View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Ghent University Academic Bibliography

Supplementary data

Supplementary data for Chapter 8: Intermezzo – Biological Text Mining

Script 1

```
use LWP::UserAgent;

open (OUT,">all_abstracts.txt");

$ua = LWP::UserAgent -> new(agent => 'Googlebot/2.X
(+http://www.googlebot.com/bot.html)');

my $utils = "http://www.ncbi.nlm.nih.gov/entrez/eutils";
my $db="Pubmed";
$retmax=500000000000000000000000000000000;

#my $query="\"promoter hypermethylation\" OR \"promoter methylation\"
OR \"DNA hypermethylation\" OR \"DNA methylation\" OR
\"hypermethylation\" OR \"hypermethylated\" OR \"methylation\" OR
\"methylated\" OR \"epigenetics\" OR \"epigenetically\" OR
\"epigenetic\" OR methylat* OR epigenetic*";
my $query="cervical cancer methylation";

my $report = "abstract";
my $esearch =
"$utils/esearch.fcgi?"."db=$db&retmax=1&usehistory=y&term=";
my $esearch_result=$ua->get($esearch.$query)->content;
$esearch_result=~m/<Count>(\d+)<\/Count>.*<QueryKey>(\d+)<\/QueryKey>.*
<WebEnv>(\S+)<\/WebEnv>/s;
my $QueryKey = $2;
my $WebEnv   = $3;
my $efetch =
"$utils/efetch.fcgi?"."rettype=$report&retmode=xml&retstart=$retstart&r
etmax=$retmax&"."db=$db&query_key=$QueryKey&WebEnv=$WebEnv";

$efetch_result = $ua->get($efetch)->content;

$efetch_result=~s/\n//igm;
$efetch_result=~s/\t//igm;
$efetch_result=~s/\s{3,}/ /igm;
$efetch_result=~s/\<\?xml version\=\"1.0\"\?\>//igm;
$efetch_result=~s/\<\!DOCTYPE PubmedArticleSet PUBLIC \"-\/\/NLM\/\/DTD
PubMedArticle, 1st January 2007\/\/EN\"
\"http\:\/\/www.ncbi.nlm.nih.gov\/entrez\/query\/DTD\/pubmed\_070101\.d
td\"\>//igm;
$efetch_result=~s/\<.?PubmedArticleSet\>//igm;
$efetch_result=~s/\<\/PubmedArticle\>/<\/PubmedArticle\>\&\&/igm;
```

```
@results=split(/\&\&/,$efetch_result);

foreach $abstract (@results){
          print OUT $abstract."\n";
}
```

Script 2

```
use LWP::UserAgent;
$||=1;

$ua = LWP::UserAgent -> new(agent => 'Googlebot/2.X
(+http://www.googlebot.com/bot.html)');

open (OUT,"> all_aliases.txt");

open (GENES,"< all_symbols.txt");
while ($line=<GENES>){
          chomp $line;
          $marker=$line;

          $main_url="http://www.genecards.org/cgi-
bin/carddisp.pl?id=$marker";
          $data= $ua->get($main_url)->content;
          $data=~s/\n//g;

          if ($data=~/\<TD ID\=\"SearchContent1\"\>\<SELECT
CLASS\=\"search\"  MULTIPLE SIZE\=\"3\"
ONCHANGE\=\"SearchOptionChange\(this\)\"\>(.+)\<\/SELECT\>\<\/TD\>(.+)S
earchContent2/igm){
                    $parser=$1;
          }

          while ($parser=~/\<OPTION VALUE\=\'([^\']+)\'/igm){
                    $alias1=$1;
                    $alias1=~s/\<[^\>]+\>//igm;
                    $alias1=~s/\([^\)]+\)//igm;
                    $alias1=~s/ $//igm;
                    push(@alias,$alias1);

          }

          foreach (@alias){
                    if ($_ ne "" && length($_)>2 && $_ ne
"gene"){$double{$_}++;}
          }

          @alias=keys %double;
          foreach (@alias){print OUT $marker."\t".$_."\n";}
          undef @alias; undef %double; undef $gene; undef $alias1; undef
$data;
}
```

Script 3a

```
use LWP::Simple;
use LWP::UserAgent;

# ALIASES #
# read in all aliases and generate variants

open (VARIANTS,"> all_variants.txt");

#common aliases

open (COMMONALIAS,"< common_aliases.txt");
while ($common=<COMMONALIAS>){
          chomp $common;
          if ($common=~/^(.+)$/igm){
                    push (@common_alias,$1);
          }
}

close (COMMONALIAS);

#aliases -> read in

open (ALIAS,"< all_aliases.txt");
while ($line=<ALIAS>){
          chomp $line;
          if ($line=~/^(.+)\t.+$/igm){
                    push(@lines,$line);
                    push(@genes,$1);
          }
}

close (ALIAS);
foreach (@genes){$genes{$_}++;}
@genes=keys %genes;

#generate variants

foreach $gene(@genes){
          foreach $line(@lines){
                    if ($line=~/^$gene\t(.+)$/igm){
                    push (@alias2,$1);
                    push(@alias2,$gene);
                    }
          }

          foreach (@alias2){
                    $double2{$_}++;
          }

          @alias2=keys %double2;

          foreach (@alias2){
                    $alias1=$_;
                    $alias1=~s/\<[^\>]+\>//igm;
                    $alias1=~s/\([^\)]+\)//igm;
                    $alias1=~s/\[[^\]]+\]//igm;
```

```
                    $alias1=~s/ $//igm;

          $alias1=~tr/AZERTYUIOPQSDFGHJKLMWXCVBN/azertyuiopqsdfghjklmwxc
vbn/;


                    foreach $common(@common_alias){
                              if ($alias1=~/^$common$/igm){
                                        $commontest{$alias1}++;
                              }
                    }

                    if (exists $pubmedcount{$alias1}){}
                    else{

                              $ua = LWP::UserAgent -> new(agent =>
'Googlebot/2.X (+http://www.googlebot.com/bot.html)');
                              $utils =
"http://www.ncbi.nlm.nih.gov/entrez/eutils";
                              $db     = "Pubmed";
                              $query  = "\"$alias1\"";
                              $report = "abstract";
                              $esearch =
"$utils/esearch.fcgi?"."db=$db&retmax=1&usehistory=y&term=";
                              $esearch_result=$ua->get($esearch.$query)-
>content;

          $esearch_result=~/<Count>(\d+)<\/Count>.*<QueryKey>(\d+)<\/Que
ryKey>.*<WebEnv>(\S+)<\/WebEnv>/s;
                              $pubmedcount{$alias1}=$1;
                    }

                    #conditions: length > 2, less than 100,000 pubmed
results
                    if ($alias1 ne "" && length($alias1)>2 && $alias1 ne
"gene" && $alias1=~/[A-Z]/igm && $pubmedcount{$alias1}<100000 &&
$commontest{$alias1}==0){push(@alias,$alias1);}

                    if ($alias1=~/^(.+)
protein$/igm){$alias2=$1;push(@alias,$alias2);}

          }

          #generation of regular expressions in Perl
          foreach (@alias){
                              $alias2=$_;
                              $alias2=~s/([a-z])(\d)/$1.?$2/igm;
                              $alias2=~s/(\d)([a-z])/$1.?$2/igm;
                              $alias2=~s/[^\w]/.?/igm;
                              $alias2=~s/(\.\?)+/.?/igm;
                              $double{$alias2}++;

                              $alias2=~s/([a-z])\.\?1$/$1.?i/igm;
                              $alias2=~s/([a-z])\.\?2$/$1.?ii/igm;
                              $alias2=~s/([a-z])\.\?3$/$1.?iii/igm;
                              $alias2=~s/([a-z])\.\?4$/$1.?iv/igm;
                              $alias2=~s/([a-z])\.\?5$/$1.?v/igm;
                              $alias2=~s/[^\w]/.?/igm;
                              $alias2=~s/(\.\?)+/.?/igm;
```

```
                                $double{$alias2}++;


        }

        #check for duplicates

        @alias=keys %double;
                foreach (@alias){
                        print VARIANTS $_."\t".$gene."\n";
                }

        undef %double; undef %double2;undef @alias; undef @alias2;
}
```

Script 3b

```
#initiation#

use LWP::Simple;
use LWP::UserAgent;
use IO::Handle;
use DBI;

#database connection with mysql @ matrix

$database="*********";
$user="*******";
$passwd="********";
$count = 0;

$dsn = "DBI:mysql:host=localhost;database=${database}";
$dbh = DBI->connect ($dsn, $user, $passwd)
or die "Cannot connect to server\n";

#table initilization ->  table data in database pubmeth

$tblcreate= "
CREATE TABLE IF NOT EXISTS data (
pkey int(11) NOT NULL auto_increment,
pmid int(11),
title_genes int(11),
title_alias int(11),
title_keyword int(11),
title_sentences int(11),
title_highlights int(11),
title_detection int(11),
title_cancer int(11),
abstract_genes int(11),
abstract_alias int(11),
abstract_keyword int(11),
abstract_sentences int(11),
abstract_highlights int(11),
abstract_detection int(11),
abstract_cancer int(11),

PRIMARY KEY (pkey)
```

```
)";

$s = $dbh->prepare($tblcreate);
$s->execute();

$tblcreate= "
CREATE TABLE IF NOT EXISTS genes (
pkey int(11) NOT NULL auto_increment,
pmid int(11),
gene text(100),
PRIMARY KEY (pkey)
)";

$s = $dbh->prepare($tblcreate);
$s->execute();

$tblcreate= "
CREATE TABLE IF NOT EXISTS cancers (
pkey int(11) NOT NULL auto_increment,
pmid int(11),
cancer text(100),
PRIMARY KEY (pkey)
)";

$s = $dbh->prepare($tblcreate);
$s->execute();

#read in all abstracts

open (ABSTRACTS,"< all_abstracts.txt");
open (COUNT,"> index.txt");
COUNT->autoflush(1);

#parse abstracts: XML -> title, abstract, year, journal, authors

while ($abstract=<ABSTRACTS>){
          chomp $abstract;
          if
($abstract=~/\<PMID\>(\d+)\<\/PMID\>/igm){$pmid=$1;$pmid{$pmid}++;}
          if
($abstract=~/\<ArticleTitle\>(.+)\<\/ArticleTitle\>/igm){$title{$pmid}=
$1;}
          if
($abstract=~/\<AbstractText\>(.+)\<\/AbstractText\>/igm){$abstract{$pmi
d}=$1;}
          while
($abstract=~/\<Author\>\s+\<LastName\>([^\<]+)\<\/LastName\>\s+\<FirstN
ame\>[^\<]+\<\/FirstName\>\s+\<Initials\>([^\<]+)\<\/Initials\>\s+\<\/A
uthor\>/igm){$authors.=$1." ".$2.". - ";}
          while ($abstract=~/\<Author ValidYN\=\"Y\"\>
\<LastName\>([^\<]+)\<\/LastName\> \<ForeName\>[^\<]+\<\/ForeName\>
\<Initials\>([^\<]+)\<\/Initials\> \<\/Author\>/igm){$authors.=$1."
".$2.". - ";}
          $authors{$pmid}=$authors;
          undef $authors;

          if ($abstract=~/\<PubDate\>
\<Year\>(.+)\<\/Year\>.+\<\/PubDate\>/igm){$year{$pmid}=$1;}
```

```
        else{if ($abstract=~/\<DateCreated\>
\<Year\>(\d+)\<\/Year\>/igm){$year{$pmid}=$1;}}
        if
($abstract=~/\<MedlineTA\>(.+)\<\/MedlineTA\>/igm){$journal{$pmid}=$1;}

}

# ALIASES #
# read in all variants & parse them

open (VARIANTS,"< all_variants.txt");
while ($line=<VARIANTS>){
        chomp $line;
        if ($line=~/^(.+)\t(.+)$/igm){$alias=$1;$gene=$2;
        $alias_temp=$alias; $alias_temp=~s/\.\?//igm;
        if (length($alias_temp)>2){
                push (@variants,$line);
                push (@aliases,$alias);
                $variants{$alias}++;
        }
        undef $alias_temp;
        }
}
#variants: regexex -> normal strigs without .?
foreach
$variants(@variants){$variants2=$variants;$variants2=~s/\[\^\\w\]\?//ig
m;push(@variants2,$variants2);}


#common words: frequently used words in English

open (COMMON,"< common_words.txt");
while ($line=<COMMON>){
        chomp $line;
        if ($line=~/^(.+)$/igm){
                $commons{$1}++;
        }
}


#KEYWORDS
#read in keywords, highlights, cancerwords, detection techs (as
regexes)

@keywords=("aberrant promoter[^\w]?hyper[^\w]?methylation","aberrant
promoter[^\w]?methylation","promoter[^\w]?hyper[^\w]?methylation","prom
oter[^\w]?methylation","aberrant
DNA[^\w]?hyper[^\w]?methylation","aberrant
DNA[^\w]?methylation","DNA[^\w]?hyper[^\w]?methylation","DNA[^\w]?methy
lation","aberrant hyper[^\w]?methylation","aberrantly
hyper[^\w]?methylated","aberrant methylation","aberrantly
methylated","hyper[^\w]?methylation","hyper[^\w]?methylated","methylati
on","methylated","epi[^\w]?genetically","epi[^\w]?genetics","epi[^\w]?g
enetic");
@highlight=("cpg","cpg[^\w]?island","expression","down[^\w]?regulation"
,"down[^\w]?regulated","cell[^\w]?line[^\w]?");
@cancer=("hepatocellular","adrenocortical","anal","bladder","brain","gl
ioma","cerebellar","astrocytoma","ependymoma","medulloblastoma","suprat
entorial","neuroectodermal","pineal","hypothalamic","breast","carcinoma
```

```
","cervical","colon","colo[^\w]?rectal","endometrial","esophageal","ext
rahepatic","bile[^\w]?duct","ewings","extra[^\w]?cranial","eye","intrao
cular","gall[^\w]?bladder","gastric","germ[^\w]?cell","extragonadal","g
estational","trophoblastic","head[^\w]?and[^\w]?neck","hypo[^\w]?pharyn
geal","islet cell","laryngeal","leukemia","acute
lymphoblastic","lymphoblastic","acute myeloid","myeloid","chronic
lymphocytic","lymphocytic","chronic myelogenous","myelogenous","hairy
cell","lip[^\w]?and[^\w]?oral[^\w]?cavity","liver","lung","non[^\w]?sma
ll[^\w]?cell","small[^\w]?cell","lymphoma","central[^\w]?nervous[^\w]?s
ystem","cutaneous[^\w]?t[^\w]?cell","hodgkins","malignant[^\w]?mesothel
ioma","mesothelioma","melanoma","merkel[^\w]?cell","merkel","metasatic[
^\w]?squamous[^\w]?neck","occult[^\w]?primary","multiple[^\w]?myeloma",
"plasma[^\w]?cell[^\w]?neoplasm[^\w]?","mycosis[^\w]?fungoides","myelod
ysplastic","myeloproliferative","nasopharyngeal","neuroblastoma","oral"
,"oro[^\w]?pharyngeal","osteosarcoma","ovarian","pancreatic","exocrine"
,"islet[^\w]?cell","islet","paranasal[^\w]?sinus","nasal[^\w]?cavity","
parathyroid","penile","pheochromocytoma","pituitary","plasma[^\w]?cell"
,"prostate","rhabdomyosarcoma","rectal","renal","renal[^\w]?cell","rena
l[^\w]?pelvis","ureter","transitional[^\w]?cell","transitional","saliva
ry[^\w]?gland","salivary","sezary","skin","cutaneous","t[^\w]?cell[^\w]
?lymphoma","kaposis","melanoma","small[^\w]?intestine","soft[^\w]?tissu
e[^\w]?sarcoma","stomach","testicular","thymoma","thyroid","urethral","
uterine","vaginal","vulvar","wilm[^\w]?s","cancer","cancers","carcinoma
","carcinomas","tumor","tumour","tumors","tumours");
@detection=("5[^\w]?aza[^\w]?2[^\w]?[^\w]?deoxycytidine","5[^\w]?aza[^\
w]?CdR","5[^\w]?aza[^\w]?dC","5[^\w]?azacytidine","Amplification[^\w]?o
f[^\w]?inter[^\w]?methylated[^\w]?sites","AP[^\w]?PCR","bisulfite","bis
ulfite[^\w]?conversion","bisulfite[^\w]?modification","bisulfite[^\w]?m
odified","Bisulfite[^\w]?sequencing","bisulfite[^\w]?treated","Bisulfit
e[^\w]?PCR[^\w]?SSCP","bisulphite","COBRA","combined[^\w]?bisulphite[^\
w]?restriction[^\w]?analysis","DAC","Differential[^\w]?Methylation[^\w]
?Hybridization","DMH","methylation[^\w]?specific[^\w]?digital[^\w]?kary
otyping","Methylation[^\w]?sensitive[^\w]?Arbitrarily[^\w]?Primed[^\w]?
PCR","Methylation[^\w]?sensitive[^\w]?single[^\w]?nucleotide[^\w]?prime
r[^\w]?extension","Methylation[^\w]?specific[^\w]?PCR","methylation[^\w
]?specific[^\w]?polymerase[^\w]?chain[^\w]?reaction","Methylation[^\w]?
specific[^\w]?single[^\w]?base[^\w]?extension","MethyLight","MLPA","MSB
E","MSDK","MSP","MS[^\w]?PCR","Ms[^\w]?SNuPE","Multiplex[^\w]?ligation[
^\w]?dependent[^\w]?probe[^\w]?amplification","Restriction[^\w]?Landmar
k[^\w]?Genomic[^\w]?Scanning","RLGS","sodium[^\w]?bisulfite","trichosta
tin[^\w]?A","TSA");


#color scheme:
#alias: #ffff66 (geel)
#keyword: #a0ffff (lichtblauw)
#highlight: #ff66ff (paars)
#detection: #FF9900 (oranje)
#cancer: #ff9999 (rood)
#cell-line:
#sentences: #66FF66 (groen)

### START LOOPING OVER ALL ABSTRACTS


foreach $pmid(keys %pmid){

#initialisation of counts
```

```
$count_genes_title{$pmid}=0;
$count_aliases_title{$pmid}=0;
$count_keyword_title{$pmid}=0;
$count_sentences_title{$pmid}=0;
$count_highlight_title{$pmid}=0;
$count_detection_title{$pmid}=0;
$count_cancer_title{$pmid}=0;
$count_genes_abstract{$pmid}=0;
$count_aliases_abstract{$pmid}=0;
$count_keyword_abstract{$pmid}=0;
$count_sentences_abstract{$pmid}=0;
$count_highlight_abstract{$pmid}=0;
$count_detection_abstract{$pmid}=0;
$count_cancer_abstract{$pmid}=0;
$count_genes_title2{$pmid}=0;
$count_genes_abstract2{$pmid}=0;

open (TEMP,">temp.txt");
TEMP->autoflush(1);

#Out: file with annotated abstract of one PMID
#Js: javascript for hover-over effects

$outfile=$pmid.".html";
open (OUT,">:utf8", "$outfile");

$jsfile="js/".$pmid.".js";
open (JS,"> $jsfile");

print JS "var FiltersEnabled = 1 // if your not going to use
transitions or filters in any of the tips set this to 0\n";


print OUT "<!DOCTYPE html PUBLIC \"-//W3C//DTD XHTML 1.0
Transitional//EN\" \"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd\">
<html xmlns=\"http://www.w3.org/1999/xhtml\">
<!-- DW6 -->
<head>
<!-- Copyright 2005 Macromedia, Inc. All rights reserved. -->
<title>GoldMine literature search</title>
<meta http-equiv=\"Content-Type\" content=\"text/html; charset=iso-
8859-1\" />
<link rel=\"stylesheet\" href=\"mm_health_nutr.css\" type=\"text/css\"
/>
<script language=\"JavaScript\" type=\"text/javascript\">
//--------------- LOCALIZEABLE GLOBALS ---------------
var d=new Date();
var monthname=new
Array(\"January\",\"February\",\"March\",\"April\",\"May\",\"June\",\"J
uly\",\"August\",\"September\",\"October\",\"November\",\"December\");
//Ensure correct for language. English is \"January 1, 2004\"
var TODAY = monthname[d.getMonth()] + \" \" + d.getDate() + \", \" +
d.getFullYear();
//--------------- END LOCALIZEABLE ---------------
</script>

<SCRIPT language=\"JavaScript1.2\" src=\"js/main.js\"
type=\"text/javascript\"></SCRIPT>
```

```
<style type=\"text/css\">

symb {
          background:#ffff66;
          font-weight: bold;
}

keyword {
          background:#a0ffff;
          font-weight: bold;
}

highlight {
          background:#ff66ff;
          font-weight: bold;
}

detection {
          background:#FF9900;
          font-weight: bold;
}

canc {
          background:#ff9999;
          font-weight: bold;
}

sentences {
          background:#66FF66;
}
</style>
</head>
<body bgcolor=\"#F4FFE4\">
<DIV id=\"TipLayer\" style=\"visibility:hidden;position:absolute;z-
index:1000;top:-100\"></DIV>

<SCRIPT language=\"JavaScript1.2\"
src=\"js/$pmid.js\"type=\"text/javascript\"></SCRIPT>
<table width=\"100%\" border=\"0\" cellspacing=\"0\" cellpadding=\"0\">
<tr bgcolor=\"#D5EDB3\">
<td width=\"382\" colspan=\"2\" rowspan=\"2\"><img
src=\"mm_health_photo.jpg\" alt=\"Header image\" width=\"382\"
height=\"101\" border=\"0\" /></td>
<td width=\"378\" height=\"50\" id=\"logo\" valign=\"bottom\"
align=\"center\" nowrap=\"nowrap\">GoldMine</td>
<td width=\"100%\"> </td>
</tr>

<tr bgcolor=\"#D5EDB3\">
<td height=\"51\" id=\"tagline\" valign=\"top\"
align=\"center\">LITERATURE REPORT </td>
          <td width=\"100%\"> </td>
</tr>

<tr>
<td colspan=\"4\" bgcolor=\"#5C743D\"><img src=\"mm_spacer.gif\"
alt=\"\" width=\"1\" height=\"2\" border=\"0\" /></td>
</tr>
```

```
<tr>
                  <td colspan=\"4\" bgcolor=\"#99CC66\"
background=\"mm_dashed_line.gif\"><img src=\"mm_dashed_line.gif\"
alt=\"line decor\" width=\"4\" height=\"3\" border=\"0\" /></td>
</tr>

<tr bgcolor=\"#99CC66\">
<td> </td>
         <td colspan=\"3\" id=\"dateformat\" height=\"20\"><script
language=\"JavaScript\" type=\"text/javascript\">
document.write(TODAY);       </script> </td>
</tr>

<tr>
<td height=\"3\" colspan=\"4\" background=\"mm_dashed_line.gif\"
bgcolor=\"#99CC66\"><img src=\"mm_dashed_line.gif\" alt=\"line decor\"
width=\"4\" height=\"3\" border=\"0\" /></td>
</tr>
<tr>
<td colspan=\"4\" bgcolor=\"#5C743D\"><img src=\"mm_spacer.gif\"
alt=\"\" width=\"1\" height=\"2\" border=\"0\" /></td>
</tr>
<tr>
<td width=\"40\"> </td>
<td colspan=\"2\" valign=\"top\">
<table width=\"100%\" border=\"0\" cellspacing=\"0\"
cellpadding=\"0\"><tr><td> </td></tr><BR><BR>";

#TITLE

$title=$title{$pmid};
print COUNT $pmid."\t";


#search aliases
          foreach (@aliases){
                    if ($title=~/([^\w])($_)([^\w])/igm |
$title=~/^($_)([^\w])/igm){
                              $alias=$_;
                              if
($commons{$alias}==0){$aliases_in{$alias}=1;}
                    }
          }

          foreach $alias1(keys %aliases_in){
                    foreach $alias2(keys %aliases_in){

          $alias_temp1=$alias1;$alias_temp1=~s/\[\^\\w\]\?//igm;

          $alias_temp2=$alias2;$alias_temp2=~s/\[\^\\w\]\?//igm;
                              if ($alias_temp1 ne $alias_temp2){
                                        if
($alias_temp1=~/$alias_temp2/igm){
                                                  #print
$alias1."\t".$alias2."\n";
                                                  $aliases_in{$alias2}=0;
                                        }

                              }
```

```
                    }
          }

#highlight all other aliases of genes

          foreach $alias1(keys %aliases_in){
                    if ($aliases_in{$alias1}>0){
                              push(@alias_highlight,$alias1);
                              #print $alias1."\n";
                              undef $alias_temp;
                              $alias_temp=$alias1;
                              $alias_temp=~s/\[\^\\w\]\?//igm;
                              foreach $variants(@variants2){while
($variants=~/^$alias_temp\t(.+)$/igm){push(@genes,$1)}}
                    }
          }

          foreach $gene(@genes){$genecount{$gene}++;}
          @genes=keys(%genecount);

          foreach (@variants){foreach $gene(@genes){if
($_=~/^(.+)\t$gene$/igm){push(@alias_highlight,$1);print TEMP
$_."\n";}}}
          foreach (@alias_highlight){$doubles{$_}++;}

          foreach $gene(@genes){$count_genes_title{$pmid}++;}
          print COUNT $count_genes_title{$pmid}."\t";

          foreach $gene(@genes){print COUNT $gene." - ";}
          print COUNT "\t";


          close(TEMP);

          @alias_highlight=keys %doubles;

          foreach $alias1(@alias_highlight){

          $title=~s/([^\w])($alias1)([^\w])/$1\<symb\>$2\<\/symb\>$3/igm
;

          $title=~s/^($alias1)([^\w])/\<symb\>$1\<\/symb\>$2/igm;
          }
          $title=~s/(\<symb\>)+([^\<]+)(\<\/symb\>)+/$1$2$3/igm;
          while
($title=~/(\<symb\>)([^\<]+)(\<\/symb\>)/igm){$count_aliases_title{$pmi
d}++;}
          print COUNT $count_aliases_title{$pmid}."\t";


          undef %aliases_in; undef %aliases_in2; undef
@alias_highlight;undef %doubles; undef @genes; undef %genecount;

#keyword highlighting

          foreach $keyword(@keywords){

          $title=~s/([^\w])($keyword)([^\w])/$1\<keyword\>$2\<\/keyword\
>$3/igm;
```

```
		$title=~s/^($keyword)([^\w])/\<keyword\>$1\<\/keyword\>$2/igm;
		}

		$title=~s/(\<keyword\>)+([^\<]+)\<keyword\>([^\<]+)(\<\/keyword\>)+/$1$2$3$4/igm;
		$title=~s/(\<keyword\>)+([^\<]+)(\<keyword\>)+/$1$2/igm;
		while ($title=~/(\<keyword\>)([^\<]+)(\<\/keyword\>)/igm){$count_keyword_title{$pmid}++;}
		print COUNT $count_keyword_title{$pmid}."\t";

#sentences


		while ($title=~/(\<symb\>[^\<]+\<\/symb\>)([^\.]+)(\<keyword\>[^\<]+\<\/keyword\>)/igm){$count_sentences_title{$pmid}++;}
		while ($title=~/(\<keyword\>[^\<]+\<\/keyword\>)([^\.]+)(\<symb\>[^\<]+\<\/symb\>)/igm){$count_sentences_title{$pmid}++;}

		$title=~s/(\<symb\>[^\<]+\<\/symb\>)([^\.]+)(\<keyword\>[^\<]+\<\/keyword\>)/$1\<sentences\>$2\<\/sentences\>$3/igm;
		$title=~s/(\<keyword\>[^\<]+\<\/keyword\>)([^\.]+)(\<symb\>[^\<]+\<\/symb\>)/$1\<sentences\>$2\<\/sentences\>$3/igm;

		print COUNT $count_sentences_title{$pmid}."\t";

#additional highlight

		foreach (@highlight){$title=~s/([^\w])($_)([^\w])/$1\<highlight\>$2\<\/highlight\>$3/igm;$title=~s/^($_)([^\w])/\<highlight\>$1\<\/highlight\>$2/igm;}
		while ($title=~/(\<highlight\>)([^\<]+)(\<\/highlight\>)/igm){$count_highlight_title{$pmid}++;}
		print COUNT $count_highlight_title{$pmid}."\t";

#highlight detection strategies


		foreach $detection(@detection){
			foreach $keyword(@keywords){
				if ($detection=~/^(.*)$keyword(.*)$/igm){
					$detection1=$1;$detection2=$2;

		$title=~s/($detection1)(\<keyword\>$keyword\<\/keyword\>)($detection2)/\<detection\>$1\<\/detection\>$2\<detection\>$3\<\/detection\>/igm;
				}
			}

		$title=~s/([^\w])($detection)([^\w])/$1\<detection\>$2\<\/detection\>$3/igm;

		$title=~s/(\>)($detection)([^\w])/$1\<detection\>$2\<\/detection\>$3/igm;
```

```
		$title=~s/^($detection)([^\w])/\<detection\>$1\<\/detection\>$
2/igm;
			$title=~s/\<detection\>\<\/detection\>//igm;


		}
		while
($title=~/(\<detection\>)([^\<]+)(\<\/detection\>)/igm){$count_detectio
n_title{$pmid}++;}
		print COUNT $count_detection_title{$pmid}."\t";

#highlight cancer-related items

		foreach
(@cancer){$title=~s/([^\w])($_)([^\w])/$1\<canc\>$2\<\/canc\>$3/igm;$ti
tle=~s/^($_)([^\w])/\<canc\>$1\<\/canc\>$2/igm;}

		$title=~s/(\<canc\>)+([^\<]+)\<\/canc\>(\s)+\<canc\>([^\<]+)(\
<\/canc\>)+/$1$2$3$4$5/igm;
		while
($title=~/(\<canc\>)([^\<]+)(\<\/canc\>)/igm){$count_cancer_title{$pmid
}++;push(@cancer_temp,$2)}
		print COUNT $count_cancer_title{$pmid}."\t";

		foreach (@cancer_temp){$cancer_double{$_}++;}
		@cancer_temp=keys %cancer_double;
		foreach (@cancer_temp){
			print COUNT $_." - ";

		}
		print COUNT "\t";

#hover-over: hover over an alias to get details of the gene, using
javascript

		while ($title=~/(\<symb\>)([^\<]+)(\<\/symb\>)/igm){

			$alias=$2;
			$alias2=$alias;
			$alias2=~s/ /\+/igm;

			open (TEMP,"< temp.txt");
			while ($line=<TEMP>){
				chomp $line;
				if
($line=~/^(.+)\t(.+)$/igm){$alias_temp=$1;$gene_temp=$2;
					if
($alias=~/^$alias_temp$/igm){$gene=$gene_temp;push(@genes2,$gene);}
				}
			}
			undef $gene_temp; undef $alias_temp;close (TEMP);

			$textcount++;

			print JS "Text[$textcount]=[\"$alias\",\"";

			foreach $gene(@genes2){
			$double_genes2{$gene}++;
```

```
                    }
                    @genes2=keys %double_genes2;
                    undef %double_genes2;
                    foreach $gene(@genes2){
                              if (exists $name{$gene}){}
                              else{
                                        $ua_alias = LWP::UserAgent ->
new(agent => 'Googlebot/2.X (+http://www.googlebot.com/bot.html)');

          $url_alias="http://www.genecards.org/cgi-
bin/carddisp.pl?gene=$gene";
                                        $aliasdata = $ua_alias-
>get($url_alias)->content;
                                        $aliasdata=~s/\n//g;
                                        $aliasdata=~s/\t//g;

                                        if ($aliasdata=~/\<FONT
size\=\'4\'\>\<b\>(.+)\<\/b\>\<\/FONT\>\<br \/\>\<I\>Symbol approved by
the/igm){
                                                  $name{$gene}=$1;
                                        }


                                        if ($aliasdata=~/Ensembl
cytogenetic band\:\<\/i\>\ \;\ \;\<b\> \<font size\=-1\>\<a
href = \"[^>]+\"\>([^<]+)\<\/a\>\<\/font\>/igm){
                                                  $chromosome{$gene}=$1;
                                        }
                              }

                              print JS "<A
href='http://www.genecards.org/cgi-bin/carddisp.pl?gene=$gene'
target='_blank'>$gene</a><BR>$name{$gene} -
$chromosome{$gene}<BR><BR>";
                    }

                    print JS "\"]\n";


                    foreach
$gene(@genes2){$genecount2{$gene}++;$genecount3{$gene}++;}

                    $alias_match=$alias;
                    $alias_match=quotemeta($alias_match);

                    $title=~s/(\<symb\>)$alias_match(\<\/symb\>)/$1\<a
href\=\"\http\:\/\/www\.genecards\.org\/cgi\-
bin\/cardsearch\.pl\?search\=$alias2\&search\_type\=kwd\&mini\=yes\&spe
ed\=fast\#MINICARDS\" target\=\"\_blank\"
onMouseOver=\"stm(Text[$textcount],Style[5])\"
onmouseout=\"htm()\"\>$alias\<\/a\>$2/igm;
                    undef $alias; undef $alias2; undef $ua_alias; undef
$url_alias; undef $aliasdata;undef $name; undef $chromosome; undef
@genes2;
          }

          @genes2=keys(%genecount2);

          undef %genecount2;
```

```
          undef @genes2;


#print title
print OUT "<h2><a
href=\"http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=pubm
ed&dopt=AbstractPlus&term=$pmid%5Bpmid%5D\" target=\"_blank\">$pmid:
</a>".$title."</h2>";

undef @authors;
@authors=split(/ - /,$authors{$pmid});

#print out other details (journal; year; authors)

print OUT "<h4>";
foreach $author(@authors){$author2=$author;$author2=~s/
/\+/igm;$author2=~s/\.//igm;print OUT "<a
href=\"http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=pubm
ed&dopt=AbstractPlus&term=%22$author2%22%5BAuthor%5D\"
target=\"_blank\">".$author." </a>";}
print OUT "</h4><h4>".$journal{$pmid}." - ";
print OUT $year{$pmid}."</h4><BR>";


#########################################################################
##############################
#sentences

$sentence=$abstract{$pmid};
open (TEMP,">> temp.txt");
TEMP->autoflush(1);

#alias highlighting

          foreach (@aliases){
                    if ($sentence=~/([^\w])($_)([^\w])/igm |
$sentence=~/^($_)([^\w])/igm){
                              $alias=$_;
                              if
($commons{$alias}==0){$aliases_in{$alias}=1;}
                    }
          }

          foreach $alias1(keys %aliases_in){
                    foreach $alias2(keys %aliases_in){

          $alias_temp1=$alias1;$alias_temp1=~s/\[\^\\w\]\?//igm;

          $alias_temp2=$alias2;$alias_temp2=~s/\[\^\\w\]\?//igm;

                              if ($alias_temp1 ne $alias_temp2){

                                        if
($alias_temp1=~/$alias_temp2/igm){
                                                  #print
$alias1."\t".$alias2."\n";
                                                  $aliases_in{$alias2}=0;
                                        }
```

```
                                        }
                                }
                        }

                        foreach $alias1(keys %aliases_in){
                                if ($aliases_in{$alias1}>0){
                                        push(@alias_highlight,$alias1);
                                        #print $alias1."\n";
                                        undef $alias_temp;
                                        $alias_temp=$alias1;
                                        $alias_temp=~s/\[\^\\w\]\?//igm;
                                        foreach $variants(@variants2){while
($variants=~/^$alias_temp\t(.+)$/igm){push(@genes,$1)}}
                                }
                        }

                        foreach $gene(@genes){$genecount{$gene}++;}
                        @genes=keys(%genecount);

                        foreach (@variants){foreach $gene(@genes){if
($_=~/^(.+)\t$gene$/igm){push(@alias_highlight,$1);print TEMP
$_."\n";}}}
                        foreach (@alias_highlight){$doubles{$_}++;}

                        foreach $gene(@genes){$count_genes_abstract{$pmid}++;}
                        print COUNT $count_genes_abstract{$pmid}."\t";

                        foreach $gene(@genes){print COUNT $gene." - ";}
                        print COUNT "\t";


                        @alias_highlight=keys %doubles;

                        foreach $alias1(@alias_highlight){

                        $sentence=~s/([^\w])($alias1)([^\w])/$1\<symb\>$2\<\/symb\>$3/
igm;

                        $sentence=~s/^($alias1)([^\w])/\<symb\>$1\<\/symb\>$2/igm;
                        }
                        $sentence=~s/(\<symb\>)+([^\<]+)(\<\/symb\>)+/$1$2$3/igm;
                        while
($sentence=~/(\<symb\>)([^\<]+)(\<\/symb\>)/igm){$count_aliases_abstrac
t{$pmid}++;}
                        print COUNT $count_aliases_abstract{$pmid}."\t";


                        close (TEMP);

                        undef %aliases_in; undef %aliases_in2; undef
@alias_highlight;undef %doubles; undef @genes;undef %genecount;

#keyword highlighting

                        foreach $keyword(@keywords){

                        $sentence=~s/([^\w])($keyword)([^\w])/$1\<keyword\>$2\<\/keywo
rd\>$3/igm;
```

```
		$sentence=~s/^($keyword)([^\w])/\<keyword\>$1\<\/keyword\>$2/i
gm;
		}

		$sentence=~s/(\<keyword\>)+([^\<]+)\<keyword\>([^\<]+)(\<\/key
word\>)+/$1$2$3$4/igm;
		$sentence=~s/(\<keyword\>)+([^\<]+)(\<keyword\>)+/$1$2/igm;
		while
($sentence=~/(\<keyword\>)([^\<]+)(\<\/keyword\>)/igm){$count_keyword_a
bstract{$pmid}++;}
		print COUNT $count_keyword_abstract{$pmid}."\t";

#sentences


		while
($sentence=~/(\<symb\>[^\<]+\<\/symb\>)([^\.]+)(\<keyword\>[^\<]+\<\/ke
yword\>)/igm){$count_sentences_abstract{$pmid}++;}
		while
($sentence=~/(\<keyword\>[^\<]+\<\/keyword\>)([^\.]+)(\<symb\>[^\<]+\<\
/symb\>)/igm){$count_sentences_abstract{$pmid}++;}


		$sentence=~s/(\<symb\>[^\<]+\<\/symb\>)([^\.]+)(\<keyword\>[^\
<]+\<\/keyword\>)/$1\<sentences\>$2\<\/sentences\>$3/igm;
		$sentence=~s/(\<keyword\>[^\<]+\<\/keyword\>)([^\.]+)(\<symb\>
[^\<]+\<\/symb\>)/$1\<sentences\>$2\<\/sentences\>$3/igm;

		print COUNT $count_sentences_abstract{$pmid}."\t";

#highlights

		foreach
(@highlight){$sentence=~s/([^\w])($_)([^\w])/$1\<highlight\>$2\<\/highl
ight\>$3/igm;$sentence=~s/^($_)([^\w])/\<highlight\>$1\<\/highlight\>$2
/igm;}
		while
($sentence=~/(\<highlight\>)([^\<]+)(\<\/highlight\>)/igm){$count_highl
ight_abstract{$pmid}++;}
		print COUNT $count_highlight_abstract{$pmid}."\t";

#detection technologies
		foreach $detection(@detection){
			foreach $keyword(@keywords){
				if ($detection=~/^(.*)$keyword(.*)$/igm){
					$detection1=$1;$detection2=$2;

		$sentence=~s/($detection1)(\<keyword\>$keyword\<\/keyword\>)($
detection2)/\<detection\>$1\<\/detection\>$2\<detection\>$3\<\/detectio
n\>/igm;
				}
			}

		$sentence=~s/([^\w])($detection)([^\w])/$1\<detection\>$2\<\/d
etection\>$3/igm;

		$sentence=~s/^($detection)([^\w])/\<detection\>$1\<\/detection
\>$2/igm;
```

```
		$sentence=~s/(\>)($detection)([^\w])/$1\<detection\>$2\<\/dete
ction\>$3/igm;
			$sentence=~s/\<detection\>\<\/detection\>//igm;

		}
		while
($sentence=~/(\<detection\>)([^\<]+)(\<\/detection\>)/igm){$count_detec
tion_abstract{$pmid}++;}
		print COUNT $count_detection_abstract{$pmid}."\t";

#highlight cancer-related keywords
		foreach
(@cancer){$sentence=~s/([^\w])($_)([^\w])/$1\<canc\>$2\<\/canc\>$3/igm;
$sentence=~s/^($_)([^\w])/\<canc\>$1\<\/canc\>$2/igm;}
		$sentence=~s/(\<canc\>)+([^\<]+)\<\/canc\>(\s)+\<canc\>([^\<]+
)(\<\/canc\>)+/$1$2$3$4$5/igm;
		while
($sentence=~/(\<canc\>)([^\<]+)(\<\/canc\>)/igm){$count_cancer_abstract
{$pmid}++;push(@cancer_temp,$2);}
		print COUNT $count_cancer_abstract{$pmid}."\t";

#which cancertypes / tissue types

		foreach (@cancer_temp){$cancer_double{$_}++;}
		@cancer_temp=keys %cancer_double;
		foreach (@cancer_temp){

		$_=~tr/AZERTYUIOPQSDFGHJKLMWXCVBN/azertyuiopqsdfghjklmwxcvbn/;
			if ($_ ne "cancer" && $_ ne "tumor" && $_ ne
"tumour" && $_ ne "tumors" && $_ ne "cancers" && $_ ne "tumours" && $_
ne "carcinoma"){
			print COUNT $_." - ";
			$count_cancer_temp2.=" - ".$_;
			$insert= "insert into cancers (pmid,cancer)
			values ($pmid,'$_')";
			$s = $dbh->prepare($insert);
			$s->execute();
			}
		}
		undef @cancer_temp; undef %cancer_double;
		print COUNT "\n";


#Hover-over effects
		while ($sentence=~/(\<symb\>)([^\<]+)(\<\/symb\>)/igm){

			$alias=$2;
			$alias2=$alias;
			$alias2=~s/ /\+/igm;

			open (TEMP,"< temp.txt");
			while ($line=<TEMP>){
				chomp $line;
				if
($line=~/^(.+)\t(.+)$/igm){$alias_temp=$1;$gene_temp=$2;
					if
($alias=~/^$alias_temp$/igm){$gene=$gene_temp;push(@genes2,$gene);}
				}
```

```
                    }
                    undef $gene_temp; undef $alias_temp;close (TEMP);

                    $textcount++;

                    print JS "Text[$textcount]=[\"$alias\",\"";

                    foreach $gene(@genes2){
                    $double_genes2{$gene}++;
                    }
                    @genes2=keys %double_genes2;
                    undef %double_genes2;
                    foreach $gene(@genes2){
                              if (exists $name{$gene}){}
                              else{
                                        $ua_alias = LWP::UserAgent ->
new(agent => 'Googlebot/2.X (+http://www.googlebot.com/bot.html)');

          $url_alias="http://www.genecards.org/cgi-
bin/carddisp.pl?gene=$gene";
                                        $aliasdata = $ua_alias-
>get($url_alias)->content;
                                        $aliasdata=~s/\n//g;
                                        $aliasdata=~s/\t//g;

                                        if ($aliasdata=~/\<FONT
size\=\'4\'\>\<b\>(.+)\<\/b\>\<\/FONT\>\<br \/\>\<I\>Symbol approved by
the/igm){
                                                  $name{$gene}=$1;
                                        }


                                        if ($aliasdata=~/Ensembl
cytogenetic band\:\<\/i\>\ \;\ \;\<b\> \<font size\=-1\>\<a
href = \"[^>]+\"\>([^<]+)\<\/a\>\<\/font\>/igm){
                                                  $chromosome{$gene}=$1;
                                        }
                              }

                              print JS "<A
href='http://www.genecards.org/cgi-bin/carddisp.pl?gene=$gene'
target='_blank'>$gene</a><BR>$name{$gene} -
$chromosome{$gene}<BR><BR>";
                    }

                    print JS "\"]\n";

                    foreach
$gene(@genes2){$genecount2{$gene}++;$genecount3{$gene}++}

                    $alias_match=$alias;
                    $alias_match=quotemeta ($alias_match);


          $sentence=~s/(\<symb\>)$alias_match(\<\/symb\>)/$1\<a
href\=\"\http\:\/\/www\.genecards\.org\/cgi\-
bin\/cardsearch\.pl\?search\=$alias2\&search\_type\=kwd\&mini\=yes\&spe
ed\=fast\#MINICARDS\" target\=\"\_blank\"
```

```
onMouseOver=\"stm(Text[$textcount],Style[5])\"
onmouseout=\"htm()\"\>$alias\<\/a\>$2/igm;
                    undef $alias; undef $alias_match; undef $alias2;
undef $ua_alias; undef $url_alias; undef $aliasdata;undef $name; undef
$chromosome; undef @genes2;
          }

          @genes2=keys(%genecount2);

          @genes3=keys(%genecount3);
          foreach $gene(@genes3){
                    $count_genes_abstract2{$pmid}++;
                    $count_gene_temp2.=" - ".$gene;
                    $insert= "insert into genes (pmid,gene)
                    values ($pmid,'$gene')";
                    $s = $dbh->prepare($insert);
                    $s->execute();
          }
          undef %genecount2;undef @genes2;undef $textcount;undef
%genecount3;undef @genes3;

          print OUT $sentence."\n<BR>";

          print OUT "</table><br /> <br /> </td>
          <td width=\"100%\"> </td>
          </tr>
          <tr>
          <td width=\"40\"> </td><td width=\"342\"> </td>
          <td width=\"378\"> </td>
          <td width=\"100%\"> </td>
          </tr>
          </table>
          </body>
          </html>";

print JS
"Style[0]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\
"\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,10,10,51,1,0,\"\",\"\"]
Style[1]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"center\",\"\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\
",\"\",\"\"]
Style[2]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"left\",\"\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\",\"\",
\"\"]
Style[3]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"float\",\"\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\"
,\"\",\"\"]
Style[4]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"fixed\",\"\",\"\",\"\",200,\"\",2,2,1,1,\"\",\"\",\"\",\"\",\
"\",\"\"]
Style[5]=[\"white\",\"black\",\"#99CC66\",\"#D5EDB3\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"\",\"\",\"sticky\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\
",\"\",\"\"]
Style[6]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"\",\"\",\"keep\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\",
\"\",\"\"]
Style[7]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"\",\"
\",\"\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,40,10,\"\",\"\",\"\",\"\"
,\"\"]
```

```
Style[8]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"
\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,10,50,\"\",\"\",\"\",\"\"
,\"\"]
Style[9]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\"
\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,10,10,51,0.5,75,\"simple\
",\"gray\"]
Style[10]=[\"white\",\"black\",\"black\",\"white\",\"\",\"\",\"right\",
\"\",\"Impact\",\"cursive\",\"center\",\"\",3,5,200,150,5,20,10,0,50,1,
80,\"complex\",\"gray\"]
Style[11]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\
"\",\"\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,10,10,51,0.5,45,\"simple
\",\"gray\"]
Style[12]=[\"white\",\"black\",\"#000099\",\"#E8E8FF\",\"\",\"\",\"\",\
"\",\"\",\"\",\"\",\"\",\"\",\"\",\"\",200,\"\",2,2,10,10,\"\",\"\",\"\",\"\
",\"\"]

applyCssFilter()";
close JS;
close OUT;

$insert= "insert into data
(pmid,title_genes,title_alias,title_keyword,title_sentences,title_highl
ights,title_detection,title_cancer,abstract_genes,abstract_alias,abstra
ct_keyword,abstract_sentences,abstract_highlights,abstract_detection,ab
stract_cancer)
values
($pmid,$count_genes_title2{$pmid},$count_aliases_title{$pmid},$count_ke
yword_title{$pmid},$count_sentences_title{$pmid},$count_highlight_title
{$pmid},$count_detection_title{$pmid},$count_cancer_title{$pmid},$count
_genes_abstract2{$pmid},$count_aliases_abstract{$pmid},$count_keyword_a
bstract{$pmid},$count_sentences_abstract{$pmid},$count_highlight_abstra
ct{$pmid},$count_detection_abstract{$pmid},$count_cancer_abstract{$pmid
})";

$s = $dbh->prepare($insert);
$s->execute();
undef $count_gene_temp; undef $count_gene_temp2; undef
$count_cancer_temp; undef $count_cancer_temp2;
}
```

Supplementary data for Chapter 13: Discovery of methylation markers in cervical cancer, using relaxation ranking

```
#Results-> affy probe ID in column 1; P-calls for primary
cancers (0-39, in column 2); for cell-lines (0-4, in column 3);
for treated cell-lines (0-15 in column 4)

count<-function(x,y,z){l<-
length(results[results$V2<=x&results$V3<=y&results$V4>=z,1]);}

#count:function that counts how many rows in results are found
with a certain condition

f<-matrix(nrow=3200,ncol=4)

#initialisation of f

i<-0;for (x in 0:39) {for (y in 0:4) { for (z in 0:15) {i<-
i+1;f[i,1]<-count(x,y,z);f[i,2]=x;f[i,3]=y;f[i,4]=z; }}}

#simulate all conditions and store -> first column: number of
genes (w), 2nd: x - 3th: y - 4th:z

ordered<-f[order(f[,1],f[,2],f[,3],-f[,4]),]

#order, first by w, followed by x, y (ascending) and z
(descending)

kiezen<-function(x) {if (length(grep(x,list))>0){}else{list<<-
append(list,x);}}

#function to find out whether a gene is already in the list, if
not, add it to the list

list<-"";counter<-1;n<-0;while(n<=3000) {counter<-
counter+1;x2<-ordered[counter,2];y2<-ordered[counter,3];z2<-
ordered[counter,4];temp<-
results[(results$V2<=x2)&(results$V3<=y2)&(results$V4>=z2),1];f
ound<-length(temp);counter2<-1;while (counter2<= found)
```

```
{kiezen(as.character(temp[counter2]));counter2<-counter2+1;n<<-
length(list);}}

#generate a list of n genes

#row per row, get x2, y2 and z2 from ordered

#check which genes are found with these criteria

#loop over these genes (names in first column of results)

#for each gene, check if already selected in previous round, if
not, add to the list

#do this for more rows, until n genes reached ->these are in
list
```

Supplementary table 1: list of primers used for BSP

Name	Forward primer (5'-3')	Reverse primer (5'-3')	Ta	Start position[1]	End position	RefSeq
DAZL	TTTGGGGGTGATGTGTGTGTTT	TCTCCCTCAACTCACCATAATA	54	-161	312	NM_001351
ADARB1[2]						NM_015834
SYCP3	AAAATTTAAAAATTGGAAGGTATTAGG	ACCTCACTAATCAAAAACAACCTCT	54	-208	+186	NM_153694
AUTS2	TTTTAAAAGTGATAAAGTTGGTTATGGT	CCCTTTTCTTTCTCCTCTCTTTCT	56	+300	-184	NM_015570
NNAT	GGTTAGGGATTGGGGAGAA	GCTAAACTTACCTACAACAACAC	54	-271	210	NM_005386
SST	GGGGTATGTGGAATTGTGTG	AAATCTCCTTACCTACTTCCC C	54	-185	+276	NM_001048
HTRA3	GTYGGTTTTGTYGTTATGTAGGY	AACTTCACTTCCTCCCTAACC	57	+190	+622	NM_053044
ZFP42	AGTAGGTGTTTGTTGAAGATAG	ACTCATAACACACATAACCATC	60	+308	+580	NM_174900
NPTX1	GGTAGTGGGGGTTTGATAG	AAATAATCTCCTTCTACTACAACAC	54	-2	+372	NM_002522
GDA	TATAGAAGGTGGAGGAAGTTGA	CACCTCCATAAAACAAATCCAAA	54	-239	+194	NM_004293
CCNA1	TATAGTTGGAGTTGGAGGGT	AAACAACTAACAAATACACTAAAA	54	-279	+146	NM_153694

[1]: +1 is transcription start site (TSS) ; [2]: Several primer pairs were tested, however, none worked.

Supplementary table 2: Overview of the 45 known methylation markers in cervical cancer selected from literature search and their position after relaxation ranking

Rank	Probe ID	Gene Name
234	205899_at	CCNA1
404	203167_at	TIMP2
651	209278_s_at	TFPI2
1242	209243_s_at	PEG3
1463	204197_s_at	RUNX3
1742	209032_s_at	IGSF4
1926	204054_at	PTENP1
2270	227345_at	TNFRSF10D
2500	201147_s_at	TIMP3
2733	216933_x_at	APC
3589	226389_s_at	RAPGEF1
3609	200824_at	GSTP1
3896	220804_s_at	TP73
4942	217561_at	CALCA
5388	203139_at	DAPK1
5566	219550_at	ROBO3
6546	202037_s_at	SFRP1
7654	222713_s_at	FANCF
7937	200665_s_at	SPARC
8699	1558797_at	ROBO1
8710	209897_s_at	SLIT2
8734	215551_at	ESR1
10590	1555271_a_at	TERT
10772	204346_s_at	RASSF1
12291	1562857_at	CDH13
12935	201130_s_at	CDH1
13595	211851_x_at	BRCA1
13943	1566484_at	FHIT
14073	240776_at	PGR
14988	204803_s_at	RRAD
16174	207109_at	POU2F3
17208	205080_at	RARB
20493	204880_at	MGMT

Rank	Probe ID	Gene Name
23251	204745_x_at	MT1G
24079	209644_x_at	CDKN2A
31063	1552925_at	PCDH10
32394	1557615_a_at	SLIT1
33526	1561574_at	SLIT3
35987	205714_s_at	ZMYND10
36278	206657_s_at	MYOD1
37007	208461_at	HIC1
48347	204121_at	GADD45G
48574	211163_s_at	TNFRSF10C
51948	230378_at	SCGB3A1
54166	206983_at	FGFR1OP

Genes selected using Textmining through NCBI E-Fetch, using GeneCards for aliases (www.genecards.org)

Supplementary table 3: Overview of published imprinted genes (Imprinted Gene Catalog), their position and gene name after relaxation ranking

Rank	Imprinted rank	Gene Name
21	1	NNAT (Evans et al., 2001)
49	2	H19 (Rachmilewitz et al., 1992; Zhang and Tycko, 1992; Zhang et al., 2006b)
59	3	HFE (Bulaj et al., 1996)
118	4	MEG3 (Miyoshi et al., 2000)
355	5	NF2 (Evans et al., 1992)
703	6	ZIM2 (Murphy et al., 2001)
769	7	PPP1R9A (Nakabayashi et al., 2004)
818	8	PON2 (Ono et al., 2003)
1241	9	PEG3 (Murphy et al., 2001)
1246	10	CPA4 (Bentley et al., 2003)
1605	11	GABRA5 (Meguro et al., 1997)
1951	12	PEG10 (Ono et al., 2003)
2001	13	L3MBTL (Li et al., 2004)
2788	14	PHLDA2 (Lee and Feinberg, 1998)
2855	15	CDKN1C (Taniguchi et al., 1997)
2948	16	GNAL (Corradi et al., 2005)
3896	17	TP73
4623	18	SLC38A4
4635	19	HTR2A
4674	20	IL4R
4843	21	SGCE
6591	22	DIRAS3
6938	23	PON3
7195	24	ASCL2
7877	25	GABRG3
8396	26	ZNF215
8722	27	PON1
8901	28	SPINK5
8943	29	NDN
9497	30	OSBPL5
9843	31	PHF11
10226	32	ASB4
10805	33	GRB10
11994	34	IGF2R
13384	35	TSPAN32
13891	36	DCN
13909	37	CTNNA3
13960	38	KCNQ1
14256	39	ZNF264
14331	40	SLC22A18
14423	41	TSSC4
14943	42	C6orf66
16963	43	DLX5
17595	44	IGF2
19031	45	DIO3

19245	46	NAP1L5
19840	47	WT1
20712	48	GABRB3
21448	49	APP
21500	50	CD81
21677	51	M6PR
22329	52	MEST
22333	53	SDHD
23653	54	BRD2
24342	55	UBE3A
29362	56	PLAGL1
31040	57	TCEB3C
31095	58	ZIM3
35914	59	DDC
36087	60	SLC22A18AS
36250	61	MKRN3
36255	62	INS
36766	63	CALCR
36894	64	MAS1
37122	65	DLK1
37350	66	RASGRF1
38039	67	ATP10A
39762	68	IGF2AS
39787	69	KCNQ1DN
39902	70	USP29
40591	71	TRPM5
42765	72	PWCR1
51829	73	GNGT1
52756	74	MS4A2

Genes were selected using the Imprinted Gene Catalogue (Morison et al., 2005)

Supplementary table 4: enriched gene ontology terms, descriptions, number of genes associated with this GO term and P-value versus all human genes in the first 3000 probes. GO terms and statistics as determined by GOStat

Gene Ontology ID	Description	Number of genes	P-value
GO:50789	regulation of biological process	470	4.69e-19
GO:50791	regulation of physiological process	421	2.81e-18
GO:50794	regulation of cellular process	443	2.81e-18
GO:50875	cellular physiological process	1168	5.85e-18
GO:51244	regulation of cellular physiological process	409	1.23e-17
GO:9653	morphogenesis	91	1.21e-16
GO:9987	cellular process	1354	3.71e-16
GO:48513	organ development	75	5.44e-11
GO:44238	primary metabolism	821	1.97e-09
GO:8219	cell death	84	1.97e-09
GO:16265	death	84	1.97e-09
GO:6366	transcription from RNA polymerase II promoter	65	1.97e-09
GO:6139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	419	2.04e-09
GO:30154	cell differentiation	75	2.33e-09
GO:7049	cell cycle	96	3.18e-09
GO:19222	regulation of metabolism	305	5.64e-09
GO:6355	regulation of transcription, DNA-dependent	265	6.32e-09
GO:6351	transcription, DNA-dependent	269	1.03e-08

Gene Ontology ID	Description	Number of genes	P-value
GO:7399	nervous system development	61	1.48e-08
GO:51242	positive regulation of cellular physiological process	66	1.63e-08
GO:48731	system development	61	2.04e-08
GO:44237	cellular metabolism	837	2.04e-08
GO:48522	positive regulation of cellular process	74	2.44e-08
GO:44255	cellular lipid metabolism	73	2.73e-08
GO:31323	regulation of cellular metabolism	294	3.5e-08
GO:6915	apoptosis	78	4.06e-08
GO:43119	positive regulation of physiological process	67	4.06e-08
GO:12501	programmed cell death	78	4.45e-08
GO:6629	lipid metabolism	89	4.45e-08
GO:6350	transcription	287	9.49e-08
GO:7417	central nervous system development	23	1.02e-07
GO:48518	positive regulation of biological process	81	1.17e-07
GO:45449	regulation of transcription	277	1.4e-07
GO:8152	metabolism	897	1.5e-07
GO:19219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	278	3.1e-07
GO:902	cell morphogenesis	41	1.02e-06
GO:51726	regulation of cell cycle	61	1.44e-06
GO:16043	cell organization and biogenesis	186	2.08e-06
GO:74	regulation of progression through cell cycle	60	2.52e-06

Gene Ontology ID	Description	Number of genes	P-value
GO:42981	regulation of apoptosis	53	2.62e-06
GO:43067	regulation of programmed cell death	53	3.43e-06
GO:1501	skeletal development	23	5.37e-06
GO:6357	regulation of transcription from RNA polymerase II promoter	37	9.74e-06

Supplementary table 5: overview of Ingenuity networks, highly represented in the top-3000 list

Genes	Score (number of genes in top-3000, associated with pathway)	Focus genes (total number of genes in pathway)	Pathway
AMID, ASPM, CABLES1, CABLES2, CCNA1, CHD3, CSTF1 (includes EG:1477), DGKA, DUSP2, GART, GBP1, GLIPR1, GSTM5, GTSE1, HSPA9B (includes EG:3313), KARCA1, KIAA1794, LPIN1, MAGEA4 (includes EG:4103), MCAM, PDRG1, PEG3 (includes EG:5178), PMS2, PODXL, PPP4R2, PTP4A1, RBBP6, RECQL4, RPRM, SNRK, TBX3, THRAP2, TP53, TRIO, VPS33A	33	35	Cancer, Cellular Compromise, DNA Replication, Recombination, and Repair
ACP1, AKAP12, CAD, CHRM2, CPE, CPEB1, DLGAP1, DUSP4, DUSP16, ELF1, FHL1, GNA11, GNAO1, GNB5, GRIK1, GRK6, INA, L1CAM (includes EG:3897), MAPK1, NEFH, NPDC1, NTS (includes EG:4922), PDE4D, PDE4DIP (includes EG:9659), PTPN7, PTPRR, RASD1, RGS7, SCLT1, SHANK2, SST, SSTR1, SYNGAP1, ULBP1, ULK1	33	35	Cell Signaling, Nucleic Acid Metabolism, Small Molecule Biochemistry
ACOX1, ACTN2, API5, CIDEC, CNTFR, COL4A6, CTNND2, CYP24A1, DYRK1A, ERBB2IP, FGF2, FGFR4, GPC1, GREM1, GRIN2C, IL6ST, KCNAB1, LDB3, LIFR, LRRC7, MGAT5, MYPN, NR2C2, OLIG2, PDZD2, PI3, PTCRA, PTRF, SCD, SQSTM1, SYNPO, TEAD4, TGFB2, TGFBR3, ZNF148	33	35	Cellular Development, Nervous System Development and Function, Reproductive System Development and Function
ANKS1B, BCL2, C1QL1, CASP9, CYCS, DCTN1, DCTN4, DNM1L, DYNLL1, ENDOG, GAST, GGA2, HK2, ITPR3, MAGEA3, NALP1, NALP2, NDE1, NNAT, NR4A1, NSF, PECAM1, PRKCE, PRKD2, PSEN2, PTGS2, RABEP1, RYR3, SLC25A4, TIMP3, TNK2, TRA@, TXNL5, UACA, WWOX	33	35	Cancer, Cell Death, Cell Morphology

Genes	Score (number of genes in top-3000, associated with pathway)	Focus genes (total number of genes in pathway)	Pathway
ATF3, BCR, BMP2, BMP6, BMPER, COL1A1, CUL3, CYP7A1, EIF2AK4, FOSL1, FOSL2, FZD8, HGF, IGFBP5, JAG1, JUND, KLF9, KRR1, LDLR, MFI2, MLLT4, NFIX, NR2F1, PAPPA, PAPPA2, PCSK9, PTHR1, PTPN2, PVR, PVRL2, RHOD, RNF7, TMEM97, TRAM2, VTN	33	35	Organismal Development, Cellular Growth and Proliferation, Cellular Movement
AACS, ACACB, ACLY, ACSL1, ALDOC, CPT1A, DGAT1, DHCR7, FOXC2, GPX2, HMGCS1, HNF4A, IRS2, KRT20, LPL, LSS, MBTPS1, MUC13, NFATC4, PDK4, PDLIM7, PEG10, PPARG, PPARGC1A, PTGES, SCAND1, SCARB1, SLC16A1, SLC27A1, SLC27A2, SREBF1, SULT1A3, TIMM22, TPM2, UGT1A9 (includes EG:54600)	33	35	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
ACD, AKAP13, ANK1, ARFGEF2, CAMK2N1, CGA, DNASE1, ECM1, H2AFX, HAND1, HMGA2, MAGED1, MEF2A, MUC5AC, NRGN, PDCD6, PRKAR1A, PRL, RECK, RHAG, S100A4, SFPQ, SH3GL2, SH3KBP1, SLC2A4RG, SND1 (includes EG:27044), SYNJ2, TDG, TERF2, THRA, TIAM1, TIMP2, TINF2, TNRC6A, XDH	33	35	Gene Expression, Organ Morphology, Reproductive System Disease
ARID4A, AYTL2, C1ORF181, CTH, DDX54, DNAJA2, EGR1, EGR4, FLI1, FLJ23584, GDF15, HDAC9, HIPK2, HSP90B1, IFRD1, IRS4, JARID1A, KLF1, LRCH4, MYEF2, NCOR1, PBX1, PEX6, PML, PRG1 (includes EG:5552), RARA, SF3A1, SF3A3, SIN3B, SMAD2, SMAD6, SMARCC2, TAT, VASN, ZBTB16	33	35	Gene Expression, Cancer, Hematological Disease
CHP, DUSP8, DUSP19, EDAR, FGF12, GTF2I, ICAM1, IKBKG, MAP2K7, MAPK9, MDFIC, NFATC3, NFKB2, PDE8A, PRRX1, PSMB9, RAB6IP2, RALBP1, RIMS2, SCIN, SCN11A, SPN, STK4, STX7, TAOK2 (includes EG:9344), TNC, TNFAIP3,	33	35	Cell Death, Connective Tissue Disorders, Cancer

Genes	Score (number of genes in top-3000, associated with pathway)	Focus genes (total number of genes in pathway)	Pathway
TNFRSF10B, TNFRSF10D, TP73L, TRAF1, TRAF2, TRAF3, TRAF4, TRIB3			
ALG3, AURKA, CBX4, CCNB2, CDK10, CHFR, COX5A, COX7B2, COX8C, CTDSPL, DHFR, EIF4A1, ELN, FDFT1, GMDS (includes EG:2762), GPX3, HCN2, ID4, LZTR1, MGA (includes EG:23269), MXD4, MYBL2, NDUFC1, NEU1, NKX2-5, PCBP2, PDCD4, PLEC1, PTEN, SEC23A, SEC24A, SMOX, TFDP1, TNFAIP1, WDR45 (includes EG:11152)	33	35	Cell Cycle, Embryonic Development, Protein Synthesis
ANP32A, ATP2B1, ATXN1, BAT4, CUGBP1, CYP2C9, DLX1, DUSP1, ELK1, ELK4, EPB41L1, EYA2, FOXG1B, GAB1 (includes EG:2549), HELLS, INSR, ITGB8, ITPR1, ITSN1, KHDRBS2, KIFAP3, MXI1, PDE3B, PLCG1, PQBP1 (includes EG:10084), PRKCB1, PTPRG, SEZ6L2, SH2B1, SIX4, SIX5, SLC30A1, THPO, UNC5C, VIL1	33	35	Gene Expression, Cancer, Cell Cycle
ATF1, CALR, CBS, CTGF, DAB2, DDAH1, DVL2, EPOR, FBLN1, GATA2, HLA-A, ITGB3, KIAA0020, LAMA4, LILRB2, MAGEA1, MAGEA2, MEF2C, MMP16, NAGA, NF1, NID2, NRG2, PCSK1, PCSK1N, PPAP2B, SDC4, SLA, TBCD, TCF3, TEAD1 (includes EG:7003), TGM2, TUBB4, TUBB2A, TWIST2	33	35	Cardiovascular System Development and Function, Organismal Survival, Cell-To-Cell Signaling and Interaction
ARHGEF9, ARID1B (includes EG:57492), CALB1, CALD1, CD44, CDC42, CDC42BPA, CHM, CHML, DMD, FAH, IGF1R, KCNJ4, KCNJ12, KRAS, MKL2, MLLT10, MST1R, NF2, PIK3C2B, PIK3CA, PIP5K1A, RAB1A, RGS12, RPL35A, RPS24, RSN, SMARCA2, SNTG1, SPTA1, SPTBN1, SRF (includes EG:6722), SS18, VAV2, VAV3	33	35	Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance

Genes	Score (number of genes in top-3000, associated with pathway)	Focus genes (total number of genes in pathway)	Pathway
C2, C3, CACNA1A, CACNA1C, CR1, CSN2, DECR2, EPAS1, ERBB4, ETS2, HEXA, HSPA5, IGL@, IL12B, LLGL1, MASP1, MTA1, NEK1, NEK8, OSMR, PCLO, PKIB, PLD2, RAB40B, RNF41, SNAI1, SNAP25, SPINK4, STAT5B, STX3, STXBP1, SYT4, SYT9, TCEB2, WISP2	33	35	Cellular Function and Maintenance, Organismal Injury and Abnormalities, Molecular Transport

Supplementary table 6: The ranking of possibly functional methylated genes from the highest ranking probe-list (TOP250). Probes were ranked according the relaxation ranking algorithm ("original ranking"). Possible functionally methylated genes were selected ("new ranking") by omitting probes that do not fulfill the following criteria: (1) probes without gene symbol (i.e. gene ID) or hypothetical genes (marked as "unknown"); (2) probes/genes without a CpG island (marked as "no CpG") because the expression of such markers is most probably reactivated upon DAC/TSA treatment indirectly via methylation-regulated transcription factors (Shi et al., 2003b); (3) genes located on chromosome X (marked as "X-located") since one of the main mechanisms of the inactivation of one copy of the X-chromosome in females is DNA methylation (see text); and (4) genes with expression that is not downregulated in less than 15 of the 39 carcinomas (marked as "untreated"); optimally no expression in all 39 cases is expected (P-call = 0), but relaxation ranking allowsgenes with varying P-calls including those that are expressed in more than 40% of all carcinomas). X, Y and Z represent the P-calls for primary tumor, untreated and treated cell lines, respectively.

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
1		216370_s_at	TKTL1	Xq28	0	0	12	X-located
2		211619_s_at	ALPP	2q37	0	3	15	No CpG
3	1	206588_at	DAZL	3p24.3	1	1	13	
4	2	203865_s_at	ADARB1	21q22.3	1	2	15	
5		210394_x_at	SSX4	Xp11.23	2	1	14	X-located
6		235773_at	ZIK1	19q13.43	16	1	15	Cancer
7		214183_s_at	TKTL1	Xq28	0	0	11	Already selected
8	3	1553599_a_at	SYCP3	12q	0	1	12	
9		220217_x_at	SPANXC	Xq27.1	1	2	14	X-located
10		243742_at	LOC388481	18q22.3	4	1	14	Unknown
11		220931_at	MGC5590	13q14.11	9	2	15	Unknown
12		1565909_at	-	-	11	0	12	Unknown
13		211670_x_at	SSX3	Xp11.23	0	0	10	X-located

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
14		204534_at	VTN	17q11	0	2	12	No CpG
15		238825_at	ACRC	Xq13.1	2	1	13	X-located
16		204584_at	L1CAM	Xq28	5	3	15	X-located
17		222712_s_at	MUC13	-	9	1	14	No CpG
18		1563118_at	-	-	10	0	11	Unknown
19	4	242721_at	AUTS2	7q11.22	12	0	12	
20		1566656_a_at	RP11-151A6.2	13q32.3	0	0	9	Unknown
21	5	204239_s_at	NNAT	20q11.2-q12	0	1	11	
22	6	213921_at	SST	3q28	1	1	12	
23	7	226944_at	HTRA3	4p16.1	6	0	10	
24	8	243161_x_at	ZFP42	4q35.2	11	1	14	
25		236817_at	-	-	15	0	12	Unknown
26		243802_at	DNHD2	3p14.3	23	2	15	Cancer
27		207493_x_at	SSX2	Xp11.23-p11.22	0	0	8	X-located
28		220178_at	C19orf28	19p13.3	0	3	12	Unknown
29	9	204684_at	NPTX1	17q25.1-q25.2	2	3	14	
30		231367_s_at	LOC647131	13q14.2	5	2	14	Unknown
31		215881_x_at	SSX2	Xp11.23	11	0	11	X-located
32	10	1569555_at	GDA	9q21.13	14	3	15	
33		1561255_at	FAM50B	6p25-pter	27	0	12	Cancer
34		214035_x_at	LOC399491	16p13.1	31	2	15	Cancer
35	11	218720_x_at	SEZ6L2	16p11.2	1	3	13	
36		229715_at	-	-	1	3	13	Unknown
37	12	207013_s_at	MMP16	8q21	2	0	8	
38		239343_at	-	-	15	3	15	Unknown
39		202015_x_at	METAP2	12q22	35	2	15	Cancer
40		1556395_at	-	-	0	1	10	Unknown

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
41	13	213711_at	KRTHB1	12q13	0	1	10	
42		233938_at	FLJ22675	11q13.2	0	2	11	No CpG
43		243594_x_at	SPIRE2	16q24	0	2	11	No CpG
44	14	1553562_at	CD8B	2p12	3	0	9	
45	15	1570360_s_at	DDX3Y	Yq11	3	0	9	
46		234516_at	-	-	3	0	9	Unknown
47		1565641_at	C16orf45	16p13.11	3	1	12	Unknown
48		228404_at	IRX2	5p15.33	4	3	14	Unknown
49		224646_x_at	H19	11p15.5	31	1	14	Cancer
50		227721_at	CPAMD8	19p13.11	31	1	14	Cancer
51		1569987_at	DLEU7	13q14.3	1	1	11	Unknown
52		206626_x_at	SSX1	Xp11.23-p11.22	1	1	11	X-located
53		1566927_at	C21orf104	-	5	2	13	Unknown
54		215595_x_at	GCNT2	6p24	11	0	10	Unknown
55		1554274_a_at	SSH1	12q24.11	17	3	15	Cancer
56		203549_s_at	LPL	8p22	26	0	11	Cancer
57		215226_at	EXPH5	11q22.3	29	1	13	Cancer
58		229883_at	GRIN2D	19q13.1-qter	35	1	14	Cancer
59		211328_x_at	HFE	6p21.3	37	2	15	Cancer
60		236486_at	-	-	37	2	15	Unknown
61	16	234347_s_at	DENR	12q24.31	8	2	13	
62	17	227711_at	FAM112B	12q13.2	14	2	14	
63		1566110_at	-	-	0	1	9	Unknown
64		211425_x_at	SSX4	Xp11.23	0	1	9	X-located
65		232069_at	KIF26A	14q32.33	0	3	11	No CpG
66	18	210021_s_at	UNG2	5p15.2-p13.1	2	3	13	
67	19	1552785_at	FLJ37549	19q13.12	5	0	9	

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
68	20	1553449_at	FLJ36701	16q24.3	5	0	9	
69		1559441_s_at	LOC651430	-	5	0	9	Unknown
70		1559303_at	-	-	5	1	12	Unknown
71		207739_s_at	GAGE1	Xp11.4-p11.2	5	1	12	X-located
72		220051_at	PRSS21	16p13.3	18	3	15	Cancer
73		244503_at	BDNF	11p13	18	3	15	Cancer
74		1566695_at	-	-	30	0	11	Cancer
75		1557518_a_at	-	-	0	0	7	Unknown
76		205749_at	CYP1A1	15q22-q24	0	0	7	No CpG
77		206549_at	INSL4	9p24	0	0	7	No CpG
78		206627_s_at	SSX1	Xp11.23-p11.22	0	0	7	No CpG
79		208020_s_at	CACNA1C	12p13.3	0	0	7	No CpG
80		210431_at	ALPPL2	2q37	0	0	7	No CpG
81		211731_x_at	SSX3	Xp11.23	2	1	11	No CpG
82	21	201095_at	DAP	5p15.2	6	3	14	
83	22	206085_s_at	CTH	1p31.1	6	3	14	
84		239801_at	-	-	7	1	12	Unknown
85		235009_at	FAM44A	4p16.1	16	2	14	Cancer
86		207064_s_at	AOC2	17q21	20	3	15	Cancer
87		244451_x_at	TDG	12q24.1	23	0	10	Cancer
88		234697_x_at	C3orf31	3p25.2	34	1	13	Cancer
89		204664_at	ALPP	2q37	0	2	10	No CpG
90		238594_x_at	-	-	0	2	10	Unknown
91	23	203548_s_at	LPL	8p22	1	0	7	
92		228713_s_at	DHRS10	19q13.33	32	0	11	Cancer
93		234401_at	-	-	32	0	11	Unknown
94		1560477_a_at	SAMD11	1p36.33	1	3	12	No CpG

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
95	24	207147_at	DLX2	2q32	1	3	12	
96		214612_x_at	MAGEA6	Xq28	4	2	12	X-located
97	25	215672_s_at	KIAA0828	7q32.1	4	2	12	
98	26	1570528_at	XYLT2	17q21.3-17q22	6	0	9	
99	27	242763_at	PKHD1L1	8q23.1-q23.2	6	0	9	
100		239153_at	FLJ41747	12q13.13	12	2	13	No CpG
101		229532_at	ZNF502	3p21.31	19	2	14	Cancer
102		242971_at	-	-	19	2	14	Unknown
103		222890_at	HSPC065	16q21	24	3	15	Cancer
104		1558356_at	UACA	15q22-q24	26	0	10	Cancer
105		215887_at	ZNF277	7q31.1	36	1	13	Cancer
106		206640_x_at	GAGE2	Xp11.23	1	1	10	X-located
107		208155_x_at	GAGE4	Xp11.4-p11.2	1	1	10	X-located
108		208235_x_at	GAGE7	Xp11.4-p11.2	1	1	10	X-located
109		1563250_at	-	-	4	1	11	Unknown
110		220021_at	TMC7	16p12.3	22	2	14	Cancer
111		1555141_a_at	LOC150159	4q24	35	0	11	Cancer
112		239582_at	PML	15q22	35	0	11	Cancer
113		205554_s_at	DNASE1L3	3p21.1-3p14.3	39	2	15	Cancer
114		214218_s_at	XIST	Xq13.2	39	2	15	Cancer
115		217767_at	C3	19p13.3-p13.2	39	2	15	Cancer
116		218541_s_at	C8orf4	8p11.2	39	2	15	Cancer
117		223551_at	PKIB	6q22.31	39	2	15	Cancer
118		226211_at	MEG3	14q32	39	2	15	Cancer
119		204637_at	CGA	6q12-q21	0	1	8	No CpG
120		211403_x_at	VCY	Xp22.32	0	1	8	X-located
121		223365_at	DHX37	12q24.31	0	1	8	No CpG

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
122		237983_at	-	-	0	1	8	Unknown
123	28	1566951_at	SIRT5	6p23	7	0	9	
124		207851_s_at	INSR	19p13.3-p13.2	7	0	9	No CpG
125	29	207069_s_at	SMAD6	15q21-q22	11	3	14	
126		204959_at	MNDA	1q22	39	0	11	Cancer
127		228988_at	ZNF6	Xq21.1-q21.2	39	1	13	X-located
128		242907_at	GBP2	1p22.2	39	1	13	Cancer
129	30	1555788_a_at	TRIB3	20p13-p12.2	0	3	10	
130	31	206290_s_at	RGS7	1q43	8	0	9	
131		1557768_at	-	-	14	1	12	Unknown
132		244378_at	C12orf51	12q24.13	25	2	14	Cancer
133		1569577_x_at	-	-	26	3	15	Unknown
134		233185_at	ITGBL1	13q33	26	3	15	Cancer
135		239911_at	-	-	26	3	15	Unknown
136	32	204717_s_at	SLC29A2	11q13	0	2	9	
137	33	219480_at	SNAI1	20q13.1-q13.2	0	2	9	
138		1562644_at	MTHFD2L	4q13.3	2	0	7	No CpG
139		1565905_at	FLJ46026	17q25.3	2	0	7	No CpG
140		206371_at	FOLR3	11q13	2	0	7	No CpG
141		220348_at	KBTBD9	2p24.1	2	0	7	No CpG
142	34	1564066_at	SPATS1	6p21.1	6	1	11	
143	35	207151_at	ADCYAP1R1	7p14	9	0	9	
144		229685_at	TBC1D14	4p16.1	13	3	14	No CpG
145		241252_at	ESCO2	8p21.1	13	3	14	No CpG
146		231023_at	FLJ12118	13q34	14	2	13	Unknown
147	36	235872_at	TERF2	16q22.1	14	2	13	
148		239984_at	SCN7A	2q21-q23	14	2	13	No CpG

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
149		232324_x_at	PPAP2B	1pter-p22.1	28	3	15	Cancer
150	37	242138_at	DLX1	2q32	2	3	12	
151		1557873_at	SPINK5L3	5q33.1	5	2	12	No CpG
152		229932_at	LOC51255	2p11.2	7	0	8	Unknown
153		1552858_at	MAGEB6	Xp21.3	15	1	12	X-located
154		214230_at	CDC42	1p36.1	15	2	13	Cancer
155		1557193_at	PTPN2	18p11.3-p11.2	28	2	14	Cancer
156		208591_s_at	PDE3B	11p15.1	28	2	14	Cancer
157		242153_at	LARP2	4q28.2	33	0	10	Cancer
158		AFFX-DapX-3_at	-	-	0	4	15	Unknown
159		AFFX-DapX-0_at	-	-	0	4	15	Unknown
160		AFFX-LysX-3_at	-	-	0	4	15	Unknown
161		AFFX-LysX-5_at	-	-	0	4	15	Unknown
162		AFFX-1heX-3_at	-	-	0	4	15	Unknown
163		AFFX-1heX-5_at	-	-	0	4	15	Unknown
164		AFFX-1heX-0_at	-	-	0	4	15	Unknown
165		AFFX-ThrX-3_at	-	-	0	4	15	Unknown
166		AFFX-ThrX-5_at	-	-	0	4	15	Unknown
167		AFFX-ThrX-0_at	-	-	0	4	15	Unknown
168		AFFX-r2-Bs-dap-3_at	-	-	0	4	15	Unknown
169		AFFX-r2-Bs-dap-5_at	-	-	0	4	15	Unknown
170		AFFX-r2-Bs-dap-0_at	-	-	0	4	15	Unknown
171		AFFX-r2-Bs-lys-5_at	-	-	0	4	15	Unknown
172		AFFX-r2-Bs-lys-0_at	-	-	0	4	15	Unknown
173		AFFX-r2-Bs-phe-5_at	-	-	0	4	15	Unknown
174		AFFX-r2-Bs-phe-0_at	-	-	0	4	15	Unknown
175		AFFX-r2-Bs-thr-	-	-	0	4	15	Unknown

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
		0_s_at						
176		1569208_a_at	DKFZP686A01247	4p13	2	2	11	No CpG
177		228066_at	LOC642705	17q12	2	2	11	No CpG
178		240983_s_at	CARS	11p15.5	2	2	11	No CpG
179		210675_s_at	PTPRR	12q15	6	3	13	No CpG
180	38	218476_at	POMT1	9q34.1	6	3	13	
181		1562311_at	-	-	7	1	11	Unknown
182		1557558_s_at	MATN1	1p35	29	2	14	Cancer
183		220187_at	STEAP4	7q21.12	29	3	15	Cancer
184		227045_at	ZNF614	19q13.33	29	3	15	Cancer
185		AFFX-r2-Bs-thr-3_s_at	-	-	0	4	13	Unknown
186		207086_x_at	GAGE2	Xp11.23	2	1	10	X-located
187		215317_at	-	-	2	1	10	Unknown
188		239369_at	LCN8	9q34.3	2	1	10	No CpG
189	39	203731_s_at	ZFP95	7q22	14	3	14	
190		206776_x_at	ACRV1	11p12-q13	24	1	12	Cancer
191		208511_at	PTTG3	8q13.1	30	2	14	Cancer
192		1553508_at	MDS2	1p36	1	1	9	No CpG
193		204927_at	RASSF7	11p15.5	1	1	9	Unknown
194		207663_x_at	GAGE3	Xp11.4-p11.2	1	1	9	X-located
195	40	229839_at	SCARA5	8p21.1	1	1	9	
196		AFFX-r2-Bs-phe-3_at	-	-	2	4	15	Unknown
197	41	223954_x_at	APBA2BP	20q11.22	3	2	11	
198	42	1568981_at	GTF2I	7q11.23	7	3	13	
199		238729_x_at	LOC646561	4q31.22	7	3	13	No CpG
200	43	1569998_at	MMD2	7p22.1	9	1	11	

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
201		1557890_at	LOC645818	6q24.3	16	2	13	Cancer
202		241368_at	-	-	30	3	15	Unknown
203		241719_at	-	-	30	3	15	Unknown
204		1554633_a_at	MYT1L	2p25.3	1	3	11	No CpG
205		206084_at	PTPRR	12q15	1	3	11	Already selected
206		1552829_at	TMEM23	10q11.2	25	1	12	Cancer
207		207059_at	PAX9	14q12-q13	25	1	12	Cancer
208		1569600_at	DLEU2	13q14.3	38	0	10	Cancer
209		244313_at	CR1	1q32	38	0	10	Cancer
210		210016_at	MYT1L	2p25.3	0	3	9	Already selected
211		219184_x_at	-	17p13	0	3	9	Unknown
212		239267_at	NEK6	9q33.3-q34.11	5	4	15	Untreated
213		1568639_a_at	-	-	8	3	13	Unknown
214		239302_s_at	-	-	9	2	12	Unknown
215		1556387_at	-	-	17	2	13	Unknown
216		216450_x_at	HSP90B1	12q24.2-q24.3	17	2	13	Cancer
217		1568830_at	RBMS1	2q24.2	26	1	12	Cancer
218		1552906_at	FMR1NB	Xq27.3-q28	0	0	6	X-located
219		1564854_at	-	-	0	0	6	Unknown
220		1565836_at	-	-	0	0	6	Unknown
221	44	1567341_at	FOXD4	9p11-q11	0	0	6	
222		220921_at	SPANXB1	Xq27.1	0	0	6	X-located
223		229895_s_at	NCK1	3q21	0	0	6	Unknown
224		231307_at	PAGE2	Xp11.21	0	0	6	X-located
225		235700_at	RP13-36C9.1	Xq26.3	0	0	6	X-located
226		237580_at	-	-	0	0	6	Unknown
227		238568_s_at	C18orf8	18q11.2	0	0	6	Unknown

Original rank	New rank	Affy ID	Gene symbol	Chromosomal location	X	Y	Z	Reason for omitting
228	45	238607_at	ZNF342	19q13.32	0	0	6	
229	46	240313_at	DMRTB1	1p32.3	0	0	6	
230		243171_at	-	-	0	0	6	Unknown
231		1561705_at	TTBK2	15q15.2	0	2	8	Unknown
232		207214_at	SPINK4	9p13.3	0	2	8	No CpG
233		1556107_at	-	-	3	0	7	Unknown
234	47	205899_at	CCNA1	13q12.3-q13	3	0	7	
235		221112_at	IL1RAPL2	Xq22.2-q22.3	3	0	7	X-located
236	48	233490_at	DCTN4	5q31-q32	3	0	7	
237		1566823_a_at	-	-	3	1	10	Unknown
238		226516_at	C19orf28	19p13.3	3	1	10	Already selected
239	49	229475_at	MAEL	1q24.1	3	1	10	
240		223575_at	KIAA1549	7q34	15	3	14	Cancer
241		233931_at	ELF1	13q13	15	3	14	Cancer
242		236646_at	C12orf59	12p13.2	15	3	14	Cancer
243		207660_at	DMD	Xp21.2	18	2	13	X-located
244		219922_s_at	LTBP3	11q12	31	2	14	Cancer
245	50	1552368_at	CTCFL	20q13.31	14	0	9	
246		1552414_at	WFDC9	20q12-q13.1	14	0	9	No CpG
247		1562674_at	-	-	14	0	9	Unknown
248		205876_at	LIFR	5p13-p12	32	3	15	Cancer
249		210233_at	IL1RAP	3q28	32	3	15	Cancer
250		215606_s_at	RAB6IP2	12p13.3	32	3	15	Cancer

[1]: Cancer means “too many cancers” positive and therefore not selected based on the set criteria.
[2]: Untreated means that in the untreated cell Lines already all cell Lines are positive and therefore not selected based on the set criteria.

Supplementary materials of Chapter 14: Exploring the cancer methylome using genome-wide promoter analysis

Supplementary table 1: Two different classes of genes are defined. Genes listed in Class A are only methylated in cancer and not in normals. Genes listed in Class B are at least partially methylated in normals.

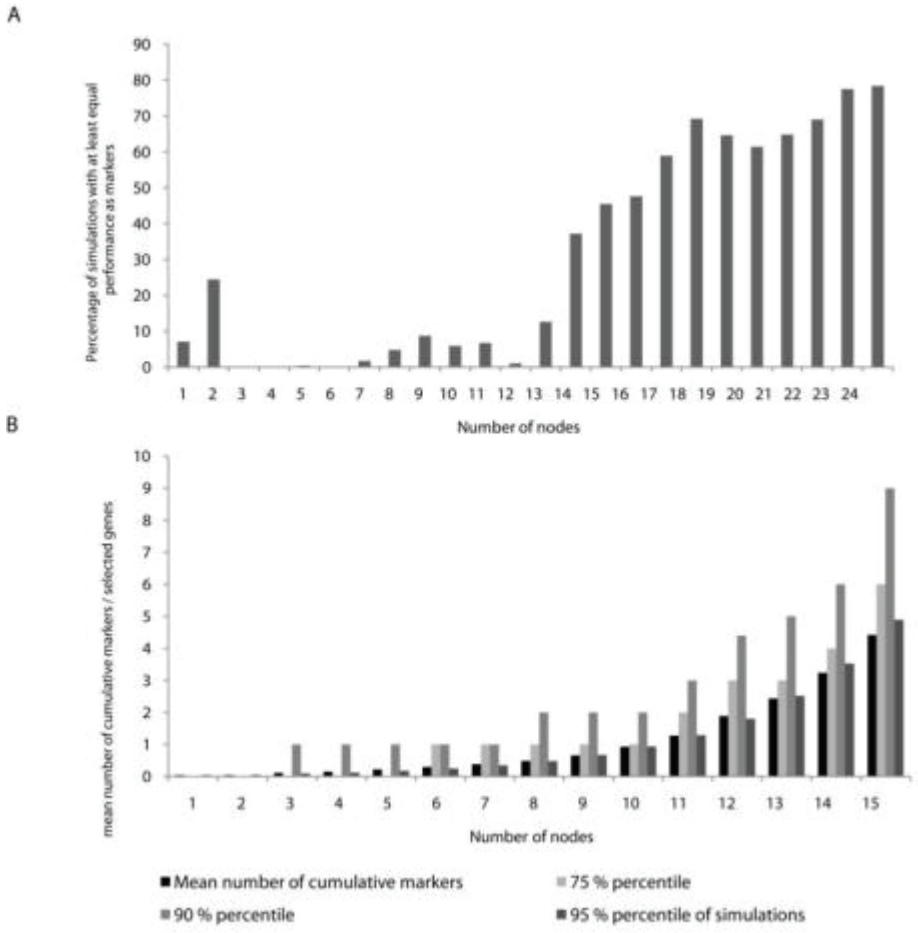
Class A: Cancer-specific methylation			Class B. Partial or complete methylation in normals		
No.	**Accession #**	**Gene Name**	**No.**	**Accession #**	**Gene Name**
1	D64137	KIP2 gene for Cdk-inhibitor p57KIP2 (KIP-2)	1	AB029041	AZI (KIAA1118)
2	NM_000038	Adenomatosis polyposis coli (APC)	2	X07732	Serine protease (Hepsin)
3	NM_001311	Cysteine-rich protein 1 (CRIP1)	3	J02611	Apolipoprotein D
4	NM_007182	Ras association (RalGDS/AF-6) domain family 1 (RASSF1)	4	L21715	Troponin I fast-twitch isoform
5	NM_002371	Mal, T-cell differentiation protein (MAL)	5	M55682	Human cartilage matrix protein (CMP) gene (Matrilin-3)
6	NM_001541	Heat shock 27kDa protein 2 (HSPB2)	6	L13203	Forkhead homolog-3 (HFH-3)
7	NM_003914	Cyclin A1 (CCNA1)	7	M84526	Human adipsin/complement factor D (Adipsin)
8	NM_004181	Ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) (PGP9.5)	8	AB022083	SOX30
9	NM_001200	Bone morphogenetic protein 2 (BMP2)	9	U19146	GAGE5
10	U27185	Human RAR-responsive gene (TIG1)	10	J04430	Tartrate-resistant acid phosphatase type 5
11	NM_001427	Engrailed homolog 2 (EN2)	11	U90842	SSX-5

12	M69199	G0S2 protein gene (G0S2)	12	AB000277	DAP-1
13	S76475	Trk-C	13	U02687	Human growth factor receptor tyrosine kinase (STK-1) (Flt-3)
14	K15306	Neurofilament heavy chain (NFH)	14	NM_001306	Claudin-3
15	NM_000834	Human N-methyl-D-aspartate receptor subunit 2B (GRIN2B) (NMDAR2B)	15	NM_006984	Claudin-10
16	NM_004613				

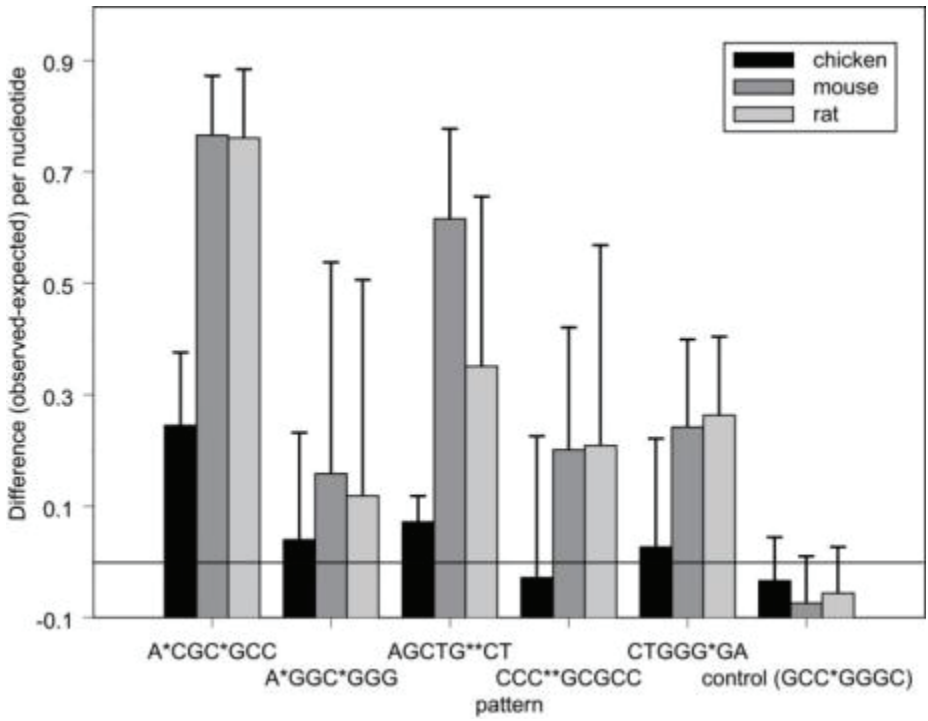
Supplementary table 2: list of 56 known markers for broad analysis

No.	Accession #	Gene Name
1	NM_000038	adenomatosis polyposis coli (APC)
2	NM_000115	endothelin receptor type B (EDNRB)
3	NM_000125	estrogen receptor 1 (ESR1)
4	NM_000249	mutL homolog 1 (MLH1)
5	NM_000320	quinoid dihydropteridine reductase (QDPR)
6	NM_000362	TIMP metallopeptidase inhibitor 3 (TIMP3)
7	NM_000378	Wilms tumor 1 (WT1)
8	NM_000474	twist homolog 1 (TWIST1)
9	NM_000797	dopamine receptor D4 (DRD4)
10	NM_000852	glutathione S-transferase pi (GSTP1)
11	NM_000963	prostaglandin-endoperoxide synthase 2 (PTGS2)
12	NM_000965	retinoic acid receptor, beta (RARB)
13	NM_001200	bone morphogenetic protein 2 (BMP2)
14	NM_001257	cadherin 13 (CDH13)
15	NM_001265	caudal type homeobox transcription factor 2 (CDX2)
16	NM_001310	cAMP responsive element binding protein-like 2 (CREBL2)
17	NM_001343	disabled homolog 2 (DAB2)
18	NM_001645	apolipoprotein C-I (APOC1)
19	NM_001647	apolipoprotein D (APOD)
20	NM_001742	calcitonin receptor (CALCR)
21	NM_001759	cyclin D2 (CCND2)
22	NM_001804	caudal type homeobox transcription factor 1 (CDX1)
23	NM_002012	fragile histidine triad gene (FHIT)
24	NM_002052	GATA binding protein 4 (GATA4)
25	NM_002055	glial fibrillary acidic protein (GFAP)
26	NM_002158	forkhead box N2 (FOXN2)
27	NM_002190	interleukin 17A (IL17A)
28	NM_002412	O-6-methylguanine-DNA methyltransferase (MGMT)
29	NM_002899	retinol binding protein 1, cellular (RBP1)
30	NM_003012	secreted frizzled-related protein 1 (SFRP1)
31	NM_003014	secreted frizzled-related protein 4 (SFRP4)
32	NM_003015	secreted frizzled-related protein 5 (SFRP5)
33	NM_003220	transcription factor AP-2 alpha (TFAP2A)
34	NM_003255	TIMP metallopeptidase inhibitor 2 (TIMP2)
35	NM_003745	suppressor of cytokine signaling 1 (SOCS1)

No.	Accession #	Gene Name
36	NM_003955	suppressor of cytokine signaling 3 (SOCS3)
37	NM_004102	fatty acid binding protein 3, muscle and heart (FABP3)
38	NM_004350	runt-related transcription factor 3 (RUNX3)
39	NM_004360	cadherin 1, type 1(CDH1)
40	NM_004385	chondroitin sulfate proteoglycan 2 (CSPG2)
41	NM_004936	cyclin-dependent kinase inhibitor 2B (CDKN2B)
42	NM_006142	stratifin (SFN)
43	NM_006497	hypermethylated in cancer 1 (HIC1)
44	NM_006680	malic enzyme 3 (ME3)
45	NM_007182	Ras association domain family 1 (RASSF1)
46	NM_007295	breast cancer 1 (BRCA1)
47	NM_014333	immunoglobulin superfamily, member 4 (IGSF4)
48	NM_016734	paired box gene 5 (PAX5)
49	NM_018223	checkpoint with forkhead and ring finger domains (CHFR)
50	NM_019102	homeobox A5 (HOXA5)
51	NM_021115	seizure related 6 homolog (SEZ6L)
52	NM_022557	growth hormone 2 (GH2)
53	NM_022725	Fanconi anemia, complementation group F (FANCF)
54	NM_022808	small nuclear ribonucleoprotein polypeptide N (SNRPN)
55	NM_058195	cyclin-dependent kinase inhibitor 2A (CDKN2A)
56	NM_080473	GATA binding protein 5 (GATA5)



Supplementary figure 1: A: number of simulations (of the complete approach) that performed at least as good as the methylation markers. B: complete approach: mean number of cumulative markers and 75 and 95 % percentiles of the distribution of the markers and compared with the 95 % percentile of the means of all simulations



Supplementary figure 2: conservation of the different patterns throughout evolution. The difference between the observed and expected score per nucleotide for chicken, mouse and rat. The higher, the better the pattern is conserved in comparison with its neighboring sequence. Error bars represent the standard error

Supplementary table 3: genes, selected in broad analysis

RefSeq	Gene Symbol
NM_002197	ACO1
NM_001100	ACTA1
NM_001102	ACTN1
NM_006988	ADAMTS1
NM_020350	AGTRAP
NM_003689	AKR7A2
NM_000034	ALDOA
NM_005165	ALDOC
NM_015365	AMMECR1
NM_001145	ANG
NM_004039	ANXA2
NM_004039	ANXA2P1
NM_004039	ANXA2P3
NM_007077	AP4S1
NM_153000	APCDD1
NM_014570	ARFGAP3
NM_006321	ARIH2
NM_012097	ARL5A
NM_015161	ARL6IP1
NM_022374	ARL6IP2
NM_024585	ARMC7
NM_004192	ASMTL
NM_001686	ATP5B
NM_000332	ATXN1
NM_004048	B2M
NM_001497	B4GALT1
NM_005504	BCAT1
NM_003921	BCL10
NM_138578	BCL2L1
NM_032735	BEST3
NM_017614	BHMT2
NM_001721	BMX
NM_030968	C1QTNF1
NM_017766	CASZ1
NM_024814	CBLL1
NM_001757	CBR1
NM_001236	CBR3
NM_152499	CCDC24
NM_001775	CD38
NM_000610	CD44
NM_015891	CDC40
NM_018719	CDCA7L
NM_018101	CDCA8
NM_006727	CDH10
NM_004064	CDKN1B
NM_005195	CEBPD
NM_004854	CHST10
NM_006825	CKAP4
NM_016129	COPS4
NM_006833	COPS6
NM_006371	CRTAP
NM_003651	CSDA
NM_015235	CSTF2T
NM_001901	CTGF
NM_001908	CTSB
NM_003591	CUL2
NM_001914	CYB5A
NM_000775	CYP2J2
NM_006716	DBF4
NM_016216	DBR1

RefSeq	Gene Symbol
NM_003472	DEK
NM_018981	DNAJC10
NM_020234	DTWD1
NM_138287	DTX3L
NM_001394	DUSP4
NM_005226	EDG3
NM_004720	EDG4
NM_032378	EEF1D
NM_023932	EGFL9
NM_003754	EIF3S5
NM_015904	EIF5B
NM_006375	ENOX2
NM_001977	ENPEP
NM_012307	EPB41L3
NM_000122	ERCC3
NM_007036	ESM1
NM_024785	FAM124B
NM_014864	FAM20B
NM_153690	FAM43A
NM_016623	FAM49B
NM_004629	FANCG
NM_018660	FBXO16
NM_004112	FGF11
NM_017946	FKBP14
NM_004293	GDA
NM_021067	GINS1
NM_000171	GLRA1
NM_001500	GMDS
NM_020895	GRAMD1A
NM_152451	GRINL1A
NM_000856	GUCY1A3

RefSeq	Gene Symbol
NM_004130	GYG1
NM_000183	HADHB
NM_018645	HES6
NM_006339	HMG20B
NM_005520	HNRPH1
NM_002147	HOXB5
NM_006665	HPSE
NM_016246	HSD17B14
NM_144617	HSPB6
NM_006547	IGF2BP3
NM_016260	IKZF2
NM_001567	INPPL1
NM_030980	ISG20L2
NM_014288	ITGB3BP
NM_144720	JAKMIP1
NM_012282	KCNE1L
NM_004982	KCNJ8
NM_005552	KLC1
NM_012316	KPNA6
NM_000526	KRT14
NM_032832	LRP11
NM_012321	LSM4
NM_138395	MARS2
NM_006086	MC1R
NM_002397	MEF2C
NM_139202	MLC1
NM_004530	MMP2
NM_023075	MPPE1
NM_019051	MRPL50
NM_000251	MSH2
NM_005955	MTF1

RefSeq	Gene Symbol
NM_019041	MTRF1L
NM_002470	MYH3
NM_053030	MYLK
NM_005380	NBL1
NM_031858	NBR1
NM_000266	NDP
NM_002489	NDUFA4
NM_006403	NEDD9
NM_002507	NGFR
NM_007184	NISCH
NM_020726	NLN
NM_006681	NMU
NM_006178	NSF
NM_014917	NTNG1
NM_031438	NUDT12
NM_018245	OGDHL
NM_020205	OTUD7B
NM_004670	PAPSS2
NM_033018	PCTK1
NM_014476	PDLIM3
NM_006474	PDPN
NM_002630	PGC
NM_002631	PGD
NM_006608	PHTF1
NM_004569	PIGH
NM_080476	PIGU
NM_014889	PITRM1
NM_032471	PKIB
NM_006227	PLTP
NM_005397	PODXL
NM_016937	POLA1

RefSeq	Gene Symbol
NM_006502	POLH
NM_017917	PPP2R3C
NM_002809	PSMD3
NM_002818	PSME2
NM_144651	PXDNL
NM_014226	RAGE
NM_000966	RARG
NM_002898	RBMS2
NM_005132	REC8
NM_007370	RFC5
NM_018339	RFK
NM_002918	RFX1
NM_005613	RGS4
NM_001145	RNASE4
NM_000978	RPL23
NM_001005	RPS3
NM_031924	RSHL2
NM_014016	SACM1L
NM_016038	SBDS
NM_001037	SCN1B
NM_018400	SCN3B
NM_004757	SCYE1
NM_006216	SERPINE2
NM_014454	SESN1
NM_024860	SETD6
NM_004630	SF1
NM_016538	SIRT7
NM_005135	SLC12A6
NM_017945	SLC35A5
NM_017767	SLC39A4
NM_080669	SLC46A1

RefSeq	Gene Symbol
NM_000343	SLC5A1
NM_003601	SMARCA5
NM_003068	SNAI2
NM_017826	SOHLH2
NM_014041	SPCS1
NM_014850	SRGAP3
NM_006280	SSR4
NM_014188	SSU72
NM_003896	ST3GAL5
NM_052902	STK11IP
NM_013442	STOML2
NM_014849	SV2A
NM_005639	SYT1
NM_005680	TAF1B
NM_006019	TCIRG1
NM_030752	TCP1
NM_007111	TFDP1
NM_006070	TFG
NM_004817	TJP2
NM_003258	TK1
NM_020123	TM9SF3
NM_030938	TMEM49
NM_033504	TMEM54
NM_014408	TRAPPC3
NM_017722	TRMT1
NM_006022	TSC22D1
NM_003594	TTF2
NM_016614	TTRAP
NM_006000	TUBA4A
NM_006086	TUBB3
NM_016261	TUBD1

RefSeq	Gene Symbol
NM_006659	TUBGCP2
NM_005783	TXNDC9
NM_003330	TXNRD1
NM_007019	UBE2C
NM_003359	UGDH
NM_006294	UQCRB
NM_018428	UTP6
NM_003378	VGF
NM_003383	VLDLR
NM_016485	VTA1
NM_004184	WARS
NM_145054	WDR16
NM_152348	WDR81
NM_003391	WNT2
NM_006522	WNT6
NM_032792	ZBTB45
NM_003413	ZIC3
NM_003453	ZMYM2
NM_021088	ZNF2
NM_003457	ZNF207
NM_152287	ZNF276
NM_032433	ZNF333
NM_018660	ZNF395
NM_020951	ZNF529
NM_152909	ZNF548
NM_024341	ZNF557
NM_152477	ZNF565
NM_014630	ZNF592
NM_014699	ZNF646
NM_145271	ZNF688

Supplementary table 4: genes, selected in the deep analysis

RefSeq	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat.6	Pat.7	Different patterns
NM_024509	2	0	1	1	1	1	1	7
NM_018092	0	0	0	4	1	1	3	9
NM_006667	2	2	0	1	2	0	1	8
NM_004104	1	0	0	2	0	2	1	6
NM_003345	1	0	1	1	0	1	2	6
NM_153613	1	1	1	0	1	1	2	7
NM_001508	0	1	1	1	1	0	2	6
NM_003787	2	0	1	1	0	0	1	5
NM_005397	1	1	0	0	2	1	1	6
NM_015160	1	0	0	1	1	1	2	6
NM_015049	1	1	1	0	1	1	0	5
NM_002143	1	1	1	0	1	1	1	6
NM_015675	0	0	1	1	1	1	1	5
NM_002489	1	0	1	1	1	0	0	4
NM_022131	1	0	1	1	1	0	0	4
NM_001782	1	0	0	1	1	0	1	4
NM_002899	1	0	0	1	0	1	1	4
NM_005738	1	0	0	0	1	1	1	4
NM_001497	1	0	2	0	0	0	3	6
NM_016472	0	2	1	0	1	0	3	7
NM_030954	3	1	0	2	0	1	1	8
NM_002661	1	1	0	1	1	0	4	8
NM_004383	1	0	0	2	1	0	3	7
NM_022343	1	2	0	2	0	1	3	9
NM_031922	3	0	0	0	1	0	2	6
NM_012446	0	1	1	2	2	0	1	7
NM_003224	3	0	1	1	0	0	1	6
NM_000553	3	1	0	1	0	0	1	6
NM_006936	2	0	0	1	0	2	1	6
NM_031462	2	0	0	2	1	0	1	6
NM_002658	1	0	0	0	2	1	2	6
NM_000382	0	1	0	2	1	1	2	7
NM_017739	2	0	0	0	1	1	2	6
NM_002371	1	0	0	0	0	2	2	5
NM_007152	1	0	2	0	0	1	1	5
NM_000725	1	2	1	0	0	0	2	6
NM_024835	2	1	1	0	0	0	1	5
NM_024619	0	1	1	1	2	0	0	5
NM_002855	2	1	1	0	0	1	1	6
NM_016201	2	1	1	0	1	0	1	6
NM_007367	1	1	1	2	1	0	0	6
NM_016732	1	1	1	2	1	0	0	6
NM_017825	1	1	1	0	0	0	2	5
NM_018380	1	0	1	0	1	0	0	3
NM_022147	1	1	1	0	1	2	0	6
NM_152643	1	1	1	0	2	0	1	6
NM_012396	0	0	1	0	1	0	1	3
NM_001666	0	0	1	0	2	0	1	4
NM_021629	0	0	1	2	0	0	1	4
NM_025158	0	0	1	0	1	2	1	5
NM_001677	1	1	0	1	0	0	1	4
NM_006004	1	1	0	2	0	1	1	6

RefSeq	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat.6	Pat.7	Different patterns
NM_003876	2	0	0	0	1	1	1	5
NM_015578	2	0	0	0	1	0	1	4
NM_006500	1	1	0	0	0	1	2	5
NM_013366	1	0	0	1	0	1	2	5
NM_015537	1	0	0	0	1	0	2	4
NM_021630	1	0	0	1	0	0	2	4
NM_022727	1	0	0	1	0	2	1	5
NM_022760	1	1	0	1	0	1	2	6
NM_024076	1	0	0	2	0	0	1	4
NM_002522	0	1	0	1	1	1	2	6
NM_003318	0	0	0	1	0	1	2	4
NM_004078	0	1	0	2	1	1	1	6
NM_005505	0	1	0	0	1	1	2	5
NM_013368	0	0	0	1	1	1	2	5
NM_147193	0	1	0	1	0	2	1	5
NM_001200	1	0	0	2	0	1	1	5
NM_001616	2	0	0	1	0	1	1	5
NM_016199	2	0	0	0	1	0	1	4
NM_000292	1	0	0	1	1	0	2	5
NM_002466	1	0	0	1	1	0	2	5
NM_003004	1	0	0	0	1	2	1	5
NM_006815	1	0	0	0	2	1	1	5
NM_007075	1	0	0	1	2	0	1	5
NM_025113	1	0	0	1	2	0	0	4
NM_032892	1	0	0	1	0	0	1	3
NM_138340	1	0	0	1	1	0	2	5
NM_144627	1	0	0	0	1	2	0	4
NM_153355	1	0	0	0	1	0	2	4
NM_005318	0	0	0	1	1	1	2	5
NM_012334	0	3	1	0	1	1	1	7
NM_005796	0	1	1	0	0	1	1	4
NM_013312	0	3	1	0	1	0	1	6
NM_000775	1	1	1	1	1	0	0	5
NM_002486	1	0	1	0	0	1	1	4
NM_003866	1	0	1	0	1	0	1	4
NM_004738	1	0	1	1	0	1	0	4
NM_005911	1	0	1	0	1	0	1	4
NM_015555	1	0	1	0	1	0	1	4
NM_018846	1	0	1	0	1	0	1	4
NM_145065	1	0	1	0	1	0	1	4
NM_000267	0	0	1	0	1	1	0	3
NM_000701	0	0	1	1	1	1	0	4
NM_001780	0	0	1	1	1	0	1	4
NM_002888	0	0	1	0	1	0	1	3
NM_004453	0	0	1	0	1	1	1	4
NM_005342	0	0	1	0	0	1	1	3
NM_006810	0	0	1	0	1	1	1	4
NM_019062	0	0	1	0	1	1	1	4
NM_019074	0	0	1	0	1	0	1	3
NM_024099	0	0	1	1	1	1	0	4
NM_024834	0	0	1	0	1	1	1	4
NM_033084	0	0	1	0	1	1	1	4
NM_002151	1	1	0	0	1	1	1	5
NM_005485	1	1	0	1	0	0	1	4
NM_152757	1	2	0	0	1	0	1	5

RefSeq	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat.6	Pat.7	Different patterns
NM_001519	1	1	0	1	1	1	0	5
NM_004252	1	1	0	1	1	0	1	5
NM_016096	1	1	0	0	1	0	1	4
NM_018103	0	1	0	1	1	1	1	5
NM_003597	0	0	0	1	1	1	1	4
NM_000305	1	0	0	1	1	1	0	4
NM_003128	1	0	0	1	0	1	0	3
NM_006111	1	1	0	1	1	0	1	5
NM_012407	1	1	0	0	1	0	1	4
NM_014154	1	0	0	1	0	1	0	3
NM_015373	1	0	0	1	1	0	1	4
NM_152499	1	1	0	0	1	1	0	4
NM_001744	0	0	0	1	1	1	1	4
NM_004939	0	1	0	1	1	0	1	4
NM_052855	0	1	0	1	1	1	1	5
NM_080655	0	0	0	1	1	0	1	3
NM_000224	1	0	0	0	1	1	1	4
NM_002014	1	0	0	0	0	1	1	3
NM_003999	1	0	0	1	1	1	0	4
NM_005720	1	0	0	0	0	1	1	3
NM_006703	1	0	0	1	1	0	0	3
NM_013286	1	0	0	1	0	1	1	4
NM_014944	1	0	0	1	0	1	1	4
NM_018144	1	0	0	1	1	0	1	4
NM_021945	1	0	0	0	1	1	1	4
NM_024544	1	0	0	0	1	0	1	3
NM_032704	1	0	0	0	1	1	1	4
NM_000732	0	0	0	0	1	1	1	3
NM_015149	0	0	0	1	1	0	1	3
NM_018177	0	0	0	1	1	0	1	3
NM_130839	0	0	0	1	0	1	1	3
NM_031892	0	0	0	2	0	0	4	6
NM_025163	1	0	0	4	0	0	2	7
NM_030927	0	0	0	1	1	0	5	7
NM_024665	2	0	0	3	0	0	2	7
NM_002247	2	0	0	2	0	0	3	7
NM_002275	0	1	3	0	2	1	0	7
NM_032047	1	0	3	0	0	0	0	4
NM_000361	0	2	0	3	0	1	2	8
NM_000034	1	1	0	0	1	0	4	7
NM_145204	4	0	0	0	0	1	1	6
NM_001362	0	0	0	0	1	0	3	4
NM_170606	2	0	0	0	0	1	3	6
NM_005160	1	0	0	0	1	0	4	6
NM_145696	2	0	0	1	0	0	3	6
NM_003307	0	0	0	0	2	1	3	6
NM_012205	2	0	0	0	2	2	0	6
NM_002221	0	0	0	0	1	0	2	3
NM_139126	0	0	0	0	1	1	4	6
NM_152384	0	0	3	0	1	1	0	5
NM_001998	0	1	2	0	1	0	2	6
NM_015393	1	0	2	0	0	0	0	3
NM_022784	0	0	2	2	0	0	1	5
NM_006031	1	1	1	0	0	0	3	6
NM_080432	0	1	1	1	0	0	3	6

RefSeq	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat.6	Pat.7	Different patterns
NM_014877	1	0	1	0	0	0	3	5
NM_018173	0	2	1	0	1	0	3	7
NM_001541	0	0	1	0	1	0	3	5
NM_017967	0	0	1	0	1	0	3	5
NM_031477	0	0	1	3	0	0	1	5
NM_001541	0	0	1	0	1	0	3	5
NM_012287	3	1	0	1	0	0	0	5
NM_014824	1	1	0	2	0	0	2	6
NM_021947	0	1	0	1	1	0	3	6
NM_002663	0	0	0	1	0	0	3	4
NM_025140	2	0	0	1	0	0	2	5
NM_016250	0	0	0	2	0	2	0	4
NM_016621	2	0	0	2	0	0	1	5
NM_000517	1	0	0	2	0	2	0	5
NM_017493	1	2	0	1	0	0	3	7
NM_003282	0	2	0	0	3	1	1	7
NM_005414	0	0	0	0	1	1	3	5
NM_016534	0	0	0	2	0	1	2	5
NM_017993	0	0	0	2	0	1	2	5
NM_032630	0	0	0	0	0	2	1	3
NM_033087	0	2	0	0	3	0	1	6
NM_032233	2	1	0	1	0	0	2	6
NM_002754	1	0	0	0	1	0	3	5
NM_014454	1	1	0	0	2	0	2	6
NM_139265	1	1	0	0	1	0	3	6
NM_002081	0	0	0	2	1	0	2	5
NM_003129	0	0	0	1	1	0	3	5
NM_004490	0	0	0	2	0	1	2	5
NM_015435	0	0	0	1	1	0	3	5
NM_015993	0	0	0	0	3	1	1	5
NM_022373	0	1	0	0	1	0	3	5
NM_032111	0	1	0	0	1	1	3	6
NM_003599	2	0	0	0	1	2	0	5
NM_004641	2	0	0	2	0	0	1	5
NM_020727	2	0	0	1	0	0	0	3
NM_080616	2	0	0	2	0	0	1	5
NM_002027	1	0	0	0	1	3	0	5
NM_002333	1	0	0	0	1	0	3	5
NM_005380	1	0	0	2	0	0	2	5
NM_022743	1	0	0	0	3	0	1	5
NM_000861	0	0	0	2	2	0	1	5
NM_007014	0	0	0	0	1	2	2	5
NM_012267	0	0	0	0	2	2	1	5
NM_016098	0	0	0	1	0	0	3	4
NM_033182	0	0	0	0	1	0	3	4
NM_138343	0	0	0	1	1	0	0	2
NM_021783	1	1	2	0	1	0	0	5
NM_003896	0	0	2	1	0	0	1	4
NM_004254	0	0	2	0	1	0	1	4
NM_005276	0	1	2	0	1	1	0	5
NM_139276	0	1	2	0	1	1	0	5
NM_001633	1	0	2	0	0	1	0	4
NM_007344	0	0	2	0	1	0	0	3
NM_017789	0	0	2	0	1	0	0	3
NM_022772	1	1	1	2	0	0	0	5

RefSeq	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat.6	Pat.7	Different patterns
NM_002357	2	1	1	0	0	1	0	5
NM_001904	0	1	1	1	0	0	2	5
NM_004905	0	0	1	1	2	0	0	4
NM_015569	2	0	1	0	0	0	1	4
NM_001275	2	0	1	0	0	0	1	4
NM_006454	2	0	1	0	1	0	0	4
NM_007216	2	0	1	0	1	0	0	4
NM_080552	1	0	1	0	0	0	2	4
NM_000155	0	0	1	0	1	0	2	4
NM_001553	0	0	1	0	2	0	1	4
NM_005566	0	0	1	0	2	0	1	4
NM_014392	0	0	1	0	1	0	2	4
NM_001627	2	0	1	0	1	0	0	4
NM_003380	2	0	1	0	0	0	1	4
NM_018243	2	0	1	1	0	0	0	4
NM_153687	2	0	1	0	0	0	1	4
NM_016167	1	0	1	0	0	0	2	4
NM_016408	1	0	1	0	0	0	2	4
NM_024959	1	0	1	0	2	0	0	4
NM_000626	0	0	1	0	2	1	0	4
NM_006401	0	0	1	0	0	1	2	4
NM_007317	0	0	1	0	1	0	2	4
NM_017803	0	0	1	0	0	1	0	2
NM_018480	0	0	1	0	1	0	2	4
NM_021820	0	0	1	0	2	0	1	4
NM_147159	0	0	1	0	1	0	2	4
NM_003647	2	1	0	1	0	0	1	5
NM_030912	1	2	0	1	0	2	0	6
NM_032211	1	2	0	0	2	0	1	6
NM_014950	0	2	0	1	2	0	0	5
NM_017503	2	2	0	1	0	0	1	6
NM_031439	2	1	0	0	1	0	1	5
NM_144999	2	2	0	0	0	1	1	6
NM_002396	1	2	0	1	0	0	2	6
NM_152255	1	1	0	2	0	0	1	5
NM_053005	0	2	0	0	2	1	1	6
NM_006516	2	1	0	0	1	0	0	4
NM_006233	1	1	0	0	0	1	2	5
NM_033637	1	1	0	0	1	0	2	5
NM_000245	0	1	0	1	1	0	2	5
NM_002346	0	1	0	1	0	1	2	5
NM_006788	0	1	0	0	1	1	2	5

Supplementary data for Chapter 15: Genome-wide promoter analysis uncovers portions of the cancer methylome

Supplementary table 1: Details cell line information used in screening study

Cell line	Growth Medium	Cell type
DU145	10%RPMI1640	Prostate Ca metastasis to brain
22Rv1	10%RPMI1640	Prostate Ca
LNCaP	10%RPMI1640	Prostate Ca metastasis to lymph node
PC3	10%RPMI1640	Prostate Adenocarcinoma metastasis to bone
HCT116-p53+/+	10% McCoy 5X	colorectal carcinoma
HCT116-p53-/-	10% McCoy 5X	colorectal carcinoma
RKO-p53+/+	10% McCoy 5X	colorectal carcinoma
RKO-E6 (fuctional p53-null)	10% McCoy 5X	colorectal carcinoma
SW480	10% McCoy 5X	colorectal adenocarcinoma
DLD1	10% McCoy 5X	colorectal adenocarcinoma
MDA-MB-231	10% Leibovitz's	breast adenocarcinoma
BT-20	10% MEM	breast carcinoma
MCF-7	10% MEM	breast carcinoma
MDA-MB-436	10% Leibovitz's L-15	breast carcinoma
HTB-59	10% Leibovitz's L-15	Lung squamous cell carcinoma
HTB-58	10% MEM	Lung squamous cell carcinoma
A549	10% Ham's F12K	Lung adenocarcinoma
H23	10% RPMI 1640	Lung adenocarcinoma
HELA	10% DMEM/Ham's F15	Cervical adenocarcinoma
SiHa	10% DMEM/Ham's F15	Cervical squamous cell carcinoma
CSCC-7	10% DMEM/Ham's F14	Nonkeratinizing large cell cervical squamous cell carcinoma
CC-8	10% DMEM/Ham's F15	Cervical adenosquamous carcinoma

Leiden*: a kind gift of Prof. GJ Fleuren (Leiden University Medical Center, Leiden, the Netherlands)

Supplementary table 2: Primers used for bisulfite sequencing

Gene	Genbank accession	Forward 5'-3'	Reverse 5'-3'	Nested sequencing primer F	Nested R	Amplicon size (nucleotide range)	Annealing temperature
ERCC3	NM_000122	TTTAAGTTTTAGAAGGATGAGGTTT	AAAACTCCCTATTATCCACACAAAA	TATTATATAGGTGTTTAATATATAGT	TACCCATAACAACTACAACAACAAA	215 bp (+794 to +1009); BS2: 274 bp (+829 to +1103)	Step down 66-56°C
GLDC	NM_000170	TGGGTAGGATTGGAGATGGTAGTTT	AATTATAACTCCACCCAAAACACCT			310 bp (-261 to +49)	Step down 66-56°C
HADHA	NM_000182	TTTGGAGTATTGTTTGATTAAA	CCTCAACTCCCACTTTCCCAAA	GATTAAAATTTAGATGTCGAGTGGT		429 bp (-492 to -63)	54°C
MET	NM_000245	TTTGGGTGGTGTTAGTTTTTATTTT	ACAAAAACCTCCAAATCACCTA			292 bp (+1008 to +716)	Step down 66-56°C
NDP	NM_000266	TGGTAGTTATTGTTTGAAAAAGAA	ACCATTACAATCATATATCAATCAAAC			296 bp (-18 to -313)	Step down 66-56°C
NF1	NM_000267	TTGGTGGTTAGATTTTGATTGTTAT	ACTCTCAACTTAAACTCTCAAAAAAA			184 bp (-726 to -543)	Step down 66-56°C
PHKA2	NM_000292	TATTTGTGTTTTTTAGGGTTTTAAA	ATTACCTCCCTAAATCCCTTCTCT			160 bp (+460 to +620)	Step down 56-48°C
PON2	NM_000305	GGGGAAATAGGTTTGTGAGGATAAA	TCCCCAACAAACCCACAACCACCAA			224 bp (+931 to +1155)	Step down 66-56°C
ATXN1	NM_000332	TTATTTAGGTAATTTGGGGGATT	AAAAACCCCTTAAAACTTTC			271 bp (-546 to -817)	Step down 66-56°C
ALDH3A2	NM_000382	GTTCATTTTATTTTTTATATTTTAG	TCAACCCACAACCACAATTCACT			233 bp (+867 to +1100)	Step down 66-56°C
GALE	NM_000403	GTTTGGTAAAGGATTTTATTTTTGA	TCTTTATCCAAAATAACTAAAAACT			269 bp	Step down 66-56°C
KRT14	NM_000526	GGATAAGAAAGTTTAAAATATTTTA	ACTACAAATAATCATAATACAAAAA			264 bp (+811 to +1075)	Step down 66-56°C
CACNA1C	NM_000719	ATGGGAAATTTAAGAAGTAGGG	GAAACCCAAAACTCTACATTTC			286 bp (+1209 to +1495)	54°C
CACNB3	NM_000725	AGTTGGGTTTTTTTTATAGGGATGT	CTACCCCAACTCCTAAACTCTTAAC	BS2F:TTAAGAGTTTAGGAGTTGGGGTAG	BS2R: TTCAACCCAACCTCTCTCCACTTC	237 bp (+406 to +643); BS2: 236 bp (+619 to +855)	Step down 56-48°C
CD3D	NM_000732	GATTTGGTGTTTTTGGTTGTTTAGT	AATATCTACTTCTCTACTTTCTAC			322 bp (-306 to +15)	Step down 66-56°C
RPS28	NM_001031	TTTAGGAGGGAGGTGAGGAGTAGGA	TCTTTCCCAACAACCCCAA			185 bp (-278 to -94)	Step down 66-56°C
BACH1	NM_001186	TTAGAAGGGAGAGGAGAGGTGGTTT	AACCACTCACCAAAACTAAAA			426 bp (+122 to +548)	Step down 66-56°C
BMP2 *	NM_001200	MSPF:TTTTAGGAATTAGCGTTCGA	MSPR:ATACGAAAAAACAACTCGCC	UF:TTTTTTTTAGGAATTAGTGTTTGA	UR:ATAATACAAAAAAACAACTCACC	M1:190 bp (-227 to -38)	Step down 66-56°C
CBR3	NM_001236	TTTTGGGGTTTTTTTAAAATAATTTTT	AAAAACCCAACTAATATCAATACC			120 bp (+841 to +961)	Step down 56-48°C
CDK6	NM_001259	GGAAGGTTTTAATTTGTATATTTTTTTG	CTCCTCACCTAAAAAACCCAATC			385 bp (+741 to +1125)	54°C
DARS	NM_001349	TATTAGGTGGTTGGTTGTAGATTTG	ACAAACCAACTACACTCCTCTATCC			282 bp (-420 to -138)	54°C

B4GALT1*	NM_001497	MF1:GGTAGAAAAAGGTTATTTAGGAAACG	MR1:AAACACAAAAATAACAACCCGTC			M1: 151 bp (-717 to -566)	60°C
GPR39	NM_001508	GTGTTTATGTGTTGAGTTAAATTTYGT	CTTACTAAAACAACCCAACCTAAA			377 bp (+455 to +831)	54°C
APOC1	NM_001645	GGGTTTGGTGGTTAAAGATTGATTT	CCCTTTCCCCACTCAAAATATAACA			355 bp (-364 to -10)	Step down 66-56°C
ATP1B1	NM_001677	GTTGGGTTGGTTTGTAAAGAAGAG	AAACTATAAATCCACCCCTCCTAAA	AGAAGAGGGAAGGGT		193 bp (-983 to -790)	54°C
CAMK4	NM_001744	TTTTGATAGAGGATGGTTGGTTAGT	AAACTTAAAAAAAACACAAAACCC			265 bp (+441 to +706)	Step down 56-48°C
CBR1	NM_001757	AGGGGTAGGGATGGTTTAGTTT	ATACCCTTATTACCTCCAATCACC			242 bp (-168 to +75)	54°C
CLU	NM_001831	GGGAATTGTGAGATTGTGTTTTG	CCCATCCRTCCTAATATAACTCTAC			445 bp (+810 to +1254)	54°C
ENPEP	NM_001977	GAAGTTTTTTAGTTAGGAGGGAGAG	CAAAAAATTATCACAACTCCCCTT			260 bp (-273 to -13)	Step down 66-56°C
FKBP4	NM_002014	TTTTTTTTAAGTAGGGAAGGGTTT	TAATTCTTACCTCCTTCTAACTACC			250 bp (-220 to +29)	
HPCA	NM_002143	GTTTAGGAGAGAGATTTTTTGGTATTATT	ACTAAAACTACCCCCACTCAC			368 bp (+687 to +1054)	54°C
HOXB5	NM_002147	TGTAGGGAGTTAGTAGGGAGGTAGT	AAATTAACTTTAACCTCTAAACAAAAC			109 bp (+673 to +782)	Step down 56-48°C
HPN	NM_002151	TGTTTTTGTTGTTTGTTTGGGGAGG	CCCCTTTCTAAATAACCACAATTCA			342 bp (-461to -120)	Step down 66-56°C
MAL	NM_002371	TTGATTAAAGTGGTTTTGTTTTGTTT	ACAAACCACTAAACAAAATACTACC			379 bp (+722 to +1101)	Step down 56-48°C
MEF2	NM_002397	TTTTTTGTAAGAGTGAAATTGATGATTT	AAAAATTAAAAAATACTAAAAAAA			245 bp (+418 to +663)	Step down 56-48°C
MYBL2	NM_002466	AGTTTTAGTTATTTGGGAGGTTGAG	ATACCTCTCCCTACTTCCTCTCCTAC			364 bp (-734 to -370)	54°C
NCBP1	NM_002486	TTTAAGTAAATAATTTAAGGAAAATGA	AACCTTAAAAAAAATCAAAACTAAC	BS2F: GTGGGAGTAGATGAGTATATATT	BS2R: ATAACCTCTCAATAATTTTCCC	237 bp (+211 to +448); BS2: 218 bp (+396 to +614)	Step down 56-48°C
PAK3	NM_002578	TTTGAGAGATTGGAAGGGTTTTAT	AACCAAAAAAAATAAAAAATCACAAC			298 bp (+687 to +985)	Step down 56-48°C
PEX13	NM_002618	TGTATGGATTTAGTTGGTAAAGTAA	GAACCCACTAAACTCACAAC			296 bp (+861 to +1156)	54°C
PLCG2	NM_002661	GGGGAATTTTTTTTGTATTTGGTTAAG	AATCCCTAAACRCCCTACCTA			444 bp (+726 to +1169)	60°C
PVRL1	NM_002855	TGGGAGTAGAGGGTTTGTAGAAGT	ACAAAAACAAACAATAACCCCTAAC	BS2F: GTTTGTAAGTTGGAGT	BS2R: ATAAAATCCAAATAAACTAA	162 bp (+156 to +318); BS2: 270 bp (+821 to +1091)	Step down 56-48°C
RBMS2	NM_002898	TTAGTATTTTGGGAGGTTGA	ACTCAAAACTTCCCCTACTTTACACT			302 bp (-764 to -462)	54°C
RNASE4	NM_002937	GTTTTTTGTTTGGGTTGTTTTTG	AAATATTCTTTTAAATCCTACTTTACATCT	BS2F: GATGGGTGGTTTTGTTAAAGAAG		190 bp (+307 to +497); BS2: 223 bp (+553 to +776)	Step down 56-48°C
SECTM1	NM_003004	TAGGAAAATGAAAGATA	AAAATCCAACTCACTCTACCC	BS2F:AGAGGGAGGGGTTAGGGAG		418 bp (-129 to +289)	54°C

TFAP2A	NM_003220	TTGGGGGAAAGTTTAAAGTTAATAAA	CTCCAAAAATACCTTCTACAAAACC			296 bp (-709 to -514)	Step down 66-56°C
UBE2I	NM_003345	GAGGGAATGAATGAATGAGTG	AAAAAAAATATAAATACTCCCTAAACC	GGGAGGGAATGAATGAATGTG		358 bp (-502 to -144)	54°C
UGDH	NM_003359	GGGTAGAATTTAGGAGGGTTTGGTT	CTTACCCACTTCCCAACAACAAA			254 bp (+1059 to +1313)	Step down 66-56°C
VGF	NM_003378	TTTTTTTATTTATTTATTTATTTTTTTT	ATTAAAAAATACCTACCTCTCCC	BS2F: TTATGAATGAATATTGA	BS2R: TTACCAACTAATATCAC	180 bp (+661 to +841); BS2: 293 bp (+907 to +1200)	Step down 56-48°C
ϕ XPO1	NM_003400						
ZNF198	NM_003453	GTTTTTTTTATTTTGGTTGATGGAG	TAAAAACTTAAAACCCCCTAAAAAC			210 bp (+10 to +220)	Step down 56-48°C
KLF11	NM_003597	GAGGAGGGGTGGTTTTT	AACTTCTATCATTCTCCCTT			478 bp (-998 to -520)	54°C
ϕ NOL4	NM_003787						
C17orf35	NM_003876	ATTTTTGAAGAAATTGAGTTTT	TCTCGACCTCCAAAATACCC	BS2F: TAGTTTGGGAGTTATAGG		381 bp (-377 to +4)	54 °C
ST3GAL5	NM_003896	AAATTTAGAAGGAGATTATAGGTATA	TAAAAATAAAAAAAATTCAAAAAAAA	BS2F: GTTTTGTTTTGGTAATGTTTAT	BS2R: CATACTAATAAAAAAAC	216 bp (+427 to +643); BS2: 358 bp (+645 to +1003)	Step down 56-48°C
OSMR	NM_003999	TAGTTTGGGTGATATAGTGAGATGT	AAAAATCCTTTAACTAAATTTTAACC			285 bp (-761 to -476)	54°C
RHOB	NM_004040	AGTTTGTTAGGAAGAGGGGTA	AAACAAAACCAACTTAATTAA	BS2F: ATTTTGAATGGGAGT		253 bp (-28 to +225)	54°C
CSRP1	NM_004078	TTAATTAGTGGGTTTTGTATATTG	AACCTACCCTAAAAAAATCTATCTTACCC	BS2F:TATTATTTTAGGGTTGATGTTATG		386 bp (-448 to -62)	54°C
ASMTL	NM_004192	GGATATTTTTATTTTTATTTTTTTT	TAAATACTTAAAATTCCTTACAAAC	BS2F:GGTTTTAGGGGAGATTAATGGG	BS2R:CTTATACAACAACTTCCCAATCAC	254 bp (+103 to +357); BS2: 328 bp; M1: 141 bp	BS1: Step down 56-48°C; BS2: 54°C; M1:60°C
SLC9A3R1	NM_004252	AGTTTTGGGATAGTATAAA	AAACTCCTAACCTCAAATAATCC	BS2F:GGATTAGGGTTAGGGTTT		275 bp (-848 to -573)	54°C
KIF1A *	NM_004321	MSPF: GCGCGATAAATTAGTTGGCGATT	MSPR: CTCGACGACTACTCTACGCTAT			140 bp (+870 to +1010)	60°C
BAG1*	NM_004323	MSPF:GTTGTTTAAGATGAATCGATTTTTC	MSPR: TATACATAATCAAATACCCACCGTA			100 bp	58°C
ETFDH	NM_004453	GTTGGGATTTGGAGTTTTTATTTT	AAAAACAAAAACTCTACTACAATT			172 bp (-158 to 14)	Step down 66-56°C
HUS1	NM_004507	GATTTTTATTAGGGTTATTTT	AACACTTAAAATAAATATCACAAAAT	BS2F:GAGGAGGTGGGGTTAGGGGGTTA		266 bp (-256 to +10)	54°C
PIGH	NM_004569	TGAGATTAATTTGGTTAATATGGTG	AAAAAACAAAACCCTATCTCAAAAC			208 bp (680 to 888)	Step down 56-48°C
PAPSS2	NM_004670	GGTGGATTGAAATAAGTTGTTGG	ACATAAACCCCCTCCATCCTACCCC	*		285 bp (-728 to -354)	54°C
EDG4	NM_004720	GGGTAGGGTTTTTTTAGGTGGT	AAACCTCACCTAAACCTCCAATC			196 bp (937 to 1133)	Step down 56-48°C
CDK5	NM_004935	TTGGATTTTAAGTTTTAGGAAGTAA	ACAATACCACCAACCAAATCAA	*		342 bp (-452 to -197)	54°C

CEBPD *	NM_005195	MSPF: AGGTTTAACGTAAGATGTCGTTATAC	MSPR: CGAAATACGAACTCTCCGAC	UMSPF:GGTTTAATGTAAGATGTTGTTATATGA	UMSPR: CTCCTCAAAATACAAACTCTCCAAC	136 bp (-118 to -253)	56°C
NBL1	NM_005380	TGGTAAGGTGTTTAATATGTGTGAA	AAATTCCAACAATTCTCAACCTCC			286 bp (-874 to -580)	Step down 66-56°C
PODXL *	NM_005397	MSPF: TTGTAGAACGCGTAGTAGTTCG	MSPR: GAATACTTCTAAATAAACCCGACGA	UMSPF: GTTGTAGAATGTGTAGTAGTTTGG	UMSPR: AAATACTTCTAAATAAACCCAACAAA	146 bp (-469 to -324)	56°C
SCARB1	NM_005505	GGTAGTTTTGGGTTTTAGGATTT	AATAACCAACCATAATAATAATCCC	*		385 bp (-523 to -138)	54°C
HMGN2	NM_005517	GTTTTTAGAGGAGTAGGGTAGTGTAG	CCATTCATTAACCCAACTAAAATC			383 bp (+624 to +1006)	54°C
KNS2	NM_005552	GTTTGTAGGTATTGGGAGGGTTT	AAATATAAACTATAAAAATAAAAAACAACC	BS2F: GGGGATAGAGGGGGGTTGTT	BS2R: ATCCCCCCAAAACCCAACTA	255 bp (+22 to +277); BS2: 292 bp (+234 to +526)	Step down 56-48°C
LGMN	NM_005606	AGGTTTTTTTAAGTGGTTTTAATGTG	CCAAAAAAACTACAAATCCCAATATAC	BS2F: GTAGATGTTTTTGGAATTAT	BS2R: CAACCTCACACCAAACACCCA	208 bp (+680 to +888); BS2: 300 bp (+743 to +1043)	Step down 56-48°C
RGS4	NM_005613	TAGAGGGAGATAGAGGAGTTGGTATT	ACAAACCTACAAACCCTTTACACAT			245 bp (+904 to +1149)	Step down 56-48°C
SGK	NM_005627	GTTTTTTTAATGGGGATAGAAT	TACCCTTAACAACCTCAATTTTCAC	BS2F: TTGGTTTTGGTTAAAAGTATAAAAA	BS2R: CTACATCTCCCCCATAAACAAC	262 bp (+823 to +1085); BS2: 201 bp (+14 to +215)	Step down 56-48°C
ARL4A	NM_005738	TTGTTTAAGGTTAATTATTGGGGGA	TCTCCTAACAACGCCAAAACAAAAT			30b bp (-392 to -89)	Step down 66-56°C
RPL10	NM_006013	GTAAAGTTTAGGGTAAGAGTTTTA	ACCTAAAAAATGTAAAAACCAACAA			364 bp (+714 to +1078)	Step down 66-56°C
TUBB4	NM_006087	TTGTATTTAATTATGGGAGGAAAAT	AACAAACATCTAAAAAATCCTACCC			266 bp (+428 to +694)	Step down 56-48°C
PLTP	NM_006227	TTTTGTTTTTTAGGTTGGAGTGTAGT	AACTCATACCTTTAATCCCAACACT			216 bp (-257 to -41)	Step down 66-56°C
ARIH2	NM_006321	ATTTTAATTTGGGTGATAGAGGGAG	ATAATCCCAACACTTTAAAAAACCA			320 bp (-446 to -765)	Step down 66-56°C
HMG20B	NM_006339	TTTTTATTTTTGGTTAGGGATTTTT	CTTCTTACCCTTAAACCAAC	BS2F: GGCCCTACCTCCATCTTT		238 bp; BS2: 184 bp	Step down 66-56°C
COVA1	NM_006375	TTATATATAGGGTGTTTATTG	CAAAATACTAAAAACTTAAC	BS2F: TTTATTGGTTATTGGTTAT		392 bp (-384 to +8)	54°C
MCAM	NM_006500	GTTAGTTAATGTAGTTTTAGGGGA	CTCCCCATACAAATACTTCTAAATC			256 bp (-338 to -594)	Step down 66-56°C
NUDT3	NM_006703	GTGGAAATTAGTATTTGAAAGGA	CATCCTAAAATTCAAACTATCAAC	TTTGAAAGGAGAGGGA		427 bp (-512 to -85)	54°C
PC4	NM_006713	TATTTTTGGGAAAATGGGTTAGA	TAAACCCTTTCCAAAACCTTAAAA	GTATTGATTTTTTTTATTATGA		278 bp (-376 to -98)	54°C
RNP24	NM_006815	GAAAGGATTAGGGAAAAATTTTTCC	ACCAAAAACACCAACAATTCAACAAA			318 bp (+739 to +1057); BS2: 342 bp (+715 to +1075)	Step down 66-56°C
WDR45 *	NM_007075	MSPF: TTTAGTGTTGTGGTTTTATGTGGAC	MSPR: AAAATATCCACTCCCCAACGTA	UMSPF: TAGTGTTGTGGTTTTATGTGGATGT	UMSPR: CTAAAAATATCCACTCCCCAACATA	129 bp (-588 to -460)	56°C

ZNF195	NM_007152	TTTTTGTTATTTAGGGTTTGAGGG	ACACTCACCATCTCCTAACCTCCAA			262 bp (+831 to +1093)	Step down 66-56°C
NISCH	NM_007184	GGTTTTTGTGTTTAGTGGTAG	AAATAAAACCTATACTTACC			254 bp (315 to -569)	Step down 56-48°C
RALY	NM_007367	GTTAGGTTGAGAGGGTGGG	CCCCCTCAACTTCTCTAAATTC			314 bp (+984 to +1297)	54°C
RRAS2	NM_012250	TATTTTTAATTGATATTTATTTTTA	CTACCAAACTAAAATACTACATATA	BS2F: TATAGATAATTAGGGTGATATATTT		174 bp (+800 to +974); BS2: 214 bp (+760 to +974)	Step down 66-56°C
KPNA6	NM_012316	TTTTATTTTTTATTTTTAGTAGTTG	ACACTCCCTAAACCCAAAAACCCTA	BS2F: TTTTGGAGTTAATTGGTTTGTTTTT		263 bp (+862 to +1125); BS2: 176 bp +949 to +1125)	Step down 66-56°C
SSBP2	NM_012446	GTTTTTTTTGTAAGAGGTGAG	AAAAAAACCTAAAACCCTCTTCTAC			155 bp (-649 to -804)	Step down 66-56°C
RBM15B	NM_013286	ATTTAGTTGTGTGTGAGGAGGATTT	ACCTTAACCCTATAAACCATATCCAA			228 bp (+883 to +1111)	Step down 56-48°C
ϕ HOOK2	NM_013312						
ANAPC2	NM_013366	TGAGGTAGTTGATAGAGAAGTTG	CTAAATTCCTCAATAAACCTAACCA			274 bp (-442 to -169)	Step down 66-56°C
ARMC8	NM_014154	GTTGATAAATTAAGTAGTTAGGAAG	TTCACTTTCATATTCATTCTCTCTC			237 bp (-427 to -191)	Step down 66-56°C
ZNF237 *	NM_014242	MSPF: TCGCCAGCCAACGAGCGCACG	MSPR: CCGGGACTGGACCTCGCGGAG			110 bp (+808 to +918)	62°C
ITGB3BP	NM_014288	GTGTTGTTTTGGGTTTAGAATTTTT	ATCTTCTCTTCTTACAAAAAACCCC	BS2F: AGGAAGGAGAAGTAGTTGTA	BS2R: AAAAATTCTAAACCCAAAACAACAC	280 bp (+875 to +1155); BS2: 360 bp (+520 to +880)	Step down 56-48°C
SESN1	NM_014454	GTTGTTAATAGTTTTTTGGGGTAGA	CTAAATCACCCTCTCACCCTCTCCTTAT			207 bp (+829 to +622)	Step down 66-56°C
PDLIM3	NM_014476	ATAGGTGTGATTTTAAGAGA	AAAAACTCCAACTTTAAAAA			140 bp (+777 to +917)	Step down 56-48°C
ARFGAP3	NM_014570	TTATATTTTAGTTTGGGGGATAGAG	ATAAAAAAATCAAAAACAAAAAAA			161 bp (+505 to +666)	Step down 56-48°C
GANAB	NM_014610	GGGTTTTTAGAGTAAGAAAAGGTTA	ACTATACTACACAATATCCTCCACC			127 bp (-492 to -365)	54°C
ZNF592	NM_014630	TTGTTAGTTTTTTATTTTTGTTTGT	CCACTAAAAAAACAAAATCTACCTA			286 bp (+190 to +476)	Step down 66-56°C
FAM20B	NM_014864	GTATGGAAGAGGTTAAGGGGGGTAGT	ACAATCCCACACCCACCCCCC	BS2F: GTTTGGTTGTTTATTTATTTAAGAA	BS2R: ACTACCCCCCTTAACCTCTTCCATAC	172 bp (+969 to +797); BS2: 219 (+604 to +823)	Step down 66-56°C
ALS2CR3	NM_015049	GTTAATGGTAGTTTTTATATAAT	ACTCTCTTCAAACTACCAA	TGGGGAAGTTAAAGGAGAGT		376 bp (-385 to -9)	54°C
SLC35D1	NM_015139	GGTTTAGTTTGGGTAGGGGT	AAAAAAACATTCAAAATTAATCTAAAC	TGGAGAGAGTTTGGAGAAG		349 bp (-557 to -208)	54°C
RGL1	NM_015149	MSPF: TTTTTGTTTTTTTACGGAGAGCGCG	MSPR: AAATCACGTAACGCAATCGAACG	(Um F): TTTTTGTTTTTTTATGGAGAGTGTG	(UmR): AAATCACATAACACAATCAAACA	238 bp (-941 to -703)	59°C
CSTF2T	NM_015235	GGGGTAAATTTTAAATGTGTTGAAA	AACTTTCTATAAAACCCCAACAACC			319 bp (-530 to -222)	Step down 66-56°C

NEDD4L	NM_015277	GTTTGTGATTGGTTTGGGGGG	TCTCCCCTCTACCTCCAAATCCCCC		463 bp (+745 to +1208)	Step down 66-56°C
PGEA1	NM_015373	TGAATGTAGGTTTTGGTTTTAGAGG	CTCTATAACCCAAAATTACCTTAAC		239 bp (-221 to +18)	Step down 66-56°C
NELF	NM_015537	ATTTGTTGTTTAGATTTGTAG	CAAATACAAACCAAACAAATAAAAC	TTTGTAGGATTATAGGGAGGT	222 bp (-510 to -290)	54°C
ZNF451	NM_015555	ATAGGATTTAGTTGGTTTTTGGTTT	CAACACCACTACTCCTACTATCCCTA		342 bp (+681 to +1023)	Step down 66-56°C
C19orf13	NM_015578	TAGGGAGTATGGTTTTGGATTTTTA	CTCAAATATTTCAACCCCTTCCTAT		177 bp (-749 to -572)	54°C
EIF5B	NM_015904	AATTGATAGTTTTATTGTTTATTTT	CACTAAAACAAAACACATATAATAT		239 bp (+804 to +1043)	Step down 66-56°C
ϕRASD1	NM_016084					
COPS4	NM_016129	TGGGATTATAGGGGTGTATTATTTA	CATTCCAATTTATCCAATACAAAAC	TTGTATTTTTAGTAG	242 bp (-611 to -369)	54°C
ING4	NM_016162	TATTGTTTAGGTTGGAGTGTAATGG	CCCAACACTTTAAAAAACTAAAATAAAC		235 bp (-749 to -982)	Step down 66-56°C
AMOTL2	NM_016201	GTTTTTGTTTTATGTTTTGTTTT	TAAATATCTTTACAATCCCTCCTCC	GAGTTGTTGGAGATAATTT	319 bp (-475 to -156)	54°C
DBR1	NM_016216	GGGATTTTAGAATTAGGTTTG	TCTCATAAATCTTATCCAACT		253 bp (+899 to +1152)	Step down 56-48°C
LEF1	NM_016269	AAGGAGGTGGTGATTGAGGGTTTTT	TTCATCTTCCTCCCTCTACCAAAAC		352 bp (-651 to -1002)	Step down 66-56°C
ϕISYNA1	NM_016368					
TUBG2	NM_016437	GATTGAAGTGGGAGGATTATTTAAGT	AAAAACTAAAAACCAACTTTTCTTC		190 bp (-415 to -225)	54°C
ARTS-1	NM_016442	AGGGTTAGGGGTATGTAGGAAAG	ATCTCCCACTATAATCATTCTCTAC		249 bp (-621 to -373)	Step down 66-56°C
SIRT7	NM_016538	TTTTGTAAGAGGTTAGGTTTTTGTT	CAAAATTAATCAATCCCTACCTACC		200 bp (-478 to -278)	54°C
TTRAP	NM_016614	AGGAAAGTTAGTGAGTAAAT	TTTCCCATACCTTTCCCAAAACTCTCTC		287 bp (-467 to -180)	54°C
POLA	NM_016937	TTTTTAAGGGGTTATTATAGGG	CCAACCAACCCAAAACTAAC		178 bp (-189 to -12)	Step down 66-56°C
FLJ20244 => new name TRMT1	NM_017722	TTTTTATTTGTGAAATGGGATTAGG	TAACTAATCCCTATAAAACCRATACC		344 bp (+797 to +1140)	54°C
FLJ20277	NM_017739	AGGTTGGAGTGTAGTGGTATTATTT	ATTCAAAACCAACCTAACCAACATA		181 bp (-716 to -535)	54°C
SLC39A4	NM_017767	GGGTTAATTATTTTAGTGGTAGATTTT	CTATCTCTACTAAAAATACAAAAATAACC	TTTTAATAGTTTGGAGTTA	386 bp (-932 to -646)	54 °C
DDX27	NM_017895	GAGATAATGGAATTTGTTTATTTAAA	CAATCAAAAAAACCCAAAAACCAAAA		365 bp (+868 to +1233)	Step down 66-56°C
SLC35A5	NM_017945	GGGTTGAAGGGAGATTTGAGGTGAGA	CCTATTACAAAACAAAAACCA		319 bp (+758 to +1077)	Step down 66-56°C
FKBP14	NM_017946	ATTGGTTATTGGTGTTGTTTTTTAG	AAATCCTAAAATTATAACATCTT	TAGGTAGTTTTTAAGTTAATGA	340 bp (-447 to -107)	54°C

NETO2	NM_018092	GGGGTAGTGAGGGTGTTTTT	AAATCCTAAAATTATAACATCTT			428 bp (-566 to -138)	54°C
OGDHL	NM_018245	TTTTTTATATATAAAGGTTGTTTGT	AAAACTACAAATCAAAAAACTA	BS2F:GTGGTTTTAGAGAGGAGTTT	BS2R:CCAAACTCAAAACTACRAACAA	313 bp (+741 to +1054); BS2: 382 bp; M1: 127 bp	BS1: Step down 66-56°C; BS2:54°C; M1: 60°C
FLJ10983	NM_018291	GTTTAAAGGGTTAATTAGGT	TCACTTACAATACTTATCCAATCT			296 bp (-455 to -160)	Step down 66-56°C
DDX28	NM_018380	TAGGGTAATGAGGAATAATTAAAGA	CCAACTACCRCTATAAAACCACT			486 bp (+654 to +1139)	54°C
SLC35C1	NM_018389	GATAGGAGTAGTTTTGGAAAGA	CCCCTTCCAAAAAATACTATTTTCC			327 bp (+104 to +430)	54°C
ZNF398	NM_018660	TTTAGGTTTGTTTTTGGGATAGGTT	AAAAACTAACAATTAACTTTCTCCCATT			312 bp (-154 to-465)	Step down 66-56°C
KLHL7	NM_018846	AAATTTGGAAATTTGTTTTGGGTA	AACCCTACCCAAAAACCTA	TTGGGTATATAAAAGAGTAG		229 bp (-228 to +1)	54°C
DNAJC10	NM_018981	TTTTTTTTAAAAGATTTTTAGGGAATTTAA	AAAAAACTACCCTAACCAAACAACC	TAAAATTTATTTTTTGGG		175 bp (-364 to -189)	54°C
DLL4 **	NM_019074	GTGTTGGGATTGTAGTAGTTAG	CTAATCCTAAAACTACAAACRAACTAA	MF1:TTTGAGTGATAGGCGGCGA	MR1:GCTACACCAAAACGACTAAAACTACGA	226 bp (+552 to +777); M1: 135 bp (+606 to +740)	54°C; M1: 60°C
LZTFL1	NM_020347	GTTATTTTTTGTAATTGTAGTTAGT	AAAATAATTAAACCACAAAAATAA			238 bp (+825 to +1063)	Step down 66-56°C
CKMT1	NM_020990	GGGTTTGGGGGAGTTTTATTTTTGG	CCAACCATAACTTAAAATCTAACTA			276 bp	Step down 66-56°C
GNB4	NM_021629	GAGAGAGAAAATAGGAAGTAGAAGTT	CTCCTACCTAAAAAAAACTCAAAC			340 bp (+905 to +1244)	BS1: 54°C
RNF123	NM_022064	AGGGTTGGGAGTATTTTTAGAGTA	CCCCAAACTCCACAAAACC			402 bp (+822 to +1223)	BS1: 54°C
CLSTN2	NM_022131	GTAGGGGGATTTGGATAGGG	CCRCCCAAACAACATCCTC			396 bp (+620 to +1015)	62°C
C9orf19 => new name GAPR1_HUMAN **	NM_022343	GATTGTAGGGGTGGGAATTG	AACTCACCCCTAAAAAACAACTC	MF1:TGTGTCGTGTTAGGTCGC	MR1:CGAATAACTACTCTACGCCGAC	443 bp (+403 to +845); M1: 122 bp (+592 to +713)	BS1: 61°C; M1: 60°C
CENTD3	NM_022481	GATTTTGAGTTGTATTTGTGTTTTTG	CCCAACTAAAAAAAAACAAAAAAAACC			163 bp (-823 to -986)	Step down 66-56°C
GDAP1L1	NM_024034	TTTTGTAGGGAGTTTGATATTGAGG	AACCACCAACTACAATTAATAAAAATCA			287 bp (+849 to +1135)	BS1: 54°C
DDX54	NM_024072	TGTATTTAGTAATTAGGGAGTAGTTAGTA	CTTCTTCCTCCACTAAACCATAAC			372 bp (+726 to +1097)	BS1: 54°C
FLJ2118 => new name NP_078813.1	NM_024537	GTGGGTTGTAGTTAGGTTCGTT	AACCCTAATTCTCCACCCTTACC			334 bp (+1161 to +828)	60°C

FLJ12875 => new name C1ORF166 **	NM_024544	GTTTAGTAATTTTAGGAAATTAGAGGG	ATACCCAAAAAAATAAACTAACACAA	MF1:GTTGGGCGATTGATTTCGA	MR1:AACCCCGATAACCGACTATAACG	444 bp (+744 to +1187); M1: 124 bp (+1004 to +1127)	BS1: 54°C; M1: 60°C
ARMC1 => should be ARMC7 **	NM_024585	TAGTAATTTGTATTTGGATATGGGATA	TTCTACAATAATTTACAACTAAAAATAAAA	MF1:ATCGTTTATTGGTTGGGGTTC	MR1:ATAAAAACAACGCGACTTCCG	404 bp (+755 to +1158); M1: 99 bp (+1050 to +952)	BS1: 54°C; M1: 60°C
FN3KRP	NM_024619	TTATTGTAGGTAATATTAATT	TCCCAAAAACCTAACAAAACATAAT	TTAATTAGGGTTGTAGTT		397 bp (-986 to -589)	54°C
CBLL1	NM_024814	TTTTGAGTAATTTTTTATTGGAGGG	CTAAAACTTCTCCTCATATACTTCA			333 bp (-552 to -219)	Step down 66-56°C
C10orf119	NM_024834	GTGTATATTTAGAGAAGGGGAATTT	ACCTCAACAAAAAACTACTATAAT			274 bp (-862 to -588)	54°C
C13orf18 **	NM_025113	ATTTGGGGTTGGGAAGTTATTT	CTCTCTAAAAACTCCTCAAAAAAAC	MF1:TTTTTAGGGAAGTAAAGCGTCG	MR1:ACGTAATACTAAACCCGAACGC	248 bp (+654 to +901); M1: 119 bp (+890 to +772)	BS1: 54°C; M1: 60°C
RUFY1	NM_025158	GGTAAGAGAGAAGGTTGGAAAAG	CGCAAATAAATAATCCCTAATCCC			399 bp (+709 to +1107)	BS1: 61°C
SIRT2	NM_030593	TTTTAGTTAGGTAAGAAAG	CCCGCCTATAACTAAACTTTACC			270 bp (-290 to -20)	54°C
NBR1	NM_031858	TTTTGGTGATATAAAAAGAAAGAGA	CTAAACTAAAAACCTCCATTAAAAC	BS2F: ATAGTTTTGTAGAATGAGGG	BS2R: CTAACTATATCTTAATTTCCTAT	240 bp (+891 to +1131); BS2: 286 bp (+541 to +827)	Step down 56-48°C
MGC15668 => new name	NM_032756	BS1:GGAAGGGAGGTTTAGTGTTTAAG	BS1:CTTCRAAACTCTCAATCCTAAAC			377 bp (+841 to +1217)	BS1: 54°C
GLOXD1 MGC15396	NM_052855	TATTGTTTAGGTTGGAGTGTAATAG	CTCATACCTATAATTCCAACACTTT			241 bp (-1741 to -1982)	Step down 66-56°C
MYLK	NM_053030	AGAGAGAGAGTAAGTGGGAGAG	ACCTAAAACCTAAATCTCCCAATATAAA			352 bp (-437 to -85)	54°C
MGC9564 => new name HCP1_HUMAN **	NM_080669	GGTTGGGAGGGAGGGTTGTA	GTAATAAACRAACCCTACAAAACCA	MF1:AGGCGTAGATAGCGTAAGTTTTACG	MR1:GCGCTCCCCTCCATATACGT	418 bp (+810 to +1227); M1: 121 bp (+993 to +1113)	BS1: 61°C; M1: 60°C
UBE3A	NM_130839	TATTTTGGGAGATATA	AAAAAATACCCAAAAAACTTCC	GGAGGATTGGTTGGG		390 bp (-381 to +9)	54°C
ABHD3	NM_138340	GGAAGGTTTTTGAGTGGTTGGTAGT	AACTTAATATTCCAAATAAAAAAA		BS2R: TCAACCCACAACCACAATTCACT	318 bp (+851 to +1169); BS2:149 bp (+851 to +1000)	Step down 66-56°C
HSPB6	NM_144617	AGGAGTTTAGATTTGGTAATTGATG	CTTCTTACTACAAAAATCCTATCC			189 bp (-366 to -157)	Step down 66-56°C

MGC35558 => new name C11ORF45	NM_145013	GATAAGGAAGGAGAATTAGTAAGTTT	CTTCTACAAAACCAAAAATAAAAATAC			392 bp (+682 to +1073)	54°C
PDCD4	NM_145341	GGAGGTAGGGATTATTAGGAAGT	ATTTATTTTTATTTTCTTCTACCCAATAAC			391 bp (+716 to +1106)	54°C
MRPL4	NM_146388	GAAGGTTTTTTGAGGGTAATAAGG	ATTACCCAAACTAAAATACAATAAC	TTTTGTTTTTAAATAAATGAAAAA		311 bp (-628 to -317)	54°C
KNDC1	NM_152643	GAGAAGTGAGGGTGGATTTTTT	CCTATTAAAATCACCCAACTCCAAA			205 bp (+705 to +909)	BS1: 61°C
TCBA1	NM_153355	GGGTTTAGTAAYGTAGGGTGTTT	CGTCCTAAAAAATAACAAACATTAAAAC			389 bp (+535 to +923)	BS1: 54°C
LOC254531	NM_153613	GAGGTTAGGTTTTTAGAGAGGAA	ACAAAACAACTACTACTACAACAAC	BS2F: TTAGTGTATTAGGGATTTTTGG	BS2R: CTCTTCCTATAAACTCCTTAACCC	283 bp (+731 to +1014); BS2: 229 bp (+318 to +547)	Step down 56-48°C

☼ Nucleotide count started from -1 of Transcriptional Start Site
* Analyzed by MSP
ϕ Failed to amplify by multiple attempts

Supplementary table 3: Primers and probes used for QMSP

Gene	Forward 5´-3´	Probe 5´-3´	Reverse 5´-3´	Genbank #	Amplicon size (Nucleotide range)	Annealing temperature
ACTB	TGGTGATGGAGGAGGTTTAGTAAGT (390-414)	ACCACCACCCAACACACAATAACAAACACA (432-461)	AACCAATAAAACCTACTCCTCCCTTAA (496-522)	Y00474	133 bp (390–522)	60 ºC
OGDHL	TCGTTAGTATCGTGGATAGC (881-900)	CGCCGTACCAATTACCTAAATCAC (983-1003)	TACAAATCAAAAAACTACGCG (1022-1042)	NM_018245	161 bp (881-1042)	55 ºC
PAK3	TTACGGTCGTCGTTATTATCG (934-955)	AACCAAAAAAAAATAAAAAATCACAACCG (954-982)	ACCGAAAATTCT ACC CTTCG (1045-1065)	NM_002578	131 bp (934-1065)	60 ºC
NISCH	TTTTTTTCGTATAGAGTTCGT (361-371)	CGCGACCCAACACGCAATAATACTC (421-446)	CTA AAC CTC TCT AAAATTCG (497-517)	NM_007184	156 bp (361-517)	56 ºC
KIF1A	GCGCGATAAATTAGTTGGCGATT (870-892)	CCTCCCGAAACGCTAATTAACTACGCG (908-934)	CTCGAC GACTACTCTACGCTAT (989-1010)	NM_004321	140 bp (870-1010)	58 ºC
OSMR	TTCGTGCGTTTTTGGTCG (329-347)	AAACCGAAAACTCGACGACCG (275-293)	CGAACTTTACGAACGAACG (205-224)	NM_003999	142 bp (224-329)	58 ºC
B4GALT1	TAGGAAACGGGTTTCGACG (583-601)	CGTTAAACAACGAAATCCAACCGAA (645-669)	CCGTCCACTTTCTTTACCG (669-688)	NM_001497	116 bp (583-688)	58 ºC
MCAM	AGAATTTAGGTCGGTTTTTATCG (448-471)	ACAATATCAAACCGACGACAACGAC (492-517)	ACGCAAAATTCTTCTCCCAAAA (537-559)	NM_006500	113 bp (448-559)	60 ºC
SSBP2	ATTTTTGCGGTCGTAGCGGT (694-714)	ATATCCAAAACGCCGCGAAACTCC (737-762)	TTCTACGACAAATCTAACGAA (763-784)	NM_012446	89 bp (694-784)	60 ºC

Supplementary table 4: Cancer-specific methylated genes and their proposed functions

Gene symbol	Locus	Name	Proposed function	Accession #
PAK3	Xq22.3-q23	cyclin-dependent kinase inhibitor 1A (p21, Cip1) ACTIVATED KINASE 3	PAK proteins are critical effectors that link Rho GTPases to cytoskeleton reorganization and nuclear signaling Point mutation in this gene has been linked to nonsyndromic X-linked mental retardation	NM_002578
NISCH	3p21.1	nischarin	Human I(1)-imidazoline receptor candidate gene, IRAS hIRAS expression in PC12 cells resulted in protection against apoptosis	NM_007184
KIF1A	2q37.3	kinesin family member 1A	Anterograde motor protein that transports membranous organelles along axonal microtubules May play a critical role in the development of axonal neuropathies	NM_004321
OGDHL	10q11.23	oxoglutarate dehydrogenase-like	oxoglutarate dehydrogenase plays a major role in the citric acid cycle, converting converts alpha-ketoglutarate to succinyl CoA	NM_018245
OSMR	5p12	oncostatin M receptor	The oncostatin M receptor/gp130 ligand murine oncostatin M induces apoptosis in adrenocortical Y-1 tumor cells	NM_003999
B4GALT1	9p21	beta 1,4- galactosyltransferase, polypeptide 1	Elevated beta1,4-galactosyltransferase I was observed in highly metastatic human lung cancer cells Impairment of skin wound healing observed in beta-1,4-galactosyltransferase-deficient mice	NM_001497
MCAM	11q23.3	melanoma cell adhesion molecule	Cell adhesion	NM_006500
SSBP2	5q13.3	single-stranded DNA-binding protein 2	DNA binding	NM_012446

Supplementary table 5: Frequency of methylation in different tissue types based on bisulfite sequencing, COBRA, quantitative MSP, or conventional MSP

Cancer type	No	Gene	Cell lines	Tumor tissues	Normal Tissues	Computational method	Validation method
				# of methylation positive / # of total cases (%)			
COLON	1	B4GALT1†	M	5/5 (100%)	0/5 (0%)	DEEP	C-MSP
	2	C10orf119	M	0/10 (0%)	0/10 (0%)	DEEP	SEQ
	3	COPS4	M	0/5 (0%)	0/5 (0%)	BROAD	SEQ
	4	CSRP1	M	3/6 (50%)	5/5 (100%)	DEEP	SEQ
	5	DARS	M	2/9 (22%)	4/10 (25%)	BROAD	SEQ
	6	FKBP14	M	0/5 (0%)	0/4 (0%)	BROAD	SEQ
	7	FN3KRP	M	5/5 (100%)	5/5 (100%)	DEEP	SEQ
	8	FLJ20277	M	5/5 (100%)	5/5 (100%)	DEEP	SEQ
	9	HUS1	M	2/9 (22%)	0/5 (0%)	BROAD	SEQ
	10	KLF11	M	6/6 (100%)	6/6 (100%)	DEEP	SEQ
	11	MYBL2	M	9/9 (100%)	5/5 (100%)	DEEP	SEQ
	12	MRPL4	M	5/5 (100%)	5/5 (100%)	BROAD	SEQ
	13	MYLK	M	1/10 (10%)	0/10 (0%)	BROAD	SEQ
	14	OSMR†	M	5/5 (100%)	0/5 (0%)	DEEP	C-MSP
	15	PAPSS2†	M	5/5 (100%)	0/5 (0%)	BROAD	SEQ, C-MSP
	16	RBMS2	M	10/10 (100%)	10/10 (100%)	BROAD	SEQ
	17	SECTM1	M	0/5 (0%)	0/5 (0%)	DEEP	SEQ
	18	SIRT7	M	4/4 (100%)	3/3 (100%)	BROAD	SEQ
	19	SLC39A4	M	6/9 (67%)	10/10 (100%)	BROAD	SEQ, C-MSP
	20	SLC9A3R1	M	5/5 (100%)	5/5 (100%)	DEEP	C-MSP
	21	TUBG2†	M	4/7 (60%)	0/5 (0%)	BROAD	COBRA, Q-MSP
	22	NTRK2†	M	5/5 (100%)	0/5 (0%)	KNOWN	C-MSP
	23	SFRP4†	M	5/5 (100%)	0/5 (0%)	KNOWN	C-MSP

Cancer type	No	Gene	Cell lines	Tumor tissues	Normal Tissues	Computational method	Validation method
LUNG	1	PAK3†	M	4/5 (80%)	1/6 (16%)	BROAD	SEQ
	2	NISCH†	M	7/9 (78%)	1/5 (20%)	BROAD	SEQ
	3	TUBB4†	M	7/9 (78%)	1/5 (20%)	BROAD	SEQ
	4	PIGH†	M	4/4 (100%)	1/5 (20%)	BROAD	SEQ
	5	PHKA2	M	3/17 (18%)	1/5 (20%)	DEEP	SEQ
	6	CBR3	M	0/15 (0%)	0/5 (0%)	BROAD	SEQ
	7	CAMK4	M	1/11 (9%)	1/8 (12.5%)	DEEP	SEQ
	8	HOXB5	M	3/8 (37.5%)	0/7 (0%)	BROAD	SEQ
	9	ZNF198	M	1/13 (8%)	0/5 (0%)	BROAD	SEQ
	10	RGS4	M	1/12 (8%)	0/5 (0%)	BROAD	SEQ
	11	RBM15B	M	15/15 (100%)	9/9 (100%)	DEEP	SEQ
	12	PDLIM3	M	2/15 (13%)	1/5 (20%)	BROAD	SEQ
	13	ARFGAP3	M	15/15 (100%)	9/9 (100%)	BROAD	SEQ
	14	VGF	M	0/5 (0%)	0/5 (0%)	BROAD	SEQ
	15	EDG4	M	1/15 (6%)	1/5 (20%)	BROAD	SEQ
BREAST	1	KIF1A†	M	8/9 (89%)	1/6 (17%)	BROAD	SEQ
	2	MAL†	M	3/6 (50%)	0/6 (0%)	DEEP	SEQ
	3	GALE	M	*	3/3 (100%)	BROAD	SEQ
	4	HMG20B	M	*	3/3 (100%)	BROAD	SEQ
	5	KRT14	M	*	3/3 (100%)	BROAD	SEQ
	6	CKMT	M	0/6 (0%)	0/3 (0%)	BROAD	SEQ

Cancer type	No	Gene	Cell lines	Tumor tissues	Normal Tissues	Computational method	Validation method
PROSTATE	1	NDP	M	5/5 (100%)	3/6 (50%)	BROAD	SEQ
	2	APOC1	M	5/5 (100%)	6/6 (100%)	KNOWN	SEQ
	3	ENPEP†	M	4/5 (80%)	1/6 (17%)	BROAD	SEQ
	4	NBL1	M	5/5 (100%)	6/6 (100%)	BROAD	SEQ
	5	PLTP	M	5/5 (100%)	6/6 (100%)	BROAD	SEQ
	6	MCAM†	M	4/5 (80%)	0/6 (0%)	DEEP	SEQ
	7	SSBP2†	M	4/5 (80%)	0/6 (0%)	DEEP	SEQ
	8	MGC15396	M	5/5 (100%)	6/6 (100%)	DEEP	SEQ
	9	ATXN1	M	0/5 (0%)	0/6 (0%)	BROAD	SEQ
	10	CD3D	M	0/5 (0%)	0/6 (0%)	DEEP	SEQ
	11	FKBP4	M	0/5 (0%)	0/6 (0%)	DEEP	SEQ
	12	TFAP2A	M	0/5 (0%)	0/6 (0%)	KNOWN	SEQ
	13	ING4	M	0/5 (0%)	0/6 (0%)	BROAD	SEQ
	14	LEF1	M	0/5 (0%)	0/6 (0%)	BROAD	SEQ
	15	CENTD3	M	0/5 (0%)	0/6 (0%)	BROAD	SEQ
	16	BMP2	M	1/5 (20%)	0/6 (0%)	DEEP	SEQ, C-MSP

Cancer type	No	Gene	Cell lines	Tumor tissues	Normal Tissues	Computational method	Validation method
CERVIX	1	PTGS2†	U	4/10 (40%)	0/2 (0%)	KNOWN	SEQ
	2	NPTX1†	M	2/7 (29%)	0/5* (0%)	KNOWN	COBRA
	3	CCNA1†	M	7/9 (78%)	0/5* (0%)	KNOWN	COBRA
	4	ASMTL†	M	5/9 (56%)	0/2 (0%)	BROAD	SEQ
	5	TFPI2†	M	8/10 (80%)	0/1 (0%)	KNOWN	SEQ
	6	OGDHL†	M	9/10 (90%)	0/2 (0%)	BROAD	SEQ
	7	ARMC7†	M	6/10 (60%)	0/2 (0%)	BROAD	SEQ
	8	C13orf18†	M	4/5 (80%)	0/1 (0%)	DEEP	SEQ
	9	DLL4†	U	6/6 (100%)	0/2 (0%)	DEEP	SEQ
	10	C9orf19†	M	6/9 (67%)	0/2 (0%)	DEEP	SEQ
	11	HCP1†	M	9/9 (100%)	0/2 (0%)	BROAD	SEQ
	12	GDAP1L1†	M	4/8 (50%)	0/2 (0%)	BROAD	SEQ
	13	C1ORF166†	U	10/10 (100%)	0/2 (0%)	DEEP	SEQ
	14	CDK6	M	0/10 (0%)	0/2 (0%)	BROAD	SEQ
	15	PEX13	M	0/10 (0%)	0/1 (0%)	BROAD	SEQ
	16	DDX28	M	0/10 (0%)	0/2 (0%)	DEEP	SEQ
	17	PLCG2	M	0/10 (0%)	0/2 (0%)	DEEP	SEQ
	18	CLSTN2	M	0/10 (0%)	0/2 (0%)	DEEP	SEQ
	19	RNF123	M	0/10 (0%)	0/2 (0%)	BROAD	SEQ
	20	DDX54	M	0/10 (0%)	0/2 (0%)	BROAD	SEQ
	21	MGC35558 (C11ORF45)	M	0/10 (0%)	0/2 (0%)	BROAD	SEQ
	22	TCBA1	M	0/10 (0%)	0/2 (0%)	DEEP	SEQ
	23	GNB4	M	0/10 (0%)	1/2 (50%)	DEEP	SEQ
	24	CLU	M	10/10 (100%)	2/2 (100%)	BROAD	SEQ
	25	CACNA1C	M	10/10 (100%)	5/5 (100%)	BROAD	COBRA

Supplementary data for Chapter 17: Predicting platinum response in ovarian cancer, using DNA methylation profiling

Supplementary table 1: Ranking of methylation candidates in ovarian cancer in general and the different ranking parameters

Unpublished and under validation – hidden data

Supplementary table 2: Ranking of methylation candidates in platinum-resistant ovarian cancer (methylated in platinum-resistant cell lines)

Unpublished and under validation – hidden data

Supplementary table 3: Ranking of methylation candidates in platinum-sensitive ovarian cancer (methylated in platinum-sensitive cell lines)

Unpublished and under validation – hidden data

Supplementary data for Chapter 20: Cervical cancer and the HPV family of viruses

Supplementary table 1: MSDK analysis results: tags with a unique location in the human chromosome, and the tag counts and statistics between stages. Green notes possible demethylation while red indicates possible methylation in the later stage

Unpublished and under validation – hidden data

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	4	16	49	0,03989	1	2,53E-11	1	8,5E-08	1	xxx	xxx
xxx	618	819	442	0,966941	-1	2,92E-07	-1	6,18E-08	-1	xxx	xxx
xxx	29	25	2	0,111022	-1	3,15E-07	-1	8,42E-05	-1	xxx	xxx
xxx	8	8	41	0,553711	-1	8,57E-07	1	3,74E-09	1	xxx	xxx
xxx	29	47	76	0,41125	1	2,23E-06	1	1,8E-05	1	xxx	xxx
xxx	19	32	54	0,42823	1	2,43E-05	1	0,000162	1	xxx	xxx
xxx	16	15	47	0,325383	-1	5,77E-05	1	1,16E-07	1	xxx	xxx
xxx	12	23	38	0,318164	1	0,000155	1	0,001949	1	xxx	xxx
xxx	17	42	46	0,028252	1	0,000172	1	0,064862	1	xxx	xxx
xxx	2	6	17	0,360524	1	0,000349	1	0,002578	1	xxx	xxx
xxx	20	39	48	0,166603	1	0,000486	1	0,017601	1	xxx	xxx
xxx	9	16	30	0,512725	1	0,000555	1	0,002029	1	xxx	xxx
xxx	0	8	10	0,012786	1	0,000886	1	0,262202	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	2	8	15	0,168403	1	0,001157	1	0,029598	1	xxx	xxx
xxx	15	36	38	0,051999	1	0,001197	1	0,124271	1	xxx	xxx
xxx	5	3	21	0,272123	-1	0,001296	1	6,52E-06	1	xxx	xxx
xxx	3	7	17	0,461617	1	0,0013	1	0,005546	1	xxx	xxx
xxx	27	43	55	0,473856	1	0,001471	1	0,006805	1	xxx	xxx
xxx	16	26	39	0,547079	1	0,001473	1	0,004682	1	xxx	xxx
xxx	0	5	9	0,068892	1	0,001788	1	0,104571	1	xxx	xxx
xxx	9	15	0	0,622584	1	0,002132	-1	0,000283	-1	xxx	xxx
xxx	9	9	0	0,5319	-1	0,002132	-1	0,007865	-1	xxx	xxx
xxx	4	11	18	0,227562	1	0,002267	1	0,034947	1	xxx	xxx
xxx	3	6	16	0,622215	1	0,002269	1	0,004613	1	xxx	xxx
xxx	1	4	11	0,380193	1	0,003096	1	0,017615	1	xxx	xxx
xxx	6	11	21	0,563795	1	0,003209	1	0,008885	1	xxx	xxx
xxx	5	14	19	0,1576	1	0,003552	1	0,081897	1	xxx	xxx
xxx	5	14	19	0,1576	1	0,003552	1	0,081897	1	xxx	xxx
xxx	0	4	8	0,120776	1	0,003607	1	0,096788	1	xxx	xxx
xxx	-	-	8	-	-	0,003607	1	0,00091	1	xxx	xxx
xxx	0	7	8	0,022415	1	0,003607	1	0,389989	1	xxx	xxx
xxx	17	26	38	0,674275	1	0,003708	1	0,006798	1	xxx	xxx
xxx	13	30	32	0,096189	1	0,003759	1	0,147114	1	xxx	xxx
xxx	13	30	31	0,096189	1	0,005526	1	0,188394	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	1	7	10	0,099451	1	0,005801	1	0,176477	1	xxx	xxx
xxx	1	8	10	0,062221	1	0,005801	1	0,262202	1	xxx	xxx
xxx	1	3	10	0,575564	1	0,005801	1	0,013728	1	xxx	xxx
xxx	1	3	10	0,575564	1	0,005801	1	0,013728	1	xxx	xxx
xxx	5	17	18	0,058836	1	0,005809	1	0,282526	1	xxx	xxx
xxx	2	2	12	0,73942	-1	0,006682	1	0,001147	1	xxx	xxx
xxx	2	1	12	0,429883	-1	0,006682	1	0,000252	1	xxx	xxx
xxx	3	12	14	0,080472	1	0,006753	1	0,241005	1	xxx	xxx
xxx	10	6	1	0,118615	-1	0,006939	-1	0,164749	-1	xxx	xxx
xxx	7	13	21	0,507299	1	0,007069	1	0,024506	1	xxx	xxx
xxx	19	34	39	0,309258	1	0,007075	1	0,060278	1	xxx	xxx
xxx	-	-	7	-	-	0,007278	1	0,00214	1	xxx	xxx
xxx	0	4	7	0,120776	1	0,007278	1	0,163204	1	xxx	xxx
xxx	0	5	7	0,068892	1	0,007278	1	0,269469	1	xxx	xxx
xxx	0	3	7	0,211733	1	0,007278	1	0,086164	1	xxx	xxx
xxx	0	1	7	0,650743	1	0,007278	1	0,011977	1	xxx	xxx
xxx	0	2	7	0,371193	1	0,007278	1	0,037421	1	xxx	xxx
xxx	7	5	0	0,283728	-1	0,008381	-1	0,07208	-1	xxx	xxx
xxx	7	11	0	0,767695	1	0,008381	-1	0,002598	-1	xxx	xxx
xxx	14	4	3	0,003124	-1	0,008403	-1	0,930071	1	xxx	xxx
xxx	7	21	20	0,057704	1	0,010998	1	0,409034	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	3	7	13	0,461617	1	0,0115	1	0,044686	1	xxx	xxx
xxx	9	9	23	0,5319	-1	0,011744	1	0,000866	1	xxx	xxx
xxx	2	4	11	0,71452	1	0,011809	1	0,017615	1	xxx	xxx
xxx	8	10	21	0,867601	-1	0,014127	1	0,004937	1	xxx	xxx
xxx	10	30	24	0,021613	1	0,014458	1	0,757637	1	xxx	xxx
xxx	0	3	6	0,211733	1	0,014685	1	0,152039	1	xxx	xxx
xxx	0	7	6	0,022415	1	0,014685	1	0,75946	1	xxx	xxx
xxx	0	1	6	0,650743	1	0,014685	1	0,025273	1	xxx	xxx
xxx	-	-	6	-	-	0,014685	1	0,005031	1	xxx	xxx
xxx	0	1	6	0,650743	1	0,014685	1	0,025273	1	xxx	xxx
xxx	0	1	6	0,650743	1	0,014685	1	0,025273	1	xxx	xxx
xxx	0	3	6	0,211733	1	0,014685	1	0,152039	1	xxx	xxx
xxx	0	4	6	0,120776	1	0,014685	1	0,267319	1	xxx	xxx
xxx	0	1	6	0,650743	1	0,014685	1	0,025273	1	xxx	xxx
xxx	0	4	6	0,120776	1	0,014685	1	0,267319	1	xxx	xxx
xxx	0	2	6	0,371193	1	0,014685	1	0,071806	1	xxx	xxx
xxx	-	-	6	-	-	0,014685	1	0,005031	1	xxx	xxx
xxx	0	1	6	0,650743	1	0,014685	1	0,025273	1	xxx	xxx
xxx	5	9	16	0,631075	1	0,015107	1	0,031687	1	xxx	xxx
xxx	6	6	0	0,602911	-1	0,016616	-1	0,041427	-1	xxx	xxx
xxx	4	8	14	0,551205	1	0,017339	1	0,047655	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	4	3	14	0,445595	-1	0,017339	1	0,00096	1	xxx	xxx
xxx	4	10	14	0,310643	1	0,017339	1	0,119262	1	xxx	xxx
xxx	9	14	22	0,745845	1	0,017559	1	0,025308	1	xxx	xxx
xxx	1	2	8	0,849571	1	0,019981	1	0,019142	1	xxx	xxx
xxx	1	1	8	0,790156	-1	0,019981	1	0,005617	1	xxx	xxx
xxx	6	11	17	0,563795	1	0,020229	1	0,053551	1	xxx	xxx
xxx	2	7	10	0,24837	1	0,020677	1	0,176477	1	xxx	xxx
xxx	8	8	1	0,553711	-1	0,023083	-1	0,066058	-1	xxx	xxx
xxx	13	27	27	0,191968	1	0,023812	1	0,262804	1	xxx	xxx
xxx	5	7	15	0,97327	1	0,023983	1	0,016237	1	xxx	xxx
xxx	5	15	15	0,114684	1	0,023983	1	0,398861	1	xxx	xxx
xxx	22	38	39	0,338249	1	0,025144	1	0,148447	1	xxx	xxx
xxx	4	8	13	0,551205	1	0,028113	1	0,075347	1	xxx	xxx
xxx	0	3	5	0,211733	1	0,029632	1	0,260469	1	xxx	xxx
xxx	0	10	5	0,00416	1	0,029632	1	0,515836	-1	xxx	xxx
xxx	0	1	5	0,650743	1	0,029632	1	0,052628	1	xxx	xxx
xxx	0	2	5	0,371193	1	0,029632	1	0,134694	1	xxx	xxx
xxx	0	1	5	0,650743	1	0,029632	1	0,052628	1	xxx	xxx
xxx	0	2	5	0,371193	1	0,029632	1	0,134694	1	xxx	xxx
xxx	0	3	5	0,211733	1	0,029632	1	0,260469	1	xxx	xxx
xxx	0	1	5	0,650743	1	0,029632	1	0,052628	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	-	-	5	-	-	0,029632	1	0,011831	1	xxx	xxx
xxx	-	-	5	-	-	0,029632	1	0,011831	1	xxx	xxx
xxx	0	2	5	0,371193	1	0,029632	1	0,134694	1	xxx	xxx
xxx	0	3	5	0,211733	1	0,029632	1	0,260469	1	xxx	xxx
xxx	0	4	5	0,120776	1	0,029632	1	0,423115	1	xxx	xxx
xxx	0	1	5	0,650743	1	0,029632	1	0,052628	1	xxx	xxx
xxx	-	-	5	-	-	0,029632	1	0,011831	1	xxx	xxx
xxx	-	-	5	-	-	0,029632	1	0,011831	1	xxx	xxx
xxx	0	7	5	0,022415	1	0,029632	1	0,998857	-1	xxx	xxx
xxx	0	6	5	0,039297	1	0,029632	1	0,807059	1	xxx	xxx
xxx	0	4	5	0,120776	1	0,029632	1	0,423115	1	xxx	xxx
xxx	0	1	5	0,650743	1	0,029632	1	0,052628	1	xxx	xxx
xxx	8	21	19	0,100785	1	0,031806	1	0,511952	1	xxx	xxx
xxx	3	4	11	0,950361	1	0,032319	1	0,017615	1	xxx	xxx
xxx	5	3	0	0,272123	-1	0,032941	-1	0,218216	-1	xxx	xxx
xxx	5	1	0	0,05559	-1	0,032941	-1	0,660631	-1	xxx	xxx
xxx	5	0	0	0,01257	-1	0,032941	-1	-	-	xxx	xxx
xxx	5	1	0	0,05559	-1	0,032941	-1	0,660631	-1	xxx	xxx
xxx	5	1	0	0,05559	-1	0,032941	-1	0,660631	-1	xxx	xxx
xxx	2	6	9	0,360524	1	0,035818	1	0,174372	1	xxx	xxx
xxx	2	3	9	0,966307	1	0,035818	1	0,025803	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	1	2	7	0,849571	1	0,036647	1	0,037421	1	xxx	xxx
xxx	5	4	14	0,439796	-1	0,037629	1	0,002757	1	xxx	xxx
xxx	7	2	1	0,040077	-1	0,04161	-1	0,86409	-1	xxx	xxx
xxx	18	41	32	0,05468	1	0,042093	1	0,80453	1	xxx	xxx
xxx	18	21	32	0,6701	-1	0,042093	1	0,009097	1	xxx	xxx
xxx	10	22	21	0,193294	1	0,044549	1	0,394259	1	xxx	xxx
xxx	4	2	12	0,255522	-1	0,045022	1	0,001147	1	xxx	xxx
xxx	4	6	12	0,903263	1	0,045022	1	0,041089	1	xxx	xxx
xxx	4	6	12	0,903263	1	0,045022	1	0,041089	1	xxx	xxx
xxx	4	5	12	0,885076	-1	0,045022	1	0,021176	1	xxx	xxx
xxx	13	37	25	0,015012	1	0,046825	1	0,7466	-1	xxx	xxx
xxx	20	43	34	0,07438	1	0,049928	1	0,75711	1	xxx	xxx
xxx	3	3	10	0,697978	-1	0,053193	1	0,013728	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	0	9	4	0,007293	1	0,059792	1	0,436534	-1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	0	6	4	0,039297	1	0,059792	1	0,92672	-1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	-	-	4	-	-	0,059792	1	0,02782	1	xxx	xxx
xxx	2	2	8	0,73942	-1	0,061277	1	0,019142	1	xxx	xxx
xxx	2	2	8	0,73942	-1	0,061277	1	0,019142	1	xxx	xxx
xxx	2	1	8	0,429883	-1	0,061277	1	0,005617	1	xxx	xxx
xxx	2	1	8	0,429883	-1	0,061277	1	0,005617	1	xxx	xxx
xxx	2	1	8	0,429883	-1	0,061277	1	0,005617	1	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	4	10	0	0,310643	1	0,065305	-1	0,00452	-1	xxx	xxx
xxx	4	10	0	0,310643	1	0,065305	-1	0,00452	-1	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	4	9	0	0,417454	1	0,065305	-1	0,007865	-1	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	4	14	0	0,082674	1	0,065305	-1	0,000493	-1	xxx	xxx
xxx	4	0	0	0,029261	-1	0,065305	-1	-	-	xxx	xxx
xxx	1	1	6	0,790156	-1	0,066538	1	0,025273	1	xxx	xxx
xxx	1	1	6	0,790156	-1	0,066538	1	0,025273	1	xxx	xxx
xxx	1	1	6	0,790156	-1	0,066538	1	0,025273	1	xxx	xxx
xxx	1	0	6	0,369089	-1	0,066538	1	0,005031	1	xxx	xxx
xxx	13	6	5	0,026987	-1	0,069242	-1	0,807059	1	xxx	xxx
xxx	4	5	11	0,885076	-1	0,071113	1	0,036784	1	xxx	xxx
xxx	6	14	1	0,264891	1	0,074257	-1	0,003639	-1	xxx	xxx
xxx	6	1	1	0,026961	-1	0,074257	-1	0,777477	1	xxx	xxx
xxx	6	9	1	0,860498	1	0,074257	-1	0,04131	-1	xxx	xxx
xxx	3	3	9	0,697978	-1	0,086288	1	0,025803	1	xxx	xxx
xxx	3	2	9	0,445146	-1	0,086288	1	0,009641	1	xxx	xxx
xxx	11	10	20	0,376664	-1	0,09949	1	0,008128	1	xxx	xxx
xxx	11	10	20	0,376664	-1	0,09949	1	0,008128	1	xxx	xxx
xxx	2	1	7	0,429883	-1	0,103309	1	0,011977	1	xxx	xxx
xxx	2	2	7	0,73942	-1	0,103309	1	0,037421	1	xxx	xxx
xxx	2	1	7	0,429883	-1	0,103309	1	0,011977	1	xxx	xxx
xxx	2	12	7	0,031645	1	0,103309	1	0,657162	-1	xxx	xxx
xxx	2	1	7	0,429883	-1	0,103309	1	0,011977	1	xxx	xxx
xxx	7	0	2	0,00232	-1	0,115715	-1	0,153823	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	7	12	2	0,629006	1	0,115715	-1	0,034319	-1	xxx	xxx
xxx	1	0	5	0,369089	-1	0,119313	1	0,011831	1	xxx	xxx
xxx	1	0	5	0,369089	-1	0,119313	1	0,011831	1	xxx	xxx
xxx	1	0	5	0,369089	-1	0,119313	1	0,011831	1	xxx	xxx
xxx	1	0	5	0,369089	-1	0,119313	1	0,011831	1	xxx	xxx
xxx	0	10	3	0,00416	1	0,120648	1	0,179042	-1	xxx	xxx
xxx	0	6	3	0,039297	1	0,120648	1	0,642171	-1	xxx	xxx
xxx	20	35	11	0,337281	1	0,121761	-1	0,010079	-1	xxx	xxx
xxx	7	26	14	0,010985	1	0,12337	1	0,354417	-1	xxx	xxx
xxx	3	6	0	0,622215	1	0,129469	-1	0,041427	-1	xxx	xxx
xxx	3	8	0	0,335656	1	0,129469	-1	0,013684	-1	xxx	xxx
xxx	5	0	1	0,01257	-1	0,130891	-1	0,361708	1	xxx	xxx
xxx	2	1	6	0,429883	-1	0,171158	1	0,025273	1	xxx	xxx
xxx	2	1	6	0,429883	-1	0,171158	1	0,025273	1	xxx	xxx
xxx	2	14	6	0,013032	1	0,171158	1	0,282335	-1	xxx	xxx
xxx	2	1	6	0,429883	-1	0,171158	1	0,025273	1	xxx	xxx
xxx	7	7	13	0,577271	-1	0,175913	1	0,044686	1	xxx	xxx
xxx	6	1	2	0,026961	-1	0,188524	-1	0,419044	1	xxx	xxx
xxx	9	3	4	0,02801	-1	0,190582	-1	0,430448	1	xxx	xxx
xxx	13	22	7	0,511723	1	0,203305	-1	0,046994	-1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	12	4	0,008913	1	0,21059	1	0,175445	-1	xxx	xxx
xxx	1	10	4	0,023819	1	0,21059	1	0,327031	-1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	1	0	4	0,369089	-1	0,21059	1	0,02782	1	xxx	xxx
xxx	3	2	7	0,445146	-1	0,215393	1	0,037421	1	xxx	xxx
xxx	4	9	1	0,417454	1	0,227129	-1	0,04131	-1	xxx	xxx
xxx	4	0	1	0,029261	-1	0,227129	-1	0,361708	1	xxx	xxx
xxx	4	11	1	0,227562	1	0,227129	-1	0,015855	-1	xxx	xxx
xxx	4	0	1	0,029261	-1	0,227129	-1	0,361708	1	xxx	xxx
xxx	4	0	1	0,029261	-1	0,227129	-1	0,361708	1	xxx	xxx
xxx	2	14	0	0,013032	1	0,256674	-1	0,000493	-1	xxx	xxx
xxx	2	0	5	0,158556	-1	0,277639	1	0,011831	1	xxx	xxx
xxx	2	0	5	0,158556	-1	0,277639	1	0,011831	1	xxx	xxx
xxx	9	5	14	0,112169	-1	0,287289	1	0,006682	1	xxx	xxx
xxx	5	0	2	0,01257	-1	0,300792	-1	0,153823	1	xxx	xxx
xxx	13	25	8	0,292053	1	0,305523	-1	0,034287	-1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	1	9	3	0,038625	1	0,364072	1	0,252611	-1	xxx	xxx
xxx	13	5	9	0,013361	-1	0,429353	-1	0,104571	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	2	0	4	0,158556	-1	0,438783	1	0,02782	1	xxx	xxx
xxx	4	0	2	0,029261	-1	0,467723	-1	0,153823	1	xxx	xxx
xxx	4	15	2	0,057698	1	0,467723	-1	0,009184	-1	xxx	xxx
xxx	4	0	2	0,029261	-1	0,467723	-1	0,153823	1	xxx	xxx
xxx	4	0	2	0,029261	-1	0,467723	-1	0,153823	1	xxx	xxx
xxx	4	0	2	0,029261	-1	0,467723	-1	0,153823	1	xxx	xxx
xxx	4	0	2	0,029261	-1	0,467723	-1	0,153823	1	xxx	xxx
xxx	0	6	1	0,039297	1	0,491219	1	0,164749	-1	xxx	xxx
xxx	0	7	1	0,022415	1	0,491219	1	0,104812	-1	xxx	xxx
xxx	0	7	1	0,022415	1	0,491219	1	0,104812	-1	xxx	xxx
xxx	0	6	1	0,039297	1	0,491219	1	0,164749	-1	xxx	xxx
xxx	1	7	0	0,099451	1	0,508859	-1	0,023809	-1	xxx	xxx
xxx	1	9	0	0,038625	1	0,508859	-1	0,007865	-1	xxx	xxx
xxx	1	7	0	0,099451	1	0,508859	-1	0,023809	-1	xxx	xxx

Tag	Count p5	Count p7	Count p30	P (5 vs 7)	5 vs 7	P (5 vs 30)	5 vs 30	P (7 vs 30)	7 vs 30	restriction site	gene
xxx	5	0	3	0,01257	-1	0,525329	-1	0,065416	1	xxx	xxx
xxx	5	2	7	0,141478	-1	0,559082	1	0,037421	1	xxx	xxx
xxx	7	1	5	0,012905	-1	0,603419	-1	0,052628	1	xxx	xxx
xxx	8	4	10	0,103442	-1	0,619885	1	0,031701	1	xxx	xxx
xxx	2	9	1	0,112652	1	0,638288	-1	0,04131	-1	xxx	xxx
xxx	4	0	3	0,029261	-1	0,74594	-1	0,065416	1	xxx	xxx
xxx	4	0	3	0,029261	-1	0,74594	-1	0,065416	1	xxx	xxx
xxx	4	0	3	0,029261	-1	0,74594	-1	0,065416	1	xxx	xxx
xxx	5	0	4	0,01257	-1	0,775705	-1	0,02782	1	xxx	xxx
xxx	4	0	4	0,029261	-1	0,978297	1	0,02782	1	xxx	xxx
xxx	4	0	4	0,029261	-1	0,978297	1	0,02782	1	xxx	xxx
xxx	1	10	1	0,023819	1	0,98677	1	0,025664	-1	xxx	xxx
xxx	0	6	-	0,039297	1	-	-	0,041427	-1	xxx	xxx
xxx	0	8	-	0,012786	1	-	-	0,013684	-1	xxx	xxx

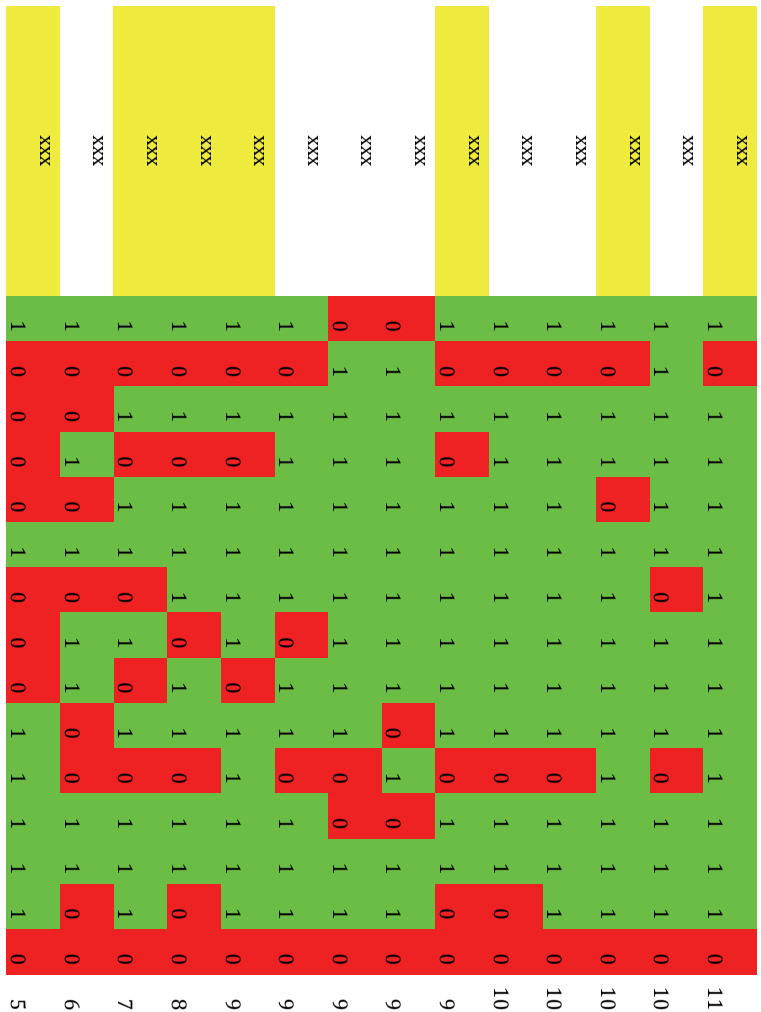
Supplementary table 2: genes, selected for validation in MethyLight experiment

Unpublished and under validation – hidden data

Supplementary table 3: MethyLight results (1=not methylated, 0=methylated)

Unpublished and under validation – hidden data

	CaSki - Cell line(3022)(153661)	EK00-12 - Cell line(3020)(153650)	EK05-2 +HPV16 - Cell line(3020)(153652)	EK07-3 - Cell line(3020)(153651)	FK16A - Cell line(3020)(153653)	FK16A - Cell line(3021)(153654)	FK16B - Cell line(3021)(153655)	FK18A - Cell line(3021)(153656)	FK18B - Cell line(3021)(153657)	FK18B - Cell line(3022)(153658)	HeLa - Cell line(3022)(153660)	SiHa - Cell line(3022)(153659)	CpGenome Universal Methylated DNA M unknown(3019)(155826)	CpGenome Universal Unmethylated DNA U unknown(3019)(155825)	NTC NTC Control(3019)(100000)	
Beta_Actin(3)(50094)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	12
xxx	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	11
xxx	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	11
xxx	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	11



CaSki - Cell line(3022)(153661)

EK00-12 - Cell line(3020)(153650)

EK05-2 +HPV16 - Cell line(3020)(153652)

EK07-3 - Cell line(3020)(153651)

FK16A - Cell line(3020)(153653)

FK16A - Cell line(3021)(153654)

FK16B - Cell line(3021)(153655)

FK18A - Cell line(3021)(153656)

FK18B - Cell line(3021)(153657)

FK18B - Cell line(3022)(153658)

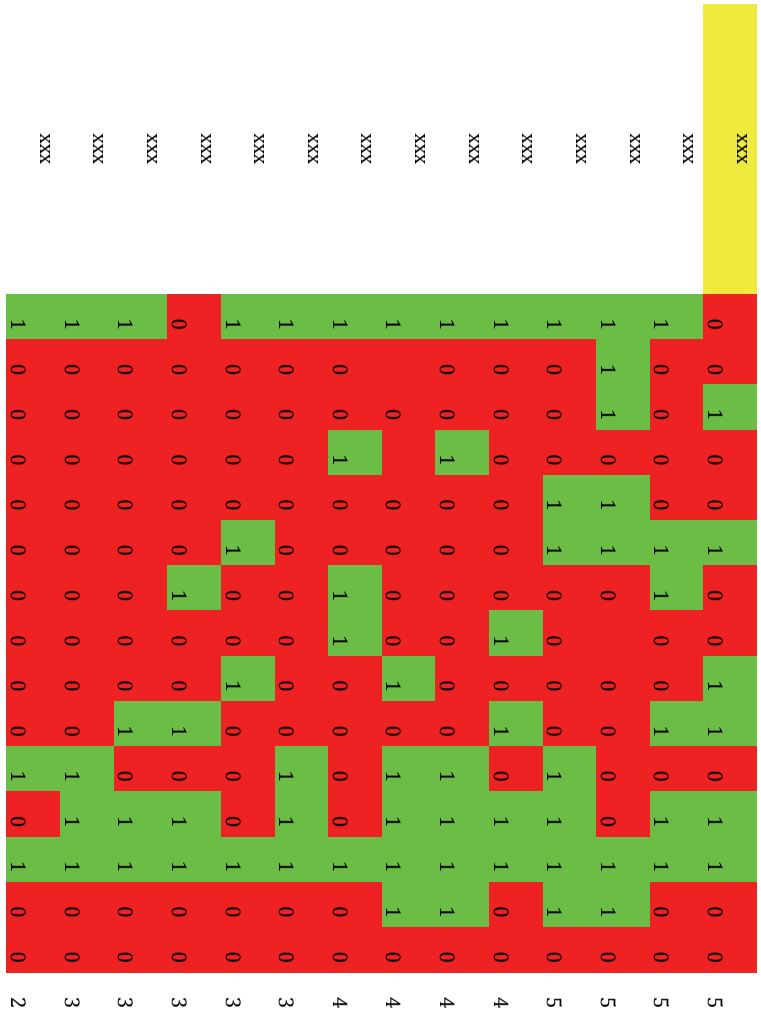
HeLa - Cell line(3022)(153660)

SiHa - Cell line(3022)(153659)

CpGenome Universal Methylated DNA M unknown(3019)(155826)

CpGenome Universal Unmethylated DNA U unknown(3019)(155825)

NTC NTC Control(3019)(100000)



CaSki - Cell line(3022)(153661)

EK00-12 - Cell line(3020)(153650)

EK05-2 +HPV16 - Cell line(3020)(153652)

EK07-3 - Cell line(3020)(153651)

FK16A - Cell line(3020)(153653)

FK16A - Cell line(3021)(153654)

FK16B - Cell line(3021)(153655)

FK18A - Cell line(3021)(153656)

FK18B - Cell line(3021)(153657)

FK18B - Cell line(3022)(153658)

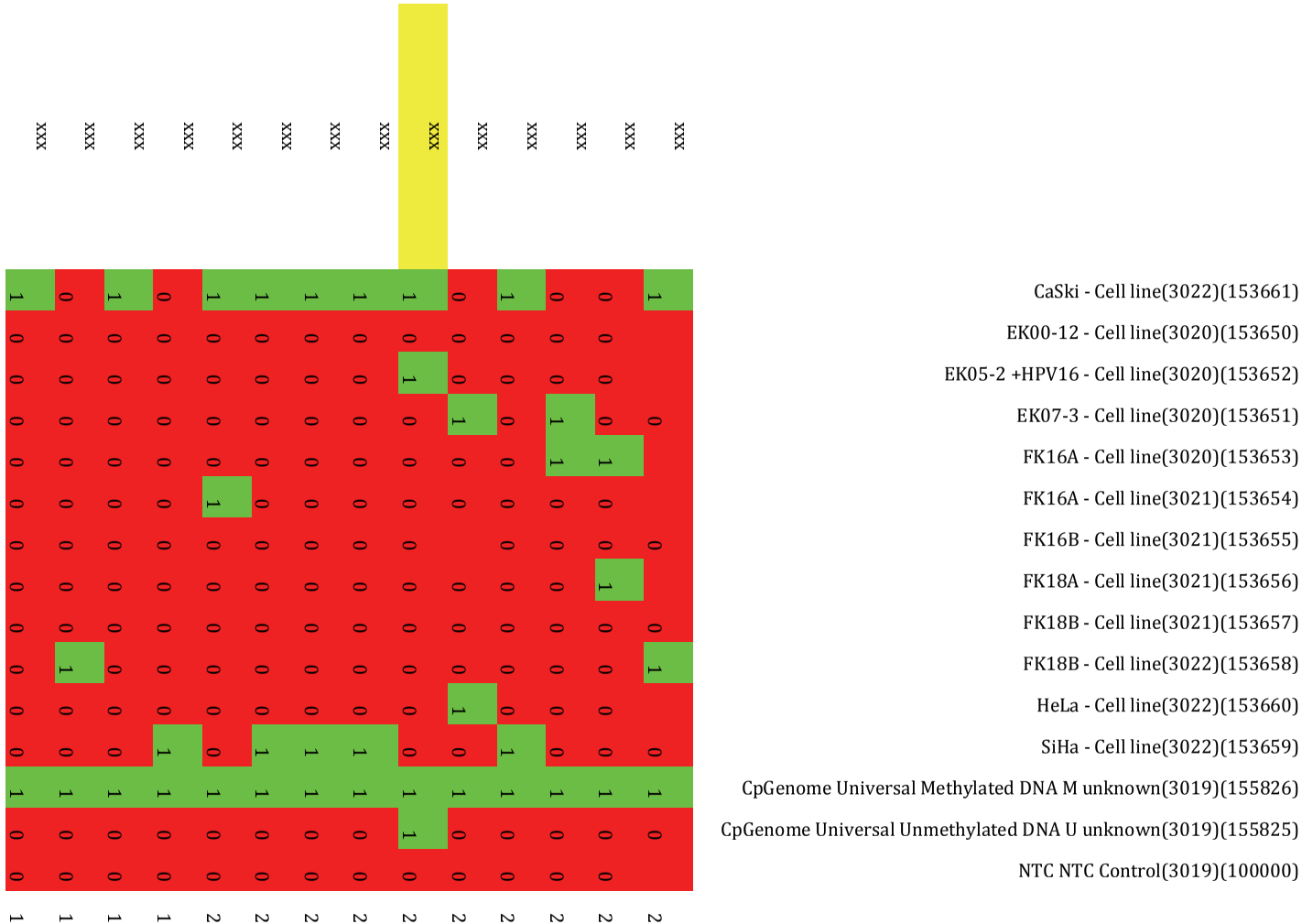
HeLa - Cell line(3022)(153660)

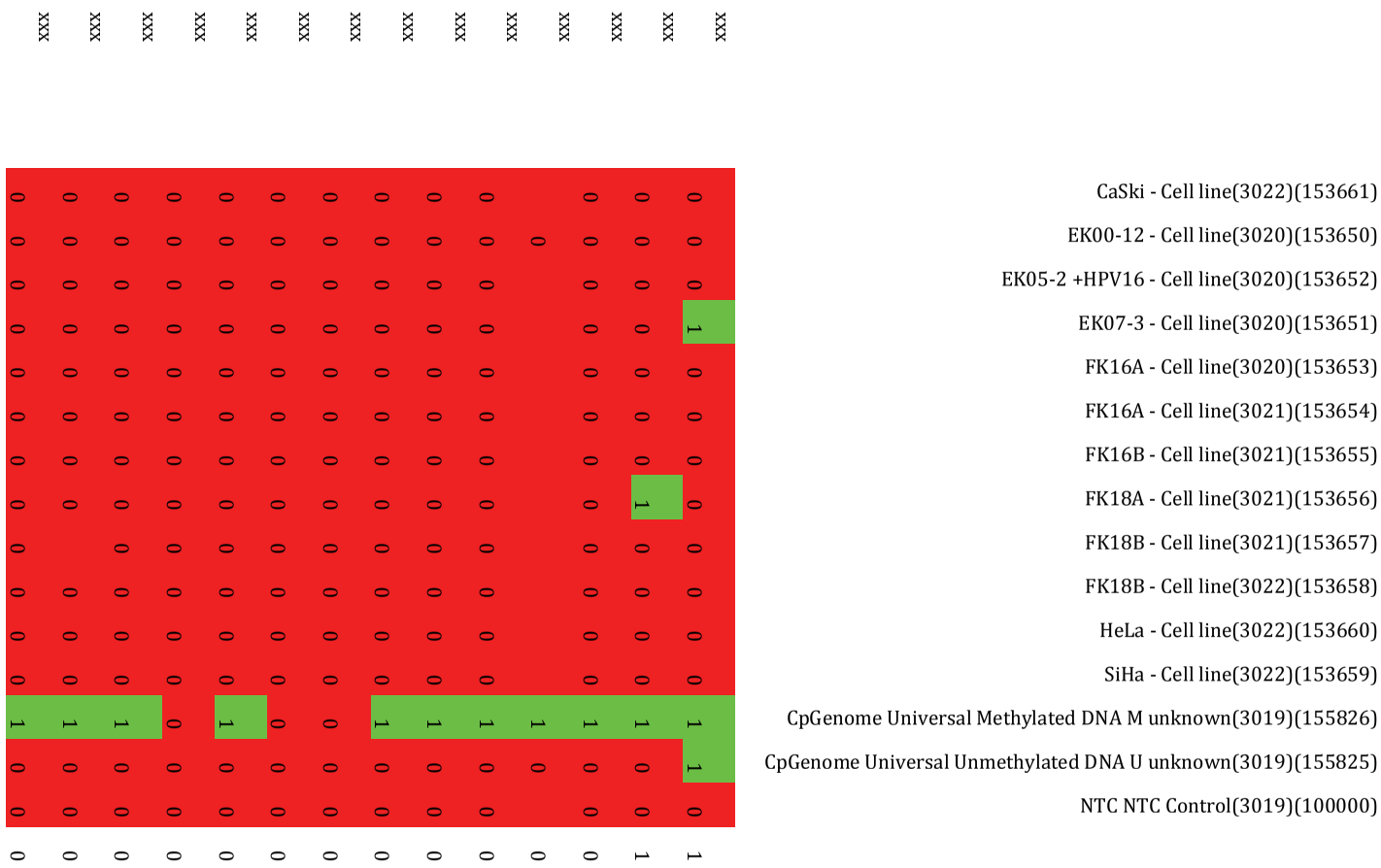
SiHa - Cell line(3022)(153659)

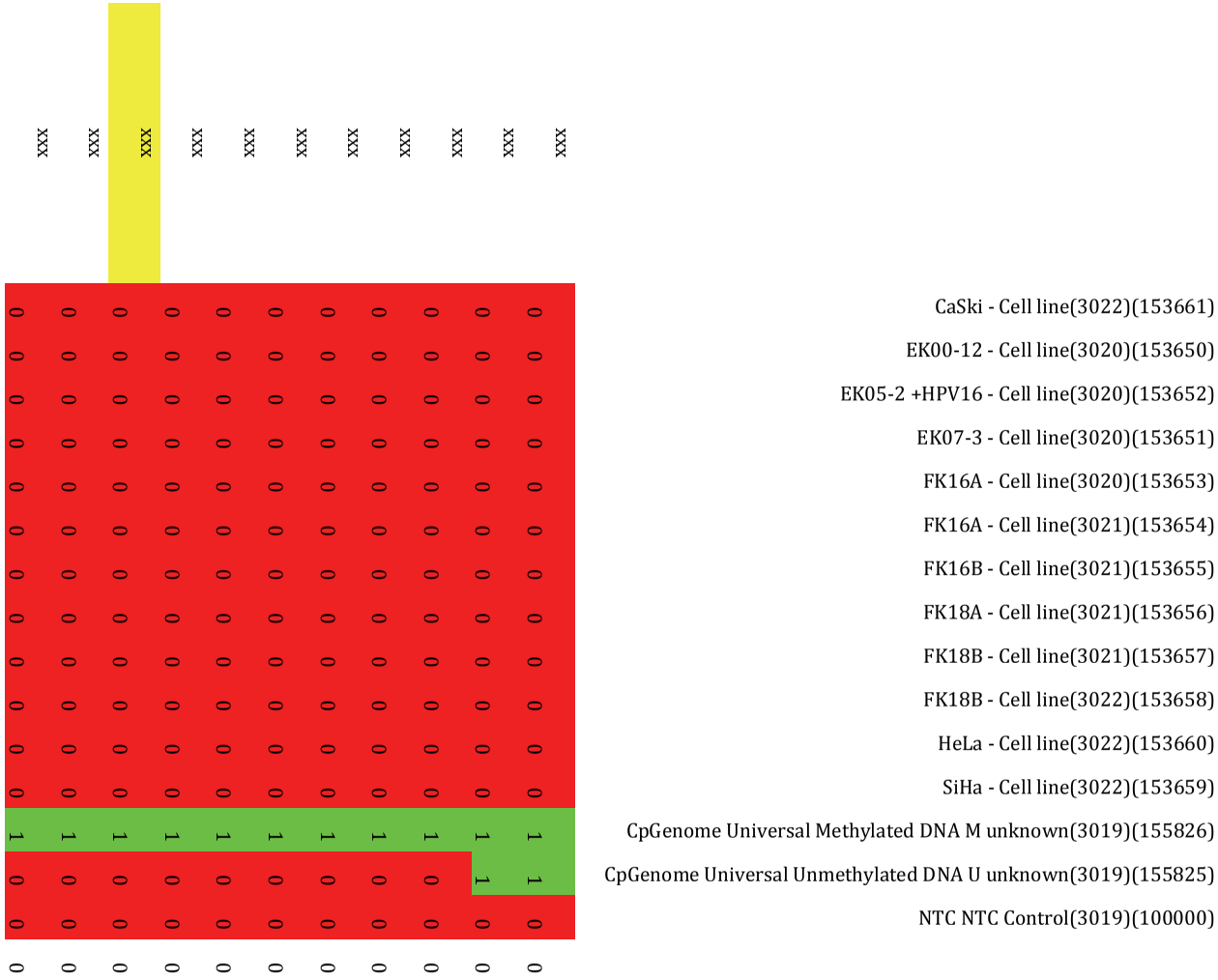
CpGenome Universal Methylated DNA M unknown(3019)(155826)

CpGenome Universal Unmethylated DNA U unknown(3019)(155825)

NTC NTC Control(3019)(100000)

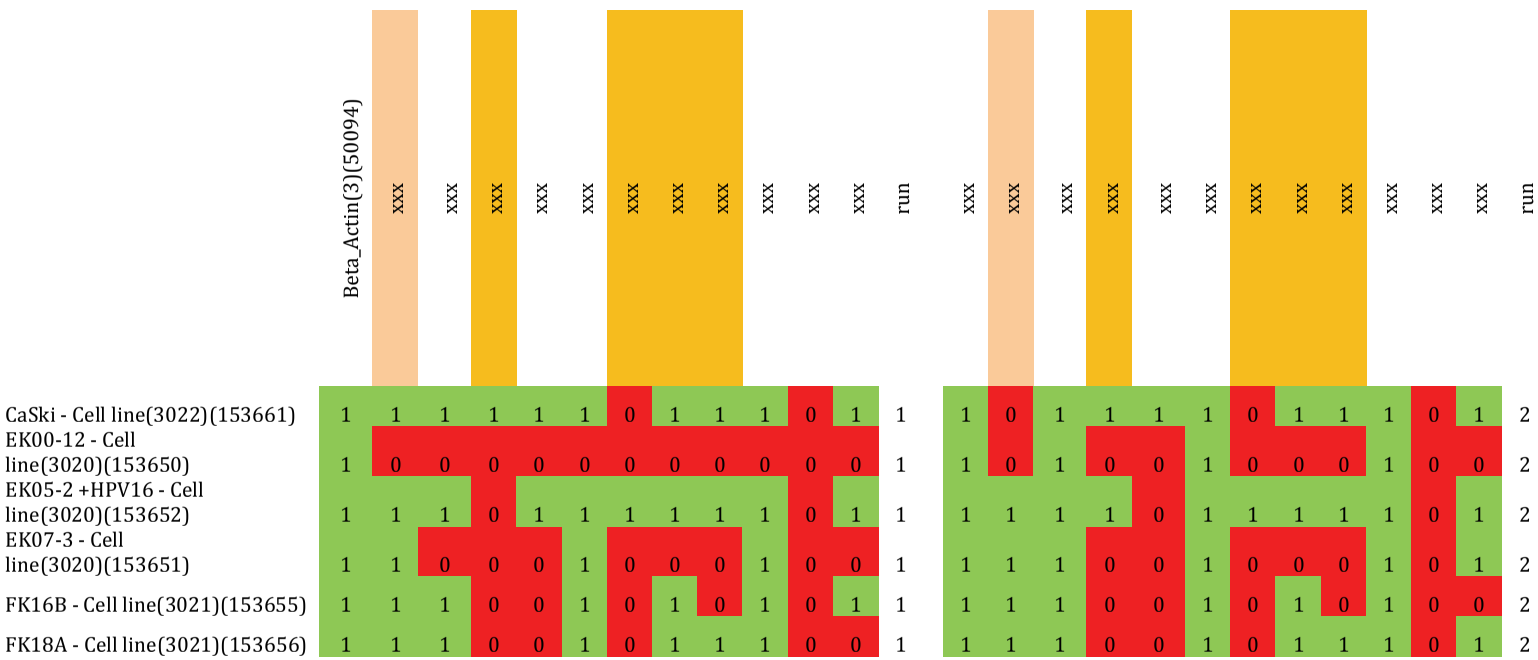

xxx
CaSki - Cell line(3022)(153661)
EK00-12 - Cell line(3020)(153650)
EK05-2 +HPV16 - Cell line(3020)(153652)
EK07-3 - Cell line(3020)(153651)
FK16A - Cell line(3020)(153653)
FK16A - Cell line(3021)(153654)
FK16B - Cell line(3021)(153655)
FK18A - Cell line(3021)(153656)
FK18B - Cell line(3021)(153657)
FK18B - Cell line(3022)(153658)
HeLa - Cell line(3022)(153660)
SiHa - Cell line(3022)(153659)
CpGenome Universal Methylated DNA M unknown(3019)(155826)
CpGenome Universal Unmethylated DNA U unknown(3019)(155825)
NTC NTC Control(3019)(100000)

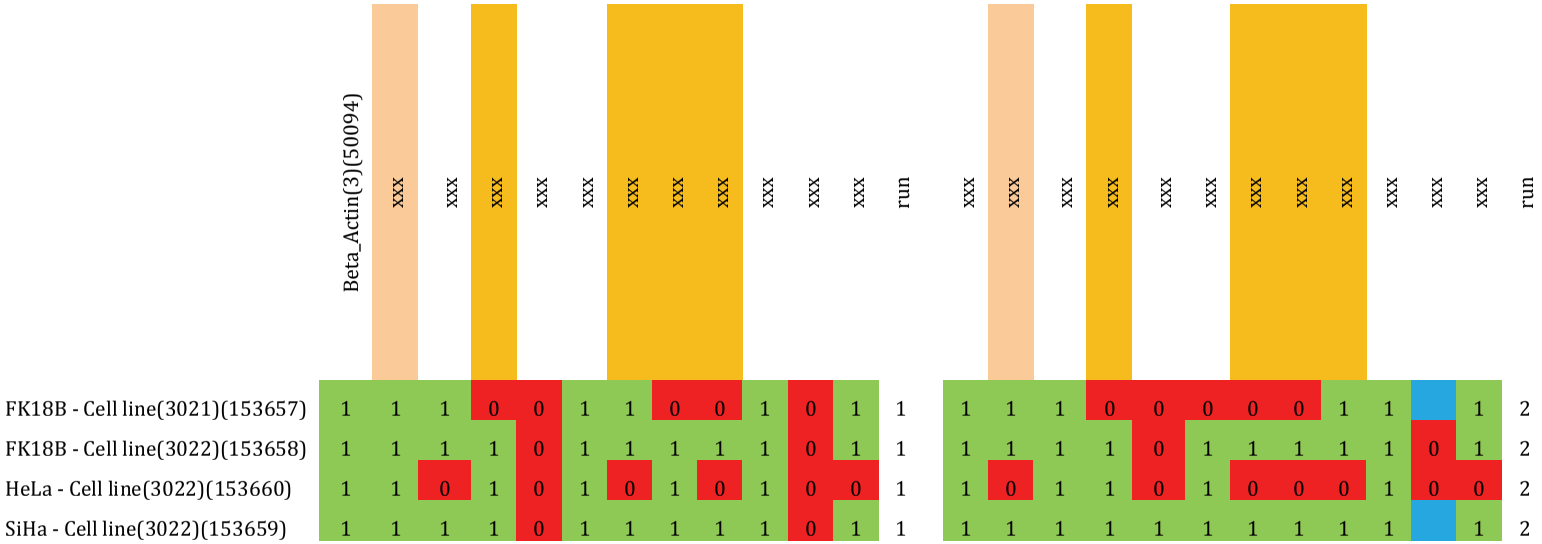

xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx
CaSki - Cell line(3022)(153661)
EK00-12 - Cell line(3020)(153650)
EK05-2 +HPV16 - Cell line(3020)(153652)
EK07-3 - Cell line(3020)(153651)
FK16A - Cell line(3020)(153653)
FK16A - Cell line(3021)(153654)
FK16B - Cell line(3021)(153655)
FK18A - Cell line(3021)(153656)
FK18B - Cell line(3021)(153657)
FK18B - Cell line(3022)(153658)
HeLa - Cell line(3022)(153660)
SiHa - Cell line(3022)(153659)
CpGenome Universal Methylated DNA M unknown(3019)(155826)
CpGenome Universal Unmethylated DNA U unknown(3019)(155825)
NTC NTC Control(3019)(100000)


XXX
XXX
XXX
XXX
XXX
XXX
XXX
XXX
XXX
XXX
XXX
CaSki - Cell line(3022)(153661)
EK00-12 - Cell line(3020)(153650)
EK05-2 +HPV16 - Cell line(3020)(153652)
EK07-3 - Cell line(3020)(153651)
FK16A - Cell line(3020)(153653)
FK16A - Cell line(3021)(153654)
FK16B - Cell line(3021)(153655)
FK18A - Cell line(3021)(153656)
FK18B - Cell line(3021)(153657)
FK18B - Cell line(3022)(153658)
HeLa - Cell line(3022)(153660)
SiHa - Cell line(3022)(153659)
CpGenome Universal Methylated DNA M unknown(3019)(155826)
CpGenome Universal Unmethylated DNA U unknown(3019)(155825)
NTC NTC Control(3019)(100000)

Supplementary table 4: genes, and the interpreted results of the additional validation in MethyLight experiments

Unpublished and under validation – hidden data




Beta_Actin(3)(50094)
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
run
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
xxx
run
FK18B - Cell line(3021)(153657) 1 1 1 0 0 1 1 0 0 1 0 1 1 1 1 1 0 0 0 0 0 1 1 1 2
FK18B - Cell line(3022)(153658) 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 1 2
HeLa - Cell line(3022)(153660) 1 1 0 1 0 1 0 1 0 1 0 0 1 1 0 1 1 0 1 0 0 0 1 0 0 2
SiHa - Cell line(3022)(153659) 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 2