

جزایر CpG در ژنوم: ساختار، نوع، عملکرد و ویژگی‌های تکاملی آنها

مصطفی محقق دولت آبادی^{۱*} و الناز محقق دولت آبادی^۲

^۱ ایران، یاسوج، دانشگاه یاسوج، گروه علوم دامی

^۲ ایران، یاسوج، دانشگاه علوم پزشکی یاسوج، دانشکده پزشکی

تاریخ دریافت: ۱۴۰۲/۱۰/۳۰ تاریخ پذیرش: ۱۴۰۳/۰۲/۲۶

چکیده

در ژنوم پستانداران، محتوای دی‌نوکلئوتیدهای CpG به دلیل دامیناسیون سیتوزین متیله کمتر بوده و این فرآیند که یک علامت سرکوب‌کننده است منجر به خاموشی طولانی مدت ژن می‌شود. با این حال، کاهش CG در تمام مناطق ژنومی رخ نمی‌دهد و ژنوم پستانداران دارای مناطقی بوده که تحت عنوان جزایر CpG (CGIs) شناخته می‌شوند. این جزایر، نسبت به کل ژنوم، از نظر محتوای GC غنی بوده و چگالی بالایی از دی‌نوکلئوتیدهای CpG دارا می‌باشند. جزایر CpG معمولاً در اکثر ژن‌ها به شکل هیپومتیله باقی می‌مانند. همچنین CGI‌ها حضور بالایی در پروموتورها و آگزون‌های اول ژن‌ها دارند و از این رو به عنوان یکی از عناصر تنظیم‌کننده بیان ژن و مکان‌های شروع رونویسی عمل می‌کنند. در واقع، اکثر CGI‌ها در مکان‌های شروع رونویسی یافت شده و حدود ۷۰ درصد از پروموتور ژن‌ها، از جمله تقریباً تمام ژن‌های خانه‌دار، و همچنین برخی از ژن‌های ویژه و تنظیم‌کننده بافتی حاوی جزایر CpG هستند. جزایر CpG معمولاً در اکثر ژن‌ها هیپومتیله باقی می‌مانند. با این حال، متیلاسیون CGI، به عنوان یک علامت اپی‌ژنتیکی وراثتی، می‌تواند در بافت‌های سوماتیکی رخ دهد که منجر به خاموش شدن ژن یا غیرفعال شدن ناحیه متیله شده مانند غیرفعال کردن کروموزوم X و نقش‌پذیری ژنومی شود. علاوه بر این، تفاوت متیلاسیون DNA بین بافت‌ها نیز مشاهده می‌شود که بیانگر نقش مهم متیلاسیون در تنظیم بیان ژن در طول ایجاد انواع متفاوت سلول می‌باشد. بر اساس اطلاعات به دست آمده از ژنوم، تقریباً نیمی از CGI‌های شناسایی شده، که به عنوان CGI‌های ارفان (oCGIs) شناخته می‌شوند، در داخل ژن‌ها (درون ژنی) یا با فاصله از ژن‌ها (بین ژن) قرار دارند. این گروه از CGI‌ها نیز می‌توانند در تنظیم رونویسی نقش اساسی داشته باشند، اما عموماً نسبت به پروموتورهای CGI کوتاه تر بوده و بیشتر مستعد متیله شدن هستند. متیلاسیون نابجای DNA در جزایر CpG در ناحیه پروموتور می‌تواند منجر به از دست دادن عملکرد ژن‌های سرکوبگر تومور در سرطان شود. مشخصات متیلاسیون جزایر CpG نشان می‌دهد که برخی از ژن‌ها متیلاسیون بیشتری نسبت به سایر ژن‌ها نشان می‌دهند و هر نوع تومور می‌تواند به مجموعه منحصر به فردی از ژن‌های متیله مربوط باشد. در این مقاله سعی خواهد شد روند تکاملی مربوط به حفظ CGI‌ها در موجودات مختلف و همچنین ویژگی‌های ساختاری، تنظیمی و عملکردی آن مورد بحث قرار گیرد.

واژه‌های کلیدی: بیان ژن، پروموتور، جزایر CpG، متیلاسیون، ویژگی تکاملی

* نویسنده مسئول، پست الکترونیکی: mmuhaghegh@yu.ac.ir

مقدمه

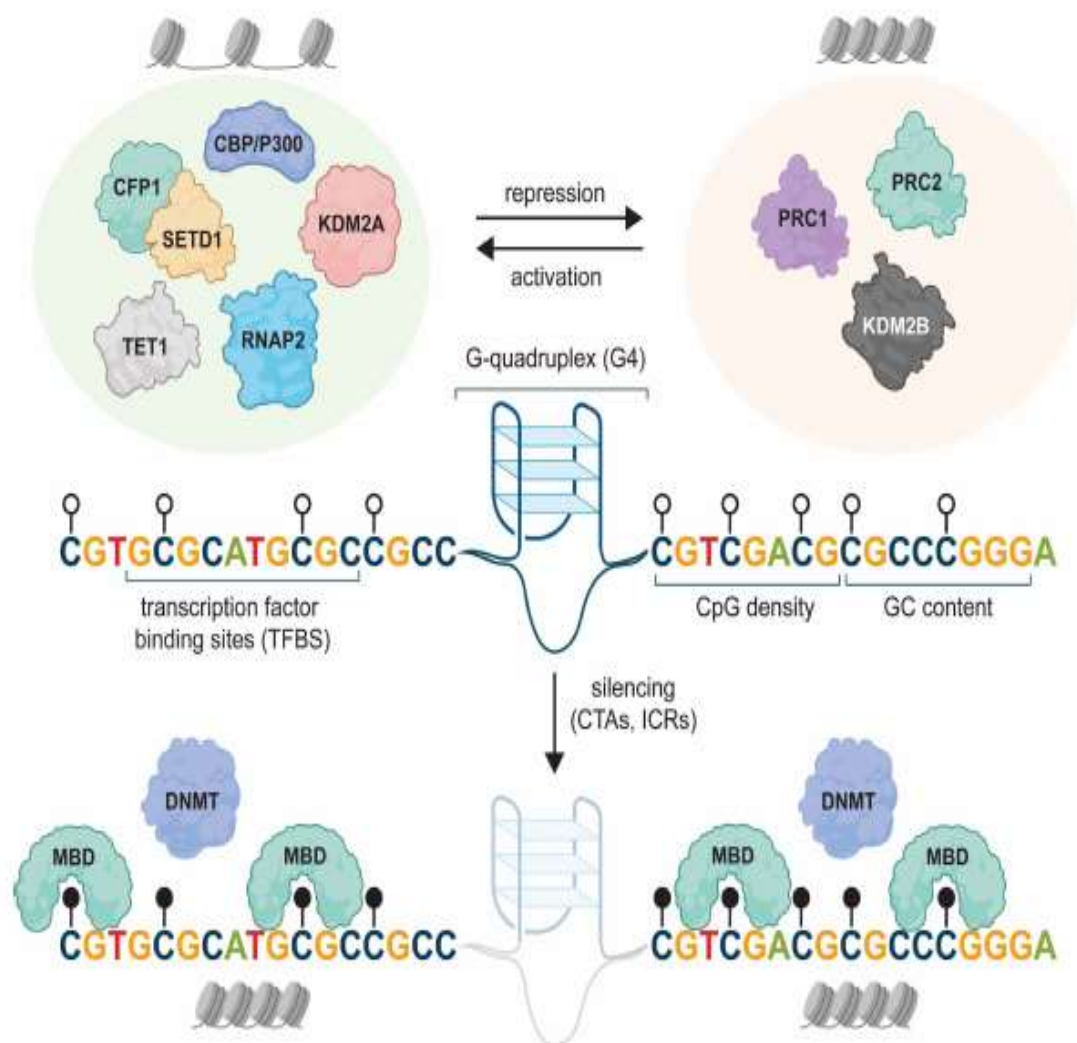
تنظیم ژن که فرآیندی به ظاهر ساده در مورد بیان یا عدم بیان یک ژن بوده و تقریباً در همه موجودات وجود دارد، پیش‌نیاز حیات می‌باشد. بطور کلی، عناصر تنظیمی (Regulatory elements: REs)، موتیف‌های دارای توالی خاص در ژنوم پستانداران هستند که بیان ژن را تنظیم می‌کنند. یکی از گروه‌های اساسی از RE، جزایر CpG (CpG islands: CGIs) هستند. CGI‌ها را می‌توان بر اساس توالی و ساختار کرماتینی آنها، که شامل محتوای بالای CpG و GC، فقدان متیلاسیون DNA (۵-متیل سیتوزین، 5mC)، وجود تری متیلاسیون در لیزین ۴ هیستون H3 (H3K4me3)، توالی‌های DNA چهارگانه G (G4) و غنی بودن از جایگاه‌های اتصال فاکتور رونویسی شناسایی کرد (شکل ۱). CpG‌ها

از طریق متیلاسیون DNA، افزودن یک گروه متیل به سیتوزین، که یک علامت اپی ژنتیکی ارثی می‌باشد، تغییر می‌کند. با این حال، هنگامی که CpGها در جزایر تجمع می‌کنند، عموماً در برابر متیلاسیون DNA محافظت می‌شوند (۱۱). در مواردی که CGIهای متیله شده در مجاورت جایگاه‌های شروع رونویسی ژن‌ها قرار گیرند با خاموشی رونویسی مرتبط هستند (۳۰). در حدود ۷۰ درصد از تمام پروموتور ژن‌های شناسایی شده در مهره‌داران، از جمله تقریباً اکثر ژن‌های خانه‌دار و همچنین زیرمجموعه‌ای از ژن‌های ویژه بافتی، مرتبط با CGI هستند. تعداد CGI گزارش شده در ژنوم انسان تقریباً ۲۶۰۰۰ مورد است (www.genome.ucsc.edu). توزیع CGIها با طول حدوداً کیلوبایت که در اکثر پروموتور ژن‌ها قرار دارد تحت عنوان پروموتورهای CpG شناخته می‌شوند (۶، ۳۰) اگرچه، CGIها معمولاً در پروموتورهای ژن مهره‌داران مورد مطالعه قرار می‌گیرند، ولی در حدود نیمی از تمام CGIهای شناسایی شده، گروه CGIهای "ارفان" (orphan CGIs) هستند که در مناطق بین ژنی و درون ژنی قرار دارند (۱۷). این گروه از CGIها نه تنها احتمال بیشتری دارد که متیله شوند، بلکه با احتمال بیشتری متیلاسیون DNA و علائم رونویسی را به روش ویژه بافتی نشان می‌دهند (۳، ۳۹). علیرغم دور بودن این گروه بزرگ از CGIهای کشف شده از جایگاه‌های شروع رونویسی، شواهدی مبنی بر تأثیر این گروه بر عملکرد پروموتور گزارش شده است (۳۹، ۷۸). این یافته‌ها بر همبستگی قوی بین CGI و شروع رونویسی تأکید می‌کنند. استفاده از روش‌های بیوشیمیایی مشخص کرد که یک چهارم کل CGIها در درون ژن‌ها بوده (intergenic CGIs: iCGIs)، و عملکردهای متعددی برای آنها گزارش شده است. CGIها می‌توانند بیان ژن را به روش‌های مختلفی، از طریق فعال بودن رونویسی خود یا از طریق تعامل با فرآیندهای بیولوژیکی در مجاورت خود، تحت تأثیر قرار دهند (۲، ۷۸).

CGIs از نظر تکامل

در مهره‌داران، CGIها به دلیل عدم متیلاسیون DNA و کمبود CpG نسبت به DNA ژنومی متمایز هستند. در موجودات بی‌مهره مانند *Drosophila melanogaster* و *Caenorhabditis elegans* و قارچ *Saccharomyces cereviceae* متیلاسیون DNA کم یا اصلاً دیده نمی‌شود و براین اساس، CpG در فراوانی مورد انتظار در سراسر ژنوم مشاهده می‌شوند. در این نوع ژنوم‌ها، CGIها قابل تشخیص نیستند زیرا به یک معنا، کل ژنوم شبیه CGI است. اگرچه، پدیده CGI در کل سلسله مهره‌داران حفظ شده است، اما نکته جالب این است که ویژگی محتوای بالای دی نوکلئوتید G+C موجود در CGI پستانداران و پرند، در برخی از گروه‌های مهره‌دارن دیده نمی‌شود (۶۹). به عنوان مثال، ماهی دارای توالی غیر متیله شده شبیه CGI در پروموتورها است، ولی این توالی‌ها به طور قابل توجهی در مقایسه با کل ژنوم غنی از G+C نیستند (۲۵، ۶۹). از آنجا که توالی‌های غنی از T در پستانداران مستعد متیلاسیون DNA هستند، از این رو می‌توان نتیجه گرفت که برخی از گروه‌های مهره‌داران ممکن است به مجموعه‌ای متفاوت از مکانیسم‌ها برای جلوگیری از متیلاسیون در CGIهای دارای کمبود از نظر G+C تکیه کنند. از سوی دیگر، بسیاری از ژنوم‌های گیاهی دارای سطح متیلاسیون بسیار بالایی هستند و تحقیقات اولیه یک بخش ژنومی شبیه CGI غیر متیله را در آنها شناسایی کردند (۳).

در گیاهان، فقدان متیلاسیون DNA آشکارا با ژن‌ها مرتبط است، زیرا جداسازی DNA غیر متیله شده نشان داد که این توالی‌ها از نظر رونویسی بسیار غنی بودند (۱۳۱). بررسی گسترده وضعیت متیلاسیون DNA در گیاهان، مناطق غیر متیله را در هر دو انتهای واحدهای رونویسی ثبت کرده است. در حال حاضر مشخص نیست که آیا این مناطق شناسایی شده را می‌توان معادل یا متمایز از CGIهای مهره‌داران در نظر گرفت (۳۳). بر خلاف مهره‌داران، اکثر حیوانات بی‌مهره ژنوم‌های متیله موزاییکی شامل دامنه‌های متیله و غیر متیله متناوب را نشان می‌دهند (۱۲، ۱۱۷). در طول تکامل، نواحی ژنومی متفاوت از نظر وضعیت متیلاسیون، باعث تقسیم توالی‌های DNA به دو بخش شده است: (الف) متیله شده و در نتیجه کمبود CpG و (ب) غیر متیله، با فراوانی مورد انتظار از نظر تعداد CpG. به نظر می‌رسد که منشاء مهره‌داران با انتقال از وضعیت موزاییک به متیلاسیون سراسری DNA، همراه با کاهش همزمان CpG در بیشتر ژنوم همزمان بوده است (۱۱۷). در میان ژنوم‌های موزاییک بی‌مهرگان، گونه دریایی *Ciona intestinalis*، که از نظر تکاملی نزدیک به مرز بی‌مهرگان و مهره‌داران است، به خوبی مطالعه شده است. ژن‌های *Ciona intestinalis* در دامنه‌های متیله‌شده گاهی با نواحی کوتاه غیر متیله شبیه CGI در ارتباط هستند که در کنار جایگاه‌های شروع رونویسی قرار دارند. بنابراین ظهور CGI ممکن است به قبل از تکامل مهره‌داران برگردد (۱۰۷).



شکل ۱. توالی و خصوصیات کروماتینی CGI

CGIها با افزایش تراکمی CpG و محتوای GC، مکان‌های اتصال فاکتور رونویسی و توالی‌های DNA چهارگانه G (G4) مشخص می‌شوند. CGIها با عناصر کلیدی تنظیم‌کننده ژن، مانند پروموتورها و تقویت‌کننده‌ها همپوشانی دارند و بسته به فعالیت ژنی که تنظیم می‌کنند، می‌توانند بین حالت‌های کروماتین فعال و سرکوب‌کننده جابجا شوند. این حالت‌ها تحت تأثیر مجموعه‌ای از پروتئین‌هایی به نام «reader»، صورت می‌گیرد که تمایل شدیدی به اتصال با CGIها دارند. از این گروه می‌توان فعال‌کننده‌های رونویسی (CBP/P300، SETD1، CFP1، TET1، KDM2A، RNAP2) و سرکوب‌کننده‌ها (PRC1، PRC2، KDM2B) را نام برد. در موارد استثنایی، مانند نواحی کنترلی ایمپرینتی (ICRs)، یا پروموتورهای ژن آنتی ژن بیضه سرطانی (CTA)، CGIها را می‌توان به طور پایدار از طریق متیلاسیون DNA (5mC) و پروتئین‌های متصل به CpG متیله شده (MBDs) خاموش کرد، حالتی که با هدف قرار دادن ثابت DNA متیل ترانسفرازها (DNMTs) تقویت می‌شود. هنوز مشخص نیست که حضور مداوم 5mC در CGIها چگونه توالی G4 را تحت تأثیر قرار می‌دهد (۱).

تا همین اواخر، مشخص نبود که آیا CGIها از نظر تعداد یا جایگاه ژنومی بین مهره‌داران مختلف در طول تکامل محافظت شده باشند. در ابتدا، CGIهای بسیار کمتری از نظر بیوانفورماتیک در ژنوم موش نسبت به ژنوم انسان پیش‌بینی شد (۱۲۰)، و این فقدان آشکار حفاظت

در طول تکامل، اهمیت نقش نظارتی CGIها را زیر سوال برد. الگوریتم‌های پیش‌بینی CGI به ضرورت از آستانه‌هایی برای تشخیص استفاده می‌کنند که تغییر آن‌ها به‌طور چشمگیری تعداد پیش‌بینی‌شده CGIها را تغییر می‌دهد (۵۰). الگوریتم‌ها همچنین قادر به در نظر گرفتن وضعیت متیلاسیون CGI نیستند. استفاده از رویکردهای بیوشیمیایی همراه با توالی یابی DNA با توان بالا تعداد تقریباً یکسانی از CGIها را در هر دو ژنوم هاپلوئید انسان (۲۵۴۹۵) و موش (۲۳۰۲۱) شناسایی کرد (۵۱). دلیل اختلاف اولیه در این بود که CGIها در موش در مقایسه با CGIها در انسان به‌طور متوسط محتوای CpG کمتری دارند. همچنین بین انسان‌ها و موش‌ها نسبت CGI مرتبط با جایگاه‌های شروع رونویسی حدود ۵۰ درصد بود. علاوه بر این، ۵۰ درصد باقیمانده CGIها به‌طور مساوی بین جایگاه‌های درون ژنی یا بین ژن‌ها در هر دو گونه توزیع شدند. به نظر می‌رسد موقعیت ژنومی بسیاری از این CGIهای اضافی از زمان واگرایی انسان و موش در ۷۵ میلیون سال پیش حفظ شده است که حاکی از اهمیت عملکردی آنها می‌باشد (۳۰).

CGIها به عنوان جایگاه شروع رونویسی

تقریباً نیمی از تمام CGIها به وضوح حاوی جایگاه‌های شروع رونویسی هستند، زیرا با پروموتورهای ژن‌ها همخوانی دارند. نیمی دیگر یا در داخل یا بین واحدهای رونویسی مشخص قرار دارند و تحت عنوان CGIهای ارفان (oCGIs) شناخته می‌شوند (۱۷، ۵۱). آیا CGIهای ارفان همبستگی بین CGIها و فعالیت رونویسی را تضعیف می‌کنند؟ آیا آنها همچنین پروموتورهای نامعلومی را مشخص می‌کنند؟ شواهد موجود نشان می‌دهد که بسیاری از CGIهای ارفان نیز محل شروع رونویسی هستند. نمونه‌های خاصی از پروموتورهای CGI درون ژنی سال‌ها قبل شناخته شده‌اند. به عنوان مثال، CGIها در انتهای ۳ ژن POMC و آگزون ۲ ژن MHC کلاس II I-Ab هر دو رونوشت‌هایی با عملکرد ناشناخته کد می‌کنند (۷۳). تجزیه و تحلیل‌های گسترده ژنوم تایید کرده است که بسیاری از CGIهای ارفان نشان‌دهنده پروموتورهای جدید هستند (۵۱، ۶۷). برای مثال شواهدی برای شروع رونویسی در ۴۰ درصد از CGIهای ارفان یافت شده است (۵۱). همچنین، CGIها معمولاً با تری متیلاسیون هیستون H3 (H3K4me3) مشخص می‌شوند که نشانه‌ای از پروموتورهای فعال است. بسیاری از پروموتورهای CGI ارفان به شیوه‌ای خاص بافتی فعال هستند (۳۹)، که نشان می‌دهد آنها به شدت تنظیم می‌شوند. از آنجایی که تاکنون تنها چند بافت مورد بررسی قرار گرفته‌اند، احتمالاً اکثر، اگر نه همه، CGIهای ارفان با یک رونوشت جدید مرتبط می‌باشند. اهمیت عملکردی این رونوشت چیست؟ برخی از CGIهای ارفان احتمالاً نشان‌دهنده پروموتورهای جایگزین ژن‌های نزدیک هستند (۷۸). برخی دیگر ممکن است آغاز کننده ncRNAهایی باشند که بیان ژن را تنظیم می‌کنند. اطلاعات بیشتری در مورد نقش رونوشت Air (نوعی RNA غیر کد کننده) در ایمپرینت شدن ژن گیرنده فاکتور رشد شبه‌انسولین ۲ (Igf2r) وجود دارد. ژن‌های ایمپرینت شده ژن‌هایی هستند که تنها یکی از آلل‌های به ارث رسیده از دو والد در یک لوکوس، بیان می‌شوند. Air یک RNA غیر کد کننده (non-coding RNA: ncRNA) است که در یک CGI در داخل ایترون ۲ ژن Igf2r قرار دارد و برای خاموش کردن آلل پدری ضروری است (۱۰۶). به‌طور مشابه، تجزیه و تحلیل یک CGI در ایترون ۱۰ ژن Kcnq1 ایمپرینت شده، آن را به عنوان منشاء رونوشت غیرکدکننده (Kcnq1ot1) شناسایی کرد که برای ایمپرینت شدن چندین ژن در داخل این دامنه مورد نیاز است (۷۶). جهت غیر فعال کردن کروموزوم X، علاوه بر Air و Kcnq1ot1، ncRNAهای Xist و Tsix نیز دخالت می‌کنند (۵۸) و گزارش شده است که RNA غیر کد کننده HOTAIR بیان ژن Hox را تنظیم می‌کند (۹۵). در مسیری متفاوت، پروموتورهای CGI در ژن‌های MHC کلاس II دقیقاً با آگزون ۲ بسیار متنوع منطبق هستند، که چندشکلی آن توسط تبدیل ژن بین اعضای کلاس II خانواده ژن ایجاد می‌شود. پیشنهاد شده است که ساختار کروماتین CGI "باز" در سلول‌های زایا تبدیل ژن را در این آگزون افزایش می‌دهد و در نتیجه برای سیستم ایمنی مفید است (۷۳). از بحث فوق واضح است که CGIها به عنوان پروموتور در ژنوم پستانداران عمل می‌کنند. این امر دو احتمال را در رابطه با اهمیت آنها باقی می‌گذارد: (i) CGIها ردپای تکاملی رویدادهای مولکولی هستند که در بسیاری از پروموتورهای یوکاریوتی رخ می‌دهند، اما فقط در موجوداتی قابل مشاهده هستند که از متیلاسیون DNA ژنومی گسترده‌ای برخوردار هستند، یا (ii) CGIها ساختارهای تنظیمی مهمی هستند که تحت انتخاب در ژنوم‌هایی که متیلاسیون DNA نقش تنظیمی ایفا می‌کند، تکامل یافته‌اند. اکنون به نظر می‌رسد وزن شواهد و مستندات گزینه دوم از این احتمالات را تأیید می‌کند (۳۰).

ویژگی‌های CGIها به عنوان پروموتور

الف) توالی‌های موتیف

معمولاً CGI‌ها با اغلب پروموتورها در ژنوم انسان و موش هم جایگاه هستند. مطالعات اولیه نشان داد که اغلب پروموتورهای CGI ممکن است فاقد جعبه‌های TATA باشند و جایگاه‌های شروع رونویسی ناهمگن را نشان دهند. تجزیه و تحلیل در مقیاس وسیع ژنومی، همبستگی گسترده بین این الگوی توزیعی شروع رونویسی، معمولاً در ناحیه ۵۰-۱۰۰ جفت باز، و حضور CGI‌ها را نشان داده است (۱۸). این مشاهدات با این ایده سازگار هستند که پروموتورهای CGI حالتی از رونویسی را اتخاذ می‌کنند که در آن شروع رونویسی می‌تواند در جایگاه‌های متعددی رخ دهد. به طور کلی، جعبه‌های TATA، همراه با سایر عناصر پروموتور اصلی (مانند BRE، DPE، و DCE)، با شروع رونویسی متمرکز مرتبط هستند، در حالی که CGI‌ها تمایل دارند این عناصر را نداشته باشند و الگوهای شروع رونویسی پراکنده‌ای را نشان دهند. با این حال، استثنایی برای این تعمیم وجود دارد. برای مثال، ژن‌های انسانی آلفا-گلوبین، MyoD1، و اریتروپویتین نه تنها دارای پروموتورهای CGI هستند، بلکه دارای جعبه‌های TATA نیز می‌باشند (۵۹). این ایده که بسیاری از پروموتورهای CGI از نظر رونویسی مجاز هستند، توسط تجزیه و تحلیل گسترده ژنوم با روش‌های ChIP و ترانسکریپتوم تایید شده‌اند. آنزیم RNA پلیمراز II به پروموتورهای CGI بسیاری از ژن‌های غیرفعال در سلول‌های بنیادی جنینی (۳۹) و در ژن‌های القایی با لیوپلی ساکارید خاموش در ماکروفاژهای تحریک‌نشده متصل می‌شود (۴۱). همچنین، تجزیه و تحلیل گسترده هسته‌ای که اجزای مولکول‌های RNAPII درگیر در رونویسی را بررسی می‌کند، رونویسی دوطرفه RNA‌های کوتاه فاقد اطلاعات پروتئینی و همچنین رونوشت‌های با طول کامل را در بسیاری از پروموتورهای CGI شناسایی کرد (۲۴، ۱۰۳). قبلاً تأکید زیادی بر استفاده از RNAPII به عنوان مرحله محدودکننده سرعت در رونویسی شده بود، اما این نتایج نشان می‌دهد که تنظیم بسیاری از پروموتورهای CGI در پایین دست از طریق اتصال RNAPII انجام می‌شود. یکی از راه‌های دستیابی به این هدف، تنظیم طول رونویسی و پردازش mRNA است. در یک مطالعه، بررسی عملکرد و اتصال RNAPII وابسته به فاکتور رونویسی Sp1 نشان داد که شکل آغازین RNAPII حتی زمانی که ژن‌ها غیرفعال بودند و سطوح پایینی از رونوشت‌های تمام طول تولید شد، به پروموتور متصل بود (۴۱). تغییر از رونویسی غیرمولد به mRNA حاوی اطلاعات پروتئینی با القا به فاکتور رونویسی وابسته به P-TEFb و متعاقب آن فسفوریلاسیون Ser2 از بخش CTD آنزیم RNAPII آغاز شد. این فرآیند منجر به تولید رونوشت‌های بالغ و پردازش شده گردید. به طور مشابه، تنظیم ژن در سطح فرآیند تولید سازی رونویسی از طریق توقف RNAPII (کنترل شده توسط فاکتورهای DSIF و NELF) و انتشار RNAPII با واسطه P-TEFb گزارش شده است که در سلول‌های بنیادی جنینی گسترده است (۹۲). به نظر می‌رسد که پروموتورهای CGI، با جذب RNAPII به نوعی درگیر رونویسی خواهند شد، مگر این که به طور فعال مهار شوند (۹۲).

ب) اتصال فاکتورهای رونویسی در CGI‌s

تنظیم ژن با واسطه آنزیم RNAPII تا حدی با اتصال فاکتورهای رونویسی (Transcription Factor: TFs) به یک هسته پروموتور به دست می‌آید که به عنوان ناحیه ± 50 جفت بازی از محل شروع رونویسی یک ژن تعریف می‌شود (۴۰). هسته پروموتورها از عناصر توالی معمولی مانند جعبه TATA یا جزایر CpG تشکیل شده‌اند. پروموتورهای حاوی جعبه TATA اغلب رونوشت‌های ویژه سلولی یا القا شده (مثلاً توسط یک هورمون) را تولید می‌کنند، در حالی که اکثر ژن‌های خانه‌دار توسط پروموتورهای دارای جزایر CpG هدایت می‌شوند (۱۰۰). CGI‌ها جدا از تراکم بالای CpG و محتوای G+C، در توالی‌های با طول بلند دارای حفاظت تکاملی پایینی هستند، و اغلب فاقد عناصر پروموتور مانند جعبه TATA هستند (۱۸، ۵۹). پس، چه ویژگی‌هایی آنها را با عملکرد پروموتور تطبیق می‌دهند؟ یک احتمال ساده این است که غنای GC احتمال اتصال عوامل رونویسی در همه جا را افزایش می‌دهد. به طور کلی در پستانداران، مکان‌های اتصال به فاکتورهای رونویسی نسبت به کل ژنوم غنی‌تر از GC هستند و بسیاری از آنها حاوی CpG در توالی تشخیص خود می‌باشند. CGI‌ها معمولاً دارای جایگاه اتصال فاکتور رونویسی عمومی Sp1 است که مشخص شده پروتئین متصل به TATA (TATA binding protein: TBP) را برای پروموتورهایی که فاقد جعبه TATA هستند، جذب می‌کند (۱۶). با بررسی فعالیت ۴۵۷۵ پروموتور انسانی مشخص شد که پروموتورهای فعال CGI در همه جا تمایل به غنی بودن برای فاکتورهای رونویسی Sp1، Nrf-1، E2F، و ETS دارند که هر کدام حاوی یک CpG هستند (۶۵). مطابق با این، مکان‌های بالقوه اتصال خانواده ETS و E2F، همانطور که در شرایط آزمایشگاهی مشخص می‌شوند، در CGI‌های موش بیش

از حد نشان داده می‌شوند (۵۶). به طور مشابه، موتیف‌های ETS، NRF1، BoxA، SP1، CRE و E-box که حاوی CpG هستند در پروموتورهای CGI ژن‌های خانه‌دار به وفور یافت شده‌اند (۹۶)، و اتصال Sp1 روی کروموزوم‌های ۲۱ و ۲۲ انسان به طور غالب بر CGI‌ها متمرکز است (۲۵). در مورد ژن آلفا-گلوبین، الگوهای اتصال فاکتور رونویسی در حالت فعال و غیر فعال ژن، متفاوت بود که نشان می‌دهد حتی در مورد یک بافت بسیار خاص، اتصال فاکتور رونویسی می‌تواند سازنده باشد (۲۷).

ویژگی‌های کروماتین در CGI‌ها

الف) نوکلئوزوم‌های ناپایدار و ساختار کروموزومی

تمام فرآیندهای مبتنی بر DNA مانند رونویسی ژن، همانندسازی و ترمیم DNA بستگی به ساختار کروماتین دارد. کروماتین اساساً از DNA و پروتئین‌های هیستونی تشکیل شده است که با یکدیگر واحد اصلی آن یعنی نوکلئوزوم را تشکیل می‌دهند. در هسته نوکلئوزوم، تقریباً ۱۴۶ جفت باز DNA به دور یک ساختار پروتئینی متشکل از دو نسخه از هر یک از هیستون‌های H2A، H3، H4 و H2B پیچیده شده‌اند. در حالی که هیستون‌های H3 و H4 ساختار تترامری به نام هسته نوکلئوزومی را تشکیل می‌دهند که در ناحیه داخلی نوکلئوزوم قرار دارد، هیستون‌های H2A و H2B بیشتر روی سطح نوکلئوزومی قرار دارند (۷۲). علاوه بر این، پیوند دهنده هیستون H1 می‌تواند با محل‌های ورود و خروج DNA نوکلئوزومی تماس داشته و در نتیجه ساختار فشرده‌تری ایجاد کند (۴۳). اصلاحات پس از ترجمه (Post translational modifications: PTMs) عمدتاً روی دم‌های هیستونی انعطاف‌پذیر و همچنین در دامنه‌های چین خورده هیستون‌ها صورت می‌گیرند. مهم‌تر از همه، انواع هیستون‌های مستقل از همانندسازی می‌توانند جایگزین هیستون‌های وابسته به همانندسازی شوند و به طور خاص در ژنوم قرار گیرند (۱۱۴، ۱۳۰). از نظر ساختار کروماتینی، شواهدی وجود دارد که CGI‌های غیر متیله در یک ساختار مشخص به گونه‌ای سازماندهی شده‌اند که آنها را مستعد فعالیت پروموتوری می‌کند. مطالعه کروماتین در ژن‌های القایی با لیپوپلی ساکارید در ماکروفاژها نشان داد که CGI‌ها نسبتاً با کمبود نوکلئوزوم مواجه هستند (۹۳). بطور کلی ژن‌های القایی دارای پاسخ اولیه به دو دسته تقسیم می‌شوند: آن‌هایی که برای فعال‌سازی خود به ترکیب بازسازی کروماتین SWI/SNF نیاز دارند و آن‌هایی که نیازمند این ترکیب نیستند. مشاهده شده که این گروه‌ها به ترتیب با پروموتورهای غیر CGI و CGI مطابقت دارند، که نشان می‌دهد DNA در کروماتین CGI ذاتاً بدون نیاز به جابجایی نوکلئوزوم وابسته به ATP قابل دسترسی است (۹۳). در ماکروفاژها، CGI‌ها تراکم کاهش یافته هیستون H3 را حتی در حالت القا نشده نشان دادند. بر این اساس، در شرایط آزمایشگاهی، مونتاژ نوکلئوزومی نشان داد که مجموعه‌ای از این CGI‌ها به طور قابل توجهی نسبت به سایر نواحی ژنوم، تمایلی به تجمع در نوکلئوزوم‌ها ندارند (۹۳). یک تفسیر جذاب از بی‌ثباتی کروماتین CGI در شرایط آزمایشگاهی این است که تضعیف این مانع امکان دسترسی بیشتر DNA زیرین به تنظیم‌کننده‌های رونویسی را فراهم می‌کند. شواهد دیگر نشان داده‌اند که کمبود نوکلئوزوم به طور کلی یکی از ویژگی‌های پروموتورهای CGI است. علاوه بر ناپایداری کروماتین، کمبود نوکلئوزومی در بدنه CGI نیز ممکن است ایجاد شود زیرا پروموتورهای CGI، به طور مشترک با تمام پروموتورهای یوکاریوتی، معمولاً دارای یک ناحیه عاری از نوکلئوزوم در اطراف جایگاه‌های شروع رونویسی هستند (۱۰۲). به عبارت دیگر، پروموتورهای فعال بنا به تعریف ممکن است دارای کمبود نوکلئوزومی باشند، چه CGI باشند یا نباشند. هنوز مشخص نیست که کمبود نوکلئوزومی در CGI عمدتاً به دلیل بی‌ثباتی ذاتی کروماتین یا حذف نوکلئوزوم به دلیل وجود کمپلکس شروع رونویسی است. البته ممکن است ترکیبی از هر دو باشد و حتی ممکن است بین CGI‌ها متفاوت باشد (۳۰).

در یک مطالعه، یک توالی DNA مصنوعی که شامل یک جزیره CpG بود به سلول‌های بنیادی موش وارد شد. در این سلول‌ها، کروماتین اطراف این جزایر CpG به راحتی با علائم شیمیایی فعال و غیرفعال تغییر یافتند. علاوه بر این، با تغییر توالی DNA مصنوعی، مشخص شد که این علائم شیمیایی تنها زمانی اضافه می‌شوند که توالی‌های DNA حاوی مقدار زیادی C و به دنبال آن G باشد. توالی‌های مصنوعی دیگر با تعداد زیاد Cs و Gs، جایی که Gها به ندرت بلافاصله پس از C یافت می‌شدند، هیچ یک از دو علامت شیمیایی روی کروماتین

را نشان ندادند (۱۱۹). از سوی دیگر، زمانی که DNA حاوی مقدار زیادی As و Ts بود، علائم شیمیایی مستقیماً به DNA (به جای کروماتین) اضافه می‌شد و این مانع از اضافه شدن علائم شیمیایی فعال و غیرفعال به کروماتین می‌شد (۱۱۹). بطور کلی، پروموتورهای دارای CGI دائماً متیله نشده باقی می‌مانند (5mC) و دسترسی کروماتین را تا حدی حفظ می‌کنند، حتی زمانی که ژن‌ها توسط عوامل کمپلکس پروتئین‌های سرکوب کننده (Polycomb repressive complex: PRC) سرکوب می‌شوند (۸). با تجمع کمپلکس پروتئین‌های سرکوب کننده پلی کمب، اجسام پلی کمب تشکیل می‌شوند و هنگامی که ژن‌ها در داخل این اجسام قرار می‌گیرند سرکوب می‌شوند (۹۰). از آنجایی که اتصال فاکتورهای رونویسی، هایپومتیلاسیون CpG و دسترسی به کروماتین از مشخصه‌های یوکروماتین هستند، این نتایج به وضوح نشان می‌دهد که ژن‌های دارای CGI عموماً با یوکروماتین مرتبط هستند. از سوی دیگر، پروموتور ژن‌های فاقد CGI عموماً متیله و غیرقابل دسترس هستند (۳۰). نتایج یک تحقیق نشان داد که نواحی اطراف پروموتورهای خاموش فاقد CGI عمدتاً با علائم هتروکروماتین در مقایسه با ژن‌های فعال فاقد CGI مرتبط هستند (۴۲). این نشان می‌دهد که ژن‌های فاقد CGI با هتروکروماتین مرتبط هستند و می‌توانند پس از فعال‌سازی به یوکروماتین تبدیل شوند. این یافته‌ها روشن می‌کنند که پروتئین‌های PRC بیان ژن‌های دارای CGI را سرکوب می‌کند در حالی که ژن‌های فاقد CGI با تشکیل هتروکروماتین غیرفعال می‌شوند (۸).

ب) تغییرات هیستونی

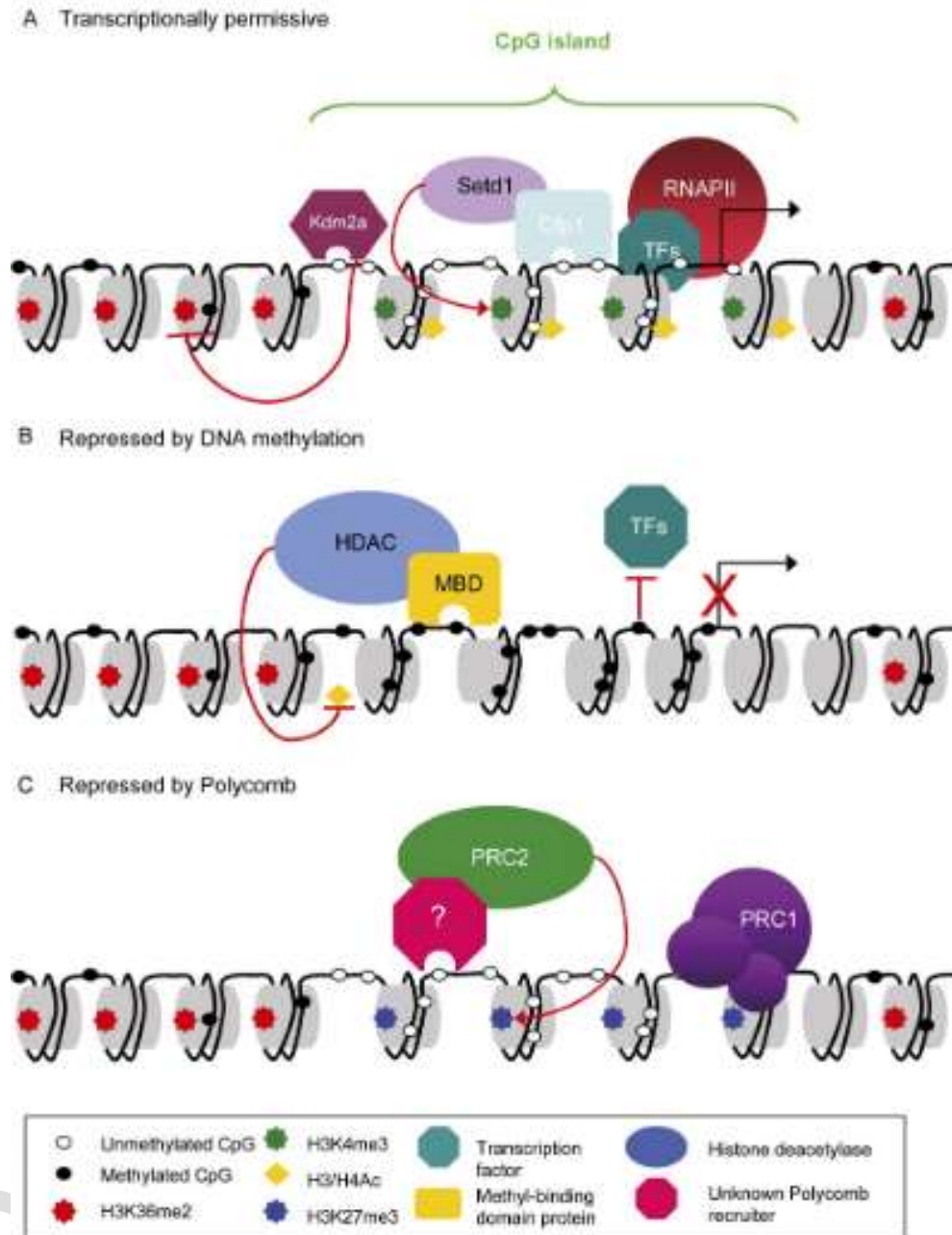
در مطالعات بیوشیمیایی، در کروماتین جدا شده از CGI سطوح بالایی از استیلاسیون هیستون H3 و H4، ویژگی کروماتین فعال رونویسی، مشاهده شد. از طرف دیگر هیستون H1 که به عنوان آنتاگونیستی با رونویسی در نظر گرفته می‌شود، در این بخش با کاهش همراه بود (۱۱۲). مطالعات گسترده ژنوم این ارتباط را با وضوح بالا تایید کرده و نشان داده اند که H3K4me3 علامت هیستونی مشخصی از پروموتورهای CGI است، که اغلب حتی زمانی که ژن مرتبط غیرفعال است، باقی می‌ماند (۳۹، ۸۲). ارتباط بیوشیمیایی بین فراوانی CpG در CGI‌ها و H3K4me3 با واسطه یک پروتئین با دامنه CXXC که به طور خاص به CpG غیر متیله متصل می‌شود، تایید شده است (۱۱۷). نشان داده شده است که خانواده‌ای از پروتئین‌ها که دامنه ZF-CxxC (zinc finger-CxxC) را کد می‌کنند، به طور خاص DNA غیر متیله را تشخیص می‌دهند و فعالیت‌های اصلاح کننده کروماتین را برای عناصر CGI به کار می‌گیرند. به عنوان مثال، CFP1 (CxxC finger protein)، KDM2A (lysine demethylase)، KDM2B و MLL (mixed lineage leukaemia protein) متیلاسیون لیزین را در دم هیستونی تنظیم می‌کنند، در حالی که TET-1 (ten-eleven translocation) و TET3 بازهای سیتوزین متیله شده را هیدروکسیله می‌کنند (۶۸). پروتئین Cfp1 به عنوان جزء جدایی ناپذیر از کمپلکس متیل ترانسفراز Setd1 H3K4، در اکثریت قریب به اتفاق CGI‌ها در ژنوم موش قرار داشته که نشان دهنده وابستگی این اصلاح هیستون به توالی DNA است (۱۱۳). مطابق با این مدل، کاهش Cfp1 باعث کاهش H3K4me3 در بسیاری از CGI‌ها می‌شود. نکته مهم این است که قرار دادن یک توالی مصنوعی DNA شبیه CGI در ژنوم منجر به جذب Cfp1 می‌شود و در غیاب آنزیم RNAPII پیک جدیدی از H3K4me3 ایجاد می‌کند (۱۱۳). بررسی رابطه بین CpG غیر متیله، Cfp1 و H3K4me3 نشان داد که تراکم CpG در CGI با سطوح H3K4me3 ارتباط مثبت دارد (۴۹، ۵۱). بطور کلی، مشخص شده که CGI‌ها توزیع H3K4me3 را تعیین کرده و به نوبه خود، H3K4me3 بر ساختار کروماتین در CGI‌ها تأثیر می‌گذارد. توانایی تراکم CpG به تنهایی برای تأثیر مستقیم بر وضعیت اصلاح کروماتین احتمالاً یک عملکرد کلیدی CGI است (۱۱۲). یکی دیگر از ویژگی‌های متمایز کروماتین CGI کاهش دی متیلاسیون هیستون H3K36 (H3K36me2) در مقایسه با پروموتورهای غیر CGI و اجسام ژنی است (۱۴). H3K36me2 هیستون دمتیلاز Kdm2a، یک پروتئین دیگری با دامنه CXXC است که مانند Cfp1، به طور خاص به CpG غیر متیله متصل می‌شود (۱۱۶). بر این اساس، Kdm2a به ۹۰ درصد از CGI‌ها در سلول‌های بنیادی جنینی موش متصل می‌شود و واسطه دمتیلاسیون H3K36me2 در این مناطق است (۱۴). این که چرا H3K36me2 باید در CGI‌ها تخلیه شود، نامشخص است، اما گزارش شده است که این اصلاح، شروع رونویسی را از طریق جذب هیستون دی استیلاز (HDAC) در مخمر مهار می‌کند (۶۷، ۱۲۸). بنابراین کاهش H3K36me2 ممکن است به یک حالت مجاز رونویسی در CGI کمک کند. پروتئین‌های MLL1 و MLL2 همچنین حاوی دامنه‌های ZF-CxxC هستند که به طور خاص به دی نوکلئوتیدهای CpG غیر متیله در شرایط آزمایشگاهی متصل می‌شوند (۲۳، ۱۲۷). MLL1 و MLL2، مانند CFP1، جهت ارتباط با CGI به دامنه‌های ZF-CxxC خود

نیاز دارند (۴۸). بطور کلی پروتئین‌های حاوی دامنه‌های ZF-CxxC دارای ویژگی‌های پروتئین‌های خاص مرتبط با CGI هستند که از تراکم CpG برای تأثیر بر اصلاح کروماتین استفاده می‌کنند.

خاموش کردن پروموتور CGI

معمولاً CGI‌ها در یک حالت غیر متیله در یک ژنوم باقی می‌مانند، حتی زمانی که ژن مربوطه از نظر رونویسی غیرفعال است. البته، نمونه‌های شناخته شده‌ای از CGI وجود دارد که در طی رشد و نمو طبیعی متیله شده و منجر به خاموش شدن پایدار پروموتور مربوطه می‌شوند (۸۴). فرآیند خاموش شدن پروموتور می‌تواند به دلیل مهار مستقیم اتصال فاکتور رونویسی توسط DNA متیله شده باشد، و یا تغییر ساختار کروماتین در DNA متیله به واسطه اتصال پروتئین‌های MBD به DNA متیله شده باشد (شکل ۲). به نظر می‌رسد متیلاسیون CGI رویداد آغازگری در خاموشی ژن نیست، بلکه در حالت خاموش عمل می‌کند. به عنوان مثال، در طی غیرفعال شدن کروموزوم X در پستانداران ماده، CGI‌های مرتبط با کروموزوم X تا پس از خاموش شدن ژن و ایجاد چندین اصلاح کروماتینی خاموش کننده مانند H3K27me3، متیله نمی‌شوند (۱۵). با این حال، متیلاسیون CGI برای حفظ غیرفعال سازی کروموزوم X ضروری است، زیرا مهار متیلاسیون DNA منجر به فعال شدن مجدد ژن در بخشی از سلول‌ها می‌شود (۲۶، ۹۹). علاوه بر این، متیلاسیون CGI نقش‌های مشخصی در فرآیند ایمپرینت شدن دارد که در آن بیان تک آلی والد منشاء توسط علائم متیلاسیون CGI کنترل می‌شود.

در موارد متعددی، CGI‌های مربوطه به عنوان پروموتورهای ncRNA عمل می‌کنند که بیان آنها توسط متیلاسیون DNA خاموش می‌شود. بیان این ncRNAها (از جمله Air و Kcnq1ot1) مسئول خاموش کردن ژن‌های مجاور است (۷۷، ۷۸، ۱۰۶). با این حال، مطالعات گسترده ژنوم با تمرکز بر CGI در جایگاه‌های شروع رونویسی ژن‌ها، موارد متعددی از متیلاسیون CGI در سلول‌های سوماتیک طبیعی را کشف کرده است. CGI‌ها در رده زایا تقریباً همیشه غیر متیله هستند، اما بخش کوچکی از آنها در بافت‌های سوماتیکی متیله می‌شوند (۱۰۱، ۱۰۵، ۱۲۱). به طور مشابه، تعداد کمی از پروموتورهای CGI در طول تمایز سلول‌های بنیادی جنین به نورون‌ها، متیلاسیون را به دست می‌آورند که بیشتر تغییرات در مراحل اولیه تمایز رخ می‌دهد. اکثر CGI‌هایی که در طول تمایز متیله می‌شوند در سلول‌های بنیادی جنینی خاموش هستند (۸۴)، که شواهد بیشتری را ارائه می‌کند که خاموش کردن ژن قبل از متیلاسیون DNA رخ می‌دهد. معمولاً ژن‌های تحت تأثیر، مانند خانواده‌ای از آنتی ژن‌های اختصاصی بیضه، فقط در رده زایا بیان می‌شوند (۲۹).



شکل ۲- وضعیت کروماتین در CGIها معمولاً در حالت مجاز برای رونویسی به صورت غیر متیله دیده می‌شوند. آنها با استیلاسیون هیستون (H3/H4Ac) و H3K4me3 که توسط پروتئین Cfp1 هدایت می‌شود مشخص می‌شوند و همچنین کاهش H3K36me2 وابسته به Kdm2a را نشان می‌دهند. کمبود نوکلئوزومی و اتصال سازنده RNAPII نیز ممکن است به این حالت مجاز رونویسی کمک کند. (B) متیلاسیون DNA با خاموش کردن طولانی مدت پایدار پروموتورهای CGI مرتبط است. این فرآیند می‌تواند توسط پروتئین‌های MBD، با جذب کمپلکس‌های مهارکننده مرتبط با فعالیت HDAC صورت گیرد و یا ممکن است به دلیل مهار مستقیم اتصال فاکتور رونویسی توسط متیلاسیون DNA باشد. (C) CGIها همچنین می‌توانند توسط پروتئین‌های گروه پلی کمب خاموش شوند و ممکن است عناصر کلیدی درگیر در بکارگیری پلی کامب باشند. یک عامل ناشناخته متصل به CGI می‌تواند مسئول جذب PRC2 به CGI باشد و سپس H3K27 را تری متیله (H3K27me3) می‌کند. این H3K27me3 توسط کمپلکس‌های PRC1 شناسایی می‌شود که مانع از تولید فرآیند رونویسی می‌شوند و در نتیجه ژن‌ها را خاموش می‌کنند. توجه داشته باشید که حالت‌های مجاز رونویسی و سرکوب‌شده چند لانه‌ای می‌توانند در CGIهای دو ظرفیتی، عمدتاً در سلول‌های جنینی همه‌توان، بطور همزمان وجود داشته باشند (۳۰).

تفاوت‌هایی در متیلاسیون CGI مرتبط با ژن بین یک بافت سوماتیک و بافت دیگر گزارش شده است، اگرچه این تفاوت‌ها در مقایسه با تفاوت‌های بین CGI‌های زایا و سوماتیکی نسبتاً نادر است (۷۹). برخلاف نادر بودن CGI‌های متیله در پروموتورهای ژن‌های شناسایی شده، CGI‌های ارفان بسیار بیشتر متیله می‌شوند. در مقایسه با ۳ درصد از CGI‌ها در پروموتورهای ژن‌های شناسایی شده، حدود ۱۷ درصد از CGI‌های ارفان در وضعیت متیله یافت شده‌اند (۴۸، ۴۹، ۷۸). با تفکیک بیشتر CGI‌های ارفان به دسته‌های درون ژنی و بین ژنی، آشکار می‌شود که CGI‌های درون ژنی به ویژه مستعد متیلاسیون هستند (۴۹، ۷۸). بر این اساس، CGI‌های واقع در بدنه ژنی، بیشترین تعداد تفاوت متیلاسیون DNA را در بین سلول‌ها و بافت‌های مختلف سوماتیک نشان می‌دهند (۳۰، ۷۸). از نظر عملکردی، می‌توان حدس زد که برخی از رونوشت‌های آغاز شده از CGI‌های داخل ژن، ncRNA‌های تنظیمی هستند که وجود یا عدم وجود آنها بر بیان ژن کدکننده پروتئین مرتبط با یک ژن مجاور تأثیر می‌گذارد (۳۱، ۸۱). احتمال دیگر این است که این جایگاه‌های غیرمعمول کروماتین و رونویسی بر اسپلایسینگ جایگزین (Alternative splicing) ژنی که در آن قرار دارند، تأثیر می‌گذارند، به نحوی که این تأثیر با وضعیت متیلاسیون متفاوت است (۶۲). همچنین ممکن است که یک CGI متیله در بدنه ژن، طویل‌سازی رونویسی را کاهش دهد، همانطور که در یک مدل سلول تراریخته گزارش شده است (۷۱).

CGI و خاموشی با واسطه پلی کمب

علاوه بر متیلاسیون DNA، پروموتورهای CGI می‌توانند توسط پروتئین‌های گروه پلی کامب (Poly comb group: PcG) خاموش شوند (شکل ۲). در پستانداران دو کمپلکس PcG وجود دارد: کمپلکس پلی کمب سرکوب کننده PRC1 و PRC2. پروتئین PRC2 واسطه H3K27me3 بوده، و این علامت توسط PRC1 شناسایی می‌شود، که تصور می‌شود با مکانیزم خاصی از مرحله طویل‌سازی رونویسی جلوگیری می‌کند (۳۲). تمایز سلول‌های بنیادی جنینی به نورون‌ها با از دست دادن و افزایش H3K27me3 در بسیاری از پروموتورها در مراحل مختلف تمایز همراه است. در مقابل، متیلاسیون DNA در تعداد نسبتاً کمی از پروموتورها فقط در مراحل اولیه تمایز به دست می‌آید و پس از آن نسبتاً پایدار باقی می‌ماند (۸۴). این نشان می‌دهد که پلی کمب یک سیستم سرکوب پویاتر از متیلاسیون DNA، حداقل در مراحل بعدی رشد و نمو است. در سلول‌های بنیادی جنین، CGI‌های خاموش شده توسط پلی کامب دارای علامت فعال H3K4me3 و همچنین H3K27me3 هستند (۵، ۱۰، ۸۲). در این پروموتورهای «دو ظرفیتی»، CGI بین دو حالت قرار دارند: رونویسی فعال یا سرکوب پایدار. پس از تمایز، آنها می‌توانند H3K27me3 را از دست بدهند و فعال شوند یا در معرض سرکوب رونویسی پایدارتری قرار بگیرند. پروموتورهای دو ظرفیتی CGI تقریباً ۲۰ درصد از پروموتورهای CGI در سلول‌های بنیادی جنین را تشکیل می‌دهند (۶۳) اما در سایر انواع سلول به میزان کمتری نیز یافت می‌شوند (۸۲). از آنجایی که اکثر CGI‌ها به طور معمول از پلی کامب استفاده نمی‌کنند، ترکیب پایه به خودی خود بسیار بعید است که تنها عامل دخیل در جذب یا حذف پلی کامب باشد. برای توضیح این گزینش پذیری، پیشنهاد شد که وجود جایگاه‌های اتصال فاکتور رونویسی در یک CGI ممکن است برای محافظت از آن در برابر بکارگیری PRC2 کافی باشد (۶۳). مطابق با این فرضیه، حذف این جایگاه‌های فعال کننده در یک پروموتور CGI که به طور معمول فاقد H3K27me3 در سلول‌های بنیادی جنین است، منجر به استفاده از PRC2 و اصلاح H3K27 شد (۸۰). این نشان می‌دهد که اتصال فاکتور رونویسی یا رونوشت تولیدی ممکن است برای محافظت در برابر خاموش شدن توسط پلی کامب کافی باشد. نتیجه این فرضیه این است که خاموشی با واسطه پلی کمب، مانند متیلاسیون DNA، به خاموش کردن ژن توسط مکانیسم‌های دیگر، فرآیندی ثانویه است. مطالعه بر روی مگس سرکه نشان می‌دهد که پروتئین‌های گروه پلی کامب بر ایجاد الگوهای بیان ژن‌های مسئول رشد تأثیر نمی‌گذارند، اما برای حفظ آن الگوها در طول زمان ضروری است. با این حال، CGI‌های خاموش از نظر رونویسی ممکن است همیشه برای دعوت پلی کامب کافی نباشند، زیرا پیشنهاد می‌شود که بکارگیری PRC2 به رونویسی ncRNA‌ها از CGI بستگی دارد (۹۵، ۱۳۳). گزارش شده است که رونوشت‌های کوتاه تولید شده از CGI‌های دو ظرفیتی در سلول‌های بنیادی جنین و سلول‌های T ساختارهای حلقه بنیادی را تشکیل می‌دهند که ممکن است در به کارگیری پلی کمب برای این CGI‌ها دخیل باشند (۵۹، ۶۰).

جزایر CpG و سرطان

مکانیسم‌های بیولوژیکی که به حفظ حالت متیله نشده CGI کمک می‌کنند، هنوز ناشناخته است، اما تغییر الگوهای متیلاسیون DNA یک ویژگی مشترک در همه انواع تومورها است. در تومورها، رویدادهای متیلاسیون در پروموتور CGI، نواحی کناری CpG، تقویت‌کننده‌ها و جداسازها رخ می‌دهند (۸۹). به طور کلی، این هایپرمتیلاسیون نابجای DNA در چنین مناطقی تنظیمی ژنوم، با سرکوب ژن‌های سرکوبگر تومور، متاستاز و ژن‌های ترمیم‌کننده DNA مرتبط است (۱۲۹). با این حال، داده‌های تجربی در مورد مشخصات وضعیت متیلاسیون بافت‌های نرمال و انواع سرطان نشان می‌دهد که هایپرمتیلاسیون نابجای DNA نشان‌دهنده یک رویداد ثانویه تثبیت‌کننده حالت غیرفعال شدن ژن است، زیرا اغلب ژن‌های هایپرمتیله نابجا در سرطان قبلاً در بافت منشا سرکوب شده‌اند (۴۵، ۱۰۷). با این وجود، هایپرمتیلاسیون DNA نشان‌دهنده مکانیسمی است که به طور مستقیم بر بیان ژن‌های مهم سرکوبگر تومور تأثیر می‌گذارد (۳۵، ۹۹). با بررسی متیلاسیون CGI‌های چندین ژن سرکوبگر تومور مخصوص سرطان مشخص شد که بسیاری از ژن‌های دخیل در اشکال خانوادگی سرطان تحت خاموشی مرتبط با متیلاسیون DNA در سرطان‌های متفاوت قرار می‌گیرند (۵۸). تصور می‌شود که این تغییرات به تکثیر کنترل نشده و در نتیجه توسعه تومور کمک می‌کند. با این حال، این که آیا متیلاسیون DNA رویداد آغازین در خاموش کردن ژن است یا در ژن‌هایی که قبلاً خاموش شده‌اند به دست می‌آید، نامشخص است. مطالعات گسترده ژنوم نشان داده است که بسیاری از CGI‌هایی که متیلاسیون غیرطبیعی را در سرطان به دست آورده‌اند با ژن‌های سرکوبگر تومور مرتبط نبودند (۵۲، ۸۳، ۹۷، ۱۲۱). برای مثال، نتایج یک مطالعه نشان داد که در میان نمونه‌های تومور سرطان بافت سنگفرشی مری، درصد متیلاسیون برای ژن‌های سرکوبگر تومور (p15، p16 و p14) به ترتیب ۹۰، ۵۳ و ۲۷ درصد بود در حالی که در هیچ یک از نمونه‌های بافت نرمال (مجاور تومور) متیلاسیونی مشاهده نشد (۸۳). پروفایل متیلاسیون DNA در نواحی کناری CGI نشان می‌دهد که الگوهای متیلاسیون خاص سرطان شبیه به آنهایی است که در بافت‌های طبیعی رخ می‌دهد (۵۳). برخلاف بافت‌های طبیعی، که معمولاً CGI‌های ارفان متیله می‌شوند، در سرطان‌های کولورکتال، نسبت CGI‌های متیله شده در جایگاه‌های شروع رونویسی، به CGI‌های ارفان متیله شده در تومور تقریباً برابر بود (۵۲).

جزایر CpG و پیری

از آنجایی که متیلاسیون DNA مکانیسم مهمی در تنظیم اپی‌ژنتیکی می‌باشد، نقش این فرآیند در مراحل اولیه در سلول‌های پیر مورد مطالعه قرار گرفته است. بررسی سطح متیل سیتوزین در DNA، هایپومتیلاسیون گسترده مرتبط با سن را در پروموتورهای ژن‌های خاصی مانند IGTAL و IL17RC گزارش شد (۲۰، ۱۲۲). همزمان با هایپومتیلاسیون عمومی و موضعی DNA در طول فرآیند پیری، پیش‌بینی می‌شود که هایپرمتیلاسیون در نواحی CpG ژنوم برای سرکوب بیان ژن‌های خاصی رخ دهد بطوری که هایپرمتیلاسیون وابسته به سن در جزایر CpG واقع در پروموتورها برای چندین ژن گزارش شده است (۴۷، ۱۲۳). اولین ژنی که برای رابطه بین پیری و متیلاسیون جزایر CpG در بافت‌های نرمال مشخص شد، ژن گیرنده استروژن بود. متیلاسیون این ژن در مخاط طبیعی روده بزرگ به طور خطی با افزایش سن افزایش می‌یابد و در سلول‌های سرطانی فراوانی آن نزدیک به ۱۰۰٪ می‌باشد. لاین‌های سلولی سرطان سینه و روده بزرگ که چنین متیلاسیونی را نشان می‌دهند، ژن گیرنده استروژن را بیان نمی‌کنند و بیان mRNA را می‌توان با درمان با مهارکننده‌های متیلاسیون بازیابی کرد. این نتایج نشان داد که هایپرمتیلاسیون پروموتور ژن گیرنده استروژن ارتباط نزدیکی با خاموش کردن رونویسی این ژن دارد (۵۴). علاوه بر این، پیری سلولی با اختلال در لامینای هسته‌ای و وضعیت هتروکروماتین مرتبط با آن مشخص می‌شود. این که چگونه این تغییرات ساختاری در هسته با افزایش سن به انحطاط موجود کمک می‌کند، هنوز بطور کامل مشخص نشده است. ژن‌های فاقد جزایر CpG معمولاً وقتی غیرفعال هستند با هتروکروماتین مرتبط می‌شوند و بیان این ژن‌ها به طور گسترده در سلول‌ها و بافت‌های مسن فعال می‌شود بطوری که این بیان نادرست ژن‌های فاقد CGI از ویژگی مشترک پیری طبیعی و پاتولوژیک در موش و انسان است (۶۶). افزایش بیان ژن‌های فاقد CGI، به ویژه برای افزایش ترشح واسطه‌های التهابی، با اختلال در معماری هسته و تغییرات اپی‌ژنتیکی در سلول‌های پیر، مستقیماً مسئول زوال فیزیولوژیکی مرتبط با سن است (۶۶، ۷۰). علاوه بر این، بیان نادرست ژن‌های فاقد CGI در سلول‌های پیر، منجر به تغییر مکانیسم‌های مولکولی زیربنایی، از جمله از دست دادن هویت عملکردی و افزایش اختلال در رونویسی می‌شوند. به طور خاص، بخش بزرگی از ژن‌های فاقد CGI که به اشتباه بیان می‌شوند، پروتئین‌هایی را کد می‌کنند، که بسیاری از آنها با کلیه‌ها و قلب‌های مسن در موش مرتبط هستند (۶۶).

جزایر CpG و کووید-۱۹

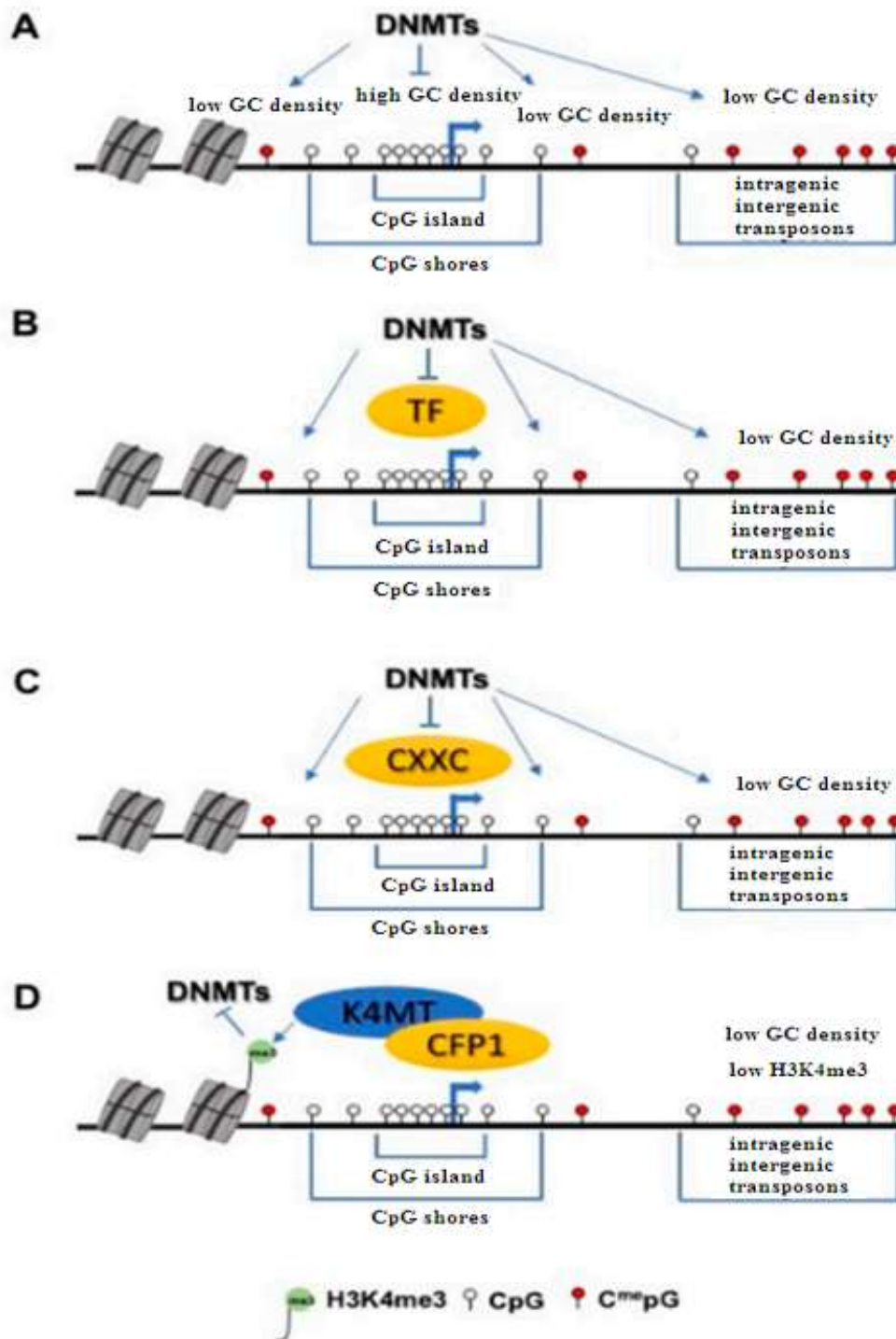
بسیاری از RNA ویروس‌ها، از جمله SARS-CoV-2، از نظر دی‌نوکلئوتیدهای CpG فقیرمی‌باشند (۱۲۶). دو نظریه اصلی برای توضیح این کمبود CpG ارائه شده است. یکی بر اساس حساسیت به جهش است، زیرا متیلاسیون سیتوزین سرعت جهش را با دامیناسیون خود به خودی ۵-متیل سیتوزین به تیمین افزایش می‌دهد (۲۱). فرضیه دیگر بر تعامل با سیستم ایمنی میزبان متمرکز است، زیرا ویروس‌ها سعی می‌کنند فراوانی‌های CpG میزبان و الگوهای متیلاسیون را با خود تطابق دهند. برای مثال، کاهش سریع فراوانی CpG در ویروس آنفلوانزا پس از انتقال از پرندگان به انسان مشاهده شده است (۳۸). در مهره داران و به ویژه در ویروس‌های آلوده‌کننده انسان، فراوانی CpG بسیار کم است (۲۱) بطوری که فراوانی بالاتر CpG با تضعیف ویروس مرتبط است (۱۱۵).

بطور کلی، متیلاسیون CpG با بیان ژن همبستگی معکوس دارد بطوری که سطوح بالای متیلاسیون در CpG‌های مرتبط با پروموتور منجر به سطوح پایین بیان ژن می‌شود. همچنین قبلاً گزارش شده است که عفونت‌های ویروسی الگوهای متیلاسیون نابجا را در سلول‌های میزبان القا می‌کنند (۴، ۸۷). تفاوت بسیاری در نواحی CpG هایپو و هایپر متیله شده در نمونه‌های به‌دست‌آمده در طی چند روز پس از عفونت حاد SARS-CoV-2، در مقایسه با نمونه‌های خون سالم جمع‌آوری شده قبل از همه‌گیری COVID-19 مشاهده شد (۷). در یک مطالعه روی بیماران مبتلا به عفونت SARS-CoV-2، تعداد ۶ جایگاه متفاوت متیله شده در ژن‌های کدکننده پروتئین‌هایی مشاهده شد که در گرانولوسازی و تمایز لنفوسیت B به گرانولوسیت شرکت می‌کنند (۹). همچنین تعداد ۴۴ جایگاه CpG در ۳۹ ژن، از جمله ژن‌های مربوط به پاسخ ایترفرون به عفونت‌های ویروسی، شناسایی شدند که به طور متفاوت در بیماران کووید ۱۹ و افراد سالم متیله شده بودند (۲۸). همچنین، از دلایل بیان بالای ACE2 در ریه‌ها، کبد و مغز، در جزایر CpG کم متیله در پروموتور ACE2 در این اندام‌ها نهفته است و آنها را در برابر آسیب‌های ناشی از SARS-CoV-2 آسیب‌پذیرتر می‌کند (۱۰۴). این فرضیه توسط مطالعه‌ی دیگری حمایت شده است که به این نتیجه رسیده است که ردپای تغییرات اپی‌ژنتیکی در مراحل مختلف پاتوژنز SARS-CoV-2، به ویژه ادغام غشای سلولی با ویروس وجود دارد (۹۱).

جزایر CpG و بیماری‌های خود ایمنی

متیلاسیون DNA و پروفایل‌های بیان ژن برای هر نوع سلول، از جمله زیر مجموعه‌های سلول‌های ایمنی، منحصر به فرد و خاص است (۱۳۴). بررسی متیلاسیون و بیان ژن در سلول‌های تک هسته‌ای خون محیطی (PBMCs) روند متفاوتی در وضعیت متیلاسیون DNA در سلول‌های T CD4+ و CD8+ نشان داد که این تفاوت عمدتاً در پایین دست پروموتورهای حاوی CGI، آگرون‌ها و اینترون‌ها مشاهده شد (۷۵). برخلاف انتظار، وضعیت متیلاسیون جایگاه‌های CpG حاوی پروموتور فراوانی‌های کمتری نسبت به کل جایگاه‌های CpG ژنوم نشان داد (۷۵). همچنین در بیماری‌های خود ایمنی روماتیسم مفصلی (Rheumatoid Arthritis: RA) و لوپوس اریتماتوز سیستمیک (Systemic Lupus Erythematosus: SLE) سطوح متیلاسیون متفاوت ژنی، عمدتاً جایگاه‌های CpG هایپومتیله شده، مشاهده شد که ممکن است با فعالیت بیماری مرتبط باشد (۳۴). در بیماری RA، تعداد ۱۸۵۹ ژن با شرایط متیلاسیون متفاوت شناسایی شدند که بیانگر نقش مهم متیلاسیون DNA در ایجاد این بیماری می‌باشد (۸۵). جایگاه‌های CpG هایپومتیله در ژن‌های کلیدی مرتبط با RA، مانند ژن STAT3 شناسایی شدند. فعالیت ژن STAT3 با افزایش سطح IL-6 پیش التهابی مرتبط است، که نقش مهمی در التهاب مزمن در بیماری‌های خودایمنی، از جمله RA دارد (۴۶). مطالعات دیگری شرایط متفاوت متیلاسیون را در سلول‌های T و B بیماران مبتلا به RA اولیه در مقایسه با افراد سالم شناسایی کرده‌اند (۳۷، ۹۴). در بیماری SLE تعداد ۷۲۴۵ جایگاه CpG با شرایط متیله متفاوت در ژنوم شناسایی شد که بیشترین تفاوت متیلاسیون در ژن‌های تنظیم شده با ایترفرون نوع I مشاهده شد که متیلاسیون کاهش یافته را در SLE نشان دادند (۶۴). در مطالعات دیگری نیز هایپومتیلاسیون ژن‌های مرتبط با سیستم ایمنی را در بیماران SLE در مقایسه با افراد سالم گزارش داده‌اند (۲۲، ۵۲). بطور کلی، بسیاری از سلول‌های ایمنی کاهش متیلاسیون DNA را در بین ژن‌های پیش التهابی در طول بیماری‌های خود ایمنی نشان دادند که ممکن است با القای بیان ژن مرتبط باشد (۱۳۵). با این حال، در مورد نقش تغییرات اپی‌ژنتیکی در بیماری‌های خودایمنی چیزهای زیادی وجود دارد.

شناسایی تغییرات اصلی اپی‌ژنتیک، استفاده از آنها را به عنوان نشانگرهای زیستی برای خطر اپی‌ژنتیکی بهبود می‌بخشد و به درمان‌های مبتنی بر اصلاح تغییرات اپی‌ژنتیک کمک می‌کند.



شکل ۳- مهم‌ترین مکانیسم‌های حفاظت از جزایر CpG در مقابل متیلاسیون (A): اثر چگالی GC و فراوانی CpG بر آنزیم‌های DNA متیل ترانسفراز (B); تاثیر فاکتورهای رونویسی بر آنزیم‌های DNA متیل ترانسفراز (C); اثر دامنه CXXC حاوی پروتئین بر آنزیم‌های DNA متیل ترانس فراز (D); نمونه‌ای

از اثر ایجاد شده توسط جذب متیل ترانسفرازهای هیستون H3K4 (K4MT) توسط پروتئین حاوی دامنه CXXC (CFP1) بر روی فعالیت آنزیم‌های DNA متیل ترانسفراز (۱۲۹).

مکانیسم‌های محافظت CGIها در مقابل متیلاسیون

CGIها عموماً دامنه‌های غنی شده با CpG غیر متیله هستند که در پس زمینه متیلاسیون DNA در سطح ژنوم و در نتیجه کاهش CpG رخ می‌دهند. چگونه آنها از امواج متیلاسیون خودبخودی DNA که در طول رشد و نمو رخ می‌دهد محافظت می‌شوند؟ یک توضیح ساده می‌تواند این باشد که CGIها ذاتاً در برابر عمل آنزیم‌های DNA متیل ترانسفرازها (DNMTs) مقاوم هستند، اما این احتمال بسیار کم است. برخی از CGIها در طول رشد و نمو طبیعی متیله می‌شوند. به عنوان مثال، صدها CGI روی یک کروموزوم X ماده به شدت متیله شده اند، اما در کروموزوم دیگر غیر متیله هستند. یک احتمال دیگر این است که CGI به طور فعال توسط یک آنزیم DNA دمتیلاز حذف می‌شود (۱۲۵)، اما عدم اطمینان در مورد هویت فعالیت‌های دی‌متیلاسیون DNA در حیوانات، آزمایش این فرضیه را دشوار کرده است.

کشف این که ۵-متیل سیتوزین (5mC) می‌تواند به ۵-هیدروکسی متیل سیتوزین (5hmC) تبدیل شود، احتمال و سوسه انگیزی را ایجاد می‌کند که هیدروکسی متیل سیتوزین (hmC) یک واسطه در فرآیند دی‌متیلاسیون است (۵۵، ۱۲۵). آنزیم دی‌متیله کننده Tet1 دارای یک دامنه مربوط به CXXC متصل به CpG است که در Cfp1 یافت می‌شود (۱۱۰)، و در تئوری می‌تواند CGIها را هدف قرار دهد. مطابق با این فرضیه، کاهش Tet1 منجر به افزایش متیلاسیون CpG در CGI سلول‌های بنیادی موش گردید (۱۲۴). یک سناریوی جذاب این است که CGIها در معرض متیلاسیون جدید پراکنده قرار می‌گیرند، اما به طور مداوم توسط مکانیزمی شامل اکسیداسیون 5mC پاک می‌شوند. نقص در چنین سیستمی ممکن است مستعد متیلاسیون جدید CGI باشد، همانطور که در بسیاری از سرطان‌ها دیده می‌شود. در واقع، جهش در ژن TET2 در لوسمی‌ها بوفور دیده می‌شود که واکنش هیدروکسیلاسیون را به خطر می‌اندازد (۶۱). البته، دیدگاه رایج این است که فراوانی دی‌نوکلئوتیدهای CpG و محتوای C و G (شکل ۳) یک ویژگی کلیدی مورد نیاز برای جلوگیری از متیلاسیون در CGI است (۱۱۲، ۱۱۹). صرف نظر از مکانیسم مولکولی دقیق، شواهدی وجود دارد که وضعیت عاری از متیلاسیون CGIها به طور علی با عملکرد آنها به عنوان پروموتور مرتبط است. ویژگی‌های توالی و اتصال فاکتور رونویسی خاص ممکن است تعیین کننده بکارگیری پروتئین‌های دیگری باشد که عملکردهای آنها را یکپارچه کنند. در این دیدگاه، تمام پروتئین‌های حاوی یک دامنه CXXC، مانند MLL1، MLL2، KDM2A، KDM2B، TET1 و TET3، که به طور خاص CGI را هدف قرار می‌دهد (شکل ۳)، ممکن است نقش احتمالی در محافظت از متیلاسیون CGI داشته باشند (۶۸). حذف یا جهش جایگاه‌های اتصال به فاکتور رونویسی Sp1 در پروموتور Aprt موش منجر به حذف حالت متیله CGI در پروموتور Aprt در آزمایشات تراریخته شد (۷۴). این نشان می‌دهد که اتصال فاکتورهای رونویسی یا فرآیند رونویسی در طول رشد و نمو اولیه برای ایجاد حالت عاری از متیلاسیون DNA مورد نیاز است (شکل ۳). این ایده که رونویسی با متیلاسیون CGI متضاد است نیز به خوبی با این مشاهدات مطابقت دارد که وجود RNAPII در CGIها، صرف نظر از فعالیت ژن، با مقاومت در برابر متیلاسیون DNA در سرطان مرتبط است (۱۱۱). فاکتورهای رونویسی اختصاصی توالی DNA نیز در جلوگیری از متیلاسیون DNA نقش دارند، زیرا اتصال فاکتورهای رونویسی Sp1، Nrf-1، و YY1 در مونوسیت‌های طبیعی، از متیلاسیون CGI در سلول‌های لوسمی حفاظت می‌کند (۳۶).

در این مورد، پروموتورهای CGI محدود به این عوامل نیز بالاترین سطح بیان را داشتند. بررسی CGIهای انسانی همبستگی را با اتصال فاکتور رونویسی تقویت کرد. زیرمجموعه‌های CGI که در طول رشد و نمو متیله یا غیرمتیله می‌شوند، برای موتیف‌های توالی DNA که ممکن است بر این تصمیم تأثیر بگذارند، غربالگری شدند. یک الگوریتم مبتنی بر این موتیف‌ها، که تعدادی از آنها محل اتصال فاکتور رونویسی بودند، قادر به پیش‌بینی این که کدام CGI از متیلاسیون DNA مصون می‌ماند، نشان می‌دهد که سوئیچ‌ها به فاکتورهای عامل سیس وابسته هستند (۱۰۸). توضیح احتمالی دیگر برای مصونیت بیشتر CGIها در برابر متیلاسیون DNA این است که علائم کروماتینی،

H3K4me3، با فعالیت آنزیم DNMTs تداخل می‌کند (شکل ۳). اتصال خودبخودی Dnmt3L، پروتئین شریک آنزیم‌های DNA متیل ترانسفرازهای Dnmt3a و Dnmt3b، به کروماتین توسط H3K4me3 مهار می‌شود (۸۶). نکته مهم این است که دامنه‌های ADD هر دو Dnmt3a و Dnmt3b نیز نمی‌توانند با H3K4me3 تعامل داشته باشند و در مقایسه با کروماتین تغییر یافته و یا اصلاح‌شده با H3K9me3 در شرایط *in vitro* بر روی کروماتین حاوی این تغییرات کمتر فعال هستند (۱۳۲). قابل توجه است که یک توالی غنی از CpG بدون پروموتور مشخص شده با اتصال به Cfp1 و H3K4me3 تنها تا حدی در برابر متیلاسیون DNA مصون بود، که نشان می‌دهد این اصلاح کروماتین به خودی خود کافی نیست (۱۱۳). ممکن است ترکیبی از عوامل، مانند عوامل شروع رونویسی، برای حذف متیلاسیون DNA از CGI مورد نیاز باشد.

نتیجه‌گیری

در پستانداران، یکی از ویژگی‌های بارز پروموتورها، وجود جزایر CpG (CGIs) در بیش از دو سوم ژن‌ها می‌باشد، در حالی که جعبه‌های TATA تنها در تعداد کمی از پروموتورها وجود دارند. بطورکلی، جزایر CpG گروهی از توالی‌های پراکنده DNA، اما منسجمی هستند که در تنظیم بیان ژن نقش مهمی بعهده دارند. عملکرد CGI‌ها به ویژگی‌های مشترک توالی DNA آنها، به ویژه، غنای CpG و محتوای G+C بالاتر از حد متوسط آنها بستگی دارد. اگرچه، نقش تنظیمی CGI‌ها به طور کامل شناخته شده نیست، اما تصور می‌شود که این فرآیند از طریق ایجاد حالت‌های مجاز کروماتین جهت رونویسی با بی‌ثبات کردن نوکلئوزوم‌ها و جذب پروتئین‌های متصل به DNA باشد. با استفاده از رویکردهای گسترده ژنومی، مشخص شده است که محتوای GC و عناصر CGI پروموتور در سلول‌های پستانداران قادر به کنترل ساختار کروماتین باز بوده و از عدم رونویسی پشتیبانی می‌کنند. در پستانداران، معمولاً اکثر CGI‌ها متیله نشده هستند و اغلب CGI‌های متیله شده در اینترون‌ها، نواحی بین ژنی و کدکننده ژن وجود داشته در حالی که کمترین تعداد CGI‌های متیله در نواحی پروموتور وجود دارد. این یافته‌ها تایید می‌کنند که متیلاسیون DNA در CGI‌ها ممکن است بیان ژن را در سطح ژنوم تنظیم کند. البته متیلاسیون خودبخودی جزایر CpG می‌تواند منجر به خاموش شدن برگشت ناپذیر رونویسی یک ژن شده، و بسته به ژن و زمان تحت تأثیر، ممکن است عواقب مخربی برای سلول یا موجود زنده داشته باشد.

منابع

1. Angeloni A. Bogdanovic O (2021). Sequence determinants, function, and evolution of CpG islands. *Biochem Soc Trans*. 49(3):1109-1119. <https://doi.org/10.1042/BST20200695>.
2. Amante SM. Montibus B. Cowley M. Barkas N. Setiadi J. Saadeh H. Gienza. J. Contreras-Castillo S. Fleischanderl K. Schulz R. Oakey RJ (2020). Transcription of intragenic CpG islands influences spatiotemporal host gene pre-mRNA processing. *Nucleic Acids Res*. 48: 8349-8359. <https://doi.org/10.1093/nar/gkaa556>.
3. Antequera F. Bird AP (1988). Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO J*. 7: 2295-229. <https://doi.org/10.1002/j.1460-2075.1988.tb03072.x>.
4. Atlante S. Mongelli A. Barbi V. Martelli F. Farsetti A. Gaetano C (2020). The epigenetic implication in coronavirus infection and therapy. *Clin Epigenetics* 12(1):156-167. <https://doi.org/10.1186/s13148-020-00946-x>.
5. Azuara V. Perry P. Sauer S. Spivakov M. Jorgensen HF. John RM. Gouti M. Casanova M. Warnes G. Merckenschlager M. Fisher AG (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*. 8: 532-538. <https://doi.org/10.1038/ncb1403>.
6. Babenko VN. Kosarev PS. Vishnevsky OV (1999). Investigating extended regulatory regions of genomic DNA sequences. *Bioinformatics*. 15(7): 644-653. <https://doi.org/10.1093/bioinformatics/15.7.644>.
7. Balnis J. Madrid A. Hogan KJ. Drake LA. Chieng HC. Tiwari A. Vincent CE. Chopra A. Vincent PA. Robek MD. Singer HA (2021). Blood DNA methylation and COVID-19 outcomes. *Clinical epigenetics*, 13(1): 118-133. <https://doi.org/10.1186/s13148-021-01102-9>.
8. Beck S. Rhee C. Song J. Lee BK. LeBlanc L. Cannon L. Kim J (2018). Implications of CpG islands on chromosomal architectures and modes of global gene regulation. *Nucleic Acids Res*. 46(9): 4382-4391. <https://doi.org/10.1093/nar/gky147>.
9. Bernardes JP. Mishra N. Tran F. Bahmer T. Best L. Blase JI. Bordoni D. Franzenburg J. Geisen U. Josephs-Spaulding J. Köhler P (2020). Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* 53(6):1296-314. <https://doi.org/10.1016/j.immuni.2020.11.017>.
10. Bernstein BE. Mikkelsen TS. Xie X. Kamal M. Huebert DJ. Cuff J. Fry B. Meissner A. Wernig M. Plath K. et al (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 125: 315-326. <https://doi.org/10.1016/j.cell.2006.02.041>.
11. Bird AP (1978). Use of restriction enzymes to study eukaryotic DNA methylation. *J Mol Biol*. 118: 49-60. [https://doi.org/10.1016/0022-2836\(78\)90243-7](https://doi.org/10.1016/0022-2836(78)90243-7).

12. Bird AP (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8: 1499–1504. <https://doi.org/10.1093/nar/8.7.1499>.
13. Bird AP, Taggart MH, Smith BA (1979). Methylated and unmethylated DNA compartments in the Sea Urchin genome. *Cell.* 17: 889–901. [https://doi.org/10.1016/0092-8674\(79\)90329-5](https://doi.org/10.1016/0092-8674(79)90329-5).
14. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ (2010). CpG islands recruit a histone H3 lysine 36 demethylases. *Mol Cell.* 38: 179–190. <https://doi.org/10.1016/j.molcel.2010.04.009>.
15. Bogdanovic O, Veenstra GJ (2009). DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma.* 118: 549–565. <https://doi.org/10.1007/s00412-009-0221-9>.
16. Butler JE, Kadonaga JT (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16: 2583–2592. <https://doi.org/10.1101/gad.1026202>.
17. Cain JA, Montibus B, Oakey RJ (2022). Intragenic CpG islands and their impact on gene regulation. *Front Cell Dev Biol.* 10: 832348. <https://doi.org/10.3389/fcell.2022.832348>.
18. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, et al (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet.* 38: 626–635. <https://doi.org/10.1038/ng1789>.
19. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell.* 116: 499–509. [https://doi.org/10.1016/S0092-8674\(04\)00127-8](https://doi.org/10.1016/S0092-8674(04)00127-8).
20. Cedar H, Bergman Y (2012) Programming of DNA methylation patterns. *Annu Rev Biochem* 81:97-117. <https://doi.org/10.1146/annurev-biochem-052610-091920>.
21. Cheng X, Virk N, Chen W, Ji S, Ji S, Sun Y, Wu X (2013) CpG usage in RNA viruses: data and hypotheses. *PLoS one* 8(9):e74109. <https://doi.org/10.1371/journal.pone.0074109>.
22. Chung SA, Nititham J, Elboudwarej E, Quach HL, Taylor KE, Barcellos LF, Criswell LA (2015) Genome-wide assessment of differential DNA methylation associated with autoantibody production in systemic lupus erythematosus. *PLoS one.* 10(7): e0129813. <https://doi.org/10.1371/journal.pone.0129813>.
23. Cierpicki T, Risner LE, Grembecka J, Lukasik SM, Popovic R, Omonkowska M, Shultis DD, Zeleznik Le, NJ, Bushweller JH (2010). Structure of the MLL CXXC domain–DNA complex and its functional role in MLL–AF9 leukemia. *Nat Struct Mol Biol.* 17(1): 62–68. <https://doi.org/10.1038/nsmb.1714>.
24. Core LJ, Waterfall JJ, Lis JT (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* 322: 1845–1848. <https://doi.org/10.1126/science.11622>.
25. Cross S, Kovarik P, Schmidtke J, Bird AP (1991). Non-methylated islands in fish genomes are GC-poor. *Nucleic Acids Res.* 19:1469–1474. <https://doi.org/10.1093/nar/19.7.1469>.
26. Csankovszki G, Nagy A, Jaenisch R (2001). Synergism of Xist RNA, DNA methylation and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol.* 153: 773–784. <https://doi.org/10.1083/jcb.153.4.773>.
27. Cuadrado M, Sacristan M, Antequera F (2001). Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep.* 2: 586–592. <https://doi.org/10.1093/embo-reports/kve131>.
28. de Moura MC, Davalos V, Planas-Serra L, Alvarez-Errico D, Arribas C, Ruiz M, Aguilera-Albesa S, Troya J, Valencia-Ramos J, Vélez-Santamaria V, Rodríguez-Palmero A (2021) Epigenome-wide association study of COVID-19 severity with respiratory failure. *EBioMedicine.* 1:66. <https://doi.org/10.1016/j.ebiom.2021.103339>.
29. De Smet C, Lurquin C, Lethe B, Martelange V, Boon T (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol.* 19: 7327–7335. <https://doi.org/10.1128/MCB.19.11.7327>.
30. Deaton AM, Bird A (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25: 1010–1022. <http://www.genesdev.org/cgi/doi/10.1101/gad.2037511>.
31. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, et al (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18: 1433–1445. <http://www.genome.org/cgi/doi/10.1101/gr.078378.108>.
32. Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A, et al (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell.* 38: 452–464. <https://doi.org/10.1016/j.molcel.2010.02.032>.
33. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci.* 107: 8689–8694. <https://doi.org/10.1073/pnas.1002720107>.
34. Funes SC, Fernández-Fierro A, Rebolledo-Zelada D, Mackern-Oberti JP, Kalergis AM (2021) Contribution of dysregulated DNA methylation to autoimmunity. *Int J Mol Sci* 22(21):11892. <https://doi.org/10.3390/ijms222111892>.
35. Galamb O, Kalmár A, Péterfia B, Csabai I, Bodor A, Ribli D, Krenács T, Patai ÁV, Wichmann B, Barták BK, et al (2016). Aberrant DNA methylation of WNT pathway genes in the development and progression of CIMP-negative colorectal cancer. *Epigenetics.* 11: 588–602. <https://doi.org/10.1080/15592294.2016.1190894>.
36. Gebhard C, Benner C, Ehrlich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreessen R, et al (2010). General transcription factor binding at CpG islands in normal cells correlates with resistance to de novo DNA methylation in cancer cells. *Cancer Res.* 70: 1398–1407. <https://doi.org/10.1158/0008-5472.CAN-09-3406>.
37. Glossop JR, Glossop JR, Emes RD, Nixon NB, Nixon NB, Haworth KE, Haworth KE, Packham JC, Packham JC, Dawes PT (2014) Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T-and B-lymphocyte populations. *Epigenetics* 9(9):1228-1237.
38. Gu H, Fan RL, Wang D, Poon LL (2019) Dinucleotide evolutionary dynamics in influenza A virus. *Virus Evol* 5(2):vez038. <http://dx.doi.org/10.4161/epi.29718>.
39. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell.* 130: 77–88. <https://doi.org/10.1016/j.cell.2007.05.042>.

40. Haberer V, Stark A (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 19: 621–637. <https://doi.org/10.1038/s41580-018-0028-8>.
41. Hargreaves DC, Horng T, Medzhitov R (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell*. 138: 129–145. <https://doi.org/10.1016/j.cell.2009.05.047>.
42. Henikoff S (2000). Heterochromatin function in complex genomes. *Biochim Biophys Acta*. 1470: 1–8.
43. Hergeth SP, Schneider R (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep*. 16(11):1439–1453. <https://doi.org/10.15252/embr.201540749>.
44. Herzog LB, Romer JT, Horn JM, Ashworth A (1997). Xist has properties of the X-chromosome inactivation centre. *Nature*. 386: 272–275.
45. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun H, van den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM, et al (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*. 2012. 22: 271–282. <http://www.genome.org/cgi/doi/10.1101/gr.117523.110>.
46. Hirano T (2010). Interleukin 6 in autoimmune and inflammatory diseases: A personal memoir. *Proc Jpn Acad Ser B Phys Biol Sci* 86: 717–730. <https://doi.org/10.2183/pjab.86.717>.
47. Horvath S (2013). DNA methylation age of human tissues and cell types. *Genome Biol* 4: R115. <http://genomebiology.com/14/10/R115>.
48. Hu D, Gao X, Cao K, Morgan MA, Mas G, Smith ER, Volk AG, Bartom ET, Crispino JD, Di Croce L, Shilatifard A (2017). Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. *Molecular Cell*. 65(3): 460–475. <https://doi.org/10.1016/j.molcel.2017.01.013>.
49. Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, et al (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol*. 6: e22. <http://doi.org/10.1371/journal.pbio.0060022>.
50. Illingworth RS, Bird AP (2009). CpG islands ‘a rough guide.’ *FEBS Lett*. 583: 1713–1720. <https://doi.org/10.1016/j.febslet.2009.04.012>.
51. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet*. 6: e1001134. <http://doi.org/10.1371/journal.pgen.1001134>.
52. Imgenberg-Kreuz J, Almlöf JC, Leonard D, Alexsson A, Nordmark G, Eloranta ML, Rantapää-Dahlqvist S, Bengtsson AA, Jönsen A, Padyukov L, Gunnarsson I (2018) DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Ann Rheum Dis* 77(5): 736–43. <https://doi.org/10.1136/annrheumdis-2017-212379>.
53. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*. 41: 178–186. <http://doi.org/10.1038/ng.298>.
54. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* 7(4):536-40. <https://doi.org/10.1038/ng0894-536>.
55. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010). Role of Tet proteins in 5mC to 5hmC conversion. ES cell self-renewal and inner cell mass specification. *Nature*. 466: 1129–1133. <http://doi.org/10.1038/nature09303>.
56. Jaeger SA, Chan ET, Berger MF, Stottmann R, Hughes TR, Bulyk ML (2010). Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics*. 95: 185–195. <https://doi.org/10.1016/j.ygeno.2010.01.002>.
57. Jones PA, Baylin SB (2002). The fundamental role of epigenetic events in cancer. *Nat Rev*. 3: 415–428. <http://doi.org/10.1038/nrg816>.
58. Jones PA, Baylin SB (2007). The epigenomics of cancer. *Cell*. 128: 683–692. <https://doi.org/10.1016/j.cell.2007.01.029>.
59. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT (2008). The RNA polymerase II core promoter—the gateway to transcription. *Curr Opin Cell Biol*. 20: 253–259. <http://doi.org/10.1016/j.ceb.2008.03.003>.
60. Kanhere A, Viiri K, Araujo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, et al (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell*. 38: 675–688. <https://doi.org/10.1016/j.molcel.2010.03.019>.
61. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, et al (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*. 468: 839–843. <http://doi.org/10.1038/nature09586>.
62. Kornblihtt AR (2006). Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol*. 13: 5–7. <http://doi.org/10.1038/nsmb0106-5>.
63. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4: e1000242. <http://doi.org/10.1371/journal.pgen.1000242>.
64. Lanata CM, Nititham J, Taylor KE, Solomon O, Chung SA, Blazer A, Trupin L, Katz P, Dall'Erà M, Yazdany J, Sirota M (2022). Dynamics of methylation of CpG sites associated with systemic lupus erythematosus subtypes in a longitudinal cohort. *Arthritis Rheumatol* 74(10):1676-86. <https://doi.org/10.1002/art.42237>.
65. Landolin JM, Johnson DS, Trinklein ND, Aldred SF, Medina C, Shulha H, Weng Z, Myers RM (2010). Sequence features that drive human promoter function and tissue specificity. *Genome Res*. 20: 890–898. <http://www.genome.org/cgi/doi/10.1101/gr.100370.109>.
66. Lee JY, Davis I, Youth EH, Kim J, Churchill G, Godwin J, Korstanje R, Beck S (2021). Misexpression of genes lacking CpG islands drives degenerative changes during aging. *Sci Adv* 7(51):eabj9111. <https://doi.org/10.1126/sciadv.abj9111>.
67. Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL, Shilatifard A (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem*. 284: 7970–7976. <https://doi.org/10.1074/jbc.M808220200>.

68. Long HK, Blackledge NP, Klose RJ (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans*. 41: 727–740. <https://doi.org/10.1042/BST20130028>.
69. Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grutzner F, Odom DT, Patient R, Ponting CP, Klose RJ (2013). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife*. 2: e00348. <http://doi/10.7554/eLife.00348>.
70. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013). The hallmarks of aging. *Cell* 153(6):1194–217. <https://doi.org/10.1016/j.cell.2013.05.039>.
71. Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol*. 11: 1068–1075. <http://doi/10.1038/nsmb840>.
72. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389(6648): 251–60.
73. Macleod D, Ali RR, Bird A (1998). An alternative promoter in the mouse major histocompatibility complex class II I-Ab gene: implications for the origin of CpG islands. *Mol Cell Biol*. 18: 4433–4443. <http://doi/10.1128/mcb.18.8.4433>.
74. Macleod D, Charlton J, Mullins J, Bird AP (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Dev*. 8: 2282–2292. <http://doi/10.1101/gad.8.19.2282>.
75. Mamrut S, Avidan N, Staun-Ram E, Ginzburg E, Truffault F, Berrih-Aknin S, Miller A (2015). Integrative analysis of methylome and transcriptome in human blood identifies extensive sex-and immune cell-specific differentially methylated regions. *Epigenetics* 10(10): 943–57. <https://doi.org/10.1080/15592294.2015.1084462>.
76. Mancini-DiNardo D, Steele SJ, Ingram RS, Tilghman SM (2003). A differentially methylated region within the gene *Kcnq1* functions as an imprinted promoter and silencer. *Hum Mol Genet*. 12: 283–294. <https://doi.org/10.1093/hmg/ddg024>.
77. Mancini-DiNardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006). Elongation of the *Kcnq1* lot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev*. 20: 1268–1282. <http://doi/10.1101/gad.1416906>.
78. Maunakea AK, Nagarajan RP, Bilienky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, et al (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 466: 253–257. <http://doi/10.1038/nature09165>.
79. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 454: 766–770. <http://doi/10.1038/nature07107>.
80. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet*. 6: e1001244. <http://doi/10.1371/journal.pgen.1001244>.
81. Mercer TR, Dinger ME, Mattick JS (2009). Long non-coding RNAs: insights into functions. *Nat Rev*. 10: 155–159.
82. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 448: 553–560. <http://doi/10.1038/nature06008>.
83. Mohammad Ganji S, Mahboudi S, Rastgar-Jazii F (2015). Investigation of the aberrant DNA methylation of p14, p15, p16 genes in patients with ESCC in Iran. *Cell Mol Res*. (Iranian Journal of Biology). 28(3): 403–412. <http://doi/10.1007/s12238-015-0109-9>. (In Persian).
84. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell*. 30: 755–766. <http://doi/10.1016/j.molcel.2008.05.007>.
85. Nakano K, Whitaker JW, Boyle DL, Wang W, Firestein GS (2013). DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis* 72(1):110–7. <https://doi.org/10.1136/annrheumdis-2012-201526>.
86. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, et al (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*. 448: 714–717. <http://doi/10.1038/nature05987>.
87. Paschos K, Allday MJ (2010). Epigenetic reprogramming of host genes in viral and microbial pathogenesis. *Trends Microbiol* 18:439–47. <https://doi.org/10.1016/j.tim.2010.07.003>.
88. Payer B, Lee JT (2008). X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet*. 42: 733–772. <https://doi.org/10.1146/annurev.genet.42.110807.091711>.
89. Pfeifer GP (2018). Defining driver DNA methylation changes in human cancer. *Int J Mol Sci*. 19: 1166. <http://doi/10.3390/ijms19041166>.
90. Pirrotta V, Li HB (2012). A view of nuclear Polycomb bodies. *Curr Opin Genet Dev*. 22(2): 101–109. <http://doi/10.1016/j.gde.2011.11.004>.
91. Pruijboom L (2020). Methylation pathways and SARS-CoV-2 lung infiltration and cell membrane-virus fusion are both subject to epigenetics. *Front Cell Infect Microbiol* 10: 550596. <https://doi.org/10.3389/fcimb.2020.00290>.
92. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010). c-Myc regulates transcriptional pause release. *Cell*. 141: 432–445. <https://doi.org/10.1016/j.cell.2010.03.030>.
93. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*. 138: 114–128. <https://doi.org/10.1016/j.cell.2009.04.020>.
94. Rhead B, Hologue C, Cole M, Shao X, Quach HL, Quach D (2017). Rheumatoid Arthritis naive T cells share hypermethylation sites with synoviocytes. *Arthritis Rheumatol* 69: 550–559. <https://doi.org/10.1002/art.39952>.
95. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 129: 1311–1323. <https://doi.org/10.1016/j.cell.2007.05.022>.
96. Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, FitzGerald PC, Vinson C (2008). All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genom*. 9: 67. <http://doi/10.1186/1471-2164-9-67>.

97. Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G (2010). Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genom*. 11: 137. <http://doi/10.1186/1471-2164-11-137>.
98. Sado T, Fenner MH, Tan SS, Tam P, Shiota T, Li E (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol*. 225: 294–303. <https://doi.org/10.1006/dbio.2000.9823>.
99. Saunderson EA, Stepper P, Gomm JJ, Hoa L, Morgan A, Allen MD, Jones JL, Gribben JG, Jurkowski TP, Ficiz G (2017). Hit-and-run epigenetic editing prevents senescence entry in primary breast cells from healthy donors. *Nat Commun*. 8: 1450. <http://doi/10.1038/s41467-017-01078-2>.
100. Saxonov S, Berg P, Brutlag DL (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA*. 103:1412–1417. <https://doi.org/10.1073/pnas.0510310103>.
101. Schilling E, Rehli M (2007). Global comparative analysis of tissue-specific promoter CpG methylation. *Genomics*. 90: 314–323. <https://doi.org/10.1016/j.ygeno.2007.04.011>.
102. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 132: 887–898. <https://doi.org/10.1016/j.cell.2008.02.022>.
103. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008). Divergent transcription from active promoters. *Science*. 322: 1849–1851. <http://doi/10.1126/science.1162253>.
104. Sen R, Garbati MR, Bryant K, Lu Y (2021). Epigenetic mechanisms influencing COVID-19. *Genome* 64(1): 1–14. <https://doi.org/10.1139/gen-2020-0135>.
105. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA, Issa JP (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet*. 3: 2023–2036. <https://doi.org/10.1371/journal.pgen.0030181>.
106. Sleutels F, Zwart R, Barlow DP (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*. 415: 810–813.
107. Sproul D, Nestor C, Culley J, Dickson J, Dixon JM, Harrison DJ, Meehan RR, Sims AH, Ramsahoye BH (2011). Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci USA*. 108: 4364–4369. <https://doi.org/10.1073/pnas.1013224108>.
108. Strausman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H (2009). Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol*. 16: 564–571. <http://doi/10.1038/nsmb.1594>.
109. Suzuki MM, Kerr AR, De Sousa D, Bird A (2007). CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res*. 17: 625–631. <http://doi/10.1101/gr.6163007>.
110. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930–935. <http://doi/10.1126/science.1170116>.
111. Takeshima H, Yamashita S, Shimazu T, Niwa T, Ushijima T (2009). The presence of RNA polymerase II active or stalled predicts epigenetic fate of promoter CpG islands. *Genome Res*. 19: 1974–1982. <http://doi/10.1101/gr.093310.109>.
112. Tazi J, Bird A (1990). Alternative chromatin structure at CpG islands. *Cell* 60: 909–920. [https://doi.org/10.1016/0092-8674\(90\)90339-G](https://doi.org/10.1016/0092-8674(90)90339-G).
113. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, et al (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 464: 1082–1086. <http://doi/10.1038/nature08924>.
114. Tropberger P, Schneider R (2010). Going global: novel histone modifications in the globular domain of H3. *Epigenetics*. 5(2):112–117. <https://doi.org/10.4161/epi.5.2.11075>.
115. Trus I, Udenze D, Berube N, Wheler C, Martel MJ, Gerdts V, Karniyuchuk U (2020). CpG-recoding in Zika virus genome causes host-age-dependent attenuation of infection with protection against lethal heterologous challenge in mice. *Front Immunol* 10:3077. <https://doi.org/10.3389/fimmu.2019.03077>.
116. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature*. 439: 811–816. <http://doi/10.1038/nature04433>.
117. Tweedie S, Charlton J, Clark V, Bird A (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol*. 17: 1469–1475. <https://doi.org/10.1128/MCB.17.3.1469>.
118. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol*. 20: 2108–2121. <http://doi/10.1128/mcb.20.6.2108-2121.2000>.
119. Wachter E, Quante T, Merusi C, Arczewska A, Stewart F, Webb S, Bird A (2014). Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *Elife*. 3: e03397. <http://doi/10.7554/eLife.03397>.
120. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*. 420: 520–562.
121. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 39: 457–466. <http://doi/10.1038/ng1990>.
122. Wei L, Liu B, Tuo J, Shen D, Chen P, Li Z, Liu X, Ni J, Dagur P, Sen HN, Jawad S (2012). Hypomethylation of the IL17RC promoter associates with age-related macular degeneration. *Cell rep* 2(5):1151–1158. <http://dx.doi.org/10.1016/j.celrep.2012.10.013>.
123. Weidner CI, Wagner W (2014). The epigenetic tracks of aging. *Biol Chem* 395:1307–1314. <https://doi.org/10.1515/hsz-2014-0180>.
124. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Eve Sun Y, Zhang Y (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*. 473: 389–393 <http://doi/10.1038/nature09934>.

125. Wu SC, Zhang Y (2010). Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol*. 11: 607–620. <http://doi/10.1038/nrm2950>.
126. Xia X (2020). Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Mol Biol Evol* 37:2699–705. <https://doi.org/10.1093/molbev/msaa094>.
127. Xu C, Liu K, Lei M, Yang A, Li Y, Hughes TR, Min J (2018). DNA sequence recognition of human CXXC domains and their structural determinants. *Structure*. 26(1): 85–95. <https://doi.org/10.1016/j.str.2017.11.022>.
128. Youdell ML, Kizer KO, Kisseleva-Romanova E, Fuchs SM, Duro E, Strahl BD, Mellor J (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Mol Cell Biol*. 28: 4915–4926. <http://doi/10.1128/MCB.00001-08>.
129. Zardo G (2021). The role of H3K4 trimethylation in CpG islands hypermethylation in cancer. *Biomolecules*. 11(2): 143. <https://doi.org/10.3390/biom11020143>.
130. Zhang T, Cooper S, Brockdorff N (2015). The interplay of histone modifications—writers that read. *EMBO Rep*. 16(11): 1467–1481. <https://doi.org/10.15252/embr.201540945>.
131. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*. 126: 1189–1201. <https://doi.org/10.1016/j.cell.2006.08.003>.
132. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W et al (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res*. 38: 4246–4253. <https://doi.org/10.1093/nar/gkq147>.
133. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322: 750–756. <http://doi/10.1126/science.1163045>.
134. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463):477–81. <https://doi.org/10.1038/nature12433>.
135. Zouali M (2020). DNA methylation signatures of autoimmune diseases in human B lymphocytes. *Clin Immunol* 222: 108622. <https://doi.org/10.1016/j.clim.2020.108622>.

CpG islands in the genome: their structure, type, function and evolutionary characteristics

Mustafa Muhagheh-Dolatabady^{1*}, Elnaz Muhagheh Dolatabady²

¹ Dept. of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Iran

² Faculty of Medicine, Yasouj University of Medical Sciences, Yasouj, Iran

Abstract

The content of CpG dinucleotides is depleted in the mammalian genome due to deamination of methylated cytosine, and this process is a repressive signal that leads to long-term gene silencing. However, CG reduction does not occur in all genomic regions. It has been reported that the mammalian genome contains regions of the DNA sequence, known as CpG islands (CGIs), with a high GC base composition and a high density of CpG dinucleotides relative to the bulk genome. The CpG islands usually remain hypomethylated in the majority of annotated genes. CGIs have a high presence in the promoters and first exons of the gene, and therefore, act as the regulatory elements of gene expression and transcription start sites. In fact, most CGIs are found in transcription start sites and approximately 70% of annotated gene promoters, including almost all housekeeping genes, as well as some tissue-specific and regulatory genes. However, CGI methylation, as a heritable epigenetic mark, can occur in somatic tissues, leading to gene silencing such as an X-inactivation and genomic imprinting. In addition, the difference in DNA methylation between tissues is also observed, which indicates the important role of methylation in regulating gene expression during the development of different types of cells.

Based on genome annotations, almost half of identified CGIs, known as orphan CGI, are located either within genes (intragenic) or at a distance from genes (intergenic). This group of CGIs can also play an essential role in regulating transcription, but they are generally shorter than CGI promoters and are more prone to methylation. Aberrant DNA methylation of CpG islands in the promoter region is a mechanism that leads to loss of the function of tumor suppressor genes in human cancer. The methylation profile of CpG islands reveals that some genes show more frequent methylation than others and, a unique set of methylated genes is related to each tumor type. In this article, it will try to discuss the evolutionary process related to the conservation of CGI in different organisms, as well as its structural, regulatory and functional characteristics.

Key words: gene expression, promoter, CpG islands, methylation, evolutionary feature