As we have identified putative tumor suppressor genes as explained in sections 3.1 - 3.3, the next question was how these genes were down-regulated in cancer cell lines. Since it is known that, several tumor suppressor genes are silenced in cancer by promoter hypermethylation; we also hypothesized that the pTSGs might be down-regulated via hypermethylation of their promoters. In order to test this hypothesis, firstly, promoters and 5' transcription start sites of the 5 selected pTSG genes were analyzed to find possible CpG islands. Two programs, CpG Island Searcher [103] and CpGPlot [104] were used. Only 2 out of the 5 genes appeared to have CpG islands matching the criteria of observed/expected CpG ratio > 0.60, percent C + G > 50.00, and length > 200. The CpG islands of these genes - SLC9A6 and ZCCHC12 - are presented in **Figure 5A-B**. ALAS2, CDR1 and GABRA3 do not have CpG islands in their promoters, but they have CpG dinucleotides that might be subject to methylation (**Figure 5C-E**).

The methylation states of the CpG dinucleotides were analyzed by sodium bisulfite sequencing. Amplified sites are indicated in Figure 5 and 6A. ALAS2, CDR1 and ZCCHC12 were shown to be heavily hypermethylated when down-regulated (Figure 6C). Even the slight difference in expression of the ZCCHC12 gene in Colo205 and Lim1215 colon cancer cell lines was reflected by differential hypermethylation; more CpG residues in the promoter were methylated in the Colo205 cell line in which expression is less (**Figure 6B**). CpG residues within the 5' region of the CDR1 gene also exhibited this difference.

In the case of ALAS2 gene, the 5' site of the 3$^{rd}$ exon - instead of TSS – was amplified because the first exon of the ALAS2 gene bears very few CpG residues. Therefore, the 3$^{rd}$ exon was chosen instead. The region around the 3$^{rd}$ exon contains 6 CpG residues, as compared to the 2 CpGs found close to the genes' transcription start site (**Figure 5**). The few CpG residues within the amplified region were mostly methylated in all cell lines tested (**Figure 6B**). However, further data emcompassing the whole gene might be required to suggest down-regulation by hypermethylation.

**A)** SLC9A6


**B)** ZCCHC12




**C)** ALAS2

**D)** CDR1


**E)** GABRA3


**Figure 5 CpG dinucleotides present in the promoter sequences of the 5 pTSGs. (A,B) SLC9A6 and ZCCHC12 genes bear classical CpG islands in their promoters. (C-E) ALAS2, CDR1 and GABRA3 do not have CpG islands in their promoters. Few CpG dinucleotides are shown in red. Gray highlights show the exons and underlined sequences show the primers used for sequencing. Transcription start sites are indicated by arrows.**

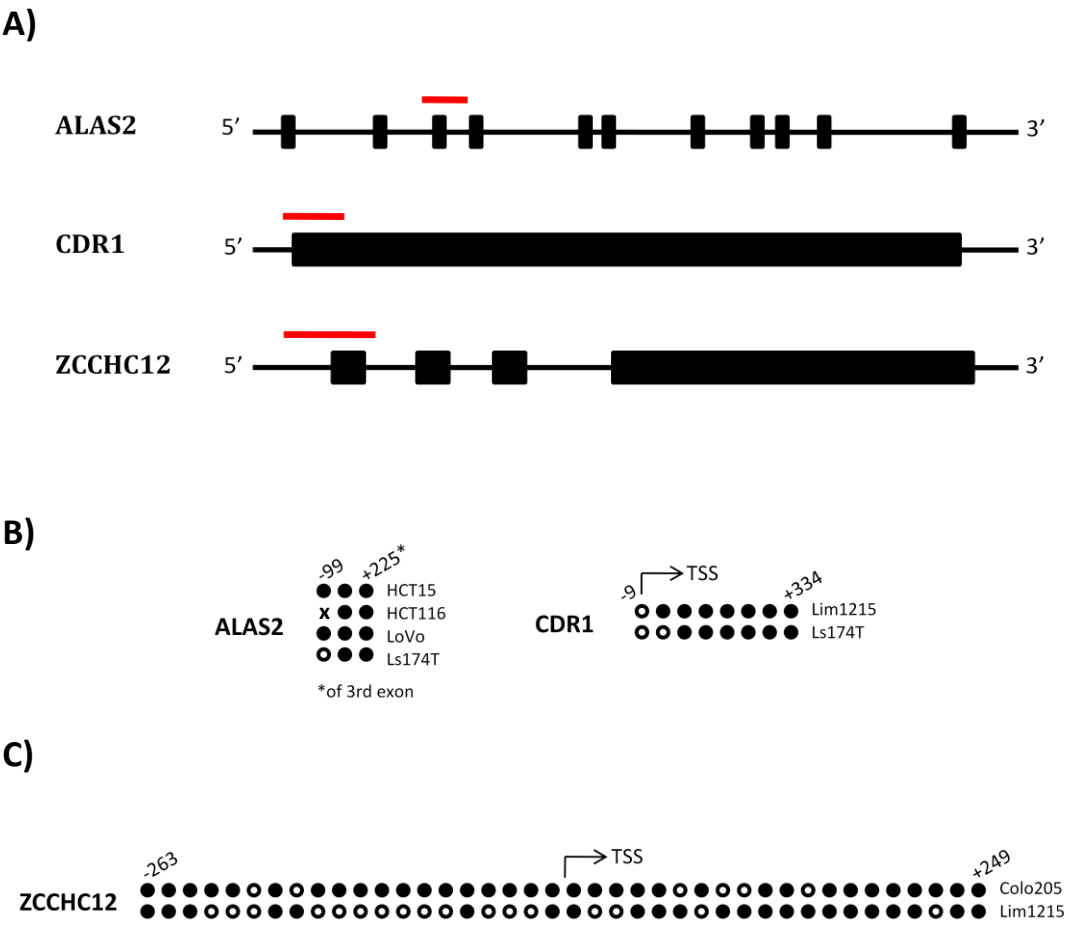



**Figure 6 Analyses of methylation states of the ALAS2, CDR1 and ZCCHC12 genes by sodium bisulfite sequencing. (A) The exon-intron structure of the genes are shown. Light and dark grey boxes denote intron and exons, respectively. Red bars indicate the sites of amplification. (B,C) Results of the assay. Tested CpG residues of the three genes are hyper-methylated in correlation with the expression levels. Each circle denotes a CpG dinucleotide. Open circle: unmethylated, closed circle: methylated. Transcription start sites are indicated by arrows. x refers to non-informative data for the specific CpG dinucleotide.**

Amplified regions of GABRA3 and SLC9A6 genes could not be sequenced successfully due to the repeated nucleotides within the regions in question. An alternative strategy might be required to analyze these two regions.

### 3.5 Analysis of Higher-Order Chromosomal Structure of CTs

Our results suggest hypermethylation of DNA as a major factor regulating expression of putative TSGs. Cancer-testis genes are already known to be regulated by hypomethylation. However, which exact mechanism(s) might result in differential DNA methylation of proximal sites is not known. In order to go one step beyond, we tried to find the underlying mechanism of CT-TSG difference in expression.

It is known that different sets of histone modifications imply different effects on transcriptional regulation, as explained in section 1.2.2. Therefore, the histone modifications present at our whole model regions should have been analyzed as a primary step. ChIP-Seq data covering the whole genome, generated by Barski and colleagues enabled us to analyze the model regions for more than twenty histone methylations together with the insulator-binding protein CTCF [47]. Since CTCF peaks were not consistent at all sites between CTs and pTSGs, we eliminated the possibility of consensus boundary element sequences in these regions. Therefore, the possibility of simple, boundary element mediated difference in expression was also eliminated.

The chromatin modification profiles reported by Barski et al. do not contain information for any of the classical CT gene regions. This observation (which is due to the fact that the Lumina mediated high throughput sequencing analysis discards information if more than one of the same DNA sequences are encountered) prompted us to focus on the "repeat regions" in which the CT genes are contained. We characterized these repeat regions for multiple CT cluster-regions utilizing the Inverted Repeats Finder program [105] and confirmed that CT genes were actually organized into large inverted repeats as a result of ampliconic duplications. Few examples of such regions are illustrated in **Figure 7**.

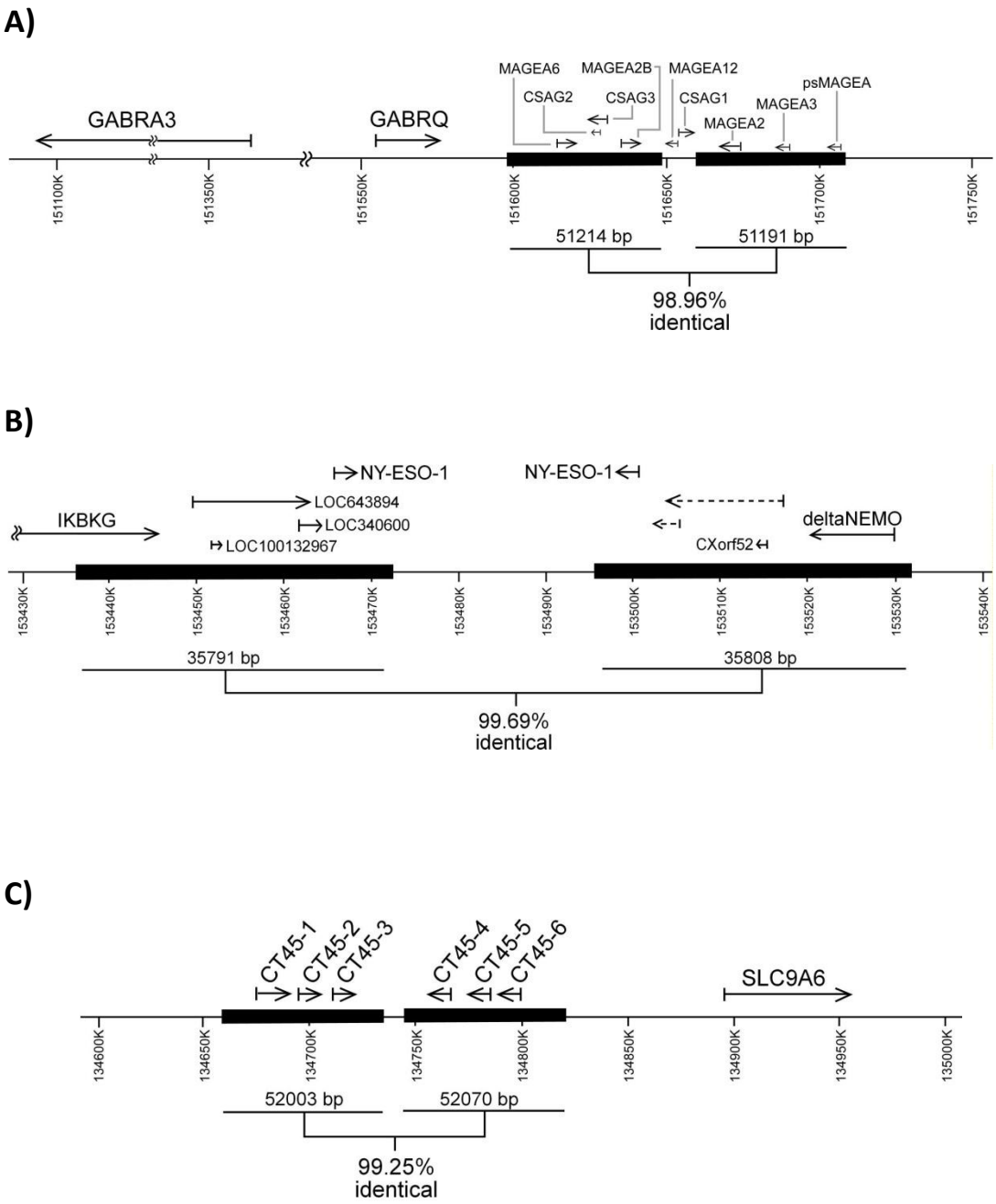



**Figure 7 Genetic structure of some CT-containing inverted repeats. (A) The 51 kb-inverted repeat in the MAGE-A cluster neighboring the GABRA3 pTSG gene. (B) The 36 kb-inverted repeat in the NY-ESO-1. (C) The 52 kb-inverted repeat in the CT45 cluster neighboring the SLC9A6 pTSG gene**

The inverted repeats overlap with the regions missing histone modifications mentioned above, which supports the assay-artifact idea. Since almost all CT genes are contained within inverted repeats, we hypothesized that this very structure itself could be part of the answer to the question of how CT and pTSG regions are separately regulated. Namely, we thought that these inverted repeats could form a self recognizing fold or loop. Therefore, in the next step, we aimed to demonstrate such a higher-order chromosomal structure formed by these regions containing CT genes. However, choosing a pair of chromosomal regions that possibly would form a fold/loop is difficult since many members of a given CT family can be spread over a very large region within the X chromosome and can consist of multiple identical repeats. Thus, we selected the simplest of these, namely the NY-ESO-1 inverted repeat region as a model to study the higher-order chromosomal structure of CT genes. We employed the chromosome conformation capture (3C) assay, which is a biochemical assay to detect intra- and inter-chromosomal interactions. Although the NY-ESO-1 region does not involve a putative TSG, the features of this region might be extrapolated to other CT regions, since CT genes are *coordinately* regulated, probably by same or similar mechanisms.

### 3.5.1 3C Analysis of the Model NY-ESO-1 Region

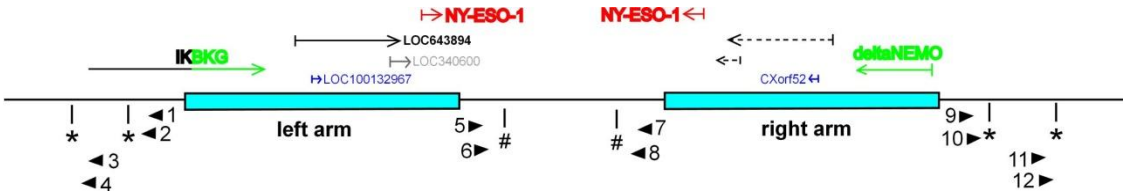
#### *a) Experimental Design*

NY-ESO-1 gene cluster is composed of two-copies of this CT gene on Xq28, and a close family member (LAGE-1) forming the third copy. Upon investigation of this chromosomal region with the Inverted Repeats Finder program, we identified that the two NY-ESO-1 genes were contained within two inverted repeats of ~ 35 kb, separated by ~ 21 kb of unrelated DNA. The two repeats are 99.69 % identical. We hypothesized that these repeats would interact with each other, possibly forming a loop as depicted in **Figure 8C**. We refer to this structure as 'looping' of DNA.

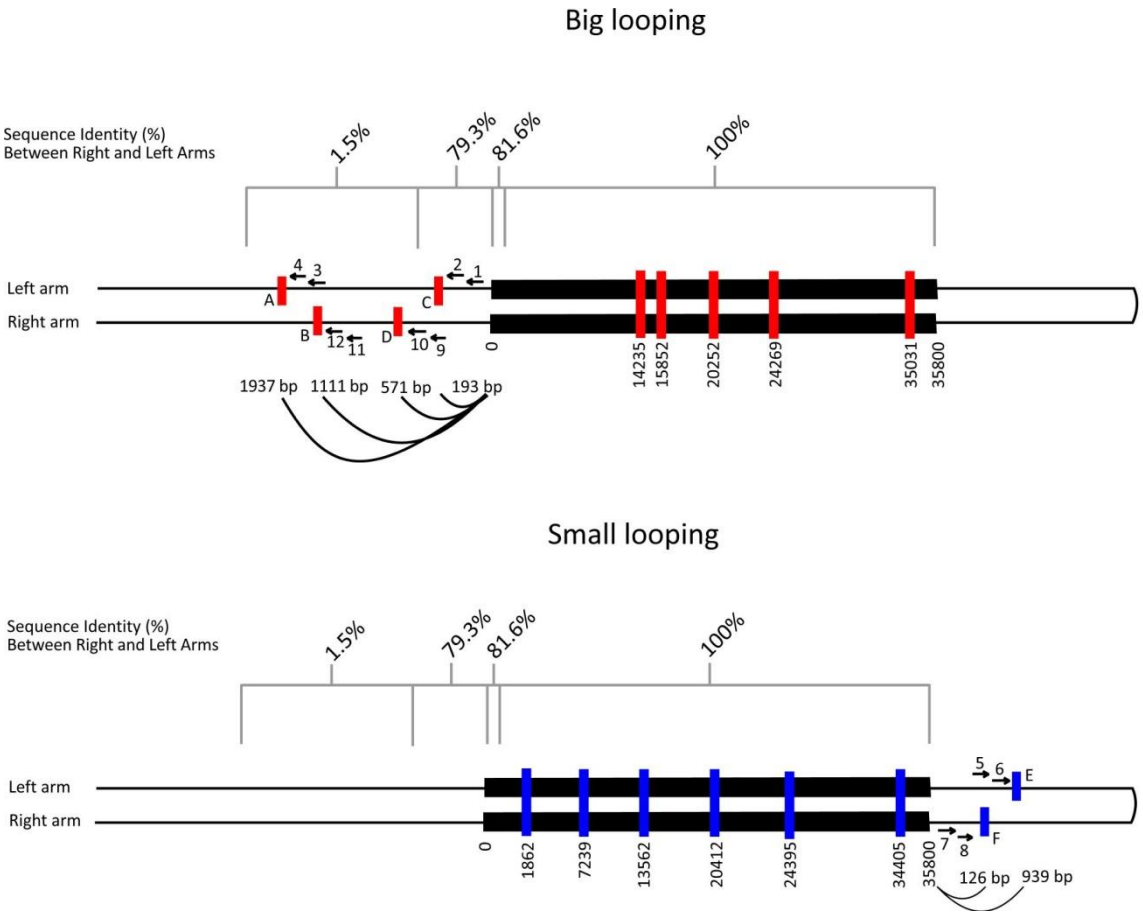
In order to test if this looping occurs or not, we performed the chromosomal conformation capture assay. Briefly, utilizing this technique, DNA is formaldehyde-crosslinked, digested by appropriate restriction endonucleases, followed by

ligation and PCR analysis. An *in vivo* interaction of DNA is thus demonstrated by PCR bands that are otherwise not detectable because distant DNA sites are not crosslinked to each other. We designed 3C primers corresponding to sequences between, as well as distant to the two inverted repeats. Primer and restriction digestion sites used for 3C are shown in **Figure 8**.

**A)**




**B)**




EcoRI restriction sites
BglII restriction sites

Distances between primers and corresponding restriction sites

| | | |
|---|---|---|
| 1-C: 105 | 5-E: 202 | 9-D: 126 |
| 2-C: 49 | 6-E: 108 | 10-D: 93 |
| 3-A: 127 | 7-F: 78 | 11-B: 108 |
| 4-A: 65 | 8-F: 14 | 12-B: 56 |

**C)**

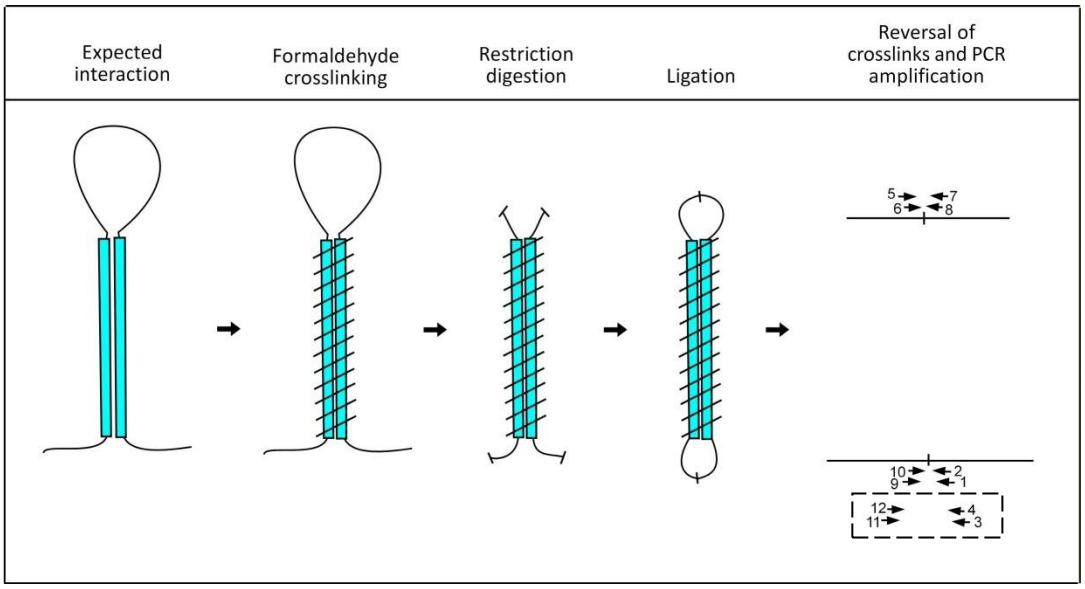



**Figure 8 Investigation of the higher-order chromosomal structure of the NY-ESO-1-containing IRs. (A) The blue boxes correspond to sequences that form the two IRs. Duplicate genes are shown in identical colors. Dashed lines indicate non-annotated transcripts that may exist. Primers used for 3C assay are indicated by arrowheads and numbers. Odd numbers show the first-round primers, even numbers show the primers for nested PCR. # and * denote restriction sites for *BglII* and *EcoRI* endonucleases, respectively. (B) Detailed view of the NY-ESO-1 IR region. *EcoRI* and *BglII* restriction sites with distances to the IR, and percent identities of the region are shown. (C) Expected intra-chromosomal interactions for this region and progress of the 3C assay for this conformation. Primers in the dashed box recognize *EcoRI* restriction sites A and B shown in (B).**

If we could show looping within this region, an equally important question would be whether the formation of this loop would occur differently between healthy and cancerous states. In order to answer these questions, we designed the 3C assay using HT29 and Colo205 colon cancer cell lines that do not express CTs as models, together with DNA obtained after 5-AZA-2'-deoxycytidine-treatment which is known to induce CT gene expression in these cells. LAGE-1 is also a CT gene proximal and similar to NY-ESO-1, therefore expression of it was also

analyzed. The 3C assay was performed also on the SK-LC-17 lung cancer cell line which is known to express most CT genes. Little amount of healthy gall bladder and thyroid tissues were available, and they were included in the investigation of big loop outside of the inverted repeats (IRs) NY-ESO-1 and LAGE-1 expression in untreated and AZA-treated HT29 and Colo205 cell lines are shown in **Figure 9**. As can be seen, AZA treatment of the cells did result in *de novo* NY-ESO-1, as well as LAGE-1 expression in both cell lines, albeit at different intensities.

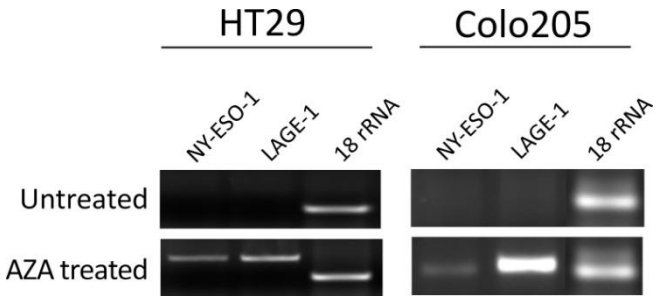



**Figure 9 CT expression levels of the colon cancer cell lines HT29 and Colo205 that were used for the 3C analysis. 5-Aza-2'-deoxycytidine treatment induces NY-ESO-1 and LAGE-1 expression.**

### *b) Chromosomal Juxtapositioning Proximal to the NY-ESO-1 IR*

Results of the 3C assay are presented in **Figure 10.** When genomic DNA was digested with EcoRI and primers # 4 & 12 (with 3 & 11) were used (as shown in **Figure 8**), we were able to observe a 174 bp band for SK-LC-17, the cell line which has strong expression for NY-ESO-1, but not for any other cell line or tissue, suggesting that this chromosomal region had adopted a conformation juxtaposing those regions complementary to the primers (Figure 10A). The same band was observed for the HT29 cell line only after it had been treated with AZA, suggesting that the formation of this chromosomal conformation occurred only upon DNA demethylation in this cell line. A second 3C experiment, aimed to demonstrate the same chromosomal conformation was performed with a different set of primers (Figure 10B). This time, we observed a band only for AZA treated HT29, but not for

SK-LC-17, most likely suggesting that the primers used for this experiment were not as efficient or that the conformation they could detect occurred less efficiently. A third set of primers (# 6 & 8) were designed that would demonstrate chromosomal juxtapositioning of the NY-ESO-1 inverted repeats regions located proximally (see Figure 8). Surprisingly, these primers were able to amplify a band, suggestive of chromosomal juxtapositioning in all cell lines

Because of the limited availability of normal tissue RNA, those were not included in this experiment. For all experiments controls included a known interaction at the gamma-actin region [106]; and the BAC clone RP11-103M23 which was used as a template (**Figure 10D**). To ensure that the obtained PCR products corresponded to the chromosomal regions we anticipated, they were gel purified and sequenced. Sequencing results are given in **Appendix F** and show that the bands correspond exactly to the regions shown in **Figure 8**.

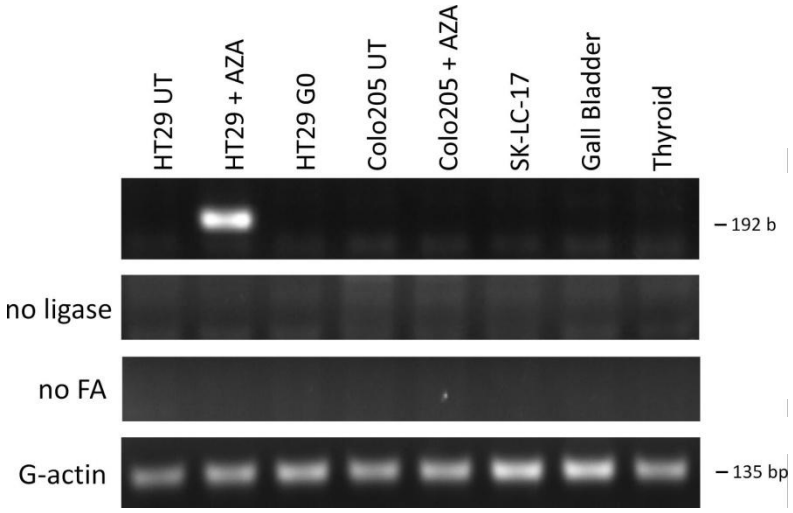
Our results, thus, demonstrate the presence of chromosomal juxtapositioning of regions neighbouring the NY-ESO-1 containing repeat regions. We believe that this juxtapositioning is due to a linear interaction between duplicated regions that abruptly ends towards both ends of the repeat due to the following reasons. Firstly, any interaction 5' to the restriction site cannot be detected by this method. Secondly, although it is difficult to tell if and if where exactly within the repeat region the above proposed interaction occurs, it has to occur within a site after (3' to) the end of the primers and before (5' to) the first EcoRI (or BglII) site downstream to the primer sites, which falls within the repeat (**Figure 8**). That two different sets of primers have both been able to amplify a product from this region strengthens this observation. Interestingly, robust bands demonstrating the juxtapositioning of the repeat regions near the opposite end of the IR, namely the ends which are separated by about 21 kb of non-repeated sequences (**Figure 8**), could be demonstrated for each sample tested. Our results, thus, demonstrate the presence of two chromosomal juxtapositioning events, one that involves both distal ends of the NY-ESO-1 IR, forming a "big loop", and one that possibly results in the looping out of the sequence between the two IRs, forming a "small loop".

### *c) Chromosomal juxtapositioning correlates with NY-ESO-1 expression*

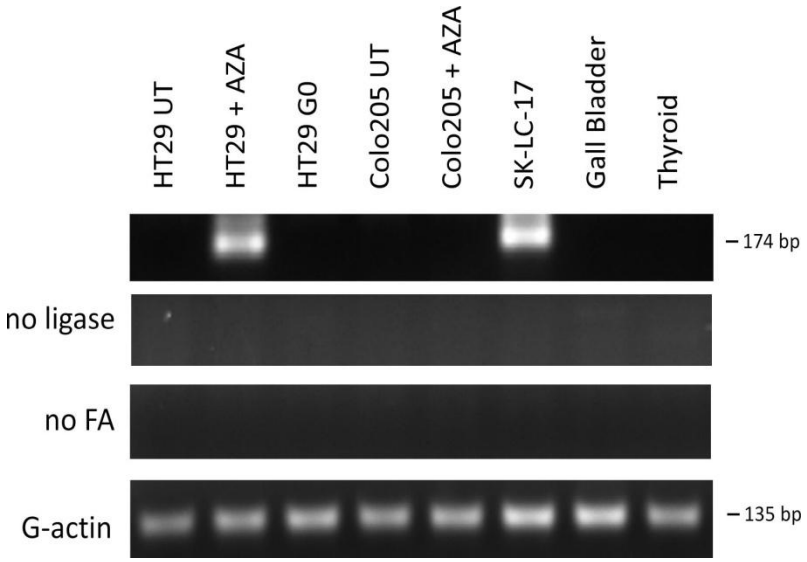
**Figure 10 A** and **B** show that formation of the "big loop" can be detected in a cell line that express NY-ESO-1 (SK-LC-17) and in HT29 cells upon AZA treatment. In cells or tissues where no NY-ESO-1 expression is detected loop formation is also not detected. This suggests a direct relation between the big loop formation and NY-ESO-1 expression. For Colo205, however, despite efficient AZA mediated NY-ESO-1 mRNA induction is observed, the presence of the big loop was never observed. As will be discussed further, this might be due to the following: 1. Loop formation correlates with NY-ESO-1 expression but is not the rule; 2. Loop formation occurs for Colo205 at lower efficiency and the primers utilized for its detection are partially effective; 3. The big loop formation contains either two NY-ESO-1 IRs or one NY-ESO-1 IR and the LAGE homologous region, and these occur at different rates in both cells.

In summary, our results for the first time suggest a correlation of CT gene expression and a particular chromosomal conformation that can respond to variations of DNA methyl content.

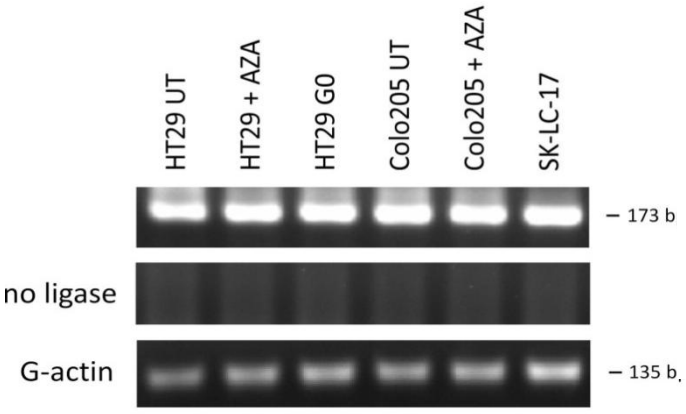
**A)** Primers 2-10



**B)** Primers 4-12




**C)** Primers 6-8




**D)** BAC clone control

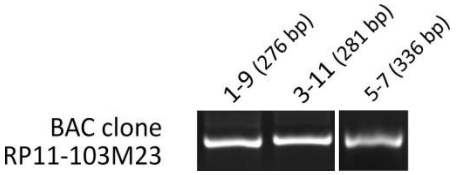



**Figure 10 3C assay for the NY-ESO-1 region. (A): PCR of the same DNA with primers # 3 & 11,reamplified using primers # 4 & 12. (B) PCR of crosslinked and EcoRI digested/ligated genomic DNA with primers # 1 & 9, reamplified using primers # 2 & 10 (see Figure 8 for primer information). (C) PCR of BglII digested/ligated DNA using primers # 6 & 8. (D) Control PCR reactions. UT: untreated; G0: cells synchronized at G0; FA: formaldehyde.**

### *d) Chromosomal Looping Correlates with Expression Originating within the Loop*

The above described data strongly supports our hypothesis that the NY-ESO-1 IRs interact linearly, forming a big and a small loop. Since the constituents of this loop are thus identical DNA sequences, fine mapping of the exact start and end points of this linear interaction utilizing 3C cannot be performed. However, should this linear interaction occur, then other transcripts that originate within these IRs would be expected to show an expression pattern similar to NY-ESO-1 and not like those genes outside the IRs. To test for this hypothesis we searched the ENCODE data present in the UCSC Genome Browser [112] for the non-coding transcripts since there is only one other annotated gene within these IRs (LOC 643894). Various annotated ncRNAs as well as the genes mapped to this region are shown in **Figure 11**. 5RACE122 and 5RACE117, for example, are ncRNA molecules which were detected in brain and colon, respectively and cover large regions. In order to analyze the whole IR region for the presence of ncRNA transcripts - either annotated or non-annotated -, 4 primer pairs were designed from within the IR and the spacer DNA (B, D, F, G primers in **Figure 12**). Primers for detecting the mRNA levels of the genes inside the repeat were also designed (A, C and E primers in **Figure 12**).

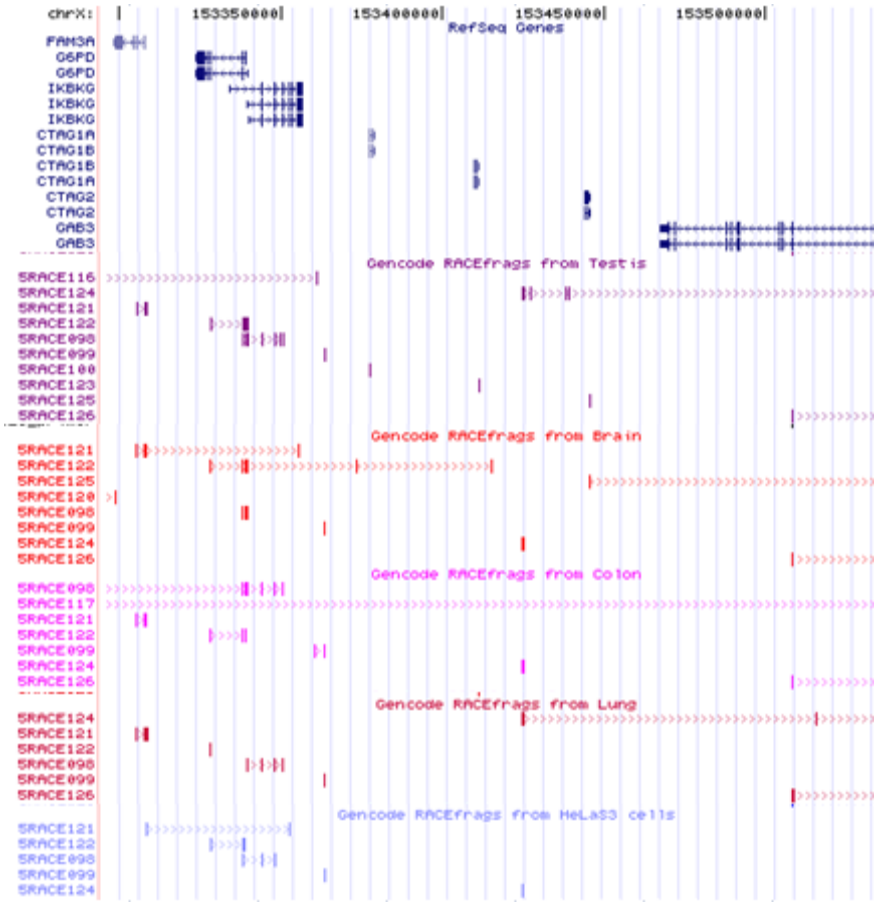



**Figure 11 Annotated transcripts in the NY-ESO-1-bearing inverted repeat region; ENCODE Project at UCSC Genome Browser. Healthy testis, brain, colon and lung tissues together with a HeLa sample are shown.**

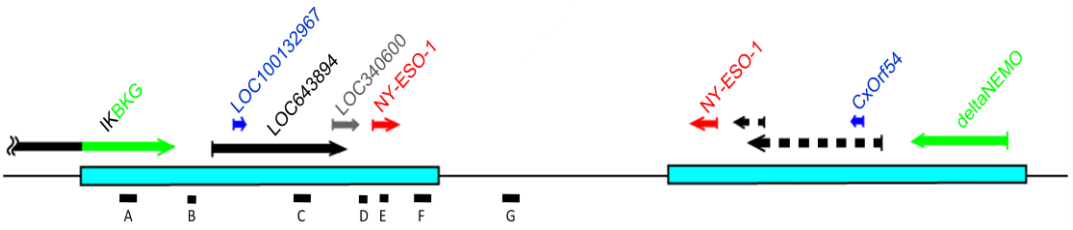



**Figure 12. Primer pairs for detecting mRNAs and non-coding RNAs from within the inverted repeat(s) (primer pairs A-F) and inside the small loop between the two repeats (primer pair G). Primer pairs A, C and E correspond to mRNAs of genes IKBKG, LOC 643894 AND NY-ESO-1. B, D and F regions correspond to non-coding RNA transcription sites. Blue boxes correspond to sequences that form the two IRs. Duplicate genes are shown in identical colors. Dashed lines indicate non-annotated transcripts that may exist.**

The same cell lines with the 3C assay were used for comparison of the results: untreated and AZA-treated HT29 and Colo205, and SK-LC-17. Normal lung, colon and testis RNAs were also used as reference. PCR results are shown in **Figure 13**. IKBKG (A) primers as well as ncRNA1 (B) primers amplified a product in all samples tested, demonstrating discordance with NY-ESO-1 transcription. We could not detect a transcript corresponding to LOC643894 (C). Primers amplifying the two ncRNAs flanking the NY-ESO-1 gene, ncRNA2 (D) and ncRNA3 (F) amplified a product most prominently in SK-LC-17, but also weakly in normal tissues, and interestingly, both ncRNAs were induced by AZA treatment in both HT29 as well as Colo205, albeit weekly. Transcript G was weakly expressed in all three cell lines tested but did not show a particular pattern. Therefore, the transcript analysis of the NY-ESO-1 regions demonstrates a close correlation of transcription patterns of ncRNAs 2, -3 and NY-ESO-1, while transcripts further upstream and downstream to these demonstrate a different expression pattern, supporting the hypothesis that a common regulatory event encompasses and extends beyond the physical boundaries of the NY-ESO-1 gene. Although this common regulatory pattern does not seem to prevail throughout the IR, as demonstrated by the behavior of transcripts A and B. However, the cumulative data suggests that there are common expression patterns within the IRs and that their boundary closely correlates with the presence of the IR, at least at one end of the repeat.

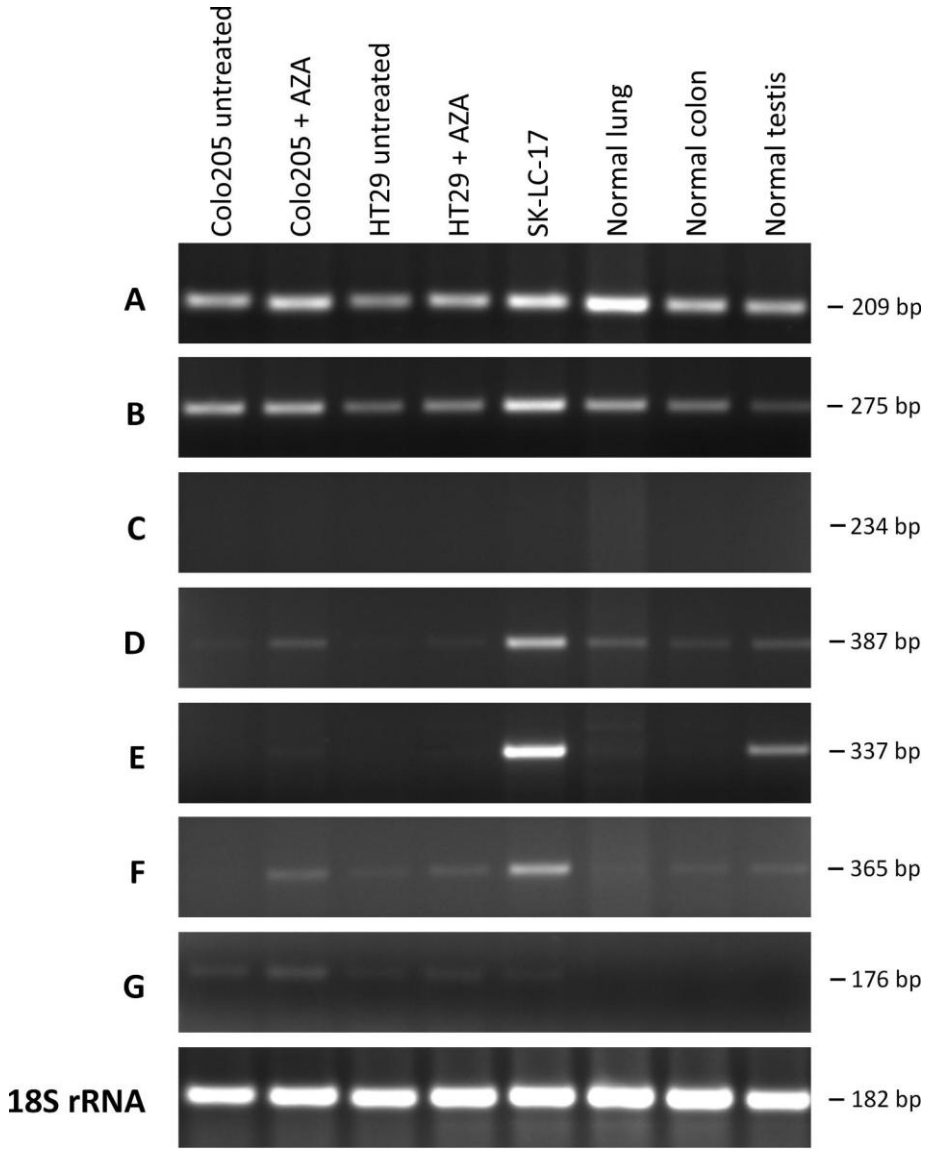



**Figure 13. mRNAs and ncRNAs expressed within the inverted repeat(s) (A-F) and within the non-repeated region between the two IRs (G). A=IKBKG; B=ncRNA1; C= LOC 643894; D=ncRNA2; E:NY-ESO-1; F=ncRNA3; G=ncRNA4.**

# 4 DISCUSSION & FUTURE PERSPECTIVES

In this study, our main aim was to understand the mechanism(s) of regulation of CT gene expression. Hypo-methylation of DNA is known be one factor reactivating CT genes in testis and during carcinogenesis. Histone modifications play unclear roles. It is of major interest to understand CT regulation since reactivation of CT genes is correlated with poor prognosis of cancer. We approached this question in a manner that combines hypo-methylated and hyper-methylated DNA sites under the same frame, so that we could observe the gradual or strict passage in between these regions. Here, tumor suppressor genes (TSGs) correspond to the hyper-methylated genes class based on previous studies supporting this idea for the cancer model. Focusing on this kind of system would also facilitate to understand ectopic and simultaneous DNA (de-)methylation that occurs in cancer.

***Newly Identified Putative Tumor Suppressor Genes***

The necessity to identify model regions containing hypo- and hyper-methylation in close proximity leaded us to detect putative TSGs neighboring CT genes on the X chromosome. As a result of *in silico* data mining and expression assays, we here suggest five X-linked putative tumor suppressor genes: ALAS2, CDR1, GABRA3, SLC9A6 and ZCCHC12. ALAS2 and CDR1 are the genes most significantly down-regulated in most of the cancer cell lines tested. They were already studied by other groups, however literature regarding these two genes is based on their associated diseases sideroblastic anemia and cerebellar degeneration [113, 114]. Interestingly, histone acetyltransferase p300 and Sp1 proteins were shown to regulate ALAS2 expression in erythroid cells [115].

Tumor suppressor genes are known to be frequently down-regulated by promoter DNA hyper-methylation in a wide variety of cancer types. $p16^{INK4a}$ and BRCA1 are

among the mostly referred examples [93]. It was likely that the same mechanism could be responsible for down-regulation of our pTSGs, therefore we have initially analyzed the methylation levels. Bisulfite sequencing data indicated hyper-methylation of DNA to be an acting mechanism for reducing expression of the ALAS2, CRD1 and ZCCHC12 genes, although only ZCCHC12 gene contained a CpG island in its promoter region. Analyzed CpG dinucleotides belonging to ALAS2 and CDR1 genes being hyper-methylated suggest aberrant methylation as a powerful regulatory mechanism even in the absence of CG-rich sequences. Further methylation analyses are though required. Firstly, T/A-cloning-wise amplification of the PCR fragments is necessary to screen the methylation rate of the cell population. Selection of many clones is an advantage of this method. Sequencing of plasmid DNA (of the clones) would also be easier than sequencing the PCR product, since faint bands are often obtained as a result of weak amplification of sodium bisulfite-treated DNA. Secondly, more DNA samples could be used for comparison. Two cell lines were used for each of CDR1 and ZCCHC12 genes (4 cell lines for ALAS2); which gives a general idea about regulation however one can make more accurate deductions if more cell lines are used. Moreover, healthy tissue DNA should better be used to compare the results to cancer cell lines, although it is not an obligation. Normal tissue samples can rarely be obtained, thus peripheral blood lymphocyte DNA is an alternative control that is easily handled. Thirdly, larger DNA sequences within the promoter and within the coding region should be screened - by multiple primers if necessary - to learn the target region of the DNA methylation machinery. The other two putative TSGs, GABRA3 and SLC9A6 genes, should also be analyzed in terms of promoter methylation levels either by bisulfite sequencing or by other means like melting curve or COBRA (Combined Bisulfite Restriction Analysis) techniques. Bisulfite sequencing of amplified regions of these two genes were unsuccessful, probably because of the short repeated elements within the regions complicating the sequencing reaction. Lastly, investigation of mRNA expression levels of the pTSGs after induction with 5-Aza-2'-deoxycytidine is necessary to show a direct link between mRNA expression and DNA hypermethylation.

One critical issue while carrying out bisulfite sequencing is the possibility of PCR bias of the designed primers. Since methylated cytosines retain as 'C's and unmethylated cytosines are converted to 'T's during PCR amplification, primers could bind to methylated sequences tighter than the unmethylated ones, which in turn would seem like the DNA is more methylated in the original sample. The resulting conclusions would then be inaccurate and misleading. In order to eliminate such situations, each primer pair should be tested to recognize different percents of unmethylated/methylated DNA mixtures and improper pairs should be avoided.

We have shown the down-regulation and promoter hyper-methylation of the newly identified pTSGs in cancer cell lines, which were necessary experiments to be done in order to determine our model sites on the X chromosome. Further experiments about understanding the regulation of proximal hypo- and hyper-methylated regions are discussed below. As we have proposed novel pTSGs, another project arises at this point. In order to suggest these genes as TSGs, further investigation should be done. Firstly, tumor suppression abilities could be assessed in an apoptosis system. Cloning of the genes into mammalian expression plasmids is necessary. However, expression from the often-used plasmids is itself toxic to cells and complicates the apoptosis assays. Instead, an inducible-plasmid system would be utilized and would give more reliable results. In this system, verification of the tumor suppression ability could be done first by simply counting the number of viable cells before and after induction; and also by analyzing the levels of apoptotic proteins activated/anti-apoptotic proteins repressed after induction. Invasion assays could also be carried out to observe any effects on the metastatic abilities.

### *Alteration of Higher-Order Chromosomal Structure Affecting NY-ESO-1 Expression*

As previously mentioned, our main focus was on understanding the regulation of CT gene expression by investigating the determined *in vivo* model system on the X chromosome. Initially, we tended to find candidate boundary elements in

between CT genes and pTSGs, as this would be an expected situation when two proximal sites are oppositely regulated. Boundary element-mediated insulation of chromosomal marks such as DNA methylation and histone modifications have been shown to be valid for many cases and can be found in literature [116, 117]. As demonstrated in **Figure 14**, this idea would perfectly fit into our model since permissive and repressive histone modifications are likely to play key roles in regulation of CTs and pTSGs in cancer, respectively. Testing of this idea was done by utilizing genome-wide chromatin immunoprecipitation data for the CTCF protein because CTCF, apart from its role as a transcription factor, binds to many insulator elements, thus could be regarded as an 'insulator protein' [118].

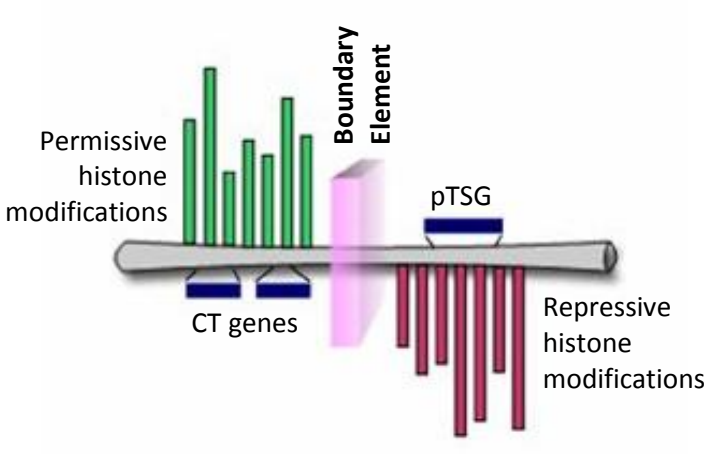



**Figure 14 Model of boundary element-mediated difference in regulation of CT and TSG gene expression**

As a result of the analysis of CTCF binding, no consensus boundary elements could be found within our model regions. This unexpected result raised questions about the types and levels of histone modifications at these sites. Analyzing the same ChIP-Seq data from Barski et al. [47], we came across other unexpected results such that none of more than twenty histone methylations were detected around CT gene clusters. Absence of histones at these sites resembled absence at sites of DNA repeats. As the next step, some sample CT clusters were investigated under the Inverted Repeats Finder program [105] and clusters of multi-copy CT genes were found to be organized into large inverted repeats as a result of ampliconic

duplications on the X chromosome. This explains the high number of pseudogenes of cancer-testis genes.

The repeat-embedded structure of the CT genes, exemplified in **Figure 7** for MAGE and CT45 clusters and NY-ESO-1, could be a candidate regulator of CT gene expression by inducing changes in chromosomal structure. It is also a candidate mechanism that leads to differences in our model CT/pTSG regions, because the pTSGs lie outside the inverted repeats. To test these hypotheses, we employed the biochemical 3C assay which is useful to gather information about the higher-order conformations of chromatin. We used untreated and 5-Aza-2'-deoxycytidin-treated HT29 colon cancer cell line, SK-LC-17 lung cancer cell line and healthy thyroid and gall bladder tissues whenever possible. As a result of all experiments, we observed changes in higher-order structure of chromatin in the NY-ESO-1-containing inverted repeat region on Xq28. As illustrated in the model (**Figure 15**), the two inverted repeats spatially converge to each other in the 5-Aza-2'-deoxycytidin-treated HT29 cell line and in the SK-LC-17 cell line, which both have high NY-ESO-1 expression. This 'large looping' is not observed in any of the untreated HT29 cell line and healthy tissues. Only the 'small loop' between the inverted repeats is observed, however this loop is observed in all cases tested (**Figure 10**), which means that it does not play a regulatory role on this system. The large looping interaction is also likely to correlate with the non-coding RNAs detected to be transcribed within the repeats when NY-ESO-1 is also transcribed (**Figure 13**).

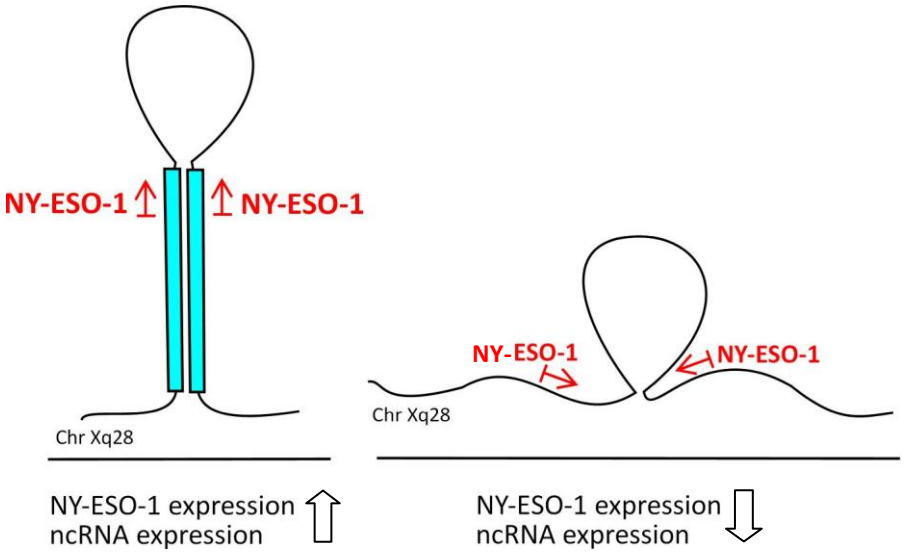


**Figure 15 Model regarding epigenetic regulation of CT gene expression: Involvement of changes in higher-order chromosomal structure**

We thus have a model regarding epigenetic regulation of CT gene expression, in which chromosomal structure is the key player for now. Working on this model by manipulating different proteins and/or pathways would produce high amount of data and contribute to the understanding of CT gene regulation.

Before passing to further experiments to be done on this system, the weaknesses of the results demonstrated here should be discussed. First of all, the 3C assay is an extremely delicate assay. It combines many techniques like formaldehyde-crosslinking, restriction digestion, ligation and purification of DNA and is followed by PCR experiments with many different primers. The complicated nature of the whole assay and the difficulty to handle *in vivo* interactions of DNA brings sensitivity problems. Even technical PCR replicates do not give the same results all the time. To overcome these problems; firstly a probe-based real-time system should be used for PCR analyses. Sybr green based real-time systems are useless for this purpose since many weak background interactions would create high degree of noise. The probe-based system would bring many advantages: Weak or delicate interactions that cannot be detected by conventional PCR can be detected by the sensitive probe-based system. For example, we could not detect any difference in the NY-ESO-1 region between untreated and AZA-treated Colo205 cell lines, which indeed have difference in CT expression. On the other hand, using a probe-based real-time PCR system would enable to calculate the interaction frequency of the region [119]. For this purpose, first of all a wider area to the outside of the inverted repeats from both sides should be screened by designing new primers corresponding to these sites. After performing real-time PCRs with the desired samples and BAC controls, sample data should be normalized to BAC data to give interaction frequency of the looping. An actual loop is the one which gives higher interaction frequency (a peak) as the primers approach to loop from both sides. Lastly to overcome the sensitivity problems, three biological replicates should be done as usual and one should pay attention at all steps. Usage of male

cells should be preferred in order to avoid complexity coming from X chromosome inactivation in female cells. The SK-LC-17 and Colo205 cell lines are already male cell lines, however HT29 is a female cell line and although the chromosome conformation changes upon induction with 5-Aza-2'-deoxycytidine, comparison of male and female cell lines might raise artifacts.

In this study, we only designed probes against the start and end points of the repeats to have a general idea if they converge or not. Therefore, we currently do not know if the arms of the inverted repeat completely touch each other or not. Answering this question is not easy by using the 3C assay since primers cannot be designed because of the high degree of identity between the arms. Therefore, another technique needs to be applied both to understand the intra-structure of the repeat arms and also to verify the 3C data. Visual techniques are the most commonly used ones when studying on DNA repeats. Two-color DNA *in situ* hybridization or FRET (Fluorescence Resonance Energy Transfer) techniques could be applied to gather *in vivo* data by fluorescent microscopy. AFM/SFM (Atomic/Scanning Force Microscopy) is another choice but would only give *in vitro* data.

Once the looping interaction is shown by performing probe-based real-time experiments, characterization of the protein complexes that mediate the looping should be done in order to fully understand the components of the system. RNAi pathway is a candidate player in this system since non-coding RNAs are transcribed from inside the repeat arms. Knocking down some critical components of the RNAi machinery, like Dicer or Argonaute, and then analyzing any alteration in looping by 3C is a choice. Moreover, ChIP technique could be employed to detect proteins bound to the loop. Antibodies against CTCF, BORIS and Sp1 proteins can be initially used because of literature suggesting their regulatory roles on NY-ESO-1 expression [24, 25]. Once the protein complex is pulled up by a known protein component, the whole complex can be characterized by co-IP, mass spectroscopy, chromatography or other proper methods.

The results shown in **Figure 10** leading to our model (**Figure 15**) were gathered from the NY-ESO-1-bearing inverted repeat. The simplicity of this region - that it bears only two copies of a CT gene - was the reason why we worked on this system. If all the above experiments are done and the acting proteins can be characterized; same assays would be extended to other, more complex CT regions also. This is easier to perform after characterization of protein components because ChIP technique would facilitate the analysis. Visual techniques are also more powerful than the 3C assay. We tried to apply the 3C assay to the MAGE and CT45 clusters shown in **Figure 7** also, however could not obtain consistent results possibly because of complex intra-chromosomal interactions of the genes embedded in the repeat. Application of same analyses to many CT regions is necessary to support the observation that CT genes are coordinately regulated [11]. If not all, many of the CT clusters are expected to be regulated by same pathways. Conclusions and hypotheses about CT regulation in general are discussed below.

***Epigenetic Regulation of Cancer-Testis Gene Expression***

Combining our data with the ones in literature provides enough background to hypothesize a general model of CT regulation. Below is a summary of known facts and our results about regulation of CT gene expression:

- CT genes are re-activated in cancer cells by selective demethylation of promoter DNA [12, 32-35].
- Histone acetylation plays role in modulating CT gene expression [32, 35].
- Acetylation-induced transcription is required for active DNA demethylation of CT genes, resulting in their reactivation [35].
- CT genes are coordinately expressed and are correlated with poor prognosis in cancer (at least in non small cell lung cancer) [11].
- Members of the MAGE-A CT family, clustered in a large inverted repeat on the X chromosome, are coordinately expressed [120].

- Non-coding RNAs are transcribed from sites of CT gene clusters; shown by us for the NY-ESO-1 region, found in literature for GAGE region and also shown for SSX cluster (unpublished data) [35].
- Expression of NY-ESO-1 genes overlaps with expression of non-coding RNAs inside the inverted repeat arms.
- Change in expression profile of NY-ESO-1 gene leads to alteration of higher-order chromosomal structure; or vice versa.
- Effects of differential chromosomal structure and hypomethylation (that induces CTs) are likely to be restricted to the inverted repeat region, since genes outside the repeats are either constantly expressed or down-regulated in cases when CT genes are up-regulated.
- Transcription factors BORIS, CTCF and Sp1 bind to promoters of CT genes and modulate their expression [24, 25].
- CT antigens are capable of interacting with each other, as MAGEC1 and NY-ESO-1 do [27].

With these information in hand, such a model about regulation of CT gene expression is proposed here:

> Improperly functioning epigenetic machinery in cancer cells leads to acetylation of histones and other permissive histone modifications at sites of CT clusters, which induces transcription of ncRNAs within the repeat area. Transcription itself or ncRNAs processed in the RNAi pathway mediate demethylation of the promoter (or the whole area), which in turn results in reactivation of CT genes. Alteration of chromosomal structure in the form of looping occurs either as a cause or as a consequence of transcription of ncRNAs; so that factors regulating CT gene expression like BORIS, CTCF and Sp1 can bind to the CT promoter and help express CTs that are clustered together. Many CT gene families are clustered into repeat regions, thus the same conditions apply, to induce their coordinate expression. Since CT genes occupy large regions on X chromosome, activation of them is regarded as an immediate sign of improper epigenetic environment within cancer cells.

Working on this model would bring important understanding in operation of the epigenetic systems within the cell, which all are indeed connected to each other through shared components. Since repeat-based and non-coding RNA-containing silencing events are not specific to CT genes and are valid for many other sites, tackling the CT problem aids in clarification of the much more general question of how epigenetic mechanisms relate to each other and why they are misregulated in cancer.

# 5 REFERENCES


1. Boel, P., et al., *BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes.* Immunity, 1995. **2**(2): p. 167-75.
2. van der Bruggen, P., et al., *A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma.* Science, 1991. **254**(5038): p. 1643-7.
3. Van den Eynde, B., et al., *A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma.* J Exp Med, 1995. **182**(3): p. 689-98.
4. Tureci, O., et al., *The SSX-2 gene, which is involved in the t(X;18) translocation of synovial sarcomas, codes for the human tumor antigen HOM-MEL-40.* Cancer Res, 1996. **56**(20): p. 4766-72.
5. Chen, Y.T., et al., *A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1914-8.
6. Pfreundschuh, M., et al., *Serological analysis of cell surface antigens of malignant human brain tumors.* Proc Natl Acad Sci U S A, 1978. **75**(10): p. 5122-6.
7. Sahin, U., et al., *Human neoplasms elicit multiple specific immune responses in the autologous host.* Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11810-3.
8. Chen, Y.T., et al., *Identification of cancer/testis-antigen genes by massively parallel signature sequencing.* Proc Natl Acad Sci U S A, 2005. **102**(22): p. 7940-5.
9. Chen, Y.T., et al., *Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6919-23.
10. Stevenson, B.J., et al., *Rapid evolution of cancer/testis genes on the X chromosome.* BMC Genomics, 2007. **8**: p. 129.
11. Gure, A.O., et al., *Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer.* Clin Cancer Res, 2005. **11**(22): p. 8055-62.
12. Qiu, G., J. Fang, and Y. He, *5' CpG island methylation analysis identifies the MAGE-A1 and MAGE-A3 genes as potential markers of HCC.* Clin Biochem, 2006. **39**(3): p. 259-66.
13. D'Arcy, V., et al., *The potential of BORIS detected in the leukocytes of breast cancer patients as an early marker of tumorigenesis.* Clin Cancer Res, 2006. **12**(20 Pt 1): p. 5978-86.
14. Suri, A., *Cancer testis antigens--their importance in immunotherapy and in the early detection of cancer.* Expert Opin Biol Ther, 2006. **6**(4): p. 379-89.

15. Simpson, A.J., et al., *Cancer/testis antigens, gametogenesis and cancer.* Nat Rev Cancer, 2005. **5**(8): p. 615-25.
16. Chen, Y., *Identification of human tumor antigens by serological expression cloning: an online review on SEREX. [updated 2004 Mar 10; cited 2008 August]* Cancer Immunity, 2004.
17. Chomez, P., et al., *An overview of the MAGE gene family with the identification of all human members of the family.* Cancer Res, 2001. **61**(14): p. 5544-51.
18. Kirkin, A.F., K.N. Dzhandzhugazyan, and J. Zeuthen, *Cancer/testis antigens: structural and immunobiological properties.* Cancer Invest, 2002. **20**(2): p. 222-36.
19. Xiao, J. and H.S. Chen, *Biological functions of melanoma-associated antigens.* World J Gastroenterol, 2004. **10**(13): p. 1849-53.
20. Cho, B., et al., *Identification and characterization of a novel cancer/testis antigen gene CAGE.* Biochem Biophys Res Commun, 2002. **292**(3): p. 715-26.
21. Nishimura, I., J.Y. Sakoda, and K. Yoshikawa, *Drosophila MAGE controls neural precursor proliferation in postembryonic neurogenesis.* Neuroscience, 2008.
22. Kondo, T., et al., *The cancer/testis antigen melanoma-associated antigen-A3/A6 is a novel target of fibroblast growth factor receptor 2-IIIb through histone H3 modifications in thyroid cancer.* Clin Cancer Res, 2007. **13**(16): p. 4713-20.
23. Vatolin, S., et al., *Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes.* Cancer Res, 2005. **65**(17): p. 7751-62.
24. Hong, J.A., et al., *Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells.* Cancer Res, 2005. **65**(17): p. 7763-74.
25. Kang, Y., et al., *Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells.* Oncogene, 2007. **26**(30): p. 4394-403.
26. Cronwright, G., et al., *Cancer/testis antigen expression in human mesenchymal stem cells: down-regulation of SSX impairs cell migration and matrix metalloproteinase 2 expression.* Cancer Res, 2005. **65**(6): p. 2207-15.
27. Cho, H.J., et al., *Physical interaction of two cancer-testis antigens, MAGE-C1 (CT7) and NY-ESO-1 (CT6).* Cancer Immun, 2006. **6**: p. 12.
28. Liu, Y., Q. Zhu, and N. Zhu, *Recent duplication and positive selection of the GAGE gene family.* Genetica, 2008. **133**(1): p. 31-5.
29. Mueller, J.L., et al., *The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression.* Nat Genet, 2008. **40**(6): p. 794-9.
30. Scanlan, M.J., A.J. Simpson, and L.J. Old, *The cancer/testis genes: review, standardization, and commentary.* Cancer Immun, 2004. **4**: p. 1.
31. Shipley, J.M., et al., *The t(X;18)(p11.2;q11.2) translocation found in human synovial sarcomas involves two distinct loci on the X chromosome.* Oncogene, 1994. **9**(5): p. 1447-53.
32. Wischnewski, F., K. Pantel, and H. Schwarzenbach, *Promoter demethylation and histone acetylation mediate gene expression of MAGE-A1, -A2, -A3, and -A12 in human cancer cells.* Mol Cancer Res, 2006. **4**(5): p. 339-49.

33. Wang, Z., et al., *SPAN-Xb expression in myeloma cells is dependent on promoter hypomethylation and can be upregulated pharmacologically.* Int J Cancer, 2006. **118**(6): p. 1436-44.
34. De Smet, C., A. Loriot, and T. Boon, *Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells.* Mol Cell Biol, 2004. **24**(11): p. 4781-90.
35. D'Alessio, A.C., I.C. Weaver, and M. Szyf, *Acetylation-induced transcription is required for active DNA demethylation in methylation-silenced genes.* Mol Cell Biol, 2007. **27**(21): p. 7462-74.
36. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. **128**(4): p. 669-81.
37. Kangaspeska, S., et al., *Transient cyclical methylation of promoter DNA.* Nature, 2008. **452**(7183): p. 112-5.
38. Bestor, T., et al., *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases.* J Mol Biol, 1988. **203**(4): p. 971-83.
39. Okano, M., S. Xie, and E. Li, *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases.* Nat Genet, 1998. **19**(3): p. 219-20.
40. Leonhardt, H., et al., *A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei.* Cell, 1992. **71**(5): p. 865-73.
41. 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.
42. Wolffe, A.P., P.L. Jones, and P.A. Wade, *DNA demethylation.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 5894-6.
43. Frommer, M., et al., *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1827-31.
44. Jones, P.A. and S.M. Taylor, *Cellular differentiation, cytidine analogs and DNA methylation.* Cell, 1980. **20**(1): p. 85-93.
45. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
46. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
47. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
48. Miao, F. and R. Natarajan, *Mapping global histone methylation patterns in the coding regions of human genes.* Mol Cell Biol, 2005. **25**(11): p. 4650-61.
49. Ausio, J., *Histone variants--the structure behind the function.* Brief Funct Genomic Proteomic, 2006. **5**(3): p. 228-43.
50. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* Genes Dev, 2005. **19**(3): p. 295-310.
51. Schuettengruber, B., et al., *Genome regulation by polycomb and trithorax proteins.* Cell, 2007. **128**(4): p. 735-45.

52. Levine, S.S., I.F. King, and R.E. Kingston, *Division of labor in polycomb group repression.* Trends Biochem Sci, 2004. **29**(9): p. 478-85.
53. Kim, D.H., et al., *Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells.* Nat Struct Mol Biol, 2006. **13**(9): p. 793-7.
54. Shao, Z., et al., *Stabilization of chromatin structure by PRC1, a Polycomb complex.* Cell, 1999. **98**(1): p. 37-46.
55. King, I.F., N.J. Francis, and R.E. Kingston, *Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro.* Mol Cell Biol, 2002. **22**(22): p. 7919-28.
56. Lavigne, M., et al., *Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin.* Mol Cell, 2004. **13**(3): p. 415-25.
57. Sanchez-Elsner, T., et al., *Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax.* Science, 2006. **311**(5764): p. 1118-23.
58. Wang, G.G., C.D. Allis, and P. Chi, *Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling.* Trends Mol Med, 2007. **13**(9): p. 373-80.
59. Roelen, B.A. and S.M. de Sousa Lopes, *Of stem cells and gametes: similarities and differences.* Curr Med Chem, 2008. **15**(13): p. 1249-56.
60. Surani, M.A., K. Hayashi, and P. Hajkova, *Genetic and epigenetic regulators of pluripotency.* Cell, 2007. **128**(4): p. 747-62.
61. Collas, P., A. Noer, and S. Timoskainen, *Programming the genome in embryonic and somatic stem cells.* J Cell Mol Med, 2007. **11**(4): p. 602-20.
62. Eilertsen, K.J., Z. Floyd, and J.M. Gimble, *The epigenetics of adult (somatic) stem cells.* Crit Rev Eukaryot Gene Expr, 2008. **18**(3): p. 189-206.
63. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
64. Croce, C.M., *Oncogenes and cancer.* N Engl J Med, 2008. **358**(5): p. 502-11.
65. Sherr, C.J., *Principles of tumor suppression.* Cell, 2004. **116**(2): p. 235-46.
66. Esteller, M., *Epigenetics provides a new generation of oncogenes and tumour-suppressor genes.* Br J Cancer, 2006. **94**(2): p. 179-83.
67. Esteller, M., *CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future.* Oncogene, 2002. **21**(35): p. 5427-40.
68. Ducasse, M. and M.A. Brown, *Epigenetic aberrations and cancer.* Mol Cancer, 2006. **5**: p. 60.
69. Feinberg, A.P. and B. Tycko, *The history of cancer epigenetics.* Nat Rev Cancer, 2004. **4**(2): p. 143-53.
70. Gronbaek, K., C. Hother, and P.A. Jones, *Epigenetic changes in cancer.* APMIS, 2007. **115**(10): p. 1039-59.
71. Hadnagy, A., R. Beaulieu, and D. Balicki, *Histone tail modifications and noncanonical functions of histones: perspectives in cancer epigenetics.* Mol Cancer Ther, 2008. **7**(4): p. 740-8.
72. Zhang, B., et al., *microRNAs as oncogenes and tumor suppressors.* Dev Biol, 2007. **302**(1): p. 1-12.
73. Tomaru, Y. and Y. Hayashizaki, *Cancer research with non-coding RNA.* Cancer Sci, 2006. **97**(12): p. 1285-90.

74. Reis, E.M., et al., *Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer.* Oncogene, 2004. **23**(39): p. 6684-92.
75. Misteli, T., *Beyond the sequence: cellular organization of genome function.* Cell, 2007. **128**(4): p. 787-800.
76. Feinberg, A.P., R. Ohlsson, and S. Henikoff, *The epigenetic progenitor origin of human cancer.* Nat Rev Genet, 2006. **7**(1): p. 21-33.
77. Ehrlich, M., *Cancer-linked DNA hypomethylation and its relationship to hypermethylation.* Curr Top Microbiol Immunol, 2006. **310**: p. 251-74.
78. Luczak, M.W. and P.P. Jagodzinski, *The role of DNA methylation in cancer development.* Folia Histochem Cytobiol, 2006. **44**(3): p. 143-54.
79. Greger, V., et al., *Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma.* Hum Genet, 1989. **83**(2): p. 155-8.
80. Gonzalez-Zulueta, M., et al., *Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing.* Cancer Res, 1995. **55**(20): p. 4531-5.
81. Herman, J.G., et al., *Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers.* Cancer Res, 1995. **55**(20): p. 4525-30.
82. Merlo, A., et al., *5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers.* Nat Med, 1995. **1**(7): p. 686-92.
83. Dobrovic, A. and D. Simpfendorfer, *Methylation of the BRCA1 gene in sporadic breast cancer.* Cancer Res, 1997. **57**(16): p. 3347-50.
84. Jarrard, D.F., et al., *Methylation of the androgen receptor promoter CpG island is associated with loss of androgen receptor expression in prostate cancer cells.* Cancer Res, 1998. **58**(23): p. 5310-4.
85. Lapidus, R.G., et al., *Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors.* Clin Cancer Res, 1996. **2**(5): p. 805-10.
86. Ottaviano, Y.L., et al., *Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells.* Cancer Res, 1994. **54**(10): p. 2552-5.
87. Virmani, A.K., et al., *Promoter methylation and silencing of the retinoic acid receptor-beta gene in lung carcinomas.* J Natl Cancer Inst, 2000. **92**(16): p. 1303-7.
88. Cairns, P., et al., *Point mutation and homozygous deletion of PTEN/MMAC1 in primary bladder cancers.* Oncogene, 1998. **16**(24): p. 3215-8.
89. Cairns, P., et al., *Frequent inactivation of PTEN/MMAC1 in primary prostate cancer.* Cancer Res, 1997. **57**(22): p. 4997-5000.
90. Collins, N., R. Wooster, and M.R. Stratton, *Absence of methylation of CpG dinucleotides within the promoter of the breast cancer susceptibility gene BRCA2 in normal tissues and in breast and ovarian cancers.* Br J Cancer, 1997. **76**(9): p. 1150-6.
91. Zhu, W.G., et al., *Increased expression of unmethylated CDKN2D by 5-aza-2'-deoxycytidine in human lung cancer cells.* Oncogene, 2001. **20**(53): p. 7787-96.

92. Esteller, M., *Cancer epigenomics: DNA methylomes and histone-modification maps.* Nat Rev Genet, 2007. **8**(4): p. 286-98.
93. Esteller, M., *Epigenetic gene silencing in cancer: the DNA hypermethylome.* Hum Mol Genet, 2007. **16 Spec No 1**: p. R50-9.
94. Esteller, M., et al., *DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis.* Hum Mol Genet, 2001. **10**(26): p. 3001-7.
95. Zochbauer-Muller, S., et al., *5' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer.* Cancer Res, 2001. **61**(9): p. 3581-5.
96. Tomizawa, Y., et al., *Inhibition of lung cancer cell growth and induction of apoptosis after reexpression of 3p21.3 candidate tumor suppressor gene SEMA3B.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13954-9.
97. Dammann, R., et al., *Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3.* Nat Genet, 2000. **25**(3): p. 315-9.
98. http://www.epigenome.org/, *The Human Epigenome Consortium Web Page*
99. http://cgap.nci.nih.gov/, *Cancer Genome Anatomy Project Web Site.*
100. Kidd, M., et al., *GeneChip, geNorm, and gastrointestinal tumors: novel reference genes for real-time PCR.* Physiol Genomics, 2007. **30**(3): p. 363-70.
101. Lee, S., et al., *Identification of novel universal housekeeping genes by statistical analysis of microarray data.* J Biochem Mol Biol, 2007. **40**(2): p. 226-31.
102. http://medgen.ugent.be/~jvdesomp/genorm/, *GeNorm Software Homepage.*
103. http://cpgislands.usc.edu/, *CpG Island Search Software.*
104. http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html, *CpG Plot Software.*
105. https://tandem.bu.edu/, *Inverted Repeats Finder Software.*
106. O'Reilly, D. and D.R. Greaves, *Cell-type-specific expression of the human CD68 gene is associated with changes in Pol II phosphorylation and short-range intrachromosomal gene looping.* Genomics, 2007. **90**(3): p. 407-15.
107. Lavrov, S.A. and M.V. Kibanov, *Noncoding RNAs and chromatin structure.* Biochemistry (Mosc), 2007. **72**(13): p. 1422-38.
108. Costa, F.F., *Non-coding RNAs, epigenetics and complexity.* Gene, 2008. **410**(1): p. 9-17.
109. Morris, K.V., *RNA-mediated transcriptional gene silencing in human cells.* Curr Top Microbiol Immunol, 2008. **320**: p. 211-24.
110. http://biobases.ibch.poznan.pl/ncRNA/, *Noncoding RNA database.*
111. Whitehouse, I., et al., *Chromatin remodelling at promoters suppresses antisense transcription.* Nature, 2007. **450**(7172): p. 1031-5.
112. http://genome.ucsc.edu/, *UCSC Genome Browser Homepage.*
113. Chen, Y.T., et al., *Cerebellar degeneration-related antigen: a highly conserved neuroectodermal marker mapped to chromosomes X in human and mouse.* Proc Natl Acad Sci U S A, 1990. **87**(8): p. 3077-81.
114. Bottomley, S.S., et al., *Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia.* J Bioenerg Biomembr, 1995. **27**(2): p. 161-8.

115. Han, L., et al., *Histone acetyltransferase p300 regulates the transcription of human erythroid-specific 5-aminolevulinate synthase gene.* Biochem Biophys Res Commun, 2006. **348**(3): p. 799-806.
116. Labrador, M. and V.G. Corces, *Setting the boundaries of chromatin domains and nuclear organization.* Cell, 2002. **111**(2): p. 151-4.
117. Zhao, H. and A. Dean, *An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene.* Nucleic Acids Res, 2004. **32**(16): p. 4903-19.
118. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.* Cell, 1999. **98**(3): p. 387-96.
119. Miele, A., et al., *Mapping chromatin interactions by chromosome conformation capture.* Curr Protoc Mol Biol, 2006. **Chapter 21**: p. Unit 21 11.
120. Bredenbeck, A., et al., *Coordinated expression of clustered cancer/testis genes encoded in a large inverted repeat DNA structure.* Gene, 2008. **415**(1-2): p. 68-73.

# 6 APPENDICES

**APPENDIX A :** CURRENT LIST OF CANCER-TESTIS GENES

**APPENDIX B :** LIST OF HISTONE-MODIFYING ENZYMES

**APPENDIX C :** X-LINKED GENES DOWNREGULATED IN CANCER ACCORDING TO ANALYSES OF SAGE/EST LIBRARIES

**APPENDIX D :** NEGATIVE-RT CONTROLS

**APPENDIX E :** REAL-TIME PCR EQUATIONS & CALCULATIONS

**APPENDIX F :** SEQUENCING RESULTS OF THE 3C ASSAY

# APPENDIX A: CURRENT LIST OF CANCER-TESTIS GENES

All human genes listed as cancer-testis (CT) genes till 2007 are presented below together with chromosomal locations and numbers of family members [10].

| CT Number | Family Name | Human Chromosome | Human Gene Number |
|---|---|---|---|
| CT1 | MAGEA | X | 13 (0) |
| CT2 | BAGE | 5, 7, 9, 18, 21 | 7 (0) |
| CT3 | MAGEB | X | 7 (1) |
| CT4 | GAGE | X | 16 (0) |
| CT5 | SSX | X | 14 (0) |
| CT6 | CTAG | X | 3 (0) |
| CT7 | MAGEC | X | 2 (0) |
| CT8 | SYCP1 | 1 | 1 (0) |
| CT9 | BRDT | 1 | 1 (0) |
| CT10 | MAGEE | X | 2 (2) |
| CT11 | SPANX | X | 11 (0) |
| CT12 | XAGE | X | 14 (0) |
| CT13 | DDX43 | 6 | 1 (0) |
| CT14 | SAGE | X | 1 (0) |
| CT15 | ADAM2 | 4, 8 | 2 (0) |
| CT16 | PAGE | X | 7 (0) |
| CT17 | LIPI | 21 | 2 (0) |
| CT21 | CTAGE | 2, 6, 7, 9, 10, 13, 14, 18 | 21 (12) |
| CT24 | CSAG | X | 4 (0) |
| CT25 | DSCR8 | 21 | 2 (0) |
| CT26 | DDX53 | X | 1 (1) |
| CT27 | CTCFL | 20 | 1 (0) |
| CT28 | LUZP4 | X | 1 (0) |
| CT29 | CASC5 | 15 | 1 (0) |
| CT30 | TFDP3 | 13, 15, X | 4 (3) |
| CT32 | LDHC | 11 | 1 (0) |
| CT33 | MORC1 | 3 | 1 (0) |
| CT34 | DKKL1 | 19, 20 | 2 (1) |
| CT35 | SPO11 | 20 | 1 (0) |
| CT36 | CRISP2 | 6 | 1 (0) |
| CT37 | FMR1NB | X | 1 (0) |
| CT38 | FTHL17 | X | 4 (4) |
| CT39 | NXF2 | X | 2 (0) |
| CT41 | TDRD | 6, 10 | 2 (0) |

| CT Number | Family Name | Human Chromosome | Human Gene Number |
|---|---|---|---|
| CT42 | TEX15 | 8 | 1 (0) |
| CT43 | FATE1 | X | 1 (0) |
| CT44 | TPTE | 13, 21, Y | 4 (0) |
| CT45 | CT45 | X | 6 (0) |
| CT46 | HORMAD1 | 1, 6 | 2 (1) |
| CT47 | LOC255313 | X | 12 (0) |
| CT48 | SLCO6A1 | 5 | 1 (0) |
| CT49 | TAG | 5 | 1 (0) |
| CT50 | LEMD1 | 1 | 1 (0) |
| CT51 | HSPB9 | 17 | 1 (1) |
| CT53 | ZNF165 | 6 | 1 (0) |
| CT54 | SPACA3 | 17 | 1 (0) |
| CT55 | CXorf48 | X | 3 (0) |
| CT56 | THEG | 19 | 1 (0) |
| CT57 | ACTL8 | 1 | 1 (0) |
| CT58 | NALP4 | 19 | 1 (0) |
| CT59 | COX6B2 | 19 | 1 (0) |
| CT60 | BC047459 | 15 | 2 (0) |
| CT61 | CCDC33 | 15 | 1 (0) |
| CT62 | BC048128 | 15 | 1 (0) |
| CT63 | PASD1 | X | 1 (0) |
| CT65 | TULP2 | 19 | 1 (0) |
| CT66 | AA884595 | 7 | 1 (1) |
| CT68 | MGC27016 | 4 | 1 (0) |
| CT69 | BC040308 | 6 | 1 (0) |
| CT71 | SPINLW1 | 20 | 1 (0) |
| CT72 | TSSK6 | 19 | 1 (1) |
| CT73 | ADAM29 | 4 | 1 (0) |
| CT74 | CCDC36 | 3 | 1 (0) |
| CT75 | BC033986 | 2 | 1 (0) |
| CT76 | SYCE1 | 10 | 1 (0) |
| CT77 | CPXCR1 | X | 1 (0) |
| CT78 | TSPY1 | Y | 14 (0) |
| CT79 | TSGA | 2, 21 | 3 (0) |
| CT81 | ARMC3 | 10 | 1 (0) |
| CTNA | PRAME | 1, 22 | 36 (0) |

Numbers in brackets denote the number of intronless gene copies, which in the case of multi-exon genes may indicate putative retrocopy genes. PRAME family has not yet been officially designated as CT.

# APPENDIX B:
# CURRENT LIST OF HISTONE-MODIFYING ENZYMES [45]

| Enzymes that modify histones | Residues modified |
|---|---|
| ***Acetyltransferases*** | |
| HAT1 | H4 (K5, K12) |
| CBP/P300 | H3 (K14, K18), H4 (K5, K8), H2A (K5), H2B (K12, K15) |
| PCAF/GCN5 | H3 (K9, K14, K18) |
| TIP60 | H4 (K5, K8, K12, K16), H3K14 |
| HB01 (ScESA1, SpMST2) | H4 (K5, K8, K12) |
| ScSAS3 | H3 (K14, K23) |
| ScSAS2 (SpMST2) | H4K16 |
| ScRTT109 | H3K56 |
| ***Deacetylases*** | |
| SirT2 (ScSir2) | H4K16 |
| ***Lysine Methyltransferases*** | |
| SUV39H1 | H3K9 |
| SUV39H2 | H3K9 |
| G9a | H3K9 |
| ESET/SETDB1 | H3K9 |
| EuHMTase/GLP | H3K9 |
| CLL8 | H3K9 |
| SpClr4 | H3K9 |
| MLL1 | H3K4 |
| MLL2 | H3K4 |
| MLL3 | H3K4 |
| MLL4 | H3K4 |
| MLL5 | H3K4 |
| SET1A | H3K4 |
| SET1B | H3K4 |
| ASH1 | H3K4 |
| Sc/Sp SET1 | H3K4 |
| SET2 (Sc/Sp SET2) | H3K36 |
| NSD1 | H3K36 |
| SYMD2 | H3K36 |
| DOT1 | H3K79 |
| Sc/Sp DOT1 | H3K79 |

| Enzymes that modify histones | Residues modified |
|---|---|
| Pr-SET 7/8 | H4K20 |
| SUV4 20H1 | H4K20 |
| SUV420H2 | H4K20 |
| SpSet 9 | H4K20 |
| EZH2 | H3K27 |
| RIZ1 | H3K9 |
| ***Lysine Demethylases*** | |
| LSD1/BHC110 | H3K4 |
| JHDM1a | H3K36 |
| JHDM1b | H3K36 |
| JHDM2a | H3K9 |
| JHDM2b | H3K9 |
| JMJD2A/JHDM3A | H3K9, H3K36 |
| JMJD2B | H3K9 |
| JMJD2C/GASC1 | H3K9, H3K36 |
| JMJD2D | H3K9 |
| ***Arginine Methyltransferases*** | |
| CARM1 | H3 (R2, R17, R26) |
| PRMT4 | H4R3 |
| PRMT5 | H3R8, H4R3 |
| ***Serine/Threonine Kinases*** | |
| Haspin | H3T3 |
| MSK1 | H3S28 |
| MSK2 | H3S28 |
| CKII | H4S1 |
| Mst1 | H2BS14 |
| ***Ubiquitilases*** | |
| Bmi/Ring1A | H2AK119 |
| RNF20/RNF40 | H2BK120 |
| ***Proline Isomerases*** | |
| ScFPR4 | H3P30, H3P38 |

# APPENDIX C:
# X-LINKED GENES DOWNREGULATED IN CANCER ACCORDING TO ANALYSES OF SAGE/EST LIBRARIES

| Gene Name | Location on Chromosome X |
|---|---|
| AFF2 | Xq28 |
| AKAP4 | Xp11.2 |
| ALAS2 | Xp11.21 |
| ARHGAP4 | Xq28 |
| ARSF | Xp22.3 |
| ATP2B3 | Xq28 |
| BGN | Xq28 |
| CACNA1F | Xp11.23 |
| CD99L2 | Xq28 |
| CDKL5 | Xp22 |
| CDR1 | Xq27.1-27.2 |
| CHRDL1 | Xq22.3 |
| CLCN4 | Xp22.3 |
| CNKSR2 | Xp22.12 |
| CYBB | Xp21.1 |
| F9 | Xq27.1-27.2 |
| FHL1 | Xq26 |
| GABRA3 | Xq28 |
| GATA1 | Xp11.23 |
| GDI1 | Xq28 |
| GPM6B | Xp22.2 |
| GPRASP2 | Xq22.1 |
| IDS | Xq28 |
| IL3RA | Xp22.3 (Yp11.3) |
| LAMP2 | Xq24 |
| LOC286411 | Xq27.1 |
| MAOB | Xp11.23 |
| MBTPS2 | Xp22.1-22.2 |
| MCF2 | Xq27 |

| Gene Name | Location on Chromosome X |
|---|---|
| MGC39900 | xq22.2 |
| MPP1 | Xq28 |
| MSL3L1 | Xp22.3 |
| OPN1LW | Xq28 |
| P2RY8 | Xp22.33 |
| PDZD4 | Xq28 |
| PEPP-2 | Xq24 |
| PLAC1 | Xq26 |
| PLP1 | Xq22 |
| PNCK | Xq28 |
| PNMA3 | Xq28 |
| PORCN | Xp11.23 |
| PRRG1 | Xp21.1 |
| RGN | Xp11.3 |
| RP11-393H10.2 | Xp22.12 |
| RP13-102H20.1 | Xq26.1 |
| RP6-166C19.1 | Xq24 |
| RRAGB | Xp11.22 |
| RS1 | Xp22.2-22.1 |
| SAT1 | Xp22.1 |
| SEPT6 | Xq24 |
| SLC9A6 | Xq26.3 |
| SLITRK2 | Xq27.3 |
| SMPX | Xp22.1 |
| SYN1 | Xp11.23 |
| SYP | Xp11.23-11.22 |
| TLR8 | Xp22 |
| TSC22D3 | Xq22.3 |
| WAS | Xp11.4-11.21 |
| ZCCHC12 | Xq24 |

# APPENDIX D:
# NEGATIVE-RT CONTROLS

In order to make the down-regulation of the intronless gene CDR1 sure, negative-RT PCRs were carried out. Indeed, the initially DNaseI-untreated RNA samples were contaminated with genomic DNA, which was eliminated after DNaseI treatment. Down-regulation was more clearly demonstrated by amplification of the DNaseI-treated samples (**Figure E1**).

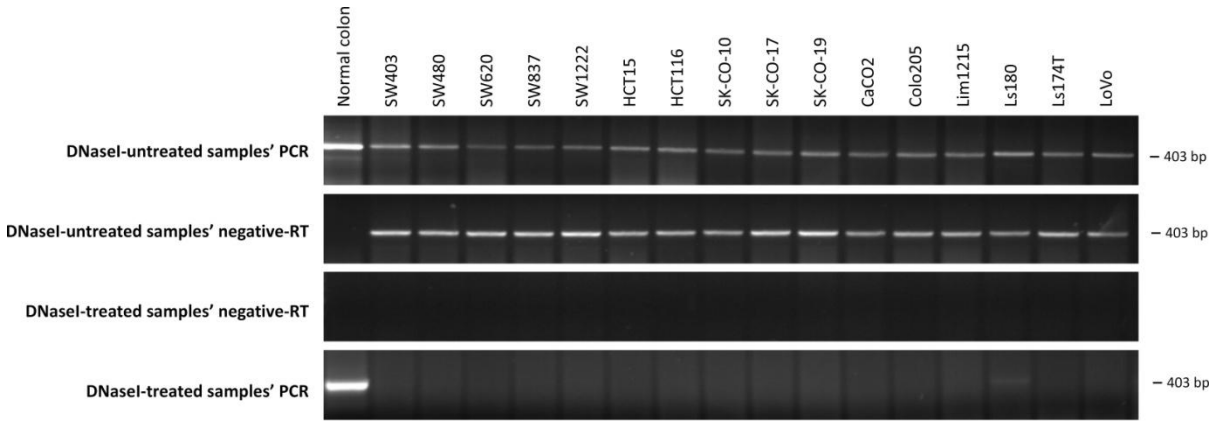



**Figure E1 Negative-RT controls and PCR amplifications of the CDR1 gene on DNaseI-treated and -untreated RNA**

Negative-RT controls were also needed for verification of non-coding RNAs that were demonstrated in **Figure 13**. As shown in Figure E1, normal colon and other normal tissue RNAs (Ambion) do not contain any genomic DNA. Negative-RT PCRs were also done for GOLGA1, TFCP2 and IKBKG mRNAs and no bands corresponding to gDNA amplification were detected (**Figure E2**). On the other hand, bands for ncRNA were seen in amplification of normal tissue RNAs (**Figure 13**), some of which were not detected in untreated Colo205 and HT29 samples. As a result, we can conclude that there is not any gDNA contamination in HT29, Colo205 and SK-LC-17 cancer cell line samples and detected bands correspond to ncRNA.

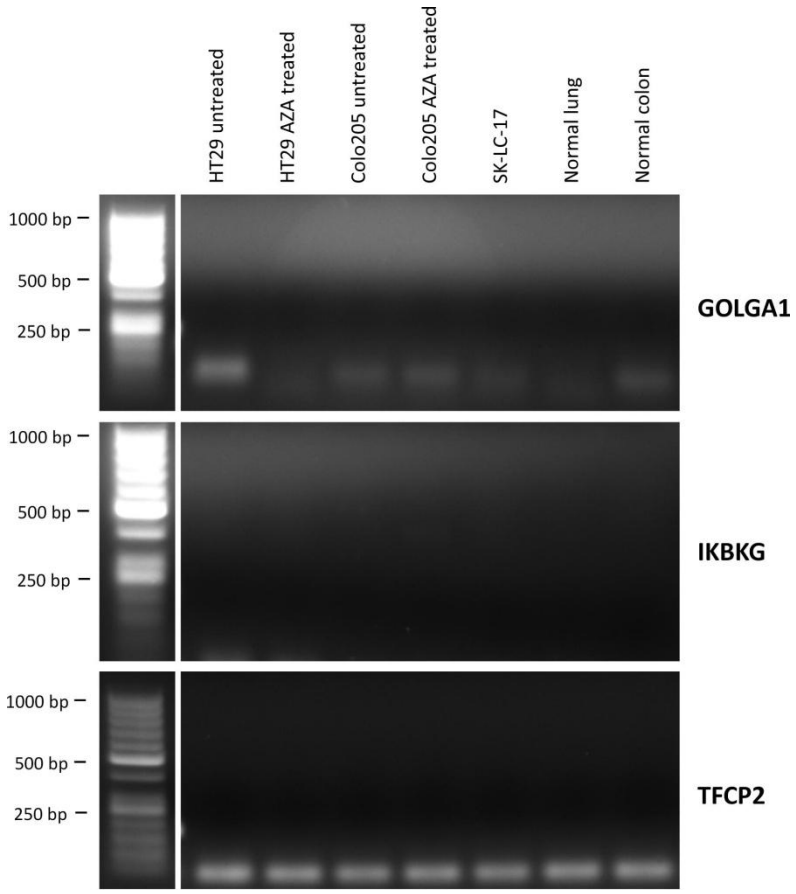


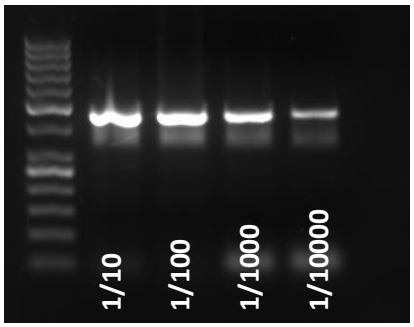

**Figure E2 Negative-RT controls of GOLGA1, TFCP2 and IKBKG genes.**

# APPENDIX E: REAL-TIME PCR EQUATIONS & CALCULATIONS

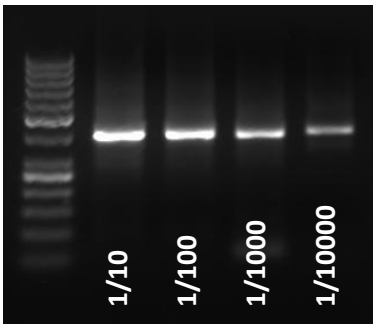
## A) CALCULATION OF EFFICIENCIES

For calculation of amplification efficiencies, 5- or 10-fold dilutions of brain cDNA were used for all genes. Samples were analyzed under same conditions by using Sybr Green dye and efficiency graphs were plotted according to the obtained data (**Figures F2 and F3**). Samples were also run on 1.5% agarose gel as shown in **Figure F1**.

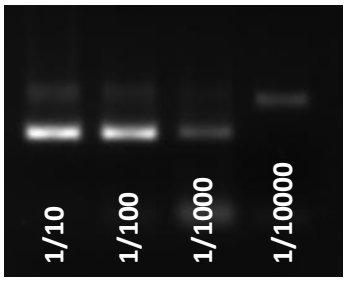
**GAPDH**



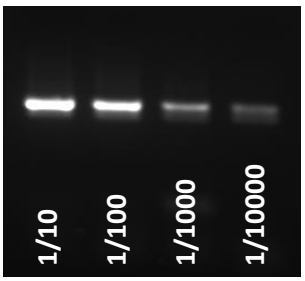

**CDR1**



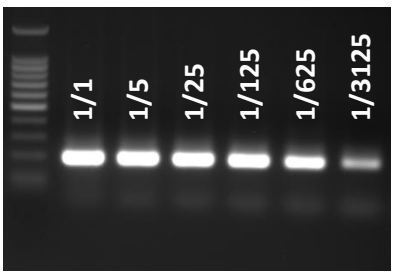

**RRAGB**



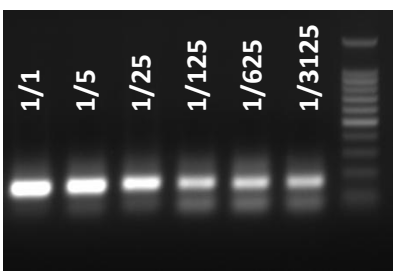

**SLC9A6**



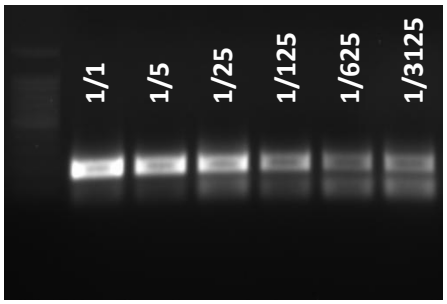

**GABRA3**



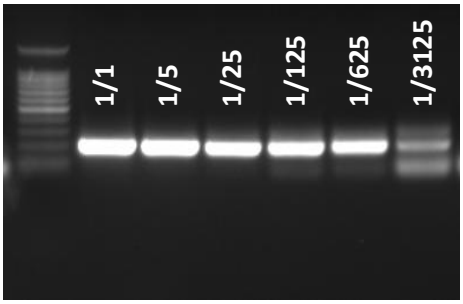

**ZCCHC12**



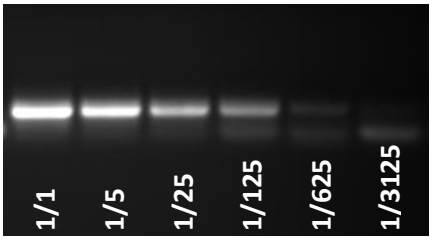
**ALAS2**




**TFCP2**




**GOLGA1**




**18S rRNA**

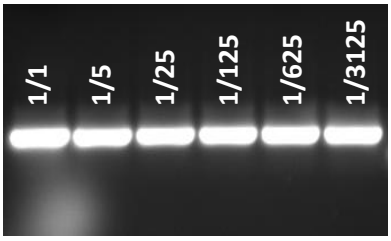
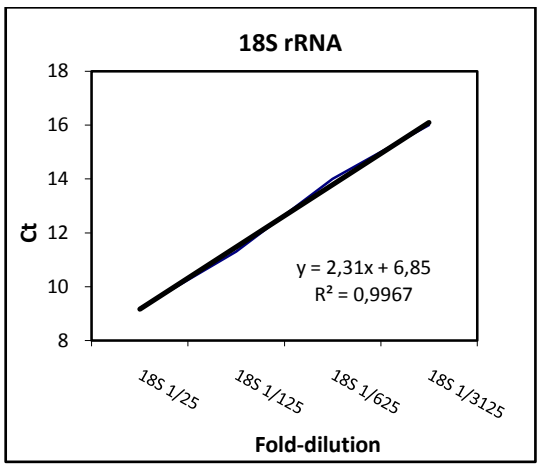
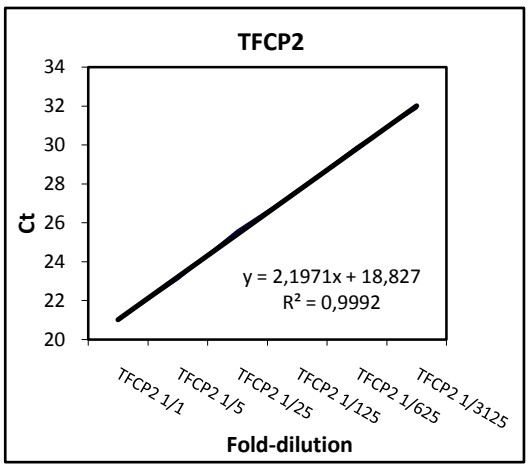



**Figure F1 PCR data for calculation of amplification efficiencies. Individual genes are indicated above the gel images.**

***Efficiency graphs of candidate reference genes and genes in question***



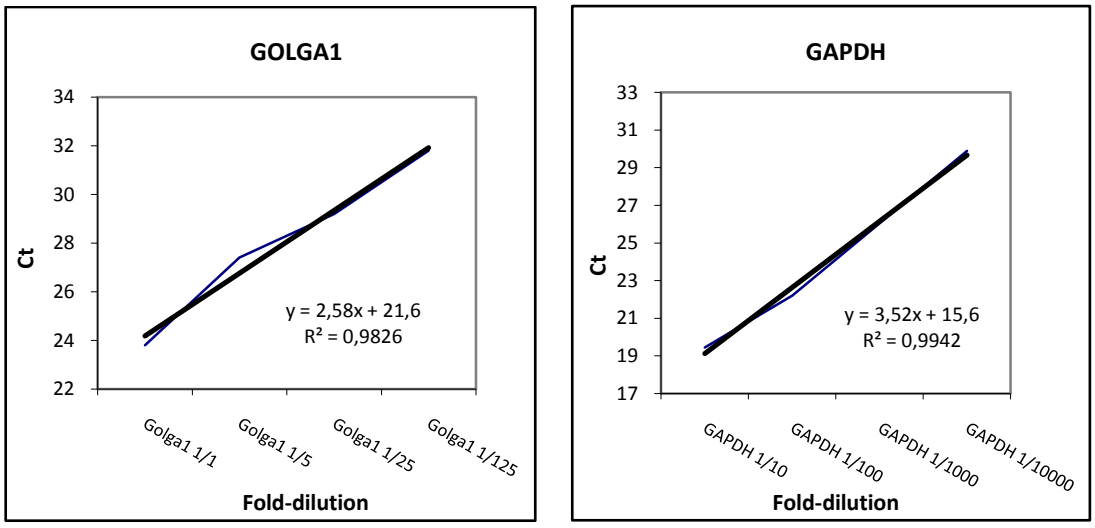
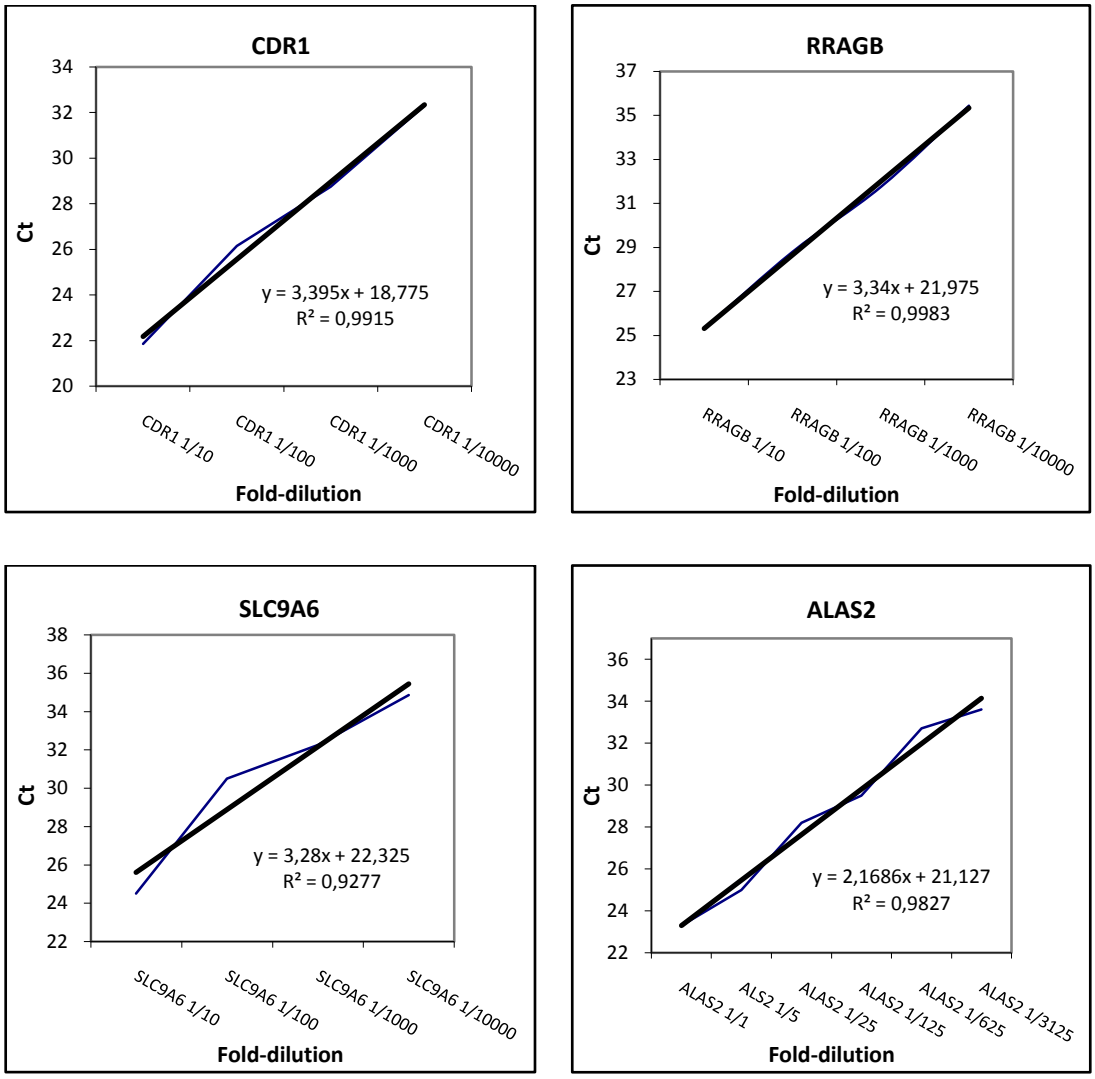



**Figure F2 Efficiency graphs of the candidate reference genes. Individual genes are indicated inside the plot area.**



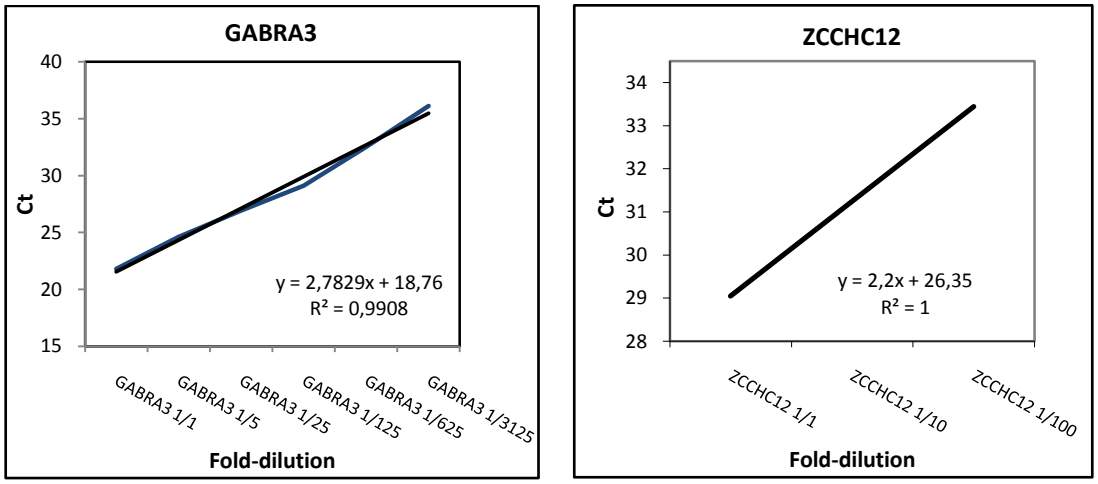



**Figure F3 Efficiency graphs of the genes in question. Individual genes are indicated inside the plot area.**

Amplification efficiencies (E) of genes were calculated according to equation (1). E≥2.0 was taken as 2.0 and results are tabulated in **Table F1**.

$$E = D^{(1/m)} \qquad (1)$$

Where D is fold-dilution of cDNA

*m* is the slope of the trend line of data

**Table F1 Amplification Efficiencies of Real-Time PCR primers of Given Genes**

| ***REFERENCE GENES*** | |
|---|---|
| **Gene Name** | **E value** |
| GAPDH | 1.9 |
| TFCP2 | 2.0 |
| 18S rRNA | 2.0 |
| GOLGA1 | 1.9 |
| ***GENES IN QUESTION*** | |
| **Gene Name** | **E value** |
| ALAS2 | 2.0 |
| CDR1 | 2.0 |
| GABRA3 | 2.0 |
| RRAGB | 2.0 |
| SLC9A6 | 2.0 |
| ZCCHC12 | 2.0 |

## B) DETERMINATION OF HOUSEKEEPING GENES

After calculating the efficiencies of the candidate reference genes, quantities of expression were calculated as in equations (2-3).

$$Q = E^{deltaCt} \quad (2)$$

$$Q = E^{(minCt - sampleCt)} \quad (3)$$

GeNorm software finds the most stable of given reference genes according to the variation in Q values. Expression of our 4 candidate reference genes, 18S rRNA, GOLGA1, GAPDH and TFCP2 were quantified in a normal tissue panel and Q values were entered to the program. Outputs are presented in **Figures F4 and F5**. TFCP2 and GOLGA1 were selected as most stable genes, followed by GAPDH and 18S rRNA. Also, usage of 2 reference genes, instead of 3, was recommended as optimum. As a result, TFCP2 and GOLGA1 were chosen as housekeeping genes to be used in our real-time PCR assays.

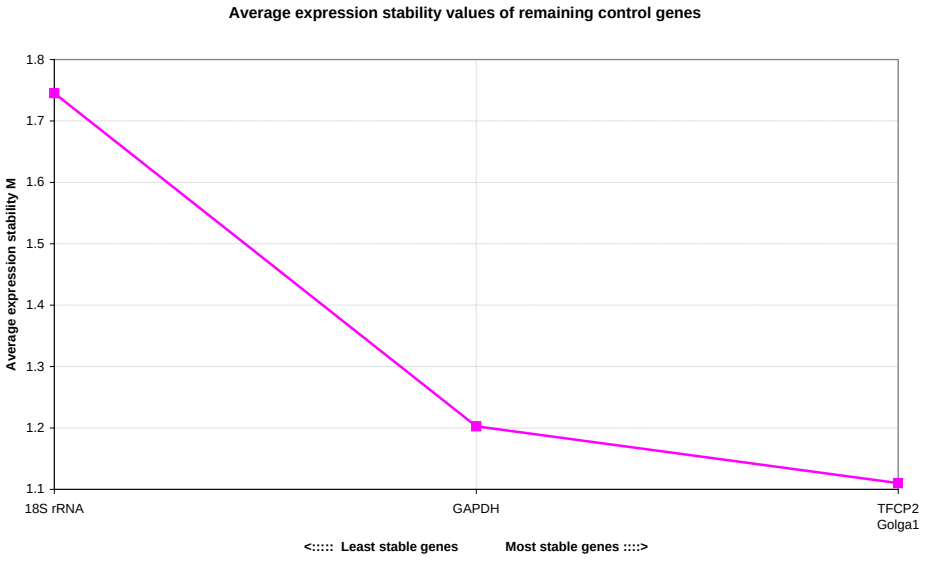



**Figure F4 Stability of the candidate reference genes within the normal tissue panel, calculated by the GeNorm software.**

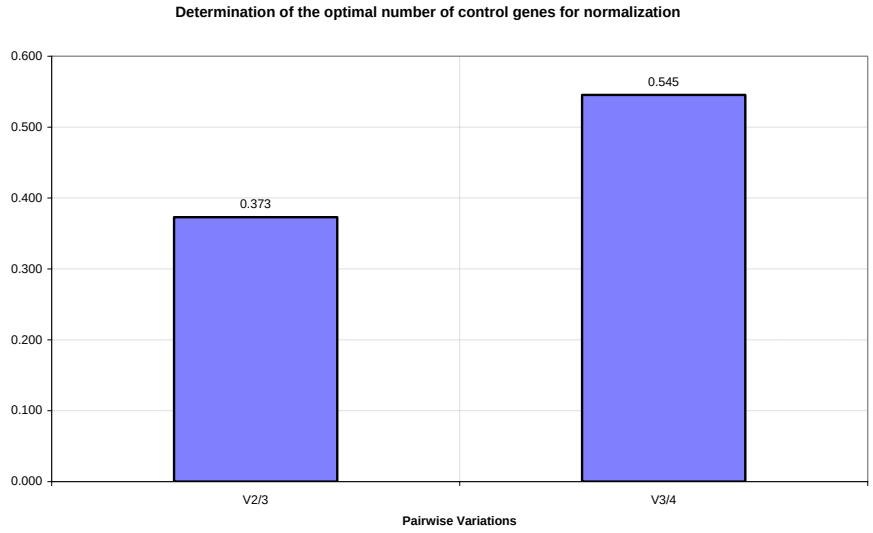



**Figure F5 Determination of the optimal number of genes for normalization by the GeNorm software**

## C) EQUATIONS USED FOR NORMALIZATION

### *Delta Ct method*

$$Q = E^{deltaCt} \quad (2)$$

$$Q = E^{(minCt - sampleCt)} \quad (3)$$

$$SD(Q) = E^{deltaCt} \times lnE \times SD(sampleCt) \quad (4)$$

where Q is the sample quantity relative to the reference sample

E is amplification efficiency (2 = 100%)

SD(*sampleCt*) is the standard deviation of the Ct values of sample replicates

### *Delta - delta Ct method*

*n* reference genes (REF) and one gene of interest (GOI) are to be measured and equations (2-4) to be applied;

Gene of interest GOI ± (SD GOI)

Housekeeper 1 $REF_1$ ± (SD $REF_1$)

Housekeeper *n* $REF_n \pm (SD\ REF_n)$

The normalization factor of *n* reference genes is calculated as the geometric mean:

$$NF_n = \sqrt[n]{REF_1 \times REF_2 \times \ldots \times REF_n} \quad (5)$$

$$SD\ NF_n = NF_n \times \sqrt{\left(\frac{SD\ REF_1}{n \times REF_1}\right)^2 + \left(\frac{SD\ REF_2}{n \times REF_2}\right)^2 + \cdots + \left(\frac{SD\ REF_n}{n \times REF_n}\right)^2} \quad (6)$$

$$\text{Normalized GOI} = GOI / NF \quad (7)$$

$$SD\ (Norm.GOI) = (Norm.GOI) \times \sqrt{\left(\frac{SD\ NF_n}{NF_n}\right)^2 + \left(\frac{SD\ GOI}{GOI}\right)^2} \quad (8)$$

$$SE = \frac{SD}{\sqrt{m}} \quad (9)$$

where SE is the standard error

m is the number of measurements

## D) CALCULATIONS

### *i) Normal Tissue Panel*

**Table F2 Detailed calculations about the real-time PCR data for normal tissues**

| | | ***Reference Genes*** | | ***Genes in Question*** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **TFCP2** | **GOLGA1** | **ALAS2** | **CDR1** | **GABRA3** | **RRAGB** | **SLC9A6** | **ZCCHC12** |
| **1** | **testis** | | | | | | | | |
| | *raw* | 23,2 | 23,1 | 33,9 | 17,4 | 25,3 | 21,8 | 21,6 | 26,4 |
| | | 24,1 | 23,6 | 33,7 | 18 | 28,1 | 23,6 | 21,5 | 26,8 |
| | | 23,5 | 23,1 | 31,1 | 18,6 | 25,3 | 23,4 | 24,4 | 26,3 |
| | *avg* | 23,60 | 23,27 | 32,90 | 18,00 | 26,23 | 22,93 | 22,50 | 26,50 |
| | *SDsample* | 0,46 | 0,29 | 1,56 | 0,60 | 1,62 | 0,99 | 1,65 | 0,26 |
| | *E^(deltaCt)* | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | *SD*{E^(deltaCt)} | 0,318 | 0,200 | 1,083 | 0,416 | 1,121 | 0,684 | 1,141 | 0,183 |
| | *SE*{E^(deltaCt)} | 0,183 | 0,116 | 0,625 | 0,240 | 0,647 | 0,395 | 0,659 | 0,106 |
| | *NF* | 1,00000 | | | | | | | |
| | *SE(NF)* | 0,10837 | | | | | | | |
| | *Norm.GOI* | | | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| | *SE(Norm.GOI)* | | | 0,63 | 0,26 | 0,66 | 0,41 | 0,67 | 0,15 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **2** | **colon** | | | | | | | | |
| | *raw* | 23,8 | 26,6 | 34 | 22,2 | 28,2 | 27,4 | 23,1 | 26,5 |
| | | 23,7 | 25,5 | 35 | 21,7 | 27,6 | 26,8 | 24,1 | 28 |
| | | 24 | 26,6 | 36,6 | 21,8 | 28,2 | 27,4 | 24,6 | 28,7 |
| | *avg* | 23,83 | 26,23 | 35,20 | 21,90 | 28,00 | 27,20 | 23,93 | 27,73 |
| | *SDsample* | 0,15 | 0,64 | 1,31 | 0,26 | 0,35 | 0,35 | 0,76 | 1,12 |
| | *E^(deltaCt)* | 0,851 | 0,157 | 0,203 | 0,067 | 0,294 | 0,052 | 0,370 | 0,425 |
| | *SD*{E^(deltaCt)} | 0,090 | 0,069 | 0,185 | 0,012 | 0,071 | 0,012 | 0,196 | 0,331 |
| | *SE*{E^(deltaCt)} | 0,052 | 0,040 | 0,107 | 0,007 | 0,041 | 0,007 | 0,113 | 0,191 |
| | *NF* | 0,36562 | | | | | | | |
| | *SE(NF)* | 0,13070 | | | | | | | |
| | *Norm.GOI* | | | 0,56 | 0,18 | 0,80 | 0,14 | 1,01 | 1,16 |
| | *SE(Norm.GOI)* | | | 0,35 | 0,07 | 0,31 | 0,05 | 0,48 | 0,67 |
| **3** | **spleen** | | | | | | | | |
| | *raw* | 30 | 32,8 | 30 | 25,6 | 28,9 | 28,7 | 32,5 | 28,3 |
| | | 29,5 | 32,2 | 30,8 | 25,4 | 28,3 | 29,3 | 29,2 | 27,3 |
| | | 29,5 | 32 | 30,7 | 25 | 27,8 | 29,8 | 29,5 | 28,1 |
| | *avg* | 29,67 | 32,33 | 30,50 | 25,33 | 28,33 | 29,27 | 30,40 | 27,90 |
| | *SDsample* | 0,29 | 0,42 | 0,44 | 0,31 | 0,55 | 0,55 | 1,82 | 0,53 |
| | *E^(deltaCt)* | 0,015 | 0,003 | 5,278 | 0,006 | 0,233 | 0,012 | 0,004 | 0,379 |
| | *SD*{E^(deltaCt)} | 0,003 | 0,001 | 1,595 | 0,001 | 0,089 | 0,005 | 0,005 | 0,139 |
| | *SE*{E^(deltaCt)} | 0,002 | 0,001 | 0,921 | 0,001 | 0,051 | 0,003 | 0,003 | 0,080 |
| | *NF* | 0,00722 | | | | | | | |
| | *SE(NF)* | 0,10137 | | | | | | | |
| | *Norm.GOI* | | | 730,68 | 0,86 | 32,29 | 1,72 | 0,58 | 52,46 |
| | *SE(Norm.GOI)* | | | 10255,09 | 12,05 | 453,24 | 24,10 | 8,14 | 736,28 |
| **4** | **pancreas** | | | | | | | | |
| | *raw* | 26,3 | 28,3 | 33,4 | 17,2 | 27,5 | 26 | 24,5 | 28,3 |
| | | 26,1 | 28,2 | 32 | 20,8 | 27,2 | 26,8 | 25,6 | 27,7 |
| | | 26,2 | 28,4 | 32,6 | 17,9 | 27,7 | 26,7 | 25,7 | 27,5 |
| | *avg* | 26,20 | 28,30 | 32,67 | 18,63 | 27,47 | 26,50 | 25,27 | 27,83 |
| | *SDsample* | 0,10 | 0,10 | 0,70 | 1,91 | 0,25 | 0,44 | 0,67 | 0,42 |
| | *E^(deltaCt)* | 0,165 | 0,043 | 1,176 | 0,645 | 0,425 | 0,084 | 0,147 | 0,397 |
| | *SD*{E^(deltaCt)} | 0,011 | 0,003 | 0,572 | 0,853 | 0,074 | 0,025 | 0,068 | 0,115 |
| | *SE*{E^(deltaCt)} | 0,007 | 0,002 | 0,330 | 0,492 | 0,043 | 0,015 | 0,039 | 0,066 |
| | *NF* | 0,08450 | | | | | | | |
| | *SE(NF)* | 0,02830 | | | | | | | |
| | *Norm.GOI* | | | 13,91 | 7,63 | 5,03 | 1,00 | 1,74 | 4,70 |
| | *SE(Norm.GOI)* | | | 6,08 | 6,36 | 1,76 | 0,38 | 0,74 | 1,76 |
| **5** | **placenta** | | | | | | | | |
| | *raw* | 25,9 | 29,1 | 26,2 | 22 | 27,3 | 25,3 | 26,1 | 27,8 |
| | | 26,2 | 29,5 | 26,7 | 21,1 | 27,9 | 26,5 | 27,6 | 26,5 |
| | | 27,6 | 29,7 | 29,9 | 22,2 | 27,4 | 27,1 | 26,4 | 29,2 |
| | *avg* | 26,57 | 29,43 | 26,45 | 21,77 | 27,53 | 26,30 | 26,70 | 27,83 |
| | *SDsample* | 0,91 | 0,31 | 2,01 | 0,59 | 0,32 | 0,92 | 0,79 | 1,35 |
| | *E^(deltaCt)* | 0,128 | 0,021 | 87,427 | 0,073 | 0,406 | 0,097 | 0,054 | 0,397 |
| | *SD*{E^(deltaCt)} | 0,080 | 0,005 | 121,653 | 0,030 | 0,090 | 0,062 | 0,030 | 0,371 |
| | *SE*{E^(deltaCt)} | 0,046 | 0,003 | 70,236 | 0,017 | 0,052 | 0,036 | 0,017 | 0,214 |
| | *NF* | 0,05226 | | | | | | | |
| | *SE(NF)* | 0,19158 | | | | | | | |
| | *Norm.GOI* | | | 1672,95 | 1,41 | 7,77 | 1,86 | 1,04 | 7,59 |
| | *SE(Norm.GOI)* | | | 6278,38 | 5,16 | 28,51 | 6,83 | 3,83 | 28,14 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | **small intestine** | | | | | | | | |
| | *raw* | 25,9 | 27,9 | 31,1 | 20,2 | 27,7 | 26,6 | 29,7 | 27,6 |
| | | 25,1 | 27,2 | 30,9 | 20,3 | 28,3 | 26,4 | 23,6 | 27,8 |
| | | 25,8 | 28,1 | 31,5 | 21 | N/A | 27,4 | 23,3 | 28,1 |
| | *avg* | 25,60 | 27,73 | 31,17 | 20,50 | 28,00 | 26,80 | 23,45 | 27,83 |
| | *SDsample* | 0,44 | 0,47 | 0,31 | 0,44 | 0,42 | 0,53 | 3,61 | 0,25 |
| | *E^(deltaCt)* | 0,250 | 0,062 | 3,325 | 0,177 | 0,294 | 0,069 | 0,518 | 0,397 |
| | *SD*{E^(deltaCt)} | 0,076 | 0,020 | 0,704 | 0,053 | 0,086 | 0,025 | 1,296 | 0,069 |
| | *SE*{E^(deltaCt)} | 0,044 | 0,012 | 0,407 | 0,031 | 0,050 | 0,015 | 0,748 | 0,040 |
| | *NF* | 0,12415 | | | | | | | |
| | *SE(NF)* | 0,12864 | | | | | | | |
| | *Norm.GOI* | | | 26,78 | 1,42 | 2,37 | 0,55 | 4,17 | 3,20 |
| | *SE(Norm.GOI)* | | | 27,95 | 1,50 | 2,49 | 0,58 | 7,42 | 3,33 |
| 7 | **liver** | | | | | | | | |
| | *raw* | 26 | 27,1 | 32,9 | 21,1 | 29,7 | 26,3 | 24,1 | 28,2 |
| | | 26,8 | 27,1 | 32,9 | 20,5 | 29 | 28 | 25,5 | 27,9 |
| | | 26,3 | 26,9 | 33,7 | 21,3 | 27,6 | 27,6 | 34 | 28,1 |
| | *avg* | 26,37 | 27,03 | 33,17 | 20,97 | 28,77 | 27,30 | 24,80 | 28,07 |
| | *SDsample* | 0,40 | 0,12 | 0,46 | 0,42 | 1,07 | 0,89 | 5,36 | 0,15 |
| | *E^(deltaCt)* | 0,147 | 0,095 | 0,831 | 0,128 | 0,173 | 0,048 | 0,203 | 0,338 |
| | *SD*{E^(deltaCt)} | 0,041 | 0,008 | 0,266 | 0,037 | 0,128 | 0,030 | 0,754 | 0,036 |
| | *SE*{E^(deltaCt)} | 0,024 | 0,004 | 0,154 | 0,021 | 0,074 | 0,017 | 0,435 | 0,021 |
| | *NF* | 0,11840 | | | | | | | |
| | *SE(NF)* | 0,08410 | | | | | | | |
| | *Norm.GOI* | | | 7,02 | 1,08 | 1,46 | 0,41 | 1,72 | 2,85 |
| | *SE(Norm.GOI)* | | | 5,15 | 0,79 | 1,21 | 0,33 | 3,87 | 2,03 |
| 8 | **breast** | | | | | | | | |
| | *raw* | 24,7 | 27,8 | 34,3 | 16,6 | 28,3 | 24,2 | 23,9 | 27,5 |
| | | 25,1 | 27,8 | 34,3 | 15,7 | 28,2 | 26,2 | 24,4 | 28,6 |
| | | 24,6 | 27,2 | 36,9 | 15,7 | 33,8 | 25,3 | 23,1 | 28,1 |
| | *avg* | 24,80 | 27,60 | 35,17 | 16,00 | 28,25 | 25,23 | 23,80 | 28,07 |
| | *SDsample* | 0,26 | 0,35 | 1,50 | 0,52 | 3,20 | 1,00 | 0,66 | 0,55 |
| | *E^(deltaCt)* | 0,435 | 0,067 | 0,208 | 4,000 | 0,247 | 0,203 | 0,406 | 0,338 |
| | *SD*{E^(deltaCt)} | 0,080 | 0,016 | 0,216 | 1,441 | 0,549 | 0,141 | 0,185 | 0,129 |
| | *SE*{E^(deltaCt)} | 0,046 | 0,009 | 0,125 | 0,832 | 0,317 | 0,081 | 0,107 | 0,074 |
| | *NF* | 0,17077 | | | | | | | |
| | *SE(NF)* | 0,08722 | | | | | | | |
| | *Norm.GOI* | | | 1,22 | 23,42 | 1,45 | 1,19 | 2,38 | 1,98 |
| | *SE(Norm.GOI)* | | | 0,96 | 12,92 | 2,00 | 0,77 | 1,37 | 1,10 |
| 9 | **kidney** | | | | | | | | |
| | *raw* | 24,4 | 26,3 | 31,8 | 21,5 | 27,9 | 23,5 | 22,5 | 27,3 |
| | | 24,5 | 26 | 31,7 | 22,2 | 27,9 | 23,9 | 22 | 27,6 |
| | | 24,4 | 25,9 | 32,6 | 22,3 | 27,7 | 23,8 | 21,3 | 27,7 |
| | *avg* | 24,43 | 26,07 | 32,03 | 22,00 | 27,83 | 23,73 | 21,93 | 27,53 |
| | *SDsample* | 0,06 | 0,21 | 0,49 | 0,44 | 0,12 | 0,21 | 0,60 | 0,21 |
| | *E^(deltaCt)* | 0,561 | 0,174 | 1,823 | 0,063 | 0,330 | 0,574 | 1,481 | 0,489 |
| | *SD*{E^(deltaCt)} | 0,022 | 0,025 | 0,623 | 0,019 | 0,026 | 0,083 | 0,619 | 0,070 |
| | *SE*{E^(deltaCt)} | 0,013 | 0,015 | 0,360 | 0,011 | 0,015 | 0,048 | 0,357 | 0,041 |
| | *NF* | 0,31282 | | | | | | | |
| | *SE(NF)* | 0,04323 | | | | | | | |
| | *Norm.GOI* | | | 5,83 | 0,20 | 1,05 | 1,84 | 4,73 | 1,56 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *SE(Norm.GOI)* | | | 1,40 | 0,04 | 0,15 | 0,30 | 1,32 | 0,25 |
| **10** | **prostate** | | | | | | | | |
| | *raw* | 26,2 | 27,9 | 30,6 | 19,9 | 28,5 | 25,9 | 26,0 | 28,0 |
| | | 26,7 | 28,4 | 30,7 | 20,9 | 28,6 | 25,6 | 25,2 | 28,5 |
| | | 26,1 | 27,8 | 30,5 | 20 | 27,6 | 25 | 27,1 | 27,4 |
| | *avg* | 26,33 | 28,03 | 30,60 | 20,27 | 28,23 | 25,50 | 26,10 | 27,97 |
| | *SDsample* | 0,32 | 0,32 | 0,10 | 0,55 | 0,55 | 0,46 | 0,95 | 0,55 |
| | *E^(deltaCt)* | 0,150 | 0,051 | 4,925 | 0,208 | 0,250 | 0,169 | 0,082 | 0,362 |
| | *SD*{E^(deltaCt)} | 0,034 | 0,011 | 0,341 | 0,079 | 0,095 | 0,054 | 0,055 | 0,138 |
| | *SE*{E^(deltaCt)} | 0,019 | 0,007 | 0,197 | 0,046 | 0,055 | 0,031 | 0,031 | 0,080 |
| | *NF* | 0,08768 | | | | | | | |
| | *SE(NF)* | 0,09096 | | | | | | | |
| | *Norm.GOI* | | | 56,16 | 2,37 | 2,85 | 1,93 | 0,94 | 4,13 |
| | *SE(Norm.GOI)* | | | 58,31 | 2,51 | 3,02 | 2,03 | 1,04 | 4,38 |
| **11** | **bladder** | | | | | | | | |
| | *raw* | 25,1 | 26,7 | 29,1 | 18,1 | 29 | 24,6 | 23,8 | 27,8 |
| | | 23,9 | 27 | 28,3 | 16,6 | 27,4 | 24,6 | 24,4 | 27,8 |
| | | 24,3 | 27,3 | 29,4 | 16,7 | 34,2 | 25,3 | 23,7 | 28,2 |
| | *avg* | 24,43 | 27,00 | 28,93 | 17,13 | 28,20 | 24,83 | 23,97 | 27,93 |
| | *SDsample* | 0,61 | 0,30 | 0,57 | 0,84 | 3,56 | 0,40 | 0,38 | 0,23 |
| | *E^(deltaCt)* | 0,561 | 0,097 | 15,635 | 1,823 | 0,256 | 0,268 | 0,362 | 0,370 |
| | *SD*{E^(deltaCt)} | 0,238 | 0,020 | 6,162 | 1,060 | 0,630 | 0,075 | 0,095 | 0,059 |
| | *SE*{E^(deltaCt)} | 0,137 | 0,012 | 3,558 | 0,612 | 0,364 | 0,043 | 0,055 | 0,034 |
| | *NF* | 0,23381 | | | | | | | |
| | *SE(NF)* | 0,13620 | | | | | | | |
| | *Norm.GOI* | | | 66,87 | 7,80 | 1,09 | 1,15 | 1,55 | 1,58 |
| | *SE(Norm.GOI)* | | | 41,82 | 5,24 | 1,68 | 0,69 | 0,93 | 0,93 |
| **12** | **ovary** | | | | | | | | |
| | *raw* | 24,9 | 27,6 | 33,8 | 18,2 | 27,8 | 24,3 | 21,8 | 26,3 |
| | | 23,9 | 28,2 | 33,8 | 18,6 | 29,5 | 23,7 | 21,6 | 26,7 |
| | | 23,8 | 28,5 | 33,6 | 19,6 | 30,1 | 24,1 | 22,4 | 26,8 |
| | *avg* | 24,20 | 28,10 | 33,73 | 18,80 | 29,13 | 24,03 | 21,93 | 26,60 |
| | *SDsample* | 0,61 | 0,46 | 0,12 | 0,72 | 1,19 | 0,31 | 0,42 | 0,26 |
| | *E^(deltaCt)* | 0,660 | 0,049 | 0,561 | 0,574 | 0,134 | 0,467 | 1,481 | 0,933 |
| | *SD*{E^(deltaCt)} | 0,278 | 0,016 | 0,045 | 0,287 | 0,111 | 0,099 | 0,427 | 0,171 |
| | *SE*{E^(deltaCt)} | 0,161 | 0,009 | 0,026 | 0,166 | 0,064 | 0,057 | 0,247 | 0,099 |
| | *NF* | 0,17988 | | | | | | | |
| | *SE(NF)* | 0,15239 | | | | | | | |
| | *Norm.GOI* | | | 3,12 | 3,19 | 0,74 | 2,59 | 8,23 | 5,19 |
| | *SE(Norm.GOI)* | | | 2,65 | 2,86 | 0,72 | 2,22 | 7,11 | 4,43 |
| **13** | **brain** | | | | | | | | |
| | *raw* | 23,9 | 26,8 | 33,3 | 12,8 | 22,6 | 25,7 | 21 | 25,4 |
| | | 23,3 | 27 | 32,5 | 13,9 | 20,1 | 24,2 | 20,6 | 24,7 |
| | | 22,9 | 27,6 | 32,1 | 12,9 | 20,1 | 23,9 | 19,7 | 23,8 |
| | *avg* | 23,37 | 27,13 | 32,63 | 13,20 | 20,93 | 24,60 | 20,43 | 24,63 |
| | *SDsample* | 0,50 | 0,42 | 0,61 | 0,61 | 1,44 | 0,96 | 0,67 | 0,80 |
| | *E^(deltaCt)* | 1,176 | 0,090 | 1,203 | 27,858 | 39,397 | 0,315 | 4,189 | 3,647 |
| | *SD*{E^(deltaCt)} | 0,410 | 0,026 | 0,510 | 11,745 | 39,415 | 0,211 | 1,933 | 2,028 |
| | *SE*{E^(deltaCt)} | 0,237 | 0,015 | 0,294 | 6,781 | 22,756 | 0,122 | 1,116 | 1,171 |
| | *NF* | 0,32461 | | | | | | | |
| | *SE(NF)* | 0,13070 | | | | | | | |
| | *Norm.GOI* | | | 3,71 | 85,82 | 121,37 | 0,97 | 12,91 | 11,23 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *SE(Norm.GOI)* | | | 1,75 | 40,38 | 85,46 | 0,54 | 6,23 | 5,79 |
| **14** | **esophagus** | | | | | | | | |
| | *raw* | 25,8 | 27 | 30,1 | 18,6 | 27,7 | 26,8 | 24,9 | 27,7 |
| | | 25,6 | 27,6 | 31,5 | 19,6 | 27,9 | 25,4 | 24,1 | 27,7 |
| | | 25,5 | 27,4 | 31,6 | 19,5 | 28,7 | 26,7 | 24,1 | 27,6 |
| | *avg* | 25,63 | 27,33 | 31,07 | 19,23 | 28,10 | 26,30 | 24,37 | 27,67 |
| | *SDsample* | 0,15 | 0,31 | 0,84 | 0,55 | 0,53 | 0,78 | 0,46 | 0,06 |
| | *E^(deltaCt)* | 0,244 | 0,079 | 3,564 | 0,425 | 0,274 | 0,097 | 0,274 | 0,445 |
| | *SD*{E^(deltaCt)} | 0,026 | 0,017 | 2,072 | 0,162 | 0,101 | 0,052 | 0,088 | 0,018 |
| | *SE*{E^(deltaCt)} | 0,015 | 0,010 | 1,196 | 0,094 | 0,058 | 0,030 | 0,051 | 0,010 |
| | *NF* | 0,13903 | | | | | | | |
| | *SE(NF)* | 0,06835 | | | | | | | |
| | *Norm.GOI* | | | 25,63 | 3,06 | 1,97 | 0,70 | 1,97 | 3,20 |
| | *SE(Norm.GOI)* | | | 15,26 | 1,65 | 1,06 | 0,41 | 1,04 | 1,58 |
| **15** | **heart** | | | | | | | | |
| | *raw* | 26 | 27,8 | 30,7 | 18,4 | 28,1 | 27,6 | 24,9 | 28,2 |
| | | 26,1 | 27,9 | 30,4 | 21,2 | 27,7 | 25,1 | 24,1 | 28,4 |
| | | 26,6 | 27,5 | 30,2 | 19,4 | 28,7 | 27 | 23,2 | 28,3 |
| | *avg* | 26,23 | 27,73 | 30,43 | 19,67 | 28,17 | 26,57 | 24,07 | 28,30 |
| | *SDsample* | 0,32 | 0,21 | 0,25 | 1,42 | 0,50 | 1,31 | 0,85 | 0,10 |
| | *E^(deltaCt)* | 0,161 | 0,062 | 5,528 | 0,315 | 0,262 | 0,081 | 0,338 | 0,287 |
| | *SD*{E^(deltaCt)} | 0,036 | 0,009 | 0,964 | 0,310 | 0,091 | 0,073 | 0,199 | 0,020 |
| | *SE*{E^(deltaCt)} | 0,021 | 0,005 | 0,557 | 0,179 | 0,053 | 0,042 | 0,115 | 0,011 |
| | *NF* | 0,09968 | | | | | | | |
| | *SE(NF)* | 0,07663 | | | | | | | |
| | *Norm.GOI* | | | 55,45 | 3,16 | 2,63 | 0,81 | 3,39 | 2,88 |
| | *SE(Norm.GOI)* | | | 43,00 | 3,02 | 2,09 | 0,75 | 2,85 | 2,22 |
| **16** | **stomach** | | | | | | | | |
| | *raw* | 26,8 | 27 | 34,1 | 21,1 | 28,4 | 26,8 | 25,3 | 27,9 |
| | | 28 | 28 | 34,3 | 19,2 | 29,4 | 28,3 | 25,9 | 28,1 |
| | | 26,2 | 28,2 | 34,8 | 18,9 | 29,3 | 28,2 | N/A | 27,6 |
| | *avg* | 27,00 | 27,73 | 34,40 | 19,73 | 29,03 | 27,77 | 25,60 | 27,87 |
| | *SDsample* | 0,92 | 0,64 | 0,36 | 1,19 | 0,55 | 0,84 | 0,42 | 0,25 |
| | *E^(deltaCt)* | 0,095 | 0,062 | 0,354 | 0,301 | 0,144 | 0,035 | 0,117 | 0,388 |
| | *SD*{E^(deltaCt)} | 0,060 | 0,027 | 0,088 | 0,249 | 0,055 | 0,020 | 0,034 | 0,068 |
| | *SE*{E^(deltaCt)} | 0,035 | 0,016 | 0,051 | 0,144 | 0,032 | 0,012 | 0,020 | 0,039 |
| | *NF* | 0,07642 | | | | | | | |
| | *SE(NF)* | 0,22401 | | | | | | | |
| | *Norm.GOI* | | | 4,63 | 3,94 | 1,88 | 0,46 | 1,53 | 5,07 |
| | *SE(Norm.GOI)* | | | 13,58 | 11,69 | 5,52 | 1,35 | 4,48 | 14,88 |
| **17** | **lung** | | | | | | | | |
| | *raw* | 26,3 | 27,9 | 25 | 20,2 | 33,5 | 26,9 | 33,7 | 27,2 |
| | | 26,8 | 28,2 | 25,9 | 19,9 | 27,1 | 27,3 | 25,5 | 28 |
| | | 26,9 | 28,4 | 26,1 | 21,6 | 30,7 | 27,1 | 27,4 | 28,1 |
| | *avg* | 26,67 | 28,17 | 25,67 | 20,57 | 30,43 | 27,10 | 28,87 | 27,77 |
| | *SDsample* | 0,32 | 0,25 | 0,59 | 0,91 | 3,21 | 0,20 | 4,29 | 0,49 |
| | *E^(deltaCt)* | 0,119 | 0,047 | 150,470 | 0,169 | 0,054 | 0,056 | 0,012 | 0,416 |
| | *SD*{E^(deltaCt)} | 0,027 | 0,008 | 61,113 | 0,106 | 0,121 | 0,008 | 0,036 | 0,142 |
| | *SE*{E^(deltaCt)} | 0,015 | 0,005 | 35,284 | 0,061 | 0,070 | 0,004 | 0,021 | 0,082 |
| | *NF* | 0,07494 | | | | | | | |
| | *SE(NF)* | 0,08169 | | | | | | | |
| | *Norm.GOI* | | | 2008,00 | 2,25 | 0,73 | 0,74 | 0,16 | 5,55 |
| | *SE(Norm.GOI)* | | | 2239,02 | 2,59 | 1,22 | 0,81 | 0,33 | 6,14 |

### *ii) Lung Cancer Panel*

**Table F3 Detailed calculations about the real-time PCR data for lung cancers**

| | | *Reference Genes* | | *Genes in Question* | | | |
|---|---|---|---|---|---|---|---|
| | | **TFCP2** | **GOLGA1** | **ALAS2** | **CDR1** | **RRAGB** | **SLC9A6** |
| **1** | **Normal lung** | | | | | | |
| | *raw* | 24,1 | 27,9 | 29 | 20,5 | 24,3 | 25,8 |
| | | 26,2 | 29,6 | 29 | 20,8 | 25,4 | 24 |
| | | 26,4 | 28 | 28,7 | 21,2 | 24,5 | 23,5 |
| | *avg* | 25,57 | 28,50 | 28,90 | 20,83 | 24,73 | 24,43 |
| | *SDsample* | 1,27 | 0,95 | 0,17 | 0,35 | 0,59 | 1,21 |
| | *E^(deltaCt)* | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | *SD*{E^(deltaCt)} | 0,883 | 0,661 | 0,120 | 0,243 | 0,406 | 0,838 |
| | *SE*{E^(deltaCt)} | 0,510 | 0,382 | 0,069 | 0,141 | 0,234 | 0,484 |
| | *NF* | 1,00000 | | | | | |
| | *SE(NF)* | 0,31848 | | | | | |
| | *Norm.GOI* | | | 1,00 | 1,00 | 1,00 | 1,00 |
| | *SE(Norm.GOI)* | | | 0,33 | 0,35 | 0,40 | 0,58 |
| **2** | **SK-LC-1** | | | | | | |
| | *raw* | 22,9 | 26,8 | 32,3 | 22,7 | 23,9 | 23,6 |
| | | 23,4 | 27,5 | 31,4 | 25,4 | 24,3 | 24,1 |
| | | 24,7 | 26,8 | 34 | 23 | 24,6 | 23,9 |
| | *avg* | 23,67 | 27,03 | 32,57 | 23,70 | 24,27 | 23,87 |
| | *SDsample* | 0,93 | 0,40 | 1,32 | 1,48 | 0,35 | 0,25 |
| | *E^(deltaCt)* | 3,732 | 2,497 | 0,079 | 0,137 | 1,382 | 1,481 |
| | *SD*{E^(deltaCt)} | 2,404 | 0,699 | 0,072 | 0,141 | 0,336 | 0,258 |
| | *SE*{E^(deltaCt)} | 1,388 | 0,404 | 0,042 | 0,081 | 0,194 | 0,149 |
| | *NF* | 3,05244 | | | | | |
| | *SE(NF)* | 0,20274 | | | | | |
| | *Norm.GOI* | | | 0,03 | 0,04 | 0,45 | 0,49 |
| | *SE(Norm.GOI)* | | | 0,01 | 0,03 | 0,07 | 0,06 |
| **3** | **SK-LC-2** | | | | | | |
| | *raw* | 25,8 | 27,8 | 33,4 | 33,6 | 24,7 | 26,2 |
| | | 24,6 | 28 | 33,9 | 32,6 | 24,3 | 26,4 |
| | | 25,9 | 29,5 | 34,4 | 33 | 24,7 | 27 |
| | *avg* | 25,43 | 28,43 | 33,90 | 33,07 | 24,57 | 26,53 |
| | *SDsample* | 0,72 | 0,93 | 0,50 | 0,50 | 0,23 | 0,42 |
| | *E^(deltaCt)* | 1,097 | 1,042 | 0,031 | 0,000 | 1,122 | 0,233 |
| | *SD*{E^(deltaCt)} | 0,550 | 0,671 | 0,011 | 0,000 | 0,180 | 0,067 |
| | *SE*{E^(deltaCt)} | 0,318 | 0,388 | 0,006 | 0,000 | 0,104 | 0,039 |
| | *NF* | 1,06930 | | | | | |
| | *SE(NF)* | 0,23562 | | | | | |
| | *Norm.GOI* | | | 0,03 | 0,00 | 1,05 | 0,22 |
| | *SE(Norm.GOI)* | | | 0,01 | 0,00 | 0,25 | 0,06 |
| **4** | **SK-LC-6** | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *raw* | 26,6 | 29 | 34,7 | 21,1 | 26,7 | 26,6 |
| | | 25,8 | 30 | 33 | 21,7 | 26,8 | 25,3 |
| | | 25,8 | 29,9 | 33,2 | 21 | 27,6 | 24,9 |
| | *avg* | 26,07 | 29,63 | 33,63 | 21,27 | 27,03 | 25,60 |
| | *SDsample* | 0,46 | 0,55 | 0,93 | 0,38 | 0,49 | 0,89 |
| | *E^(deltaCt)* | 0,707 | 0,493 | 0,038 | 0,741 | 0,203 | 0,445 |
| | *SD*{E^(deltaCt)} | 0,226 | 0,188 | 0,024 | 0,194 | 0,069 | 0,274 |
| | *SE*{E^(deltaCt)} | 0,131 | 0,109 | 0,014 | 0,112 | 0,040 | 0,158 |
| | *NF* | 0,59051 | | | | | |
| | *SE(NF)* | 0,14383 | | | | | |
| | *Norm.GOI* | | | 0,06 | 1,25 | 0,34 | 0,75 |
| | *SE(Norm.GOI)* | | | 0,03 | 0,36 | 0,11 | 0,33 |
| **5** | **SK-LC-7** | | | | | | |
| | *raw* | 24,1 | 24,9 | 33,7 | 32,2 | 22,6 | 21,3 |
| | | 22,4 | 27 | 36,2 | 32,5 | 22,3 | 21,7 |
| | | 25,2 | 27,3 | 35 | 31,8 | 21,8 | 22,2 |
| | *avg* | 23,90 | 27,15 | 34,97 | 32,17 | 22,23 | 21,73 |
| | *SDsample* | 1,41 | 1,31 | 1,25 | 0,35 | 0,40 | 0,45 |
| | *E^(deltaCt)* | 3,175 | 2,321 | 0,015 | 0,000 | 5,657 | 6,498 |
| | *SD*{E^(deltaCt)} | 3,104 | 2,104 | 0,013 | 0,000 | 1,585 | 2,031 |
| | *SE*{E^(deltaCt)} | 1,792 | 1,215 | 0,007 | 0,000 | 0,915 | 1,173 |
| | *NF* | 2,71471 | | | | | |
| | *SE(NF)* | 0,38489 | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,00 | 2,08 | 2,39 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,45 | 0,55 |
| **6** | **SK-LC-8** | | | | | | |
| | *raw* | 23,6 | 25,2 | 34 | 30,7 | 23,4 | 23,3 |
| | | 23,3 | 26,8 | 34,1 | 31,6 | 23,7 | 23,7 |
| | | 24,2 | 27,5 | 34 | 32,5 | 23,9 | 24,3 |
| | *avg* | 23,70 | 26,50 | 34,03 | 31,60 | 23,67 | 23,77 |
| | *SDsample* | 0,46 | 1,18 | 0,06 | 0,90 | 0,25 | 0,50 |
| | *E^(deltaCt)* | 3,647 | 3,482 | 0,028 | 0,001 | 2,095 | 1,587 |
| | *SD*{E^(deltaCt)} | 1,158 | 2,845 | 0,001 | 0,000 | 0,365 | 0,554 |
| | *SE*{E^(deltaCt)} | 0,669 | 1,643 | 0,001 | 0,000 | 0,211 | 0,320 |
| | *NF* | 3,56347 | | | | | |
| | *SE(NF)* | 0,25310 | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,00 | 0,59 | 0,45 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,07 | 0,10 |
| **7** | **SK-LC-13** | | | | | | |
| | *raw* | 22,6 | 26,4 | 36,3 | 28,4 | 22,2 | 22,6 |
| | | 22,5 | 28,5 | 33,6 | 28,2 | 22,6 | 21,4 |
| | | 22,5 | 27 | 35 | 29,2 | 23,1 | 22,3 |
| | *avg* | 22,53 | 27,30 | 34,97 | 28,60 | 22,63 | 22,10 |
| | *SDsample* | 0,06 | 1,08 | 1,35 | 0,53 | 0,45 | 0,62 |
| | *E^(deltaCt)* | 8,187 | 2,114 | 0,015 | 0,005 | 4,287 | 5,040 |
| | *SD*{E^(deltaCt)} | 0,328 | 1,585 | 0,014 | 0,002 | 1,340 | 2,182 |
| | *SE*{E^(deltaCt)} | 0,189 | 0,915 | 0,008 | 0,001 | 0,774 | 1,260 |
| | *NF* | 4,16015 | | | | | |
| | *SE(NF)* | 0,21674 | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,00 | 1,03 | 1,21 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,19 | 0,31 |
| **8** | **SK-LC-14** | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *raw* | 23,2 | 25,4 | 33 | 19,7 | 22,6 | 22 |
| | | 23,7 | 26,9 | 36,8 | 20 | 23,3 | 22,3 |
| | | 26,1 | 25,6 | 32,9 | 20,2 | 23,1 | 21,7 |
| | *avg* | 24,33 | 25,97 | 34,23 | 19,97 | 23,00 | 22,00 |
| | *SDsample* | 1,55 | 0,81 | 2,22 | 0,25 | 0,36 | 0,30 |
| | *E^(deltaCt)* | 2,351 | 4,856 | 0,025 | 1,823 | 3,325 | 5,401 |
| | *SD*{E^(deltaCt)} | 2,526 | 2,742 | 0,038 | 0,318 | 0,831 | 1,123 |
| | *SE*{E^(deltaCt)} | 1,459 | 1,583 | 0,022 | 0,184 | 0,480 | 0,648 |
| | *NF* | 3,37901 | | | | | |
| | *SE(NF)* | 0,35040 | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,54 | 0,98 | 1,60 |
| | *SE(Norm.GOI)* | | | 0,01 | 0,08 | 0,17 | 0,25 |
| **9** | **SK-LC-17** | | | | | | |
| | *raw* | 22,1 | 24,7 | 35,6 | 29,8 | 24,0 | 24,1 |
| | | 23,1 | 27,7 | 34 | 29,2 | 24 | 24,1 |
| | | 22,4 | 26,6 | 33,5 | 29,8 | 24,2 | 24,6 |
| | *avg* | 22,53 | 26,33 | 34,37 | 29,60 | 24,07 | 24,27 |
| | *SDsample* | 0,51 | 1,52 | 1,10 | 0,35 | 0,12 | 0,29 |
| | *E^(deltaCt)* | 8,187 | 3,863 | 0,023 | 0,002 | 1,587 | 1,122 |
| | *SD*{E^(deltaCt)} | 2,912 | 4,064 | 0,017 | 0,001 | 0,127 | 0,225 |
| | *SE*{E^(deltaCt)} | 1,681 | 2,346 | 0,010 | 0,000 | 0,073 | 0,130 |
| | *NF* | 5,62406 | | | | | |
| | *SE(NF)* | 0,32057 | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,00 | 0,28 | 0,20 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,02 | 0,03 |
| **10** | **NCI-H69** | | | | | | |
| | *raw* | 21,7 | 27,8 | 35,4 | 20,5 | 23,1 | 24,0 |
| | | 21,7 | 28,4 | 34,7 | 20,8 | 22,9 | 24,2 |
| | | 22,4 | 28,4 | 33 | 21,1 | 23,6 | 23,7 |
| | *avg* | 21,93 | 28,20 | 34,37 | 20,80 | 23,20 | 23,97 |
| | *SDsample* | 0,40 | 0,35 | 1,23 | 0,30 | 0,36 | 0,25 |
| | *E^(deltaCt)* | 12,409 | 1,206 | 0,023 | 1,023 | 2,895 | 1,382 |
| | *SD*{E^(deltaCt)} | 3,476 | 0,290 | 0,019 | 0,213 | 0,723 | 0,241 |
| | *SE*{E^(deltaCt)} | 2,007 | 0,167 | 0,011 | 0,123 | 0,418 | 0,139 |
| | *NF* | 3,86819 | | | | | |
| | *SE(NF)* | 0,10651 | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,26 | 0,75 | 0,36 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,03 | 0,11 | 0,04 |
| **11** | **NCI-H82** | | | | | | |
| | *raw* | 22,6 | 26,5 | 32,8 | 34,8 | 25,5 | 22,1 |
| | | 23,6 | 28,5 | 34 | 33 | 25,8 | 22,7 |
| | | 22,3 | 28,4 | 32,9 | 33,3 | 25,1 | 22,5 |
| | *avg* | 22,83 | 27,80 | 33,23 | 33,70 | 25,47 | 22,43 |
| | *SDsample* | 0,68 | 1,13 | 0,67 | 0,96 | 0,35 | 0,31 |
| | *E^(deltaCt)* | 6,650 | 1,548 | 0,050 | 0,000 | 0,602 | 4,000 |
| | *SD*{E^(deltaCt)} | 3,138 | 1,209 | 0,023 | 0,000 | 0,146 | 0,847 |
| | *SE*{E^(deltaCt)} | 1,811 | 0,698 | 0,013 | 0,000 | 0,085 | 0,489 |
| | *NF* | 3,20794 | | | | | |
| | *SE(NF)* | 0,26344 | | | | | |
| | *Norm.GOI* | | | 0,02 | 0,00 | 0,19 | 1,25 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,03 | 0,18 |
| **12** | **NCI-H128** | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *raw* | 22 | 26,3 | 33,2 | 32,7 | 24,7 | 23,3 |
| | | 21,9 | 26,4 | 35 | 32,1 | 22,9 | 22,5 |
| | | 22,2 | 27,5 | 35,5 | 34,4 | 23,4 | 24,3 |
| | *avg* | 22,03 | 26,73 | 34,57 | 33,07 | 23,67 | 23,37 |
| | *SDsample* | 0,15 | 0,67 | 1,21 | 1,19 | 0,93 | 0,90 |
| | *E^(deltaCt)* | 11,578 | 3,010 | 0,020 | 0,000 | 2,095 | 2,095 |
| | *SD*{E^(deltaCt)} | 1,226 | 1,389 | 0,017 | 0,000 | 1,349 | 1,309 |
| | *SE*{E^(deltaCt)} | 0,708 | 0,802 | 0,010 | 0,000 | 0,779 | 0,756 |
| | *NF* | 5,90371 | | | | | |
| | *SE(NF)* | 0,13669 | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,00 | 0,35 | 0,35 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,13 | 0,13 |
| **13** | **NCI-H146** | | | | | | |
| | *raw* | 23,3 | 28,5 | 33,3 | 35,7 | 25,5 | 22,8 |
| | | 24,2 | 28,2 | 33,7 | 36,4 | 23,2 | 23 |
| | | 22,6 | 28,5 | 32,3 | 33,6 | 23 | 24,9 |
| | *avg* | 23,37 | 28,40 | 33,10 | 35,23 | 23,90 | 23,57 |
| | *SDsample* | 0,80 | 0,17 | 0,72 | 1,46 | 1,39 | 1,16 |
| | *E^(deltaCt)* | 4,595 | 1,064 | 0,054 | 0,000 | 1,782 | 1,823 |
| | *SD*{E^(deltaCt)} | 2,555 | 0,128 | 0,027 | 0,000 | 1,716 | 1,465 |
| | *SE*{E^(deltaCt)} | 1,475 | 0,074 | 0,016 | 0,000 | 0,991 | 0,846 |
| | *NF* | 2,21146 | | | | | |
| | *SE(NF)* | 0,16419 | | | | | |
| | *Norm.GOI* | | | 0,02 | 0,00 | 0,81 | 0,82 |
| | *SE(Norm.GOI)* | | | 0,01 | 0,00 | 0,45 | 0,39 |
| **14** | **NCI-H209** | | | | | | |
| | *raw* | 23,8 | 27 | 35,7 | 29,5 | 24,1 | 23,6 |
| | | 24 | 26,9 | 36,1 | 28,8 | 25,5 | 23,4 |
| | | 23,6 | 25,7 | 34 | 28,9 | 24,2 | 23,1 |
| | *avg* | 23,80 | 26,53 | 35,27 | 29,07 | 24,60 | 23,37 |
| | *SDsample* | 0,20 | 0,72 | 1,12 | 0,38 | 0,78 | 0,25 |
| | *E^(deltaCt)* | 3,403 | 3,410 | 0,012 | 0,003 | 1,097 | 2,095 |
| | *SD*{E^(deltaCt)} | 0,472 | 1,710 | 0,009 | 0,001 | 0,594 | 0,365 |
| | *SE*{E^(deltaCt)} | 0,272 | 0,987 | 0,005 | 0,001 | 0,343 | 0,211 |
| | *NF* | 3,40648 | | | | | |
| | *SE(NF)* | 0,15018 | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,00 | 0,32 | 0,61 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,10 | 0,07 |
| **15** | **NCI-H345** | | | | | | |
| | *raw* | 24,8 | 28 | 33,5 | 22,7 | 21,5 | 21,6 |
| | | 25 | 27,3 | 36,3 | 23,1 | 22,1 | 21,6 |
| | | 25,3 | 29,1 | 34,3 | 23,4 | 21,9 | 22 |
| | *avg* | 25,03 | 28,13 | 34,70 | 23,07 | 21,83 | 21,73 |
| | *SDsample* | 0,25 | 0,91 | 1,44 | 0,35 | 0,31 | 0,23 |
| | *E^(deltaCt)* | 1,447 | 1,257 | 0,018 | 0,213 | 7,464 | 6,498 |
| | *SD*{E^(deltaCt)} | 0,252 | 0,791 | 0,018 | 0,052 | 1,581 | 1,040 |
| | *SE*{E^(deltaCt)} | 0,146 | 0,456 | 0,010 | 0,030 | 0,913 | 0,601 |
| | *NF* | 1,34878 | | | | | |
| | *SE(NF)* | 0,18841 | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,16 | 5,53 | 4,82 |
| | *SE(Norm.GOI)* | | | 0,01 | 0,03 | 1,03 | 0,81 |
| **16** | **NCI-H378** | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *raw* | 25,5 | 28 | 34,5 | 19,3 | 23,6 | 23,5 |
| | | 24,3 | 28,9 | 35,9 | 20,6 | 24,3 | 23,4 |
| | | 26,9 | 29 | 31,9 | 19,3 | 27,1 | 24,2 |
| | *avg* | 25,57 | 28,63 | 34,10 | 19,73 | 25,00 | 23,70 |
| | *SDsample* | 1,30 | 0,55 | 2,03 | 0,75 | 1,85 | 0,44 |
| | *E^(deltaCt)* | 1,000 | 0,920 | 0,027 | 2,144 | 0,831 | 1,662 |
| | *SD*{E^(deltaCt)} | 0,902 | 0,351 | 0,038 | 1,115 | 1,067 | 0,502 |
| | *SE*{E^(deltaCt)} | 0,521 | 0,203 | 0,022 | 0,644 | 0,616 | 0,290 |
| | *NF* | 0,95927 | | | | | |
| | *SE(NF)* | 0,28274 | | | | | |
| | *Norm.GOI* | | | 0,03 | 2,23 | 0,87 | 1,73 |
| | *SE(Norm.GOI)* | | | 0,02 | 0,94 | 0,69 | 0,59 |
| **17** | **KNS62** | | | | | | |
| | *raw* | 23,9 | 27,5 | 33,3 | 30,5 | 24,1 | 23,3 |
| | | 23,7 | 28,4 | 32,9 | 30,7 | 23,8 | 25,1 |
| | | 23,9 | 25,9 | 33,4 | 31,4 | 24,6 | 26,8 |
| | *avg* | 23,83 | 27,27 | 33,20 | 30,87 | 24,17 | 25,07 |
| | *SDsample* | 0,12 | 1,27 | 0,26 | 0,47 | 0,40 | 1,75 |
| | *E^(deltaCt)* | 3,325 | 2,158 | 0,051 | 0,001 | 1,481 | 0,645 |
| | *SD*{E^(deltaCt)} | 0,266 | 1,894 | 0,009 | 0,000 | 0,415 | 0,782 |
| | *SE*{E^(deltaCt)} | 0,154 | 1,094 | 0,005 | 0,000 | 0,240 | 0,452 |
| | *NF* | 2,67889 | | | | | |
| | *SE(NF)* | 0,25442 | | | | | |
| | *Norm.GOI* | | | 0,02 | 0,00 | 0,55 | 0,24 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,10 | 0,17 |
| **18** | **NCI-H889** | | | | | | |
| | *raw* | 21,8 | 27,7 | 34,1 | 33,6 | 23,4 | 22,1 |
| | | 22,3 | 28,5 | 33,7 | 33 | 24,5 | 22,2 |
| | | 22,2 | 28,2 | 36,2 | 31,7 | 23,7 | 22 |
| | *avg* | 22,10 | 28,13 | 34,67 | 32,77 | 23,87 | 22,10 |
| | *SDsample* | 0,26 | 0,40 | 1,34 | 0,97 | 0,57 | 0,10 |
| | *E^(deltaCt)* | 11,055 | 1,257 | 0,018 | 0,000 | 1,823 | 5,040 |
| | *SD*{E^(deltaCt)} | 2,027 | 0,352 | 0,017 | 0,000 | 0,719 | 0,349 |
| | *SE*{E^(deltaCt)} | 1,171 | 0,203 | 0,010 | 0,000 | 0,415 | 0,202 |
| | *NF* | 3,72780 | | | | | |
| | *SE(NF)* | 0,09665 | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,00 | 0,49 | 1,35 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,11 | 0,06 |
| **19** | **Calu1** | | | | | | |
| | *raw* | 25,2 | 28 | 32,2 | 28,1 | 24,9 | 25,5 |
| | | 25,4 | 27,5 | 32,6 | 28 | 24,8 | 24,8 |
| | | 25,8 | 27,9 | 31,1 | 28,4 | 25,5 | 25,7 |
| | *avg* | 25,47 | 27,80 | 31,97 | 28,17 | 25,07 | 25,33 |
| | *SDsample* | 0,31 | 0,26 | 0,78 | 0,21 | 0,38 | 0,47 |
| | *E^(deltaCt)* | 1,072 | 1,548 | 0,119 | 0,006 | 0,794 | 0,536 |
| | *SD*{E^(deltaCt)} | 0,227 | 0,284 | 0,064 | 0,001 | 0,208 | 0,176 |
| | *SE*{E^(deltaCt)} | 0,131 | 0,164 | 0,037 | 0,001 | 0,120 | 0,101 |
| | *NF* | 1,28787 | | | | | |
| | *SE(NF)* | 0,08087 | | | | | |
| | *Norm.GOI* | | | 0,09 | 0,00 | 0,62 | 0,42 |
| | *SE(Norm.GOI)* | | | 0,03 | 0,00 | 0,10 | 0,08 |
| **20** | **Calu3** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| *raw* | 26 | 28,4 | 30,9 | 35,6 | 25,2 | 23,9 |
| | 26,3 | 28,6 | 30,7 | 34,9 | 25 | 23,1 |
| | 26,8 | 28 | 31,1 | 34 | 24,9 | 23,9 |
| *avg* | 26,37 | 28,33 | 30,90 | 34,83 | 25,03 | 23,63 |
| *SDsample* | 0,40 | 0,31 | 0,20 | 0,80 | 0,15 | 0,46 |
| *E^(deltaCt)* | 0,574 | 1,110 | 0,250 | 0,000 | 0,812 | 1,741 |
| *SD*{E^(deltaCt)} | 0,161 | 0,235 | 0,035 | 0,000 | 0,086 | 0,557 |
| *SE*{E^(deltaCt)} | 0,093 | 0,136 | 0,020 | 0,000 | 0,050 | 0,322 |
| NF | 0,79830 | | | | | |
| *SE(NF)* | 0,10137 | | | | | |
| *Norm.GOI* | | | 0,31 | 0,00 | 1,02 | 2,18 |
| *SE(Norm.GOI)* | | | 0,05 | 0,00 | 0,14 | 0,49 |

### *iii) Colon Cancer Panel*

**Table F4 Detailed calculations about the real-time PCR data for colon cancers**

| | | ***Reference Genes*** | | ***Genes in Question*** | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **TFCP2** | **GOLGA1** | **CDR1** | **GABRA3** | **RRAGB** | **SLC9A6** | **ZCCHC12** |
| **1** | **Normal colon** | | | | | | | |
| | *raw* | 26,8 | 24,10 | 21,4 | 30,9 | 22,9 | 23,2 | 30,5 |
| | | 24,4 | 24,50 | 22,6 | 30,8 | 23 | 23 | 30,3 |
| | | 21,9 | 25,90 | 21,1 | 32,1 | 23,3 | 23,4 | 29,4 |
| | *avg* | 24,37 | 24,83 | 21,70 | 31,27 | 23,07 | 23,20 | 30,07 |
| | *SDsample* | 2,45 | 0,95 | 0,79 | 0,72 | 0,21 | 0,20 | 0,59 |
| | *E^(deltaCt)* | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | *SD*{E^(deltaCt)} | 1,698 | 0,655 | 0,550 | 0,501 | 0,144 | 0,139 | 0,406 |
| | *SE*{E^(deltaCt)} | 0,981 | 0,378 | 0,318 | 0,290 | 0,083 | 0,080 | 0,234 |
| | *NF* | 1,00000 | | | | | | |
| | *SE(NF)* | 0,52548 | | | | | | |
| | *Norm.GOI* | | | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| | *SE(Norm.GOI)* | | | 0,61 | 0,60 | 0,53 | 0,53 | 0,58 |
| **2** | **SW403** | | | | | | | |
| | *raw* | 22,3 | 25,70 | 27,7 | 22,6 | 24,7 | 22,3 | 33,6 |
| | | 23,3 | 28,80 | 27,4 | 24,6 | 26,4 | 23,1 | 34,7 |
| | | 23,8 | 27,30 | 28,5 | 24,2 | 26,3 | 23,6 | 34,1 |
| | *avg* | 23,13 | 27,27 | 27,87 | 23,80 | 25,80 | 23,00 | 34,13 |
| | *SDsample* | 0,76 | 1,55 | 0,57 | 1,06 | 0,95 | 0,66 | 0,55 |
| | *E^(deltaCt)* | 2,351 | 0,219 | 0,014 | 176,885 | 0,150 | 1,149 | 0,060 |
| | *SD*{E^(deltaCt)} | 1,245 | 0,236 | 0,005 | 129,755 | 0,099 | 0,522 | 0,023 |
| | *SE*{E^(deltaCt)} | 0,719 | 0,136 | 0,003 | 74,914 | 0,057 | 0,301 | 0,013 |
| | *NF* | 0,71784 | | | | | | |
| | *SE(NF)* | 0,34580 | | | | | | |
| | *Norm.GOI* | | | 0,02 | 246,41 | 0,21 | 1,60 | 0,08 |
| | *SE(Norm.GOI)* | | | 0,01 | 158,06 | 0,13 | 0,88 | 0,04 |
| **3** | **SW480** | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *raw* | 23,5 | 27,80 | 29 | 28,6 | 23,6 | 22,3 | 28,7 |
| | | 25,2 | 28,50 | 29,1 | 29 | 26,4 | 23,2 | 28,7 |
| | | 25,2 | 28,20 | 30,6 | 29,3 | 24,7 | 23,5 | 31,4 |
| | *avg* | 24,63 | 28,17 | 29,57 | 28,97 | 24,90 | 23,00 | 29,60 |
| | *SDsample* | 0,98 | 0,35 | 0,90 | 0,35 | 1,41 | 0,62 | 1,56 |
| | *E^(deltaCt)* | 0,831 | 0,125 | 0,004 | 4,925 | 0,281 | 1,149 | 1,382 |
| | *SD*{E^(deltaCt)} | 0,566 | 0,030 | 0,003 | 1,199 | 0,274 | 0,497 | 1,493 |
| | *SE*{E^(deltaCt)} | 0,326 | 0,018 | 0,002 | 0,692 | 0,158 | 0,287 | 0,862 |
| | *NF* | 0,32236 | | | | | | |
| | *SE(NF)* | 0,20858 | | | | | | |
| | *Norm.GOI* | | | 0,01 | 15,28 | 0,87 | 3,56 | 4,29 |
| | *SE(Norm.GOI)* | | | 0,01 | 10,12 | 0,75 | 2,47 | 3,85 |
| **4** | **SW620** | | | | | | | |
| | *raw* | 26,9 | 29,60 | 29,7 | 29,8 | 28 | 27,2 | 37,4 |
| | | 27,6 | 29,90 | 31,8 | 32,1 | 30,4 | 25 | 33,7 |
| | | 25,7 | 28,20 | 32,6 | 29,4 | 26 | 26,2 | 33,5 |
| | *avg* | 26,73 | 29,23 | 31,37 | 30,43 | 28,13 | 26,13 | 34,87 |
| | *SDsample* | 0,96 | 0,91 | 1,50 | 1,46 | 2,20 | 1,10 | 2,20 |
| | *E^(deltaCt)* | 0,194 | 0,064 | 0,001 | 1,782 | 0,030 | 0,131 | 0,036 |
| | *SD*{E^(deltaCt)} | 0,129 | 0,040 | 0,001 | 1,800 | 0,046 | 0,100 | 0,055 |
| | *SE*{E^(deltaCt)} | 0,075 | 0,023 | 0,001 | 1,039 | 0,026 | 0,058 | 0,032 |
| | *NF* | 0,11163 | | | | | | |
| | *SE(NF)* | 0,26445 | | | | | | |
| | *Norm.GOI* | | | 0,01 | 15,96 | 0,27 | 1,17 | 0,32 |
| | *SE(Norm.GOI)* | | | 0,03 | 38,94 | 0,68 | 2,83 | 0,81 |
| **5** | **SW837** | | | | | | | |
| | *raw* | 22,7 | 26,50 | 30,8 | 22,8 | 24,8 | 21,4 | 28,9 |
| | | 23,3 | 25,20 | 30,5 | 23,2 | 24 | 21,8 | 28,6 |
| | | 21,7 | 24,80 | 30,7 | 23,3 | 25,9 | 22 | 28,9 |
| | *avg* | 22,57 | 25,50 | 30,67 | 23,10 | 24,90 | 21,73 | 28,80 |
| | *SDsample* | 0,81 | 0,89 | 0,15 | 0,26 | 0,95 | 0,31 | 0,17 |
| | *E^(deltaCt)* | 3,482 | 0,660 | 0,002 | 287,350 | 0,281 | 2,764 | 2,406 |
| | *SD*{E^(deltaCt)} | 1,951 | 0,406 | 0,000 | 52,697 | 0,186 | 0,585 | 0,289 |
| | *SE*{E^(deltaCt)} | 1,126 | 0,235 | 0,000 | 30,425 | 0,107 | 0,338 | 0,167 |
| | *NF* | 1,51573 | | | | | | |
| | *SE(NF)* | 0,24039 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 189,58 | 0,19 | 1,82 | 1,59 |
| | *SE(Norm.GOI)* | | | 0,00 | 36,15 | 0,08 | 0,37 | 0,27 |
| **6** | **SW1222** | | | | | | | |
| | *raw* | 25,3 | 27,60 | 33 | 33,1 | 24,8 | 22,3 | 33,5 |
| | | 25,2 | 27,50 | 30,6 | 32,1 | 24,2 | 23,1 | 32,9 |
| | | 24,6 | 26,70 | 31,6 | 33,1 | 24,5 | 23 | 33,7 |
| | *avg* | 25,03 | 27,27 | 31,73 | 32,77 | 24,50 | 22,80 | 33,37 |
| | *SDsample* | 0,38 | 0,49 | 1,21 | 0,58 | 0,30 | 0,44 | 0,42 |
| | *E^(deltaCt)* | 0,630 | 0,219 | 0,001 | 0,354 | 0,370 | 1,320 | 0,102 |
| | *SD*{E^(deltaCt)} | 0,165 | 0,075 | 0,001 | 0,141 | 0,077 | 0,399 | 0,029 |
| | *SE*{E^(deltaCt)} | 0,095 | 0,043 | 0,000 | 0,082 | 0,044 | 0,230 | 0,017 |
| | *NF* | 0,37158 | | | | | | |
| | *SE(NF)* | 0,12442 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,95 | 1,00 | 3,55 | 0,27 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,39 | 0,35 | 1,34 | 0,10 |
| **7** | **HCT15** | | | | | | | |
| | *raw* | 22,8 | 25,50 | 26,2 | 25,2 | 23,7 | 20,6 | 29 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 23,2 | 23,90 | 27,3 | 26,1 | 22,5 | 21,1 | 28,2 |
| | | 25,3 | 26,80 | 27,7 | 26,4 | 23,5 | 21,5 | 29,1 |
| | *avg* | 23,77 | 25,40 | 27,07 | 25,90 | 23,23 | 21,07 | 28,77 |
| | *SDsample* | 1,34 | 1,45 | 0,78 | 0,62 | 0,64 | 0,45 | 0,49 |
| | *E^(deltaCt)* | 1,516 | 0,702 | 0,024 | 41,260 | 0,891 | 4,387 | 2,462 |
| | *SD*{E^(deltaCt)} | 1,411 | 0,707 | 0,013 | 17,860 | 0,397 | 1,371 | 0,842 |
| | *SE*{E^(deltaCt)} | 0,815 | 0,408 | 0,008 | 10,312 | 0,229 | 0,792 | 0,486 |
| | *NF* | 1,03169 | | | | | | |
| | *SE(NF)* | 0,39583 | | | | | | |
| | *Norm.GOI* | | | 0,02 | 39,99 | 0,86 | 4,25 | 2,39 |
| | *SE(Norm.GOI)* | | | 0,01 | 18,31 | 0,40 | 1,80 | 1,03 |
| **8** | **HCT116** | | | | | | | |
| | *raw* | 24,8 | 29,40 | 30,3 | 37,2 | 24,3 | 22,9 | 29,2 |
| | | 25,3 | 27,60 | 31,3 | 37,2 | 24,8 | 23 | 30,4 |
| | | 24,4 | 27,40 | 30,8 | 37,2 | 24,5 | 23,1 | 29,3 |
| | *avg* | 24,83 | 28,13 | 30,80 | 37,20 | 24,53 | 23,00 | 29,63 |
| | *SDsample* | 0,45 | 1,10 | 0,50 | 0,00 | 0,25 | 0,10 | 0,67 |
| | *E^(deltaCt)* | 0,724 | 0,128 | 0,002 | 0,016 | 0,362 | 1,149 | 1,350 |
| | *SD*{E^(deltaCt)} | 0,226 | 0,097 | 0,001 | 0,000 | 0,063 | 0,080 | 0,623 |
| | *SE*{E^(deltaCt)} | 0,131 | 0,056 | 0,000 | 0,000 | 0,036 | 0,046 | 0,360 |
| | *NF* | 0,30392 | | | | | | |
| | *SE(NF)* | 0,23816 | | | | | | |
| | *Norm.GOI* | | | 0,01 | 0,05 | 1,19 | 3,78 | 4,44 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,04 | 0,94 | 2,97 | 3,68 |
| **9** | **SK-CO-10** | | | | | | | |
| | *raw* | 22,9 | 27,60 | 27,9 | 23,2 | 24,8 | 20,7 | 32,7 |
| | | 22,4 | 27,90 | 28,7 | 23,8 | 25,8 | 21,9 | 33,7 |
| | | 22,5 | 25,20 | 31 | 22,3 | 26,4 | 20,7 | 31,3 |
| | *avg* | 22,60 | 26,90 | 29,20 | 23,10 | 25,67 | 21,10 | 32,57 |
| | *SDsample* | 0,26 | 1,48 | 1,61 | 0,75 | 0,81 | 0,69 | 1,21 |
| | *E^(deltaCt)* | 3,403 | 0,275 | 0,006 | 287,350 | 0,165 | 4,287 | 0,177 |
| | *SD*{E^(deltaCt)} | 0,624 | 0,283 | 0,006 | 150,375 | 0,092 | 2,059 | 0,148 |
| | *SE*{E^(deltaCt)} | 0,360 | 0,163 | 0,004 | 86,819 | 0,053 | 1,189 | 0,085 |
| | *NF* | 0,96821 | | | | | | |
| | *SE(NF)* | 0,30081 | | | | | | |
| | *Norm.GOI* | | | 0,01 | 296,79 | 0,17 | 4,43 | 0,18 |
| | *SE(Norm.GOI)* | | | 0,00 | 128,62 | 0,08 | 1,84 | 0,10 |
| **10** | **SK-CO-17** | | | | | | | |
| | *raw* | 22,9 | 25,00 | 31,0 | 25,7 | 27,2 | 23,2 | 32,5 |
| | | 23,4 | 26,50 | 29,3 | 26,6 | 25,4 | 22,9 | 31,1 |
| | | 24,5 | 28,00 | 33,2 | 23,9 | 25,6 | 23,7 | 33,3 |
| | *avg* | 23,60 | 26,50 | 31,17 | 25,40 | 26,07 | 23,27 | 32,30 |
| | *SDsample* | 0,82 | 1,50 | 1,96 | 1,37 | 0,99 | 0,40 | 1,11 |
| | *E^(deltaCt)* | 1,701 | 0,354 | 0,001 | 58,350 | 0,125 | 0,955 | 0,213 |
| | *SD*{E^(deltaCt)} | 0,965 | 0,368 | 0,002 | 55,603 | 0,085 | 0,267 | 0,164 |
| | *SE*{E^(deltaCt)} | 0,557 | 0,212 | 0,001 | 32,102 | 0,049 | 0,154 | 0,095 |
| | *NF* | 0,77560 | | | | | | |
| | *SE(NF)* | 0,34192 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 75,23 | 0,16 | 1,23 | 0,27 |
| | *SE(Norm.GOI)* | | | 0,00 | 53,04 | 0,10 | 0,58 | 0,17 |
| **11** | **SK-CO-19** | | | | | | | |
| | *raw* | 23,2 | 27,90 | 31,8 | 26 | 24,3 | 24,3 | 29,8 |
| | | 24,3 | 28,60 | 33,9 | 25,8 | 25 | 23,4 | 30,8 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 23,9 | 27,90 | 34 | 27,4 | 24,7 | 24,1 | 29,8 |
| | *avg* | 23,80 | 28,13 | 33,23 | 26,40 | 24,67 | 23,93 | 30,13 |
| | *SDsample* | 0,56 | 0,40 | 1,24 | 0,87 | 0,35 | 0,47 | 0,58 |
| | *E^(deltaCt)* | 1,481 | 0,128 | 0,000 | 29,175 | 0,330 | 0,602 | 0,955 |
| | *SD*{E^(deltaCt)} | 0,572 | 0,036 | 0,000 | 17,630 | 0,080 | 0,197 | 0,382 |
| | *SE*{E^(deltaCt)} | 0,330 | 0,021 | 0,000 | 10,179 | 0,046 | 0,114 | 0,221 |
| | *NF* | 0,43480 | | | | | | |
| | *SE(NF)* | 0,13766 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 67,10 | 0,76 | 1,38 | 2,20 |
| | *SE(Norm.GOI)* | | | 0,00 | 31,61 | 0,26 | 0,51 | 0,86 |
| **12** | **CaCO2** | | | | | | | |
| | *raw* | 21,7 | 24,10 | 32,8 | 37 | 22,9 | 22,3 | 32,1 |
| | | 21,9 | 24,20 | 31,6 | 35,9 | 22,8 | 21,5 | 32,1 |
| | | 22,1 | 24,10 | 30,5 | 36 | 23,7 | 22,6 | 33,4 |
| | *avg* | 21,90 | 24,13 | 31,63 | 36,30 | 23,13 | 22,13 | 32,53 |
| | *SDsample* | 0,20 | 0,06 | 1,15 | 0,61 | 0,49 | 0,57 | 0,75 |
| | *E^(deltaCt)* | 5,528 | 1,548 | 0,001 | 0,031 | 0,955 | 2,095 | 0,181 |
| | *SD*{E^(deltaCt)} | 0,766 | 0,062 | 0,001 | 0,013 | 0,326 | 0,826 | 0,094 |
| | *SE*{E^(deltaCt)} | 0,442 | 0,036 | 0,000 | 0,007 | 0,188 | 0,477 | 0,054 |
| | *NF* | 2,92475 | | | | | | |
| | *SE(NF)* | 0,04165 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,01 | 0,33 | 0,72 | 0,06 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,00 | 0,06 | 0,16 | 0,02 |
| **13** | **Colo205** | | | | | | | |
| | *raw* | 24 | 23,60 | 31 | 28,7 | 25 | 23,2 | 36,7 |
| | | 24,9 | 26,80 | 29 | 29,3 | 25,5 | 23,4 | 38,3 |
| | | 27 | 25,30 | 32,2 | 30,6 | 25,4 | 23,3 | 35,4 |
| | *avg* | 25,30 | 25,23 | 30,73 | 29,53 | 25,30 | 23,30 | 36,80 |
| | *SDsample* | 1,54 | 1,60 | 1,62 | 0,97 | 0,26 | 0,10 | 1,45 |
| | *E^(deltaCt)* | 0,524 | 0,779 | 0,002 | 3,325 | 0,213 | 0,933 | 0,009 |
| | *SD*{E^(deltaCt)} | 0,559 | 0,865 | 0,002 | 2,238 | 0,039 | 0,065 | 0,009 |
| | *SE*{E^(deltaCt)} | 0,323 | 0,499 | 0,001 | 1,292 | 0,023 | 0,037 | 0,005 |
| | *NF* | 0,63876 | | | | | | |
| | *SE(NF)* | 0,44443 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 5,21 | 0,33 | 1,46 | 0,01 |
| | *SE(Norm.GOI)* | | | 0,00 | 4,15 | 0,23 | 1,02 | 0,01 |
| **14** | **Lim1215** | | | | | | | |
| | *raw* | 23,4 | 25,80 | 33,1 | 36,9 | 24,5 | 23 | 34,5 |
| | | 25,8 | 26,40 | 32,9 | 37,3 | 25,6 | 21,9 | 33,1 |
| | | 23 | 27,10 | 31,9 | 37 | 24,7 | 22,2 | 32,9 |
| | *avg* | 24,07 | 26,43 | 32,63 | 37,07 | 24,93 | 22,37 | 33,50 |
| | *SDsample* | 1,51 | 0,65 | 0,64 | 0,21 | 0,59 | 0,57 | 0,87 |
| | *E^(deltaCt)* | 1,231 | 0,369 | 0,001 | 0,018 | 0,274 | 1,782 | 0,093 |
| | *SD*{E^(deltaCt)} | 1,292 | 0,166 | 0,000 | 0,003 | 0,111 | 0,702 | 0,056 |
| | *SE*{E^(deltaCt)} | 0,746 | 0,096 | 0,000 | 0,001 | 0,064 | 0,405 | 0,032 |
| | *NF* | 0,67364 | | | | | | |
| | *SE(NF)* | 0,32980 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,03 | 0,41 | 2,65 | 0,14 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,01 | 0,22 | 1,43 | 0,08 |
| **15** | **Ls180** | | | | | | | |
| | *raw* | 23,2 | 25,10 | 28,9 | 27,2 | 24 | 20,9 | 33,6 |
| | | 23,7 | 24,60 | 28,8 | 29,4 | 25,1 | 22,8 | 33,2 |
| | | 23,5 | 25,60 | 29 | 27,4 | 26,1 | 23,1 | 33,6 |

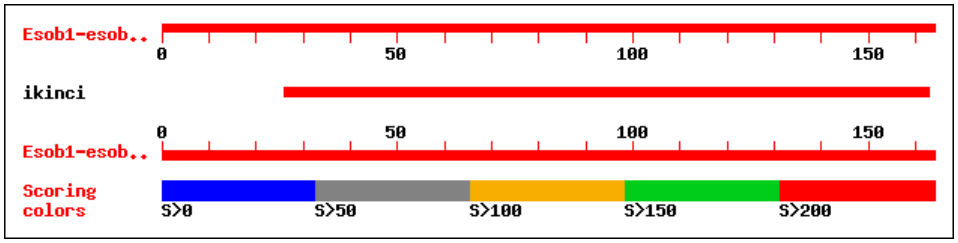
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *avg* | 23,47 | 25,10 | 28,90 | 28,00 | 25,07 | 22,27 | 33,47 |
| | *SDsample* | 0,25 | 0,50 | 0,10 | 1,22 | 1,05 | 1,19 | 0,23 |
| | *E^(deltaCt)* | 1,866 | 0,847 | 0,007 | 9,624 | 0,250 | 1,910 | 0,095 |
| | *SD*{E^(deltaCt)} | 0,326 | 0,293 | 0,000 | 8,116 | 0,182 | 1,579 | 0,015 |
| | *SE*{E^(deltaCt)} | 0,188 | 0,169 | 0,000 | 4,686 | 0,105 | 0,912 | 0,009 |
| | *NF* | 1,25702 | | | | | | |
| | *SE(NF)* | 0,11201 | | | | | | |
| | *Norm.GOI* | | | 0,01 | 7,66 | 0,20 | 1,52 | 0,08 |
| | *SE(Norm.GOI)* | | | 0,00 | 3,79 | 0,09 | 0,74 | 0,01 |
| **16** | **Ls174-T** | | | | | | | |
| | *raw* | 29,3 | 28,50 | 30,3 | 28 | 26,7 | 27,2 | 34,9 |
| | | 29,5 | 28,40 | 31,4 | 30,2 | 25,9 | 31 | 36,5 |
| | | 29,6 | 30,00 | 29,9 | 29,6 | 26,1 | 30,9 | 33,7 |
| | *avg* | 29,47 | 28,97 | 30,53 | 29,27 | 26,23 | 29,70 | 35,03 |
| | *SDsample* | 0,15 | 0,90 | 0,78 | 1,14 | 0,42 | 2,17 | 1,40 |
| | *E^(deltaCt)* | 0,029 | 0,076 | 0,002 | 4,000 | 0,111 | 0,011 | 0,032 |
| | *SD*{E^(deltaCt)} | 0,003 | 0,047 | 0,001 | 3,153 | 0,032 | 0,017 | 0,031 |
| | *SE*{E^(deltaCt)} | 0,002 | 0,027 | 0,001 | 1,820 | 0,019 | 0,010 | 0,018 |
| | *NF* | 0,04704 | | | | | | |
| | *SE(NF)* | 0,18193 | | | | | | |
| | *Norm.GOI* | | | 0,05 | 85,03 | 2,37 | 0,23 | 0,68 |
| | *SE(Norm.GOI)* | | | 0,18 | 331,11 | 9,16 | 0,93 | 2,66 |
| **17** | **LoVo** | | | | | | | |
| | *raw* | 22,6 | 26,90 | 34,2 | 34,2 | 26,3 | 22,5 | 35,1 |
| | | 23,1 | 24,60 | 32,6 | 32,2 | 25,9 | 22,7 | 32,8 |
| | | 21,9 | 25,00 | 34,1 | 33,1 | 25 | 22,7 | 34,2 |
| | *avg* | 22,53 | 25,50 | 33,63 | 33,17 | 25,73 | 22,63 | 34,03 |
| | *SDsample* | 0,60 | 1,23 | 0,90 | 1,00 | 0,67 | 0,12 | 1,16 |
| | *E^(deltaCt)* | 3,564 | 0,660 | 0,000 | 0,268 | 0,157 | 1,481 | 0,064 |
| | *SD*{E^(deltaCt)} | 1,489 | 0,562 | 0,000 | 0,186 | 0,073 | 0,119 | 0,051 |
| | *SE*{E^(deltaCt)} | 0,860 | 0,324 | 0,000 | 0,107 | 0,042 | 0,068 | 0,030 |
| | *NF* | 1,53335 | | | | | | |
| | *SE(NF)* | 0,27387 | | | | | | |
| | *Norm.GOI* | | | 0,00 | 0,17 | 0,10 | 0,97 | 0,04 |
| | *SE(Norm.GOI)* | | | 0,00 | 0,08 | 0,03 | 0,18 | 0,02 |
| **18** | **DLD-1** | | | | | | | |
| | *raw* | 24,7 | 25,10 | 26,4 | 28,7 | 22,8 | 21 | 29,8 |
| | | 25,7 | 25,20 | 29,3 | 31 | 23,6 | 21,5 | 29,2 |
| | | 25,1 | 26,40 | 28,9 | 29,5 | 23,9 | 21,3 | 31,7 |
| | *avg* | 25,17 | 25,57 | 28,20 | 29,73 | 23,43 | 21,27 | 30,23 |
| | *SDsample* | 0,50 | 0,72 | 1,57 | 1,17 | 0,57 | 0,25 | 1,31 |
| | *E^(deltaCt)* | 0,574 | 0,633 | 0,011 | 2,895 | 0,776 | 3,819 | 0,891 |
| | *SD*{E^(deltaCt)} | 0,200 | 0,317 | 0,012 | 2,343 | 0,306 | 0,666 | 0,806 |
| | *SE*{E^(deltaCt)} | 0,116 | 0,183 | 0,007 | 1,353 | 0,176 | 0,385 | 0,465 |
| | *NF* | 0,60291 | | | | | | |
| | *SE(NF)* | 0,17634 | | | | | | |
| | *Norm.GOI* | | | 0,02 | 4,80 | 1,29 | 6,33 | 1,48 |
| | *SE(Norm.GOI)* | | | 0,01 | 2,65 | 0,48 | 1,96 | 0,88 |
| **19** | **Snu-C2B** | | | | | | | |
| | *raw* | 25,3 | 29,10 | 30,8 | 29,7 | 23,8 | 21,4 | 33,2 |
| | | 25,9 | 29,10 | 29,1 | 33,7 | 23,7 | 22,1 | 33,3 |
| | | 24,8 | 28,80 | 31,3 | 31,2 | 26,1 | 23,2 | 34,6 |
| | *avg* | 25,33 | 29,00 | 30,40 | 31,53 | 24,53 | 22,23 | 33,70 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *SDsample* | 0,55 | 0,17 | 1,15 | 2,02 | 1,36 | 0,91 | 0,78 |
| *E^(deltaCt)* | 0,512 | 0,074 | 0,002 | 0,831 | 0,362 | 1,954 | 0,081 |
| *SD*{E^(deltaCt)} | 0,195 | 0,009 | 0,002 | 1,164 | 0,340 | 1,229 | 0,044 |
| *SE*{E^(deltaCt)} | 0,113 | 0,005 | 0,001 | 0,672 | 0,197 | 0,710 | 0,025 |
| *NF* | 0,19503 | | | | | | |
| *SE(NF)* | 0,11552 | | | | | | |
| *Norm.GOI* | | | 0,01 | 4,26 | 1,86 | 10,02 | 0,41 |
| *SE(Norm.GOI)* | | | 0,01 | 4,27 | 1,49 | 6,96 | 0,28 |

# APPENDIX F: SEQUENCING RESULTS OF THE 3C ASSAY

Figures presented below show the identity of alignments between the sequencing results and expected sequence when DNA is digested from indicated restriction sites (yellow in the sequence alignment) and ligated by T4 DNA ligase. Details of alignments are also shown below the figures. The gaps at the beginning of the alignments are due to imperfections in the sequencing experiments for the first 20 to 25 bases.

***Primers 2-10***



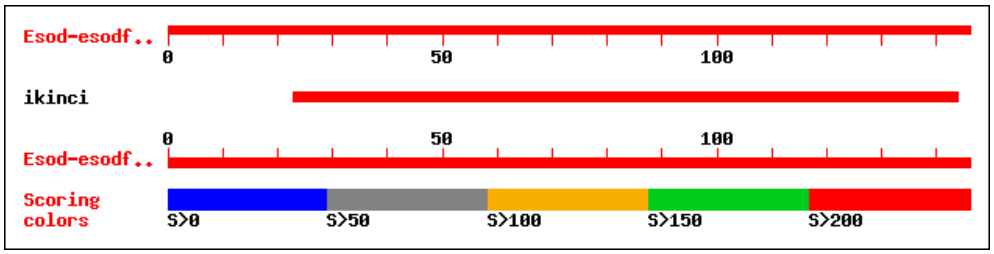

```
Score =  274 bits (138), Expect = 9e-79
Identities = 138/138 (100%)
Strand = Plus / Plus

Query: 26 gggaatgaggccccagaggacagatgagcatcatgggcagagcggggtcagattccccgt 85
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 55 gggaatgaggccccagaggacagatgagcatcatgggcagagcggggtcagattccccgt 114

Query: 86 gaattcgagtcacttacaaacaaagtggcataaggccaggcacagtggcccatgcctata 145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 115gaattcgagtcacttacaaacaaagtggcataaggccaggcacagtggcccatgcctata 174

Query: 146 atcccagcactttcggag 163
           ||||||||||||||||||
Sbjct: 175 atcccagcactttcggag 192
```

***Primers 4-12***



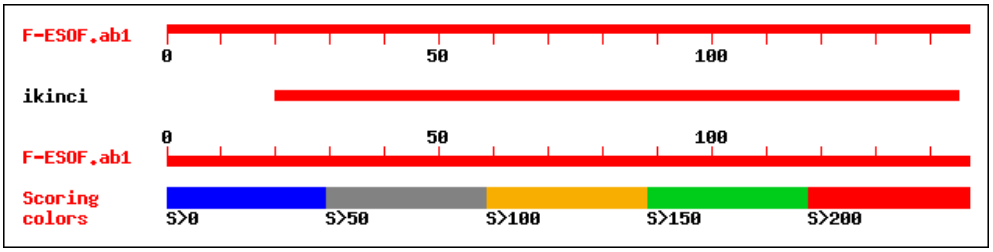
```
Score =  242 bits (122), Expect = 3e-69
Identities = 122/122 (100%)
Strand = Plus / Plus


Query: 23 agtggactcttcagacttgacttggaattctcttgtacgttgtagggtgtggagcagagt 82
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 53 agtggactcttcagacttgacttggaattctcttgtacgttgtagggtgtggagcagagt 112


Query: 83 ccctggcctgtacctgctagatgccagtagcacctgtcctgttgtgacaactaacaatat 142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 113ccctggcctgtacctgctagatgccagtagcacctgtcctgttgtgacaactaacaatat 172


Query: 143 cc 144
           ||
Sbjct: 173 cc 174
```

***Primers 6-8***




```
Score =  236 bits (119), Expect = 2e-67
Identities = 126/127 (99%), Gaps = 1/127 (0%)
Strand = Plus / Plus


Query: 20 catgggaggg-cctggtgggaggtaactgaatcatgggggcggatgtttcccgtgctgtt 78
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 47 catgggagggacctggtgggaggtaactgaatcatgggggcggatgtttcccgtgctgtt 106


Query: 79 ctcgtgatagtgaataagtctcatgagatctgtattttgaacaatgatgcctgttggaga 138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 107ctcgtgatagtgaataagtctcatgagatctgtattttgaacaatgatgcctgttggaga 166


Query: 139 atgtgca 145
           |||||||
Sbjct: 167 atgtgca 173
```