

## Alguns aspectos funcionais do epigenoma, genoma e transcriptoma nos animais (Some functional properties of epigenome, genome and transcriptome in animals)

**Dias Correia, J. H. R.:** CIISA, Departamento de Morfologia e Função, Faculdade de Medicina Veterinária de Lisboa, Alameda da Universidade Técnica, Pólo Universitário, Alto da Ajuda, 1300-477 Lisboa, Portugal e-mail [jhrdcorreia@fmv.utl.pt](mailto:jhrdcorreia@fmv.utl.pt) | **Dias Correia, A.A.:** CIISA, Departamento de Morfologia e Função, Faculdade de Medicina Veterinária de Lisboa, Alameda da Universidade Técnica, Pólo Universitário, Alto da Ajuda, 1300-477 Lisboa, Portugal

### REDVET: 2007, Vol. VIII Nº 10

Recibido: 16.07.2007 / Referencia: 100704\_REDVET / Aceptado: 10.09.2007 / Publicado: 01.10.2007

Este artículo está disponible en <http://www.veterinaria.org/revistas/redvet/n101007.html> concretamente en <http://www.veterinaria.org/revistas/redvet/n101007/100704.pdf>

REDVET® Revista Electrónica de Veterinaria está editada por Veterinaria Organización®.  
Se autoriza la difusión y reenvío siempre que enlace con Veterinaria.org® <http://www.veterinaria.org> y con REDVET® - <http://www.veterinaria.org/revistas/redvet>

### Resumo

O estudo epigenético das mudanças na expressão dos genes que são hereditárias e que não envolvem uma mutação, deu novas perspectivas aos fenómenos da hereditariedade.

Os mecanismos epigenéticos tais como a metilação ou as modificações covalentes das histonas e os RNA de interferência estão envolvidos em toda esta problemática.

Neste âmbito os diferentes estados de metilação das citosinas do DNA modelam uma série de processos biológicos.

Por outro lado as alterações epigenéticas ocorridas inclusive desencadeadas pelo meio envolvente podem levar à transmissão hereditária dessas novas características.

O conhecimento da sequência total dos genomas de algumas espécies animais permitiu, através de diversas metodologias, averiguar a sua operacionalidade inclusive de uma e de outra cadeia do DNA e as suas nuances de transcrição consoante as células, tecidos, órgãos, factores envolventes, etc.

Foi assim possível conhecer muito mais profundamente os transcriptomas correspondentes ao conjunto total de genes transcritos e começar a perceber as formas como podem ser regulados os diversíssimos genes codificadores de proteínas ou não codificadores de proteínas mas activos, condicionando e orquestrando todo o funcionamento da máquina biológica.

**Palavras chave:** Epigenoma | metilação | genoma | transcriptoma | complementaridade.

## Summary

Epigenetics studies on changes in gene expression with heredity type and without mutation give new horizon to this subject.

Epigenetics mechanism like methylation or covalent histone changes and RNA i are involved at this level.

Different levels of methylation from cytosines DNA shaped several biological process.

Epigenetics changes breaking out by surrounding can carry this new signals to lineage.

The acknowledge of all genomic sequence from some animals allow with different methodologies, search your functionality inclusive from one or another DNA strand and transcription nuances according cells, tissues, organs, surrounding factors, etc.

All that enlarge connection of all transcripts genes transcriptome and began to understand the coding genes regulation or non-coding genes regulation but actives, who determine whole biological operation.

**Keywords:** epigenome | methylation | genome | transcriptome | complementarity.

---

## Nota Preliminar

A anteceder alguns aspectos da funcionalidade do epigenoma, genoma e transcriptoma justifica-se que se introduza a evolução do conceito de gene.

O velho conceito clássico de que um gene corresponderia a uma porção do genoma responsável por um fenótipo ou função, produzindo essa porção genómica um mRNA poliadenilado que codificava uma determinada proteína, está actualmente ultrapassado na medida em que se sabe que são produzidos outros transcritos não codificadores de proteínas e que por vezes estes se sobrepõem aos genes codificadores de proteínas da mesma cadeia (sense) ou da cadeia oposta (antisense), ou mesmo localizados em regiões genómicas entre os genes (Gingeras, 2007).

Avulta, por outro lado, neste contexto que a complexidade da transcrição possível de um gene é hoje conhecida tal como se representa esquematicamente na última figura incluída neste texto.

## 1- Epigenoma

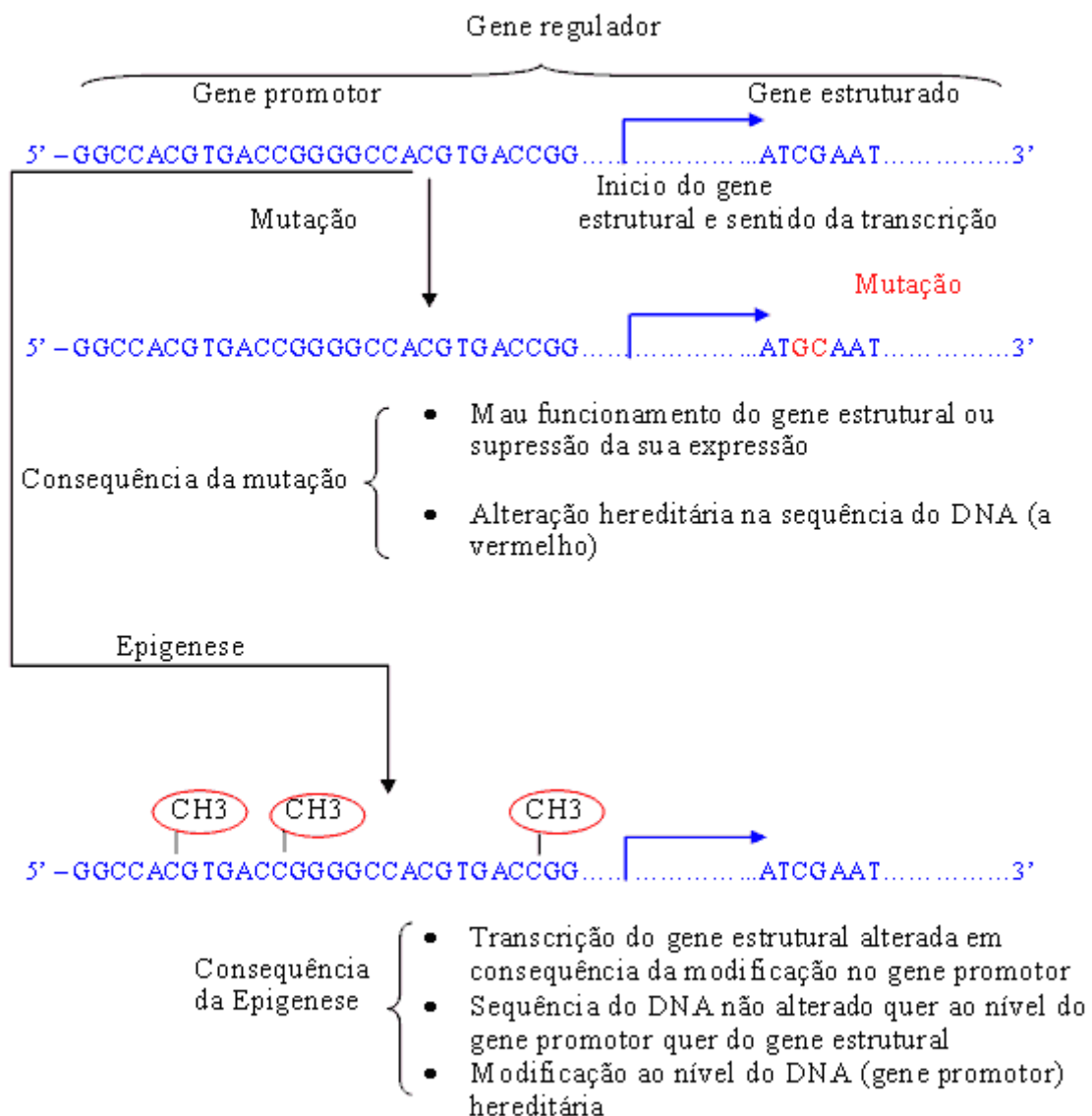
### 1.1- Definição. 1.2- Epigenoma v/s mutação.

A expressão epigenética foi cunhada por Conrad Waddington em 1942 para descrever o "estudo do processo pelo qual o genótipo dá origem ao fenótipo"(in Novik et al., 2002).

A expressão evoluiu ao longo dos tempos entendendo-se hoje como epigenético "o estudo das mudanças na expressão dos genes que são hereditárias através da mitose e/ou meiose e que não envolvem uma mudança na sequência do DNA ou seja uma mutação ( Wu and Morris, 2001 in Novik

A epigenética pode elucidar a forma como os genomas trabalham, pois combina a genética com o meio envolvente (Novik, et al., 2002).

Figura 1  
Comparação da mutação com a epigenese



Adaptação:

<http://www.hesiglobal/NR/rdonlyres/D69199BE-9F1F-4745-BAAD-C3247CBF3B10/O/EI03Goodman.pdf>

Os estudos epigenéticos deram novas perspectivas aos fenómenos da hereditariedade pois permitiram ir muito mais além do que as simples análises das sequências dos DNA.

Na ausência de alterações da sequência do DNA ocorrem variações hereditárias no funcionamento dos genes que têm sido profundamente investigadas nos estudos epigenéticos dos últimos anos.

“Mecanismos epigenéticos, tais como a metilação do DNA, a acetilação e outras modificações covalentes das histonas, e os RNA de interferência (RNA i), modelam a activação dos genes e a sua inactivação sendo poderosos reguladores da actividade dos genes, da transcrição do RNA e da homeostasia das proteínas”(Armandola,2004).

Foi mesmo criado um Projecto do Epigenoma Humano (HEP) cujos objectivos são os de identificar, catalogar e interpretar os desenhos da metilação do DNA, ao longo do genoma, de todos os genes humanos em todos os principais tecidos do organismo.

Esta metilação, que pode mudar o funcionamento do genoma, pode ser influenciada exogenamente e desempenhar um papel decisivo em todas as patologias dos seres humanos e animais.

Esta metilação do DNA ocorre naturalmente nas citosinas que se encontram nas ilhotas CpG ( Figuras 1 e 2) ao longo do genoma e controlam a expressão correcta dos genes respectivos.

Num dado tecido normal do organismo e em situações patológicas estas citosinas são metiladas diferentemente (numa extensão maior ou menor) imprimindo um desenho distinto e específico em cada situação.

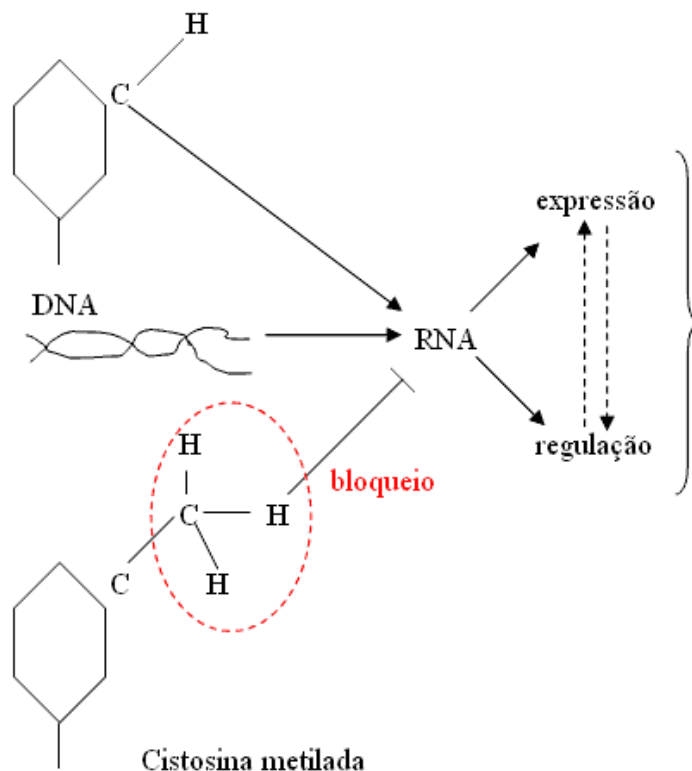
As posições variáveis destas metilações (MVP) constituem pois marcas epigenéticas e ao lado dos polimorfismos de um único nucleótido (SNP) dentro dos genomas, podem contribuir para melhor compreender e diagnosticar diversas situações patológicas.

A metilação do DNA modifica também a expressão dos genes durante o processo de desenvolvimento dos animais e em resposta a alterações que ocorrem no seu meio envolvente.

Figura 2

Diferentes estados de metilação da citosina e sua modelação dos processos biológicos

Citosina não metilada



### *Diferentes estados de metilação da citosina e sua modelação dos processos biológicos*

Para conhecimento do grau de metilação genómica ocorrida, metodologias relativamente simples permitem alcançar esse objectivo. Assim o tratamento do DNA genómico com bissulfito de sódio converte as citosinas não metiladas em uracilo, mas não afecta as citosinas metiladas. Procedendo depois à amplificação por PCR e sequenciação dos ampliados seleccionados a partir do DNA tratado com bissulfito, o grau de metilação pode ser determinado comparando as proporções dos sinais respostas correspondentes aos dinucleotidos CpG, com o dos locais onde predomina a metilação do DNA. (<http://www.sanger.ac.uk/PostGenomics/epigenome/>).

#### **1.3- Transmissão de marcas epigenéticas**

As alterações epigenéticas ocorridas podem levar à transmissão de geração para geração, dessas características genéticas específicas.

No entanto, para isto é necessário uma reprogramação permanente das linhas celulares germinativas.

Durante o desenvolvimento natural das linhas germinativas dos mamíferos o genoma é reprogramado. Assim quando as células germinativas primordiais (PGC) migram para a ponte genital (genital ridge) começa uma desmetilação do genoma que se completa quando da colonização na gonada inicial (Morgan, H. D.,2005).

Aqui nas gónadas as células germinativas sofrem remetilação genómica de forma específica consoante o sexo, durante a determinação gonadal do sexo (Anway, M.D. 2005), parecendo esta remetilação depender também das células somáticas existentes nas gónadas.

No entanto há factores ambientais que podem reprogramar as linhas germinativas e promover estados transmissíveis de geração em geração quer através de transmissão materna quer paterna (Rakyan V. K.,2003).

Certas marcas epigenéticas nos alelos (epialelos) parecem resistir à desmetilação durante a gametogénese. Durante a post-fertilização essas marcas podem não ser completamente erosionadas podendo constituir como que uma memória do estado epigenético que existia nos gâmetas dos progenitores.

Como é sabido durante o desenvolvimento dos animais as diferentes células e tecidos adquirem diferentes programas de expressão dos seus genes, sendo essa regulação atribuída a modificações epigenéticas tais como a metilação do DNA, a modificação das caudas das histonas e ainda à intervenção de RNA não codificante de proteínas (nc RNA)(Morgan, H. D., 2005).

A metilação do DNA parece silenciar permanentemente uma grande porção do DNA que os anglo-saxónicos denominam "junk",ou sejam, as sequências repetitivas do genoma com funções ainda não identificadas e que correspondem a cerca de 97% do genoma humano, inclusive a maioria das sequências dentro dos próprios intrões e o DNA intergenes (Allegrucci, C., 2005). Há ainda quem designe o DNA não codificante como DNA junk embora ele possa conter transposões que codificam proteínas ([http://en.wikipedia.org/wiki/Junk\\_DNA](http://en.wikipedia.org/wiki/Junk_DNA)).

Cada tipo celular tem pois a sua própria assinatura epigenética em consequência do seu genotipo e de todas as influências do seu meio envolvente que condicionam assim o fenotipo final.

Para a maioria das células dos organismos, as marcas epigenéticas tornam-se fixas após as células se diferenciarem ou em saindo do seu ciclo celular.

No entanto no desenvolvimento normal dos organismos animais ou em situações patológicas, algumas células sofrem uma reprogramação epigenética que envolve a remoção de marcações epigenéticas no núcleo seguindo-se depois um conjunto de marcações diferentes.

Em condições normais com a fertilização são removidas diversas marcas genéticas dos gâmetas e substituídas por marcações embrionicas para proverem os primeiros desenvolvimentos embrionicos e a toti e pluripotência celulares.

No desenvolvimento das linhas germinativas dos animais, os processos epigenéticos afectam este desenvolvimento, e os factores ambientais podem influenciar os mecanismos epigenéticos (Armandola,E.,2004).

Curiosamente Rakyan e Beck em 2006 citando Weaver referem que ratinhas mães que diminuem os cuidados maternos para com a sua prole reduzem, nesta prole descendente, a metilação do DNA e a acetilação das histonas ao nível do promotor do gene receptor glucocorticóide situado no hipocampo, o que desencadeia nessa prole, mais tarde, respostas aumentadas para o stress. Curiosamente ainda a administração central nessa prole descendente menos cuidada, de inibidores das desacetilases das histonas reverte o estado epigenético, passando esses animais quando adultos a ter uma reacção ao stress dentro do normal. O estado epigenético pode nestas circunstâncias ser revertido por condições influenciadoras envolventes diferentes, inclusive pela dieta.

Os cuidados maternos podem pois originar a transmissão de respostas adaptativas através das gerações (Weaver,I. C. G.;,2004).

Como é sabido durante o desenvolvimento normal dos mamíferos os perfis epigenéticos do genoma são reprogramados durante a gametogenese e depois a seguir à fertilização.

No entanto, a hereditariedade epigenética é menos estável que a hereditariedade baseada nas mutações no DNA havendo naquela umas marcações mais estáveis do que outras.

No ratinho parece que a metilação do DNA e epialelos hereditáveis não é apagada após a gametogénese ou no início da embriogénese, sugerindo que a metilação do DNA é uma marca herdada: mas há sugestões ainda de que não é apenas a metilação do DNA que é transmitida à geração seguinte mas possivelmente uma combinação da metilação do DNA, modificações das histonas e RNA (Rakyan e Beck, 2006).

Podem colocar-se de uma forma geral e neste contexto diversas questões tais como, que acontecimentos no meio envolvente podem desencadear estas mudanças epigenéticas, e como actuam eles ao nível da metilação do DNA e se são eles específicos para cada tipo de tecido ,etc( Colvis,C.M.,2005)?

A expressão epigenético envolve pois " uma mudança no estado de expressão de um gene que não envolve uma mutação, mas que é contudo hereditária (após divisão celular) na ausência do sinal ( acontecimento ou factor) que desencadeia esta mudança" (Ringrose e Paro,2007).

Alguns sistemas de transmissão epigenética hereditáveis, referenciados estão incluídos na figura 3

Figura 3  
Transmissão epigenética hereditária  
Sistemas de transmissão epigenética hereditáveis (EIS) referenciados

<u>Tipos de sistemas de transcrição epigenética hereditáveis(EIS)</u>	<u>Evidência da sua transmissão através da mitose</u>	<u>Evidência da sua transmissão entre gerações</u>	<u>Transmissão através da mãe ou do pai ou de ambos</u>	<u>Transmissão horizontal e vertical</u>
<ul style="list-style-type: none"> <li>• <u>Marcas na cromatina</u> <ul style="list-style-type: none"> <li>○ Metilação do DNA</li> <li>○ Marcas nas proteínas</li> <li>○ Modificação nas histonas</li> </ul> </li> </ul>	+ (20, 21, 51)	+ (9, 45, 48)	+ (40)	Improvável
	+ (22)	+ (9, 45)	+ (4)	
	+ (23)	?	?	
• <u>RNAi</u>	+ (26)	+ (27, 29)	Mais possivelmente pelo pai (28, 24)	+
• Comportamento dos seres vivos	Possível se envolver mudança no EIS	+ (31 – 34)	Fêmea e macho (31, 39)	+, quando adoptado por pais com diferentes hábitos (34)

+ = assinalada positivamente  
? = não assinalada

Números entre parêntesis indicam a respectiva fonte bibliográfica citada em Jablonka, Ev. 2004.

Uma das características da hereditariedade das marcas epigenéticas celulares, é que estas tendem a ser reversíveis.

(adaptado de Jablonka, Ev. 2004)

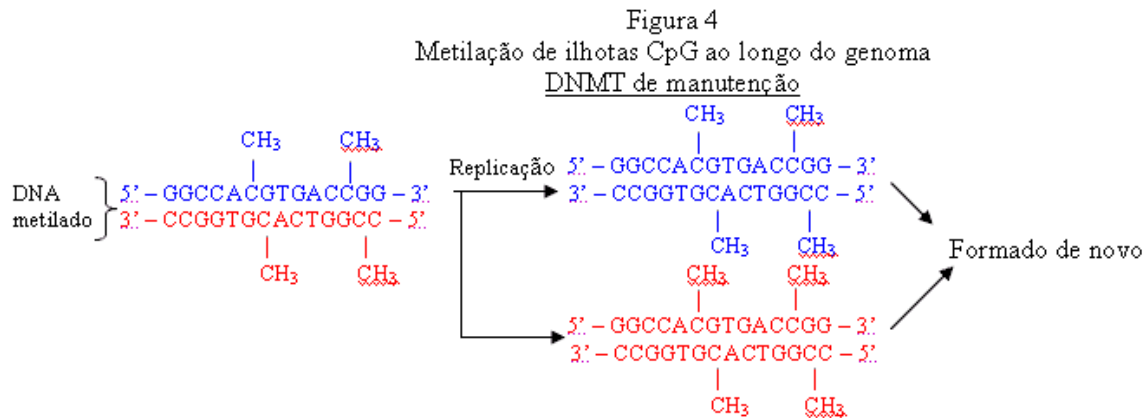
A transcrição pode ser activada ou inibida ao nível dos promotores e “enhancers” por acção de variações nas concentrações de activadores e repressores celulares que interactuam com o DNA com afinidades específicas com os seus locais de interacção. Como são feitas as memórias celulares que sobrevivem à replicação do DNA e à mitose é em grande parte ignorado. No entanto há dados que apontam para que as modificações covalentes das histonas da cromatina ao nível do promotor reflectem o silenciamento ou a activação, mas não se sabe se são eles os possuidores da informação epigenética hereditária, embora se saiba que também os nc RNA (RNA não codificador) intervêm neste fenómeno da memória epigenética.

Com efeito, estão assinaladas marcações epigenéticas hereditárias associadas com a transferência pelo zigoto de moléculas de RNA (Rassoulzadegan, M., 2006). A presença de RNA

nos espermatozoides coloca a possibilidade de que a sua transferência para o ovo fertilizado possa ser um sinal para essa transferência epigenética.

#### 1.4- **Metilação do DNA (Novik, et al., 2002)**

O maior alvo para metilação no genoma dos mamíferos é a citosina residente em ilhotas CpG ao longo do genoma (Figura 4)



A enzima DNMT (metiltransferases) de manutenção mantém, após a replicação do DNA original, as cadeias filhas formadas, também metiladas.

Nos seres humanos sabe-se que existem quase 29.000 ilhotas CpG em todo o genoma, possuindo a maioria dos cromossomas 5-15 ilhotas por cada Mb (megabase).

As ilhotas CpG muitas vezes têm sede em promotores e nos primeiros exões dos genes e a sua metilação pode originar o silenciamento da transcrição.

Os sinais de metilação no DNA atraem por seu turno proteínas com capacidade para interagir com esses grupos metilo, recrutando deacetilases (HDCA) das histonas e iniciando uma remodelação nesse local, da cromatina e a sua condensação.

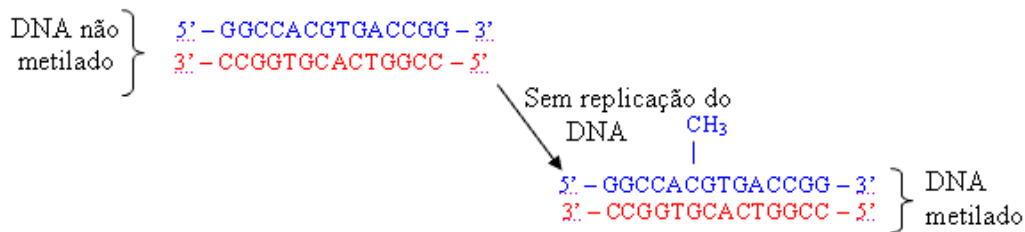
Os grupos metilo são adicionados ao DNA por enzimas metiltransferases (DNM) sendo conhecidas quatro nos mamíferos, as DNMT 1, DNMT 2, DNMT 3 A e DNMT 3 B.

Há contudo duas classes principais de DNMT nos mamíferos e que são as DNMT "de novo" e as DNMT "de manutenção", estas últimas assegurando que se uma porção do genoma for metilada numa célula em divisão, então esta porção será também metilada nas células filhas produzidas pela divisão celular (Gene expression: epigenetics in memory II), uma vez que a DNMT de manutenção metila também a nova cadeia de DNA sintetizada a partir do sinal do par nucleotídico da cadeia inicial com que emparelha (vide figura 4).

As DNMT "de novo" efectuam novas metilações em locais do DNA onde não existe qualquer sinal mesmo na hélice de DNA oposta, para iniciar essa metilação. Estas DNMT "de novo" intervêm sobretudo nas metilações que ocorrem durante a gametogenese e no embrião em pré-implantação, parecendo relativamente inactivas no adulto plenamente desenvolvido (vide figura 5)



Figura 5  
DNMT "de novo"



A enzima DNMT "de novo" metila ilhotas CpG no DNA com graus de metilação variáveis o que se pode reflectir na taxa de transcrição (vide figura seguinte).

Por ora duvida-se que existam actividades de enzimas desmetilantes nos mamíferos.

A DNMT 1 promove a metilação "de novo" e assegura os sinais de metilação, assegurando ainda que o perfil de metilação pré-replicação seja restaurado na nova hélice de DNA gerada quando da replicação.

O papel da DNMT 2 é pior conhecido e as DNMT 3 A e B têm actividades de metilases "de novo" sendo importantes durante as primeiras fases da embriogénese.

A existência de uma desmetilase do DNA foi recentemente assinalada mas há controvérsia sobre a sua existência (Wolffe *et al.*, 1999).

Na formação de tumores malignos a hipermetilação dos genes supressores de tumores pode levar à inactivação destes genes e consequentemente à tumorigénese.

A metilação do DNA é altamente dependente dos dadores de grupos metilo e de co-factores na dieta o que tem ligado a epigenética com a nutrição pré-natal e post-natal e com doenças no estado adulto como as doenças cardiovasculares, diabetes de tipo 2, obesidade e cancro (Armandola, E., 2004).

Também hoje se sabe que alguns genes são regulados por metilação ( vide figura 6) em resposta a aprendizagens realizadas (Gene expression: epigenetics in memory II).

## 2- Genoma e transcriptoma.

### 2.1- Definições.

Começamos por definir o que correntemente se entende por genoma e transcriptoma.

Assim genoma é o conteúdo genético total no conjunto haplóide de cromossomas nos eucariotes, ou num único cromossoma nas bactérias, ou no DNA ou RNA dos vírus.

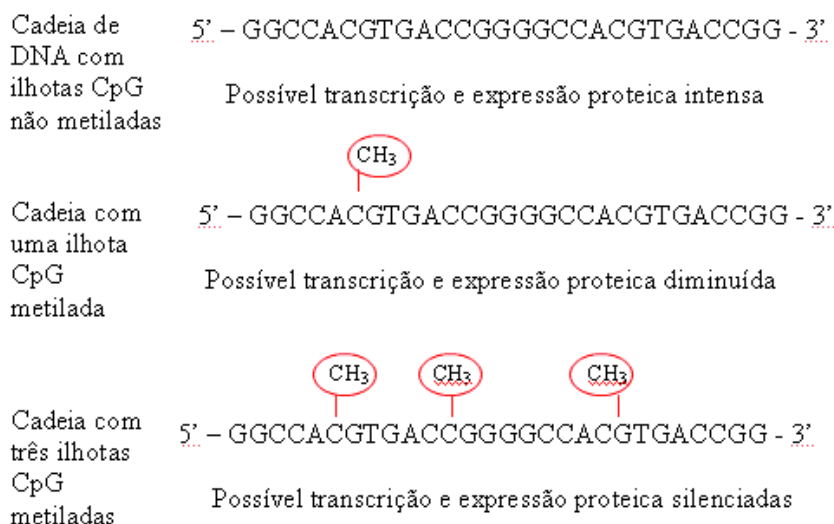
Os animais superiores actualmente têm dois genomas que juntos formam o genoma total: um genoma cromossomal, dentro do núcleo da célula formando os cromossomas e um genoma mitocondrial no citoplasma da célula habitualmente numa forma de um cromossoma circular (cromossoma mitocondrial).

Noutra definição mais simples considera-se que o genoma humano e dos animais superiores corresponde a todo o DNA que um organismo possui.

Figura 6

A metilação progressiva cumulativa do gene promotor diminui o nível de expressão (tradução) proteica podendo a hipermetilação silenciar o gene.

A metilação altera o nível de transcrição dos genes



Quanto ao transcriptoma corresponde ao conjunto total de genes activados RNA ou transcritos num dado tecido e num dado momento, ou seja, o conjunto completo de transcritos RNA produzidos pelo genoma num dado momento.

O transcriptoma é dinâmico e varia consoante os diferentes desenhos da expressão dos genes.

**2.2- Funcionalidade.**

Nos transcriptomas, ou seja, no conjunto de RNA transcritos do DNA genómico das células eucariotas dos mamíferos terá que se atender não só aos genes codificadores de proteínas mas também às sequências de DNA transcritas não codificadoras de proteínas (nc RNA), genes codificadores de RNA e respectivos transcritos.

Quadro 1

Estimativas das percentagens dentro dos respectivos genomas das regiões codificadoras de proteínas (A), regiões não codificadas de proteína mas transcritas (B), e regiões não transcritas (C).

	A	B	C
Bactérias (E.coli)	88%	1%	11%
Leveduras(Sac.cerevisiae)	68%	10%	22%
Nemátodos (Caen.elegans)	24%	30%	46%
Mamíferos (seres humanos)	2%	43%	55%

(adaptado de Shabalina e Spiridonov, 2004)

Calcula-se que nos genomas dos seres humanos (isto nos cromossomas 6, 7, 14, 20, e 22 (ref. 5 in Shabalina e Spiridonov, 2004) as sequências codificadoras de proteínas representam 2%

da sua totalidade enquanto as regiões transcritas mas não codificadoras representam 43% e as regiões não transcritas 55% (vide quadro 1).

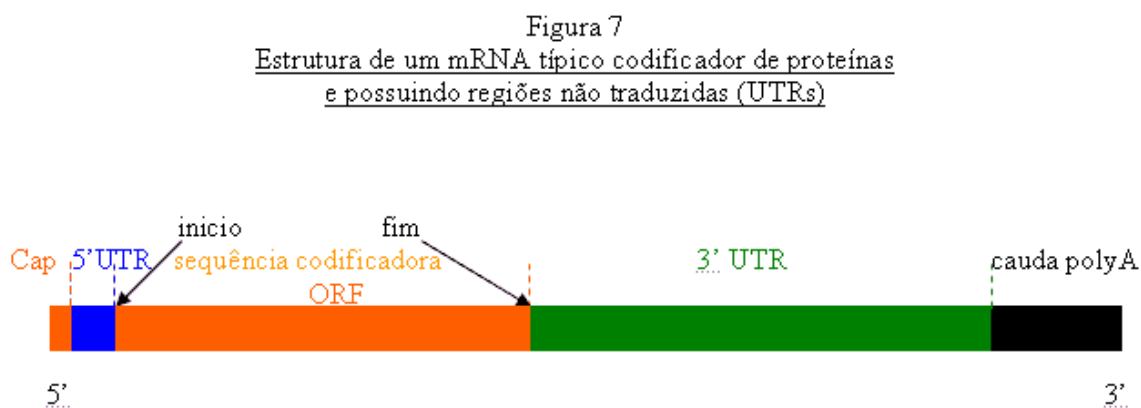
Tem sido sugerido (ref.16 e 17 *in* Shabalina e Spiridonov, 2004) que os transcritos não codificadores de proteínas (nc RNA) representam metade ou mais de todos os transcritos dos genes eucariotas.

Diversas destas sequências transcritas mas não codificadoras de proteínas, têm importantes papéis reguladores sobre o funcionamento das células em geral.

A maior parte do genoma dos mamíferos expressa-se através de transcritos no núcleo, quer de uma cadeia de DNA quer da outra cadeia de DNA, e isto não só em embriões de rato e a maior parte destes transcritos primários do núcleo correspondem a sequências nucleotídicas únicas transcritas de 32,8% do DNA e sequências moderadamente repetitivas transcritas de 32,9% do DNA.(Shabalina e Spiridonov, 2004).

Sabe-se que a expressão dos genes nas células eucariotas é controlada a diversos níveis.

Um desses controlos feito após a transcrição é crítico e é feito por elementos cis-reguladores codificados nas regiões não traduzidas 5' e 3' UTR nos mRNA. (vide Figura 7).



Diversas sequências reguladoras encontram-se na extremidade 3' UTR tais como:

- Um sinal de poliadenilação, AAUAAA ou ligeiramente diferente, onde ocorre a cisão cerca de 30 pares de bases abaixo seguindo-se a adição de varias centenas de adeninas(cauda polyA).
- Locais para interacção com proteínas, que afectam a estabilidade e localização do mRNA como por exemplo elementos SECIS que dirigem a utilização na tradução de selenocisteinas através de UGA em vez deste ser codão de paragem.
- Locais para interacção com mi RNA.

(adaptado de These prime untranslated region Wikipedia)

Normalmente estas UTR 5' e 3' nos RNA mensageiros (RNAm) encontram-se menos conservadas nas suas sequencias do que as sequências que codificam proteínas, mas mesmo assim encontram-se mais conservadas do que as sequências não transcritas.

Mas têm sido encontrados blocos de nucleótidos altamente conservados nas regiões 5' UTR e sobretudo nas regiões 3' UTR de genes ortólogos entre mamíferos, aves e peixes (ref. 25 e 26

in Shabalina e Spiridonov,2004), chegando mesmo essa conservação a ultrapassar a conservação observada nas correspondentes regiões codificadoras.

Estas regiões conservadas das sequências UTR são locais para interacção com proteínas ou com RNA anti-sense o que regula uma série de actividades subsequentes como o transporte, localização, tradução e estabilidade dos RNAm(ref.28-31 in Shabalina e Spiridonov,2004).

Para efeitos puramente comparativos refere-se a composição em RNA ribossomal (RNA r), RNA de transferência (RNA t) e RNA mensageiro (RNAm) em *E. coli* (Quadro 2).

Quadro 2  
Para efeitos comparativos refere-se a composição em rRNA, t, e mRNA de *E. coli*

Tipos de moléculas de RNA na *E. coli*

Tipo	Quantidade relativa	Coefficiente de sedimentação	Massa Kd	Número de nucleótidos
rRNA	80%	23	$1,2 \times 10^5$	3.700
		16	$0,55 \times 10^3$	1.700
		5	$3,6 \times 10^1$	120
tRNA	15%	4	$2,5 \times 10^1$	75
mRNA	5%	Variável		média cerca de 10.000 nas células eucariotas

Adaptado de Stryer, p. 96

Também as regiões 5'UTR contêm locais para interacção com componentes dos complexos de transcrição e participam no recrutamento da sub-unidade 40S ribossomal e no início da tradução.

A região 3'UTR tem ainda um papel fundamental na cisão do transcrito e na regulação da transcrição e estabilidade dos transcritos.

Nos genes que codificam proteínas nos mamíferos os intrões representam ao todo cerca de 95% de toda a sequência correspondente.

Os intrões estão envolvidos na formação dos nucleossomas e na organização da cromatina, sendo neste aspecto mais importantes que os exões ou repetições Alu(ref 52 in Shabalina e Spiridonov,2004).

Também os intrões nos mamíferos têm regiões que interactuam com o esqueleto/matriz (scaffold/matriz ou S/MAR) do núcleo estando talvez envolvidos na docagem das ansas de cromatina à matriz nuclear e ao esqueleto (scaffold) dos cromossomas (ref.52 e 54 in Shabalina e Spiridonov,2004).

Nos mamíferos, nos genes que codificam proteínas as sequências dos intrões contêm pequenas unidades independentes para transcrição, tais como pequenos genes RNA (small RNA) e elementos repetitivos(ref. 57 e 3 in Shabalina e Spiridonov,2004).

Estes elementos repetitivos encontram-se em sequências transcritas, como é o caso dos intrões e UTR, mas também em sequências não transcritas entre genes.

Os transcritos dos genes RNA não codificadores de proteínas (ncRNA) não são pois traduzidos em proteínas e funcionam como moléculas reguladoras ou catalíticas, não se sabendo contudo quantos ncRNA existem nos genomas dos mamíferos, admitindo-se no entanto que no rato

representem mais de um terço de todos os transcritos identificados (ref 4, in Shabalina e Spiridonov, 2004).

Grande numero dos transcritos dos ncRNA possuem intrões.

Estes genes de ncRNA são pequenos e com múltiplas cópias, e não possuindo ORF (vide ORF na fig. 7).O seu tamanho pode oscilar entre 20 nucleotidos (microRNA ou miRNA) até milhares de nucleótidos (ncRNA envolvido no splicing de genes)(ref.59 in Shabalina e Spiridonov,2004).

Os nc RNA estão envolvidos na regulação da transcrição e também na replicação de cromossomas, no imprinting, nos processamentos, etc.(ref. 59-67 in Shabalina e Spiridonov, 2004).

Os genes de ncRNA encontram-se pois dentro dos genomas em extensas regiões conservadas ( dentro de regiões ortologas de genomas relacionados), em regiões entre genes e em sequências intronicas com muita GC.

Como repressores da tradução, estão assinalados miRNA e pequenos RNA temporários(stRNA) que inibem a tradução do mRNA alvo, pequenos RNA nucleares (snRNA) que funcionam nos spliceossomas e pequenos RNA nucleolares (snoRNA) que modificam a estrutura química do RNA.

A ribonuclease tem uma actividade catalítica marcada sobre as moléculas de ncRNA.

A inibição e o silenciamento dos genes através de moléculas de ncRNA baseia-se nas possibilidades de interações e complementaridade entre essas moléculas(genes---RNA) formando longos duplexes perfeitos, ou formando curtos duplexes imperfeitos (miRNA e siRNA).

Na metilação especifica no local de RNA estrutural (rRNA, tRNA e snRNA) pode estar envolvido snoRNA.

A função do DNA não codificador está muito mal elucidada, mas essas regiões não codificadoras são menos conservadas que as partes dos genes codificadores de proteínas.Sabe-se, por outro lado, que a complexidade estrutural e fisiológica dos seres vivos está correlacionada com um aumento das seguintes características:

- transcritos não traduzidos (nos seres humanos calcula-se que são 1,2 biliões de nucleótidos),
- comprimento e numero de intrões dos genes codificadores de proteínas,
- número e complexidade dos elementos de cis-controlo e múltiplos promotores para cada gene,
- número de genes ncRNA e genes codificadores de proteínas,
- complexidade dos UTR e comprimento das 3' UTR,
- número de factores de transcrição.

O tamanho e diversidade do transcriptoma é fundamental e o RNA com uma única cadeia tem propriedades únicas para assumir estas funções reguladoras.

### **2.3- Transcrição a partir das duas cadeias de DNA genómico**

O mapeamento genómico do transcriptoma tem revelado que a transcrição ocorre sobre as duas cadeias do DNA genómico com potenciais sobreposições de alguns transcritos (Fantom Consortium and Riken genome.,2005).

Segundo alguns autores a maioria do genoma dos mamíferos é transcrito habitualmente a partir das duas cadeias do seu DNA.

Por outro lado, parece suceder, pelo menos nos seres humanos, que o número de transcritos saído do genoma é pelo menos dez vezes superior ao número de genes conhecidos.

Nesses transcritos as extremidades 5' e 3' revelam uma extensa variedade na sua sequência nucleotídica em consequência da utilização de promotores alternativos, locais de corte e locais de poliadenilação variáveis.

Diversos ncRNA começam nos locais de iniciação de regiões 3' não traduzidas (3' UTR) de loci codificadores de proteínas.

Os diversos ncRNA encontram-se menos conservados nas suas sequências, em média, do que as regiões 5' ou 3' UTR embora as regiões promotoras dos ncRNA se encontrem geralmente mais conservadas do que os promotores dos mRNA codificadores de proteínas, isto tanto nos seres humanos como nos ratinhos e até nos frangos, contendo no entanto locais para interacção com factores de transcrição conhecidos ((Fantom Consortium and Riken genome., 2005).

Quanto aos mRNA poliadenilados que são processados e exportados para o citoplasma, sabe-se hoje que o conjunto de RNA nucleares não poliadenilados pode ser muito grande e que muitos desses transcritos provêm de regiões intergenicas.

Também se conhece que mRNA com ou sem poly (A) podem ter estabilidades diferentes parecendo os sem poly (A) serem menos estáveis o que é contudo controverso (Functions of poly (A), stability of mRNA, translation).

No entanto os cap aumentam a tradução dos mRNA cerca de 300 vezes e a presença de poly (A) aumenta o nível da sua tradução cerca de 20 vezes.

#### **2.4- Transcrição sense-antisense.**

Sense (sentido) e antisense (anti-sentido) são expressões muito utilizadas em biologia molecular e servem para comparar a polaridade das moléculas dos ácidos nucleicos, sobretudo do RNA, com outras moléculas de ácidos nucleicos e, por vezes, consoante os contextos, têm significados um tanto dispares o que nos leva às considerações seguintes.

As moléculas antisense interactuam com cadeias complementares de ácidos nucleicos modificando a expressão dos genes.

Por complementaridade entende-se duas cadeias polinucleotídicas que podem emparelhar as bases dos nucleótidos entre si para formar uma molécula com duas cadeias.

Uma molécula de RNA antisense é pois uma molécula que é o complemento reverso de uma molécula de mRNA podendo a interacção entre elas condicionar a respectiva tradução do mRNA.

As células podem produzir em condições naturais moléculas de RNA antisense com as consequências respectivas, tal como podem, moléculas desse tipo, ser produzidas "in vitro" e introduzidas nas células ou nos organismos com finalidades terapêuticas várias.

Não devem confundir-se os efeitos desencadeados pelo RNA antisense com os efeitos do fenómeno do RNA de interferência (RNAi), que abordaremos noutra obra (Correia, 2007) e que é um processo relacionado mas diferente. Em plantas transgénicas tem sido verificado por engenharia genética, que a activação da expressão RNA antisense ou, em vez disso, a activação da via RNAi, diferem na magnitude das respostas embora ambos os processos possam silenciar os genes.

Um numero significativo de genes sense (S) dos genomas dos mamíferos possuem em condições naturais transcritos antisense (AS) ou abreviadamente (NAT) -portanto com sequências complementares para outro transcrito - supondo-se que estes pares S-AS estejam implicados numa série de acontecimentos biológicos tais como, a expressão dos genes, RNA de interferência, regulação da tradução, splicing alternativo, inativação do cromossoma X, modelação da cromatina, etc (Galante,P. AF.et al.2007;Lavorgna, G. et al.,2004).

A transcrição antisense (ou seja a transcrição a partir da cadeia oposta da cadeia sense ou seja da codificadora de proteínas (Katayama,S. et al.,2005) é pois um factor a considerar.

Uma grande parte do transcriptoma global evidencia que grande parte do genoma pode produzir transcritos das duas cadeias do DNA.

Os pares transcritos sense-antisense potencialmente podem emparelhar, formando híbridos RNA-RNA .

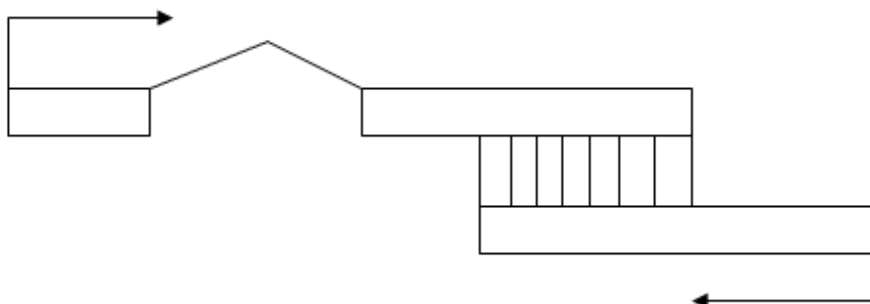
Os transcritos antisense naturais (NAT) são pois moléculas de RNA com sequencias complementares de outros RNA endógenos, podendo esses transcritos NAT ser da forma cis quando provêm de cadeias de DNA opostas (Figura 8) mas do mesmo locus genómico, que emparelham de forma perfeita e numa extensão razoável, ou da forma trans quando provêm de locus genómicos separados e diferentes e emparelham de forma imperfeita numa pequena extensão (Lavorgna,G.,2004).

Estudos computacionais de 2005 sugerem que 15 a 25% dos genes dos mamíferos podem sobrepor-se( emparelhar) originando RNA sense e antisense (Werner e Berdal,2005), interferindo estes transcritos antisense endógenos naturais com o processamento a desenvolver o que equivale a dizer que são reguladores endógenos dos genes.

A maioria dos genes ocupa os seus locais distintos dentro do genoma dos mamíferos, mas alguns estão localizados em regiões complexas (complex loci) onde partilham o seu território com outros genes utilizando as cadeias opostas do DNA (Engstrom,P.G,et al.,2006).

Os complex loci são pois regiões genómicas nas quais diversos genes partilham regiões transcritas numa orientação antisense e/ou com promotores "core" bidireccionais.

Figura 8  
Par cis – antisense (NAT)

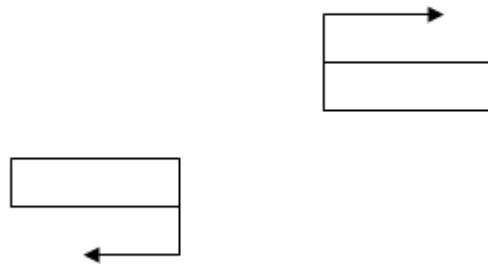


(adaptado de Engström et al. 2006)

Estes genes ou partilham regiões expressas como mRNA( formando pares –cis-antisense) ou começam numa região do genoma (chamada promotor bidireccional) onde a transcrição pode iniciar-se em ambas as direcções ao longo do DNA (ver figuras 8, 9,10 e 11).

O par cis-antisense dentro do mesmo loci possui duas unidades que são transcritas em sentido oposto (ver sentido das setas dentro da figura 8) emparelhando numa forma perfeita pelo menos 20 pares de bases da sequência dos exões.

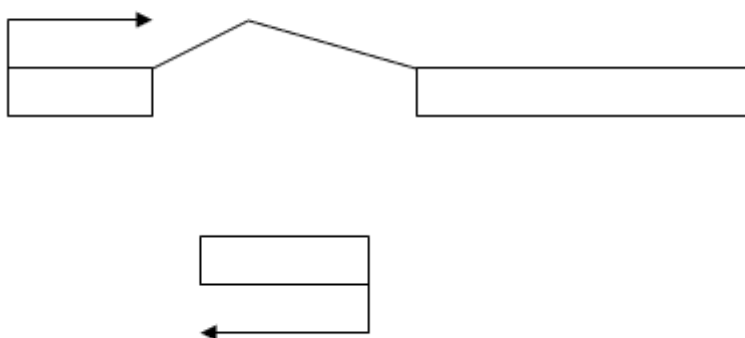
Figura 9  
Par bidirecional promotor



(adaptado de Engström et al 2006)

O par bidireccional promotor tem duas unidades transcritas divergentemente (figura 9), afastadas menos de 1.000 pares de bases e emparelham em menos de 20 pares de bases

Figura 10  
Par antisense não emparelhando exões (NOT)



(adaptado de Engström et al 2006)

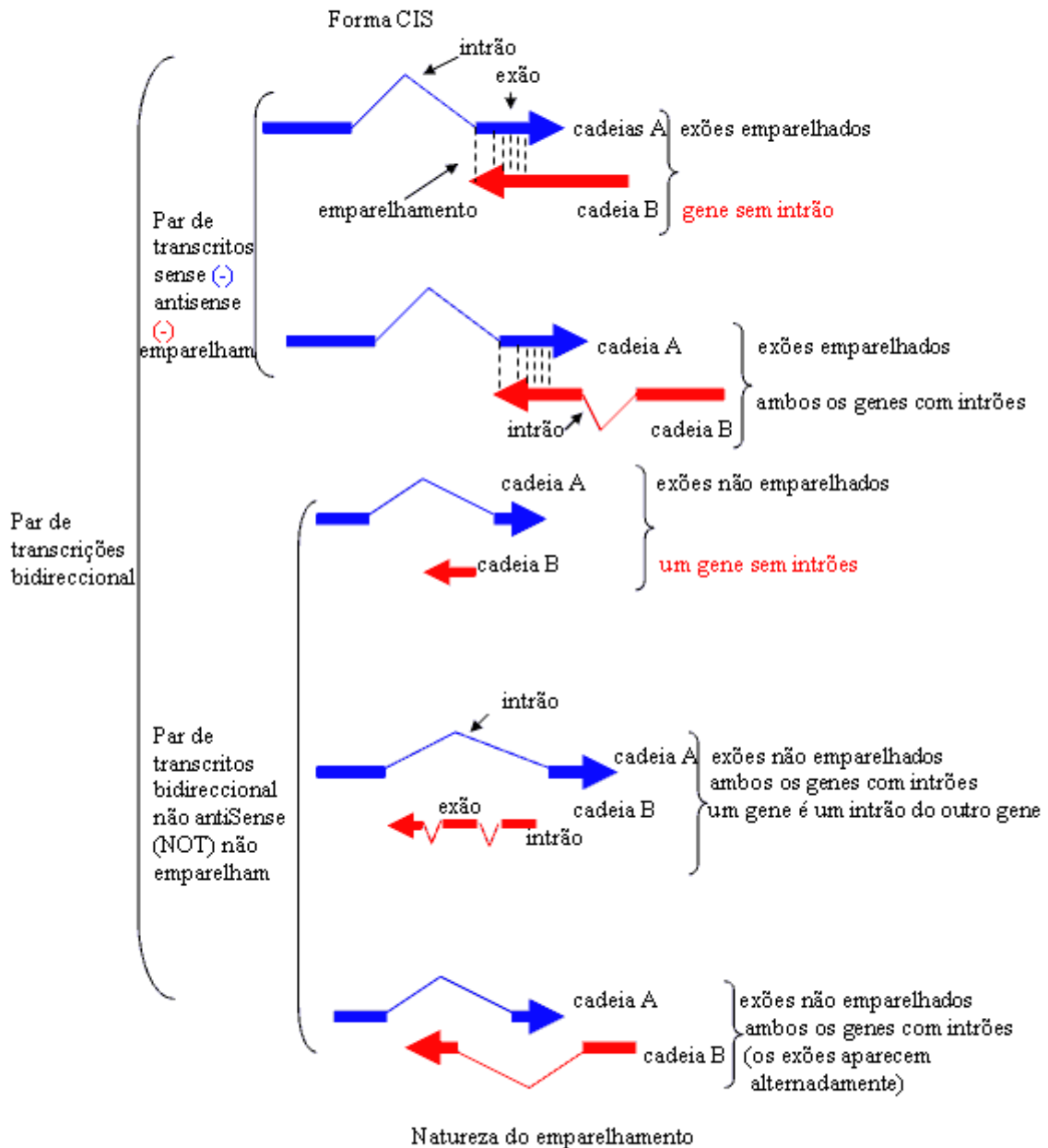
Neste par antisense não emparelhando exões os dois transcritos( figura 10) emparelham pelo menos 20 pares de bases mas não exões.

Parece que cerca de 25% dos genes conhecidos de ratinho são pares cis-antisense, mas admite-se que o valor total possa atingir os 40% (Engstrom ,2006) o que prova portanto a abundância destes loci complexos com transcritos emparelhados provenientes das duas cadeias de DNA, indicando que os pares cis-antisense consistem mais frequentemente de um transcrito codificante e um transcrito não codificante.



Figura 11  
Genoma de ratinho  
Classificação dos transcritos NAT e NOT

(Provenientes do mesmo locus mas de cadeias de DNA opostas)



(adaptado de Kiyosawa et al. 2003)

Referem-se seguidamente transcritos naturais antisense (NAT) identificados e que potencialmente hibridam com o transcrito seu cognato após processamento do RNA e os NOT que são transcritos antisense não emparelhados e que não produzem emparelhamentos nos exões (Quadro 3).

Quadro 3  
Número de NAT e NOT nos genomas de eucariotas

	NAT	NOT
Humanos	2.940	1.460
Ratos	2.481	899
Drosophila	1.027	

(Adaptado de Werner e Berdal, 2005)

Nat = transcritos antisense naturais potencialmente hibridantes  
Not = transcritos antisense não emparelhados

Mattick e Makunin referem no entanto que cerca de 98% do output transcriptional nos humanos e noutros mamíferos consiste de ncRNA a partir de intrões de genes codificadores de proteínas e de exões e de intrões de genes não codificadores de proteínas incluindo vários que são antisense para emparelhar com genes codificadores de proteínas (ref.1, 2, 3-5 in Mattick e Makunin, 2005).

O termo exão não se refere apenas a uma parte da ORF que codifica uma determinada porção da proteína completa pois são conhecidos diversos exões não codificantes em genes humanos (Zhang, 1998).

Por outro lado, os exões podem incluir uma sequência que codifica para ácidos aminados e ao mesmo tempo sequências que não são traduzidas.

Alguns exões podem fazer parte no todo ou em parte da região 5' UTR ou de regiões 3' UTR de cada transcrito.

Uma significativa proporção destes ncRNA parecem ser estáveis nas células eucariotas com semi-vidas comparáveis com as dos mRNA.

O número de genes ncRNA funcionais conhecidos excluindo tRNA, rRNA e snRNA são já mais de 800 (Mattick e Makunin, 2006).

A análise dos transcriptomas de mamíferos tem sugerido que mais de 20% dos transcritos podiam contribuir para emparelhamentos sense-antisense (S/AS), mas estudos em larga escala no projecto FANTOM 3 ( Katayama,S. Tomaru, Y. et al..2005) sugerem que este tipo de transcrição é muito mais amplo. Mattick e Makunin em 2006 indicam que 60 a 70% do genoma humano é transcrito sobre uma ou sobre as duas cadeias de DNA.

No transcriptoma de ratinho há indicações de que 72% deste interactuariam com transcritos provenientes da cadeia oposta( ref. 82 in Mattick,J.S. e Makunin, I.V. ,2006).

De realçar que os ncRNA são expressos a níveis muito mais baixos que os mRNA, alguns sendo mesmo raros (Quadro 4).

Também hoje se considera incorrecto considerar que quase todos os transcritos sejam RNA poliadenilados a ser exportados para o citoplasma para tradução.

Quadro 4  
Representação da transcrição nos mamíferos

Genoma	100%
Transcrição total ou global do genoma	72% de todo o genoma
Transcrição das duas cadeias do genoma	24%
UTRs no genoma	2%
Sequências codificadoras de proteínas no genoma	1,2%
mi RNA no genoma	0,012%
sno RNA no genoma	0,012%

(adaptado de Mattick e Makunin, 2006)

## 2.5- Transcrição codificadora e não codificadora de proteínas.

No quadro 5 podem verificar-se nos humanos e nos ratinhos uma série de características genómicas inclusive referentes aos genes codificadores e não codificadores de proteínas.

Quadro 5

	Genoma total		Porção do genoma não repetitivo	
	Humano	Ratinho	Humano	Ratinho
Numero de genes codificadores de proteínas	20.000 a 25.000	20.000 a 25.000		
Tamanho do genoma (Mb)	2.851	2.490	1.455	1.422
Sequências codificantes	Mb	34	33	29
	%	1,2	2,3	2
Sequências UTR	Mb	32	26	22
	%	1,1	1,1	1,6
Totalidade de sequências transcritas não codificadoras	Mb	1.619	867	811
	%	57	54	60
Relação sequências não codificadoras / codificadoras	47/1	43/1	27/1	28/1

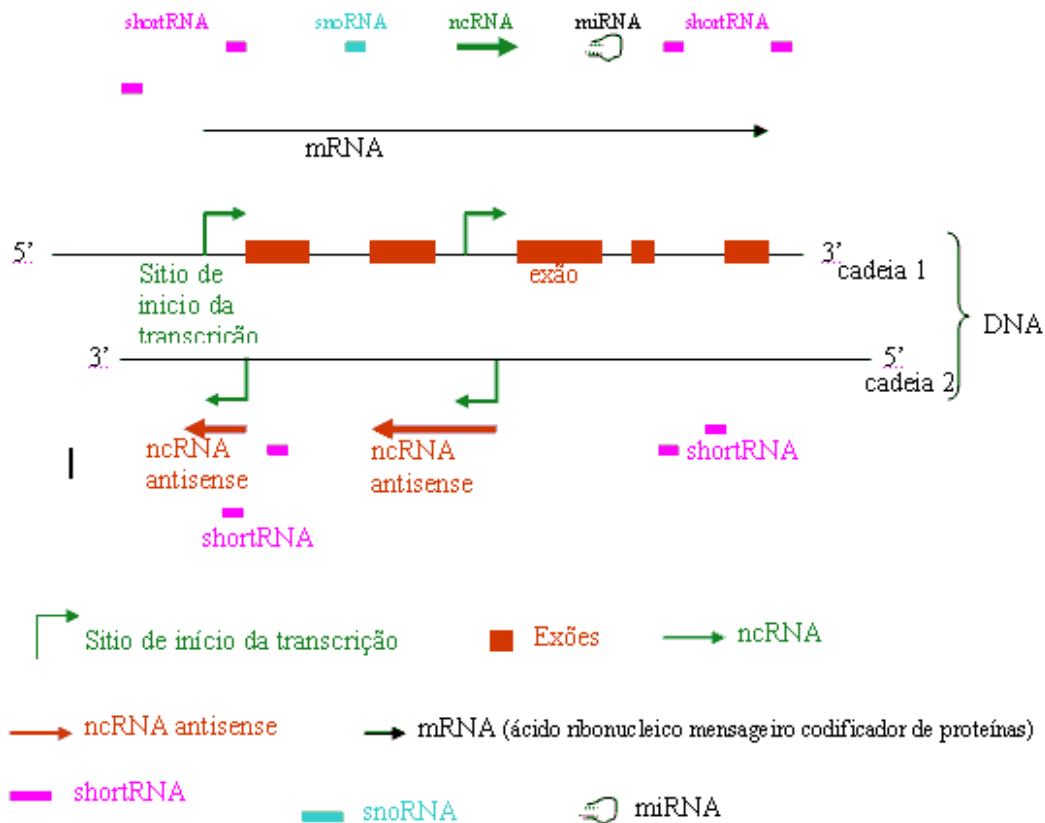
(adaptado de Frith M.C., Pheasant, M. e Mattick, J.S. 2005)

Boa parte do genoma dos mamíferos é pois transcrito de ambas as cadeias de DNA formando pares sense-antisense, sendo a maioria dos transcritos antisense ncRNA (ref. 80 in Mattick e Makunin, 2006).

Ocorre ainda que dentro de um gene podem ocorrer diversos sítios ou locais para início da transcrição de transcritos codificadores de proteínas (mRNA) ou não codificadores de proteínas (ncRNA, ncRNA antisense, short RNA, snoRNA, miRNA) intervalados tal como se esquematiza na figura 12.

Figura 12

Modelo em que um gene tem diversos sítios para início da transcrição e diversas regiões de transcritos codificadores de proteínas (mRNA) e não codificadores (ncRNA, shortRNA, snRNA, miRNA) intervalados



(adaptado de Gingeras, T.R. 2007, Origin of phenotypes: Genes and transcripts. *Genome Research*, 17: 682-684)

Sabe-se também que os genes de ncRNA não têm sinais de transcrição fortes ao contrário dos mRNA (Thomassen, 2004), embora nestes ocorram nuances.

Diversas classes de transcritos ncRNA interatuam com mRNA encontrando-se quanto á sua localização e organização embebidos dentro ou próximo destes próprios mRNA, o que desvanece ou pelos menos obscurece os limites físicos dos genes e demonstra a complexidade criada para definir quais as sequências que desempenham uma determinada função.

Por exemplo os mesmos nucleótidos que se encontram num dado transcrito mRNA podem também ser parte de um transcrito ncRNA podendo este ser um regulador desse mRNA ou de outro.



5. Correia, J.H.R. e Correia, A.A.D..2007.Funcionalidades dos RNA não codificantes (ncRNA) e pequenos RNA reguladores, nos mamíferos.
6. Engstrom,P.G. et al. ...Lipovich,L. .2006.Complex loci in human and mouse genomes.PLoS Genet.,2(4): e47.
7. FANTOM consortium and RIKEN genome exploration research group and genome science group.2005.The transcriptional landscape of the mammalian genome.Science,309 :1559-1563.
8. Frith,M.C, Pheasant,M. e Mattick,J.S..2005.The amazing complexity of the human transcriptome. European Journal of Human Genetics, 13:894-897.
9. Galante,P.A.F.,Vidal,D.O.,Souza,J.E., Camargo,A.A. e souza,S.J..2007.Sense-antisense pairs in mammals:functional and evolutionary considerations.Genome Biology,8:R 40
10. Gene expression: epigenetics in memory II-  
<http://www.gnXP.com/blog/2007/04/epigenetics-in-memory-ii.php> )
11. Gerstein,M. B. et al..2007.What is a gene, post-ENCODE? History and updated definition. Genome Research, 17 (2):669-681
12. Gingeras,T.R..2007. Origin of phenotypes: Genes and transcripts. Genome Research, 17 (2):682-689.
13. HEP. Human epigenome consortium. <http://www.epigenome.org/index.php?page=project>
14. Jablonka,E..2004.Epigenetic epidemiology. International Journal of Epidemiology, 33:929-935.
15. Katayama,S. Tomaru, Y. et al..2005.Antisense transcription in the mammalian transcriptome. Science ,309(5740):1564-1566.
16. Kiyosawa,H. ....e Hayashizaki,Y.. 2003.Antisense transcripts with FANTOM 2 clone set and their implications for gene re3gulation. Genome Research, 13:1324-1334.
17. Lavorgna,P. Dahary,D. Lehner,B., Sorek,R.,Sanderson,C.M. e Casari,G..2004.In search of antisense.Trends in Biochemical Sciences,29 (2):88-94.
18. Mattick,J.S. e Makunin, I.V. ,2006.Non-coding RNA.Human Molecular Genetcs,15,Review Issue 1:R17-R29
19. Mattick,J.S. e Makunin,I.V..2005.Small regulatory RNAs in mammals.Human Molecular Genetics,14,Review Issue 1 :R121-R132
20. Morgan, Hugh D.,Santos, Fátima, Green,K.,Dean,W.eReik,W..2005.Epigenetic reprogramming in mammals.Human Molecular Genetics,14,Review Issue :R47-R58.
21. Novik,N. L.,Nimmrich,I., Genc, B., Maier, S., Piepenbrock, C., Olek, A., e Beck, S.,2002. Epigenomics:genome-wide study of methylation phenomena. Curr. Issues Mol. Biol., 4:111-128.
22. Rakyán V. K., Chong, S., Champ, M. E., Cuthbert, P. C.,Morgan, H. D., Luu, K.V.K. e Whitelaw,E.,2003.Transgenerational inheritance of epigenetic states at murine *Axin* fu allele occurs after maternal and paternal transmission.PNAS,100(5):2538-2543
23. Rakyán, V.K. e Beck, S..2006.Epigenetic variation and inheritance in mammals.Current Opinion in Genetics & Development,16:573-577.
24. Rassoulzadegan,M.,Grandjean,V., Gounon,P.,Vincent,S.,Gillot,I. e Cuzin,F..2006.RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 441 :169-474.
25. Ringrose,L. e Paro,R.,2007.Polycomb/trithorax response elements and epigenetic memory of cell identity.Development 134:223-232
26. Shabalina, S.A. e Spiridonov,N.A..2004.The mammalian transcriptome and the function of non-coding DNA sequence.Genome Biology,5 (4):105.1-105.8
27. Thomassen, G..2004.Detection of non-coding RNA genes by searching for transcription signals in intergenic regions. University of Oslo. Department of Informatics.
28. Weaver, I. C. G., Cervoni,N., Champagne, F.A., D'Álessio,A.C.,Sharma,S.,SEckl,J.R.,Dymov,S., Szyf,M., e Meaney,M.J.. 2004.Epigenetic programming by maternal behaviour. Nature Neuroscience, 7(8):847-854.
29. Werner,A. e Berdal,A..2005.Natural antisense transcripts:sound or silence?.Physiol. Genomics 23:125-131.
30. Wolffe, A.P.,Jones, P.L. e Wade, P.A..1999.DNA demethylation. PNAS,96:5894-5896
31. Zhang,M.Q..1998.Statistical features of human exons and their flanking regions.Human Molecular Genetics, 7 (5):917-932.