Beyond genome wide association studies in celiac disease by exploring the non-coding genome

de Almeida, Rodrigo

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease

Biochim Biophys Acta. 2014; pii: S0925-4439(14)00071-4

Barbara Hrdlickova*, Rodrigo Coutinho de Almeida*, Zuzanna Borek, and Sebo Withoff

Department of Genetics, University of Groningen, University Medical Center Groningen, the Netherlands

*These authors contributed equally.
Abstract
It has been found that the majority of disease-associated genetic variants identified by genome-wide association studies are located outside of protein-coding regions, where they seem to affect regions that control transcription (promoters, enhancers) and non-coding RNAs that also can influence gene expression. In this review, we focus on two classes of non-coding RNAs that are currently a major focus of interest: micro-RNAs and long non-coding RNAs. We describe their biogenesis, suggested mechanism of action, and discuss how these non-coding RNAs might be affected by disease-associated genetic alterations. The discovery of these alterations has already contributed to a better understanding of the etiopathology of human diseases and yielded insight into the function of these non-coding RNAs. We also provide an overview of available databases, bioinformatics tools, and high-throughput techniques that can be used to study the mechanism of action of individual non-coding RNAs.

Introduction
Genome-wide association studies (GWAS) have discovered thousands of single-nucleotide polymorphisms (SNPs) that are associated with multifactorial diseases and quantitative traits. At the time of writing of this review (October 2013), the GWAS catalog [1] described 11,680 SNPs associated with diverse phenotypes and quantitative traits. Despite the wealth of information that GWAS provide, it can be difficult to interpret the results because of the limited resolution of the genome-wide chips used in the initial genotyping screen. Moreover, even state-of-the-art technology and approaches, such as using specialized high-resolution genotyping platforms (for example, the Immunochip [2] and Metabochip [3]), performing genotype imputation [4], or eQTL analysis [5], is often not enough to pinpoint the functional SNP and/or causative gene. What is emerging from these GWAS, however, is that >90% of disease-associated SNPs are located in non-coding regions of the genome for example in promoter regions, enhancers, or even in non-coding RNA genes [1,6]. This indicates that these SNPs might be regulatory. The results from the Encyclopedia of the DNA Elements (ENCODE) project suggest that 80% of the human genome is involved in at least one biochemical RNA- and/or chromatin-associated event [7]. While fewer than 10% of the GWAS SNPs affect coding sequences, most non-coding variants are concentrated in DNA stretches marked by deoxyribonuclease I (DNase I) hypersensitive sites, where they seem to perturb transcription factor binding sites or alter allelic chromatin states [8]. A small percentage seems to disrupt or create micro-RNA (miRNA) binding sites in the 3’ untranslated region (3’-UTR) of genes. All these events
Genetic variation in the non-coding genome

Genetic variation in the non-coding genome will affect the expression level of the genes regulated by these functional elements and can thereby contribute to deregulation of pathways that control healthy cell function.

The recent discovery of approximately 13,500 long non-coding RNAs (lncRNAs) has changed the view on the human genome (Figure 1). It has been estimated that approximately 7% of SNPs associated with autoimmune diseases seem to annotate to long intergenic non-coding RNAs (lincRNAs), a subclass of IncRNAs [6], and some GWAS SNPs have been demonstrated to have eQTL-effects on these lincRNAs [9]. It is thought that the majority of lncRNAs are somehow involved in regulating the expression of protein-coding genes and, therefore, SNPs associated with these ncRNAs may indirectly influence the expression of proteins involved in disease. In this review we will describe how two classes of regulatory ncRNAs, the miRNAs and the lncRNAs, regulate gene expression. We will also describe how SNPs and other types of genetic variation that affect these ncRNAs can contribute to disease phenotypes.

miRNAs

miRNAs are short regulatory RNAs (approximately 19-24 nucleotides long) involved in post-transcriptional gene regulation. The first miRNA, lin-4, was identified in 1993 in a screen for genes required for post-embryonic development in C. elegans [10], but it took another seven years to discover the second one (let-7) [11]. Since then, the number of miRNAs has increased steadily. At the time of writing, the number of human mature miRNAs described in miRBase V20 is more than...
miRNAs are estimated to regulate the translation of up to 60% of protein-coding genes [13]. A single vertebrate miRNA has been described as targeting 200 messenger RNAs (mRNAs) on average, although some miRNAs regulate only a few targets [14]. Conversely, some protein-coding genes are regulated by only a single miRNA, while others are regulated by many miRNAs [15]. The importance of miRNAs as fine-regulators of gene expression has become clear since it was discovered that they play important roles in pivotal biological processes, such as development, cell proliferation, cell differentiation, and cell death [16–19].

**miRNA biogenesis and mechanism of action**

The process of miRNA biogenesis is quite characteristic for this subclass of ncRNAs. The primary miRNA transcript (pri-miRNA) is characterized by one or many hairpins that encompass the functional mature miRNA in their stem (Figure 2). Upon recognition by two nuclear enzymes, Drosha and DGCR8, the pri-miRNA is processed into one or several hairpins approximately 70 nucleotide long; these are called precursor miRNAs (pre-miRNAs). They are exported into the cytoplasm by the nuclear export protein Exportin 5 (XPO5) [20]. In the cytoplasm, the pre-miRNA can be recognized and are then processed by the RNase III enzyme Dicer, which removes the loop of the hairpin, resulting in a ~20 bp dsRNA molecule. One of the strands will be incorporated into the RNA-induced silencing complex (RISC) containing the Argonaute protein 2 (AGO2) and the GW182 [21]. The RISC complex will target a mRNA transcript, based on sequence complimentarity between the miRNA sequence and nucleotides in the 3'-UTR of the target [22]. It is thought that binding of the RISC complex to this target leads to deadenylation of the mRNA target, which will ultimately result in degradation if the homology between the entire miRNA sequence and the target is extensive [21,23]. Efficient targeting requires continuous base-pairing of the miRNA seed region (stretches of 6-8 nucleotides between positions 1-8 of the mature miRNA) with its target [10,23,24]. Computational target prediction approaches make use of this proposed rule to predict the targets of miRNA-induced silencing. These algorithms are based on searching for perfect Watson-Crick pairing between the miRNAs' seed-sequence and the target mRNA sequence (most algorithms focus only on the 3'-UTR of genes) alone, or in combination with other rules, such as evolutionary conservation criteria to predict miRNA target sites [13,23]. Evolutionary conservation was important in defining the first identified miRNAs [25], but it is now becoming clear that many miRNAs are species-specific. In contrast, more than 60% of human protein-coding genes have been under selective pressure to maintain pairing to miRNAs [26]. miRNAs that appear
Genetic variation in the non-coding genome

to have a common ancestor, and differ only in a few nucleotides, are grouped into the same miRNA family [27]. Until recently, it was assumed that miRNAs mainly target the 3’-UTRs of mRNAs [21], but it has now been shown that miRNA target sites can also be located in the 5’-UTRs of target mRNAs, or even in the coding region of these RNAs [21,28].

Much attention has recently been paid to miRNA as potential biomarkers in circulation. Cell-free miRNAs have been described in multiple human body fluids, such as serum [29,30], saliva [31], cerebrospinal fluid (CSF) [32], and urine [33]. Most importantly, the disease-specific or even disease-stage-specific nature of circulating miRNA profiles [34,35] implies that circulating miRNAs might potentially be used as novel biomarkers to evaluate health status or disease progression.

A role for miRNAs in disease

miRNAs have been shown to be involved in cancer and in neurodegenerative, cardiovascular and autoimmune diseases [15]. Changes in the amount of specific miRNAs will result in downregulation or upregulation of their targets, leading to deregulation of the pathways in which those targets are involved. This
deregulation of miRNA levels in human diseases can occur in different ways. Firstly, altered functions of the enzymes involved in the miRNA biogenesis pathway can lead to altered miRNA profiles. Haploinsufficiency of DGCR8 accounts for over 90% of cases of DiGeorge syndrome. This dominantly inherited disorder is caused by the presence of hemizygous chromosome 22q11.2 deletions, which lead to various phenotypic defects including immunodeficiency and autoimmunity [36]. However, DGCR8 haploinsufficiency does not lead to an overall decrease in miRNA levels [37]. The presence of the XPO5 inactive mutant traps miRNAs in the nucleus in a subset of human tumors [38]. Defects in Dicer have also been associated with disease, for example, recurrent somatic missense mutations in DICER1 have been identified in non-epithelial ovarian cancers [39]. In conclusion, defects in several members of the miRNA processing machinery have been reported. However, these changes never lead to dramatic overall changes of miRNA levels in the cell, which is consistent with the notion that miRNAs are essential for cell survival.

Secondly, as pri-miRNA expression is regulated by RNA polymerase II and by the transcription factors that regulate the expression of protein-coding genes, the same epigenetic control mechanisms are involved in regulating miRNA expression. Transcriptional repression of miRNAs by promoter hypermethylation was found in many human tumors [40], for example, the miR-200 family is involved in the control of the epithelial-mesenchymal transition (EMT). In EMT, epithelial cells lose their adherence and polarity, and start to migrate. The miR-200 family downregulates ZEB1 (zinc finger E-box-binding homeobox 1) and ZEB2 (zinc finger E-box-binding homeobox 1), two important transcriptional repressors of genes involved in cell adherence (E-cadherin) and polarity (CRB3 (crumbs protein homolog 3) and LGL2 (lethal giant larva)). Thus hypermethylation of miR-200 family members in cancer leads to upregulation of ZEB1 and ZEB2, leading to decreased adherence and polarity [41]. Similarly, histone modifications might also affect miRNA expression levels. For instance, SIRT1 (sirtuin 1), a NAD-dependent histone deacetylase involved in control of axon growth and degeneration, was recently found to directly suppress the expression of miR-138 in response to peripheral nerve injury [42].

Thirdly, different types of genetic alterations to miRNA genes or to their regulatory motifs can have deleterious consequences. In fact, the first example of the involvement of miRNAs in cancer was the description of a deletion of chromosome 13q14 in chronic lymphocytic leukemia patients. The deleted area contains the miR-15a and miR-16-1 genes that target the anti-apoptotic/pro-survival gene BCL-2 (B-cell lymphoma 2) [43] and thus deletion of this region contributes
Genetic variation in the non-coding genome

to the greater survival characteristics of cancerous cells. For the purpose of this review, we will focus on the most common type of genetic variants, SNPs, and show how they affect the function of miRNAs.

**Genetic variation and miRNA function**

There are many ways in which disease-associated SNPS can affect miRNA levels. Firstly, mutations in miRNA biogenesis genes (Figure 2) can affect miRNA processing, which, in turn, can contribute to disease. For instance, the homozygous presence of SNP rs2073778, located in DGCR8, was found to be associated with a 4-fold increased risk of non-muscle bladder cancer progression [44]. Another example is the presence of SNP rs3742330 (A>G) in the Dicer gene, which is associated with increased survival of T cell lymphoma patients. Homozygous individuals carrying the GG genotype had a significantly increased overall survival [45].

Secondly, SNPs in pri-miRNA and pre-miRNA can affect miRNA maturation efficiency (Figure 3A). For example, SNP rs11671784 in the miR-27a gene reduces gastric cancer risk by impairing the processing of pre-miR-27a to mature miR-27a. It was hypothesized that downregulated levels of miR-27a lead to a significantly
increased expression level of its target gene HOXA10 (homeobox A10), and that changes in this gene have been associated with oncogenesis [46].

Thirdly, SNPs affecting the promoters of miRNAs can affect expression of the miRNA in question. SNP rs57095329, which confers risk of systemic lupus erythematosus (SLE), is located in the miR-146a promoter (Figure 3B). Individuals carrying the risk allele in the promoter showed lower expression levels of miR-146a [47].

Fourthly, SNPs can alter the binding efficiency of miRNAs to mRNA targets (Figure 3C) and SNPs in both miRNAs or in target mRNAs can affect miRNA-target interaction. An example of such an event is SNP rs3853839, which is associated with SLE susceptibility in Eastern Asians [48]. This SNP is located in the 3' UTR of the TLR7 gene and potentially has a negative effect on the binding of miR-3148. Risk allele carriers of this SNP display an increased TLR7 mRNA half-life, resulting in increased TLR7 (Toll-like receptor 7) expression levels [49]. SNPs in miRNAs or target mRNAs can also confer the potential binding to different miRNA-target combinations (Figure 3C). A SNP in the 3' UTR may create a sequence match to the seed of a miRNA that was not previously associated with the given mRNA [50]. Gong et al. predicted that 52% of SNPs in the dbSNP database (release 132) would be able to create novel miRNA binding sites [51].

The identification of GWAS SNPs in miRNA target sites helps with prioritizing functional variants. The first GWAS signal that was explained by polymorphic miRNA targeting was the synonymous SNP (c.313C>T) in the 3' UTR of IRGM (immunity-related GTPase family M protein), a GTPase involved in regulating immunity. This SNP confers risk to Crohn's disease by decreasing the binding of miR-196 [52]. Another example is the presence of SNP rs1625579, which is associated with schizophrenia and located in the intron of a putative primary transcript for the mir137 gene. This SNP alters the seed sequence of miR-137, which is known to regulate neuronal development. Interestingly, four other genes associated with schizophrenia (TCF4 (transcription factor 4), CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit), CSMD1 (CUB and Sushi multiple domains 1) and C10orf26 (chromosome 10 open reading frame 26)) contain predicted target-binding sites for miR-137, suggesting that the expression levels of these four genes might be affected by multiple mechanisms [53]. As a last example, it was suggested that SNP rs13702, which is located in the 3'-UTR of lipoprotein lipase and potentially disrupts the binding site of miR-410, could modulate the effect of diet on plasma lipid levels [54].

The strongest proof for the functionality of a SNP is an expression quantitative trait locus (eQTL) effect of the SNP on a transcript. In 2012 Gamazon et al. showed that 25% of
European 3'-UTR SNPs and 18% of African 3'-UTR SNPs predicted to alter miRNA-binding sites did indeed have a cis-eQTL effect [55]. In another study using publically available eQTL datasets, 26 eQTL SNPs were predicted to create, disrupt, or change the target site of miRNAs in genes encoding xenobiotic metabolizing enzymes [56].

**Bioinformatic and high-throughput methods for studying miRNA function**

Novel genomics techniques, such as next-generation sequencing, and analysis methods, such as eQTL analyses, yield a wealth of data and hypotheses. Concurrent with the increase in the amount of data, web-based applications and online databases have been developed that can be used to analyze miRNA data (Table 1). The main catalog of microRNA sequences is miRBase, which containing 1,872 human precursor miRNAs and 2,578 human mature miRNA sequences (miRBase V.20) [12]. Algorithms to predict miRNA-mRNA interaction, such as TargetScan [26], miRanda [57], TarBase [58] and PicTar [14], use miRBase to perform their miRNA-target predictions.

Several useful algorithms have now been developed to identify SNPs that potentially affect miRNA-binding sites, such as Patrocles [59], PolymiRTs [60], miRSNP [61], microSNiPer [62], miRdSNP [63], and miRNASNP [51]. Most of these applications use the above target prediction algorithms for their predictions. MiRNASNP [51] goes a step further and also investigates whether SNPs affect pre-miRNAs and miRNA seed sequences. As mentioned above, the most powerful indication of a SNP’s functionality is to find an eQTL effect. In the study by Gamazon et al., described above, the authors examined SNPs from the HapMap Consortium [64] located in the 3’ UTRs of genes with cis-eQTL effects using TargetScan, miRBase, Pictar, TarBase, Patrocles and PolymiRTs. This detailed analysis identified 32 3’ UTR eQTL SNPs that potentially affect miRNA binding [55]. Since miRNAs may have many targets, high-throughput methodologies are needed to analyze the data. In the miRNA field, promising approaches involve cross-linking RNA (mRNA and miRNAs) to the RISC complex, immunoprecipitation of these complexes by capturing Argonaute proteins (most often antibodies against Argonaute are applied or cell lines are used that express tagged Argonaute proteins), and sequencing of the mRNA targets in the complex. Examples of such techniques are AGO2 HITS-CLIP [65], PAR-CLIP [66], and CLIP-seq [67]. The disadvantage of these assays is that they do not yield information on specific miRNAs binding to specific targets and, therefore, the miRNA of interest is often overexpressed to enrich for complexes with this specific miRNA. However, in a recent paper by
<table>
<thead>
<tr>
<th>Database</th>
<th>Link</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rfam 11.0</td>
<td><a href="http://rfam.sanger.ac.uk/">http://rfam.sanger.ac.uk/</a></td>
<td>Database of non-coding RNA families, primarily RNAs with a conserved RNA secondary structure and related protein coding RNAs. Non-RNA sequences and their related ncRNAs (exon-exon junctions, PRESENT, conserved RNA cis-regulatory elements) are also included.</td>
</tr>
<tr>
<td>UCSC Genome Browser for functional RNA</td>
<td><a href="http://www.ncrna.org/glocal/cgi-bin/hgGateway">http://www.ncrna.org/glocal/cgi-bin/hgGateway</a></td>
<td>UCSC Genome Browser mirror with large inclusion of functional RNA related custom tracks and several enhancements for RNA secondary structure support.</td>
</tr>
<tr>
<td>NONCODE database</td>
<td><a href="http://www.noncode.org/NONCODERv3/guide.htm">http://www.noncode.org/NONCODERv3/guide.htm</a></td>
<td>Database includes almost all the types of ncRNAs (except transfer RNAs and ribosomal RNAs), ncRNA sequences and their related information (e.g., function, cellular role, cellular location, chromosomal information, etc.).</td>
</tr>
</tbody>
</table>

Table 1. Publicly available databases and bioinformatics tools for non-coding RNAs.
Continuation Table 1. Publicly available databases and bioinformatics tools for non-coding RNAs

<table>
<thead>
<tr>
<th>Database</th>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MirBase</td>
<td><a href="http://www.mirbase.org/">http://www.mirbase.org/</a></td>
<td>Online repository for all microRNA sequences and annotation</td>
</tr>
<tr>
<td>Target Scan Human</td>
<td><a href="http://www.targetscan.org/">http://www.targetscan.org/</a></td>
<td>Prediction of microRNAs targets (exist versions for mouse, worm, drosophila, zebrafish)</td>
</tr>
<tr>
<td><a href="http://www.microrna.gr/miRPathv2">http://www.microrna.gr/miRPathv2</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starbase</td>
<td><a href="http://starbase.sysu.edu.cn/">http://starbase.sysu.edu.cn/</a></td>
<td>miRNA targets and protein RNA interaction from CLIP seq</td>
</tr>
<tr>
<td>miRmap</td>
<td><a href="http://mirmap.ezlab.org/">http://mirmap.ezlab.org/</a></td>
<td>Prediction of miRNA targets that combines different known approaches</td>
</tr>
<tr>
<td>mirRanda</td>
<td><a href="http://www.microrna.org/microrna/home.do">http://www.microrna.org/microrna/home.do</a></td>
<td>Is a comprehensive resource of microRNA target predictions and expression profiles</td>
</tr>
<tr>
<td>microSNiPer</td>
<td><a href="http://epicenter.ie-freiburg.mpg.de/services/microsniper/">http://epicenter.ie-freiburg.mpg.de/services/microsniper/</a></td>
<td>Predicts the impact of a SNP on putative microRNA targets</td>
</tr>
<tr>
<td>mirdSNP</td>
<td><a href="http://mirdsnps.ccr.buffalo.edu/">http://mirdsnps.ccr.buffalo.edu/</a></td>
<td>Database of disease-associated SNPs and microRNA target sites on 3' UTRs of human genes</td>
</tr>
<tr>
<td>mirwalk</td>
<td><a href="http://www.uni-heidelberg.de/apps/zmf/mirwalk/predictedmirnagene.html">http://www.uni-heidelberg.de/apps/zmf/mirwalk/predictedmirnagene.html</a></td>
<td>Database of predicted and validate miRNA targets</td>
</tr>
<tr>
<td>PolimiRTS</td>
<td><a href="http://compbio.uthsc.edu/miRSNP/">http://compbio.uthsc.edu/miRSNP/</a></td>
<td>Database of naturally occurring DNA variations in predicted and experimentally identified miRNA target sites</td>
</tr>
<tr>
<td>Patrocles</td>
<td><a href="http://www.patrocles.org/Patrocles.htm">http://www.patrocles.org/Patrocles.htm</a></td>
<td>Database of DNA polymorphism predict to disturb miRNA target</td>
</tr>
<tr>
<td>miRSNP</td>
<td><a href="http://cmbibjnu.edu.cn/mirsnp">http://cmbibjnu.edu.cn/mirsnp</a></td>
<td>Collection of human SNPs in predicted miRNA-mRNA binding sites.</td>
</tr>
<tr>
<td>Segal Lab. of Computational Biology</td>
<td><a href="http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html">http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html</a></td>
<td>Online microRNA prediction tool-PITA algorithm on the choice of UTRs and microRNAs</td>
</tr>
<tr>
<td>Database</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>LncIPedia</td>
<td>Database for annotated human lncRNA transcript sequences and structures, contains 32,183 human annotated lncRNAs</td>
<td></td>
</tr>
<tr>
<td>LncRNA Database</td>
<td>Database providing comprehensive annotations of eukaryotic lncRNAs</td>
<td></td>
</tr>
<tr>
<td>Human lincRNA catalog</td>
<td>Reference catalog of over 8,000 human lincRNAs</td>
<td></td>
</tr>
<tr>
<td>Functional lncRNA Database</td>
<td>Database containing 204 lncRNAs and their splicing variants, analysis of the lncRNAs and their comparison to protein-coding transcripts</td>
<td></td>
</tr>
<tr>
<td>ncRNA Expression database (NRED)</td>
<td>Database provides gene expression information for thousands of lncRNAs in human and mouse, contains both microarray and in situ hybridization data</td>
<td></td>
</tr>
<tr>
<td>NONCODE v3.0</td>
<td>Integrative annotation of long noncoding RNAs</td>
<td></td>
</tr>
<tr>
<td>ncFANs</td>
<td>Function annotation and function enrichment of lncRNAs, includes 2 organisms (human, mouse) and 3 microarray platforms</td>
<td></td>
</tr>
<tr>
<td>Linc2GO</td>
<td>Web server providing comprehensive functional annotations of human lincRNAs and lncRNAs in other species</td>
<td></td>
</tr>
<tr>
<td>LncRNADisease</td>
<td>Database containing experimentally supported lncRNA-disease associations, also computes gene expression and functional enrichment of lncRNAs</td>
<td></td>
</tr>
<tr>
<td>LncBase</td>
<td>Database containing experimentally verified and predicted miRNA-lncRNA interactions</td>
<td></td>
</tr>
</tbody>
</table>

Continuation Table 1. Publicly available databases and bioinformatics tools for non-coding RNAs.
Helwak et al. CLASH (cross-linking, ligation, and sequencing of hybrids) was described [68]. In this method, loaded RISC complexes are first cross-linked, isolated by Argonaute immunoprecipitation, and then the RNA in these complexes is ligated forming hybrid miRNA-target mRNA sequences that are sequenced. This provides a high-throughput method to identify the targets to which the miRNAs bind and gives an idea of which fraction of the RISC complexes in the cell is occupied by specific miRNAs.

Long non-coding RNAs

Long non-coding RNAs (lncRNAs) are a heterogeneous group defined as transcripts more than 200 nucleotide (nt) in length that exhibit no coding potential [69–71]. They can be isolated from nuclear as well as cytosolic fractions, may or may not be polyadenylated, and >95% of them display alternative splice isoforms [70,72]. In comparison with protein-coding genes, lncRNAs have longer, but fewer, exons [70,73]. LncRNA promoter regions are similarly conserved between vertebrates as promoters of protein-coding genes. In contrast lncRNA exons are less well conserved between these species [69,74]. Like miRNAs, their expression is not regulated by RNA species-specific transcription factors or RNA polymerase molecules. It is becoming clear that lncRNAs exhibit cell-type-specific expression profiles [69] and that in their specific cellular background they can be expressed at levels, similar to the levels of protein-coding RNAs. The origin of lncRNAs is still under debate, but a recent study [75] has reported that more than two-thirds of mature lncRNA transcripts contain transposable elements (TEs), whereas only 4% of protein-coding genes contain these ‘jumping genes’ [76]. This observation led to the postulation that the majority of lncRNAs have arisen via insertion of TEs. For example, the rodent-specific brain cytoplasmic RNA 1 (BC1), the anthropoid primate-specific brain cytoplasmic RNA 200-nucleotide (BC200), and the strepsirhini primate-specific G22 lncRNAs, form a family of lncRNAs which originate independently from insertion of TEs, resulting in lncRNAs that locate to dendrites in different mammalian species [77–80].

The first lncRNAs (H19 and Xist (X-inactive specific transcript)) were discovered using traditional gene mapping approaches in the early 1990s [81–83] and were considered to be rare exceptions to the then central dogma of molecular biology. Using tiling arrays HOTAIR (HOX antisense intergenic RNA) and HOTTIP (HOXA transcript at the distal tip) were discovered in the homeobox gene regions (HOX clusters) [84,85]. In 2009, Guttman et al. was the first to describe a genome-wide approach for discovering lncRNAs that yielded 1,600 novel mouse lncRNAs. In this study gene expression data and the presence of chromatin marks for
promoter regions and gene bodies was integrated with the known annotations of coding transcripts to identify lncRNAs [74]. Since then, thousands of lncRNAs have been identified using similar approaches in the mouse and human genomes [69,86]. When used in combination with next-generation RNA sequencing, additional information can be obtained about lncRNA exon-intron structure as well as about the abundance of these transcripts [87–93]. Cabili et al. combined chromatin marks and RNA-sequencing (RNA-seq) data to generate the human lincRNA catalog, containing more than 8,000 lincRNAs determined across 24 different human cell types and tissues [73]. Although more than 13,500 human lncRNAs have been annotated by ENCODE, only a few dozen have been studied in more detail so far. The challenge is now to elucidate the function of these lincRNAs.

Classes of lncRNAs and mechanism of action

Classification of lncRNAs based on genomic location
The size limit of >200 nucleotides used to define lncRNAs is an arbitrary cut-off based on RNA isolation protocols and their size exclusion limit in the past, which potentially leads to the capture of a heterogeneous group of different transcripts with respect to function. Although different nomenclatures are used, in this review we will adhere to the classification based on the lncRNAs location relative to the nearest known protein-coding gene as described in GENCODE [70]. This subclassification leads to four broad categories (Figure 4). The long intergenic non-coding RNAs (lincRNAs) are the largest group of lncRNAs, accounting for approximately 5,000 human genes according to the GENCODE dataset V17 [70] or approximately 8,000 according to the human lincRNA catalog [73]. LincRNA genes do not overlap or lie next to protein-coding genes [74,94]. The second most prevalent class of lncRNAs is the antisense lncRNAs that is transcribed from the strand opposite of the protein-coding genes, which they are overlapping. Based on their complete or incomplete overlap, antisense lncRNAs can be subdivided into various subclasses, for example, intronic antisense lncRNAs when the lncRNA transcript falls completely within the boundaries of an opposing coding intron, or natural antisense transcripts (NATs) with partial overlap, mainly around the promoter or terminator site of the coding gene [95,96]. For antisense transcripts it holds true that the sense-antisense pairs are often co-expressed together, that they share a similar pattern of evolutionary conservation [97], and that the antisense transcript modulates the expression of the sense transcript by the formation of a sense-antisense RNA duplex [96]. The third subclass of lncRNAs comprises the sense lncRNA transcripts. Such transcripts are located on the same strand and
Genetic variation in the non-coding genome

transcribed in the same direction as a protein-coding gene. This organization is much less prevalent. In total, fewer than 1,000 sense lncRNAs have been identified, overlapping completely or partially with protein-coding genes. To date, this subclass is poorly characterized compared with the other lncRNAs. The fourth subclass of lncRNAs is the bi-directional or divergent group. These transcripts are located on the antisense strand (opposite of the protein-coding gene); they have their transcription start site (TSS) close to the TSS of the protein-coding gene, but are transcribed in the opposite direction. The majority of these bi-directional pairs are co-expressed together and conserved between human and mouse [98,99].

**Classification of lncRNAs based on molecular mechanism (lncRNA archetypes)**

LncRNAs can interact with DNA or RNA as well as proteins. Although the detailed mechanism of action is still unknown for most of the annotated lncRNAs, the few examples available show the complexity of lncRNA biology. LncRNAs control multiple mechanisms that converge upon the
gene expression process. They have been implicated in post-transcriptional gene regulation by controlling processes, like protein synthesis, RNA maturation, and RNA transport, and have been shown to control transcriptional gene silencing via epigenetic regulation and chromatin remodeling [90,100–104].

In 2011, Wang and Chang [101] proposed dividing IncRNAs into four archetypes based on their molecular mechanisms, in an attempt to simplify the complexity of action (Figure 5). The mechanism 1, the signaling archetype, is carried out by IncRNAs that act as molecular signals and may activate or silence other genes without the need for translation (Figure 5A). This mode of action is important for fast, targeted regulation, for example, carried out by IncRNAs involved in embryonic development (HOTAIR and HOTTIP, both regulating homeobox genes) [84,85,105], in the DNA damage response (e.g. lincRNA-p21 and PANDA (p21-associated ncRNA DNA damage activated IncRNA) [106,107], in stress responses (e.g. COLDAIR (cold-assisted intronic non-coding).
Genetic variation in the non-coding genome

RNA), COOLAIR (cold-induced long antisense intragenic RNA)) [108,109], and in somatic cell reprogramming (e.g. lincRNA-ROR (regulator of reprogramming)) [110,111].

The second mechanism of action is the decoy mechanism (Figure 5B). LncRNAs exploiting this strategy (e.g. Gas5 (growth arrest-specific transcript 5), MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), TERRA (telomeric repeat-containing RNA) and PANDA) can act as a decoy that binds to and intervene with the function of other RNAs or proteins, such as miRNAs, transcription factors, or RNA-binding proteins. These lncRNAs, known as “sponges”, can compete with another sequences or structures for binding and are considered as negative regulators. GAS5 acts as a decoy glucocorticoid-response element (GRE) and competes with DNA GREs for binding to the glucocorticoid receptor [112]. PANDA binds to the transcription factor NF-YA and prevents the activation of NF-YA-induced pro-apoptotic targets [107,113].

The third lncRNA mechanism is the ability to act as a guide archetype, for example, by binding proteins and transporting the created complex to specific targets, for example, chromatin modification enzymes to DNA, where the interaction may be directly between this complex and the DNA, or indirectly with heteroduplex protein-DNA (Figure 5-3). These lncRNAs may interact as activators or repressors with neighboring (cis) or distant (trans) genes. Examples of lncRNAs employing this mechanism are HOTAIR, lincRNA-p21, Xist, COLDAIR and Jpx (just proximal to XIST). COLDAIR of Arabidopsis thaliana binds to Polycomb repressive complex 2 (PRC2) and transfers it to chromatin of the flowering locus C repressor, which induces repression of this locus through trimethylation of histone H3 lysine 27 (H3K27), to inhibit flowering during unfavorably low temperatures [108].

The fourth mechanism is acting as a scaffold (Figure 5D), for instance by bringing bound proteins into a complex or in spatial proximity. Examples of lncRNAs exploiting this strategy are ANRIL (antisense ncRNA in the INK4 locus), which functions as a scaffold for the chromatin remodeling complex PRC1 (polycomb repressive complex 1) [114], HOTAIR (scaffold for PRC2 binding it to the LSD1 (lysine-specific demethylase 1A) complex) [84,105,115], and TERC (telomerase RNA component) that scaffolds the telomerase complex [116]. From the above examples, it follows that several lncRNAs can exercise more than one archetypal action, thereby adding to the complexity of the lncRNA world.

LncRNAs in human disease

The lncRNAs that have been studied in most detail have emerged as important regulators of gene expression. In many diseases the expression of protein-coding genes is deregulated and evidence is now accumulating that
altered lncRNA function might be one of the causes involved. Below we will describe examples of lncRNAs known or suggested to be involved in the etiopathology of different disorders.

Most evidence for a role of lncRNAs in disease has been found in cancer. Many lncRNAs have been described that exhibit altered expression levels in cancer cells compared to healthy tissue of the same origin [117]. LncRNA MALAT1, also known as NEAT2 (nuclear-enriched abundant transcript 2), was initially discovered as a predictive biomarker for metastasis development in lung cancer [118,119], but only almost ten years later, its mechanism of action was clarified. MALAT1 acts by inducing the expression of metastasis-associated genes [120] and it was recently reported that the in vitro metastasis of human lung cancer (EBC-1) cells can be inhibited by inhibiting MALAT1 using antisense oligonucleotides [120,121]. It is disappointing that three independently derived Malat1 knockout (Malat1-KO) mouse models exhibit no obvious phenotype or histological abnormalities [54–56].

Another lncRNA involved in metastasis is HOTAIR, which interacts with PRC2 and alters chromatin to a metastasis-promoting state [122]. In approximately one-quarter of human breast cancers, HOTAIR is highly induced, while its elevated levels are also predictive of metastasis and disease progression in other cancers, such as colon, colorectal, gastrointestinal, pancreatic and liver cancer [115,123–126]. LncRNAs are also connected to other processes involved in human cancer development and progression. ANRIL, GAS5 and lincRNA-p21 are involved in the escape of growth suppression by regulating tumor suppressor genes (ANRIL) or apoptosis regulators (GAS5, lincRNA-p21). TERRA and TERC (telomerase RNA component) regulate replicative immortality [127,128]. MALAT1 and HOTAIR activate cancer invasion and metastasis by regulating cell motility-related genes (MALAT1) or retargeting of PRC2 complex and changing it to a pattern similar to the one of embryonic fibroblasts, causing increase in cancer cell invasiveness and their ability to metastase (HOTAIR) [115,120,129]. The lncRNAs HIF (antisense to hypoxia inducible factor (HIF)) and tie-1AS (tyrosine kinase containing immunoglobulin and epidermal growth factor homology domain-1 antisense) induce angiogenesis [94,130]. PCGEM1 (prostate-specific transcript 1), UCA1 (urothelial cancer associated 1, also known as CUDR, cancer upregulated drug resistant), SPRY4-IT1 (SPRY4 intronic transcript 1), and PANDA are involved in suppressing apoptosis [107,131–133]. Some lncRNAs have, as yet, only been associated with one specific type of cancer. PCGEM1, PCA3 (prostate cancer antigen 3, known also as DD3, differential display code 3) and PCNCR1 (prostate cancer ncRNA 1) are involved in prostate cancer, while HULC (highly
up-regulated in liver cancer) is involved with liver cancer [131,134–138]. In contrast, other lincRNAs (e.g. HOTAIR and MALAT1) appear to be broadly deregulated in carcinogenesis [118,124,125,139,140].

More recently, data has been generated that shows that lncRNAs also have roles in other diseases. Some human pathological phenotypes are caused by epigenetic changes and imprinted lncRNA gene clusters have been associated with these diseases [141]. Examples of these are the imprinting-related, neurogenetic Angelman syndrome and Beckwith-Wiedemann syndrome (BWS) [142]. Angelman syndrome is caused by a loss-of-function of ubiquitin-protein ligase E3A (UBE3A), also known as E6AP [143]. Although in the majority of human tissues, both copies of the UBE3A gene are expressed, in neurons one copy is silenced by UBE3A-AS1 (ubiquitin-protein ligase E3A antisense RNA 1) [144]. In patients suffering from Angelman syndrome, the other (active) allele has either been deleted or inactivated [144]. In more than half of the cases of Beckwith-Wiedemann syndrome, maternal methylation at DMR2 (differentially methylated imprinting center 2) has been lost [145]. This center contains the two protein-coding genes CDKN1C (cyclin-dependent kinase inhibitor 1C) and KCNQ1 (potassium voltage-gated channel member 1) and antisense IncRNA KCNQ1OT1 (KCNQ1 overlapping transcript 1) [146]. In healthy individuals the maternal KCNQ1OT1 IncRNA is silenced via methylated KCNQ1 and CDKN1C is transcribed to its protein p57, a negative regulator of cell proliferation [146]. In 60% of patients with BWS, the KCNQ1 gene is not methylated, KCNQ1OT1 IncRNA is transcribed from both alleles, leading to cis repression of CDKN1C and disease [145–147].

LncRNAs have also been associated with other neurological disorders, such as BACE1-AS or BC200 in Alzheimer disease, HAR1 (human accelerated region 1 lncRNA) in Huntington disease, and ATXN8OS (Ataxin 8 opposite strand lncRNA) in spinocerebellar ataxia type 8 [142,148,149]. In Alzheimer disease, the protein-coding gene BACE-1 (α-site amyloid precursor protein-cleaving enzyme) cleaves amyloid precursor protein (APP) to α-amyloid peptide (A), the accumulation of which (amyloid plaques) is associated with disease. LncRNA BACE1-AS, located on the antisense strand to BACE1, binds complementarily to BACE1 mRNA, increases its stability, regulates BACE1 translation, and thereby the production of A [150]. BACE1-AS also prevents miRNA-induced repression of BACE-1 by masking the binding site for miR-485-5p, by competing for the same region on exon 6 of BACE-1 [151]. Two- to three-fold increased levels of BACE1-AS and a smaller increase (1.5x) of BACE-1 have been documented post mortem in the brains of patients with Alzheimer disease [150].
Besides their actions in cancer and neurological diseases, IncRNAs also exhibit aberrant expression in other disease states, such as facioscapulohumeral muscular dystrophy (FSHD). FSHD is a common, progressive, genetic disease of skeletal muscle caused by deletions that reduce the number of D4Z4 repeats in the FSHD locus at chromosome 4q35 [142,152]. Under physiological conditions, D4Z4 repeats recruit Polycomb complexes resulting in chromatin reorganization and causing repression of 4q35 genes [153]. In contrast, in FSHD patients, a deletion of D4Z4 repeats results in cis production of the DBE-T IncRNA that binds to protein complexes, reorganizes the chromatin state of the FSHD locus, and reactivates the repressed 4q35 genes [153].

**Genetic variants and IncRNA function**

As yet, how far IncRNAs are affected by genetic alterations related to the disease phenotype is unknown. The larger alterations, like chromosomal rearrangements (translocations, amplifications, or deletions) do affect the expression of IncRNAs that are involved in disease phenotypes. For example, two different balanced translocations (t(8;12)(q13;p11.2) and t(4;12)(q13.2-13.3;p11.2)) affecting chromosome 12p have been associated with the human brachydactyly type E phenotype [154,155]. The IncRNA DA125942 in the affected 12p region was recently described as interacting in cis with Parathyroid hormone-like hormone (PTHLH, which is a regulator of endochondral bone development) and in trans with SOX9 (Sex determining region Y-box 9, which acts during chondrocyte differentiation) [111,154]. Another balanced translocation (t(1;11)(q42.1;q14.3)) has been associated with schizophrenia and other psychiatric disorders [156,157]. This particular translocation affects two genes, the protein-coding disrupted in schizophrenia 1 gene (DISC1) and the antisense IncRNA disrupted in the schizophrenia 2 gene (DISC2) [157,158]. The same DISC1/DISC2 region was also associated with a 1q42 deletion in an autism spectrum disorder [159].

The interactions of IncRNAs with other molecules are probably governed by IncRNA structure, rather than by sequence. As it is difficult to predict the structure of larger RNA molecules, it is a challenge to understand how small mutations (small insertions/deletions or SNPs) are involved in disease etiology. Here we describe a few scenarios that can be envisioned. Firstly, SNPs may affect the expression level of IncRNAs. SNPs can be located in promoter sequence and directly alter the expression of IncRNAs (Figure 6A) or change the binding of inhibitory complexes, thereby allowing expressing of IncRNAs that are not expressed under physiological circumstances (Figure 6B). Secondly, SNPs within IncRNA genes may cause
Genetic variation in the non-coding genome

alternative splicing of the transcript (Figure 6C) or affect its secondary structure (Figure 6D). This can lead to an altered function of the lncRNAs. For instance, the chromosomal region 9p21 (previously described as a "gene desert") harbors the ANRIL lncRNA (also known as CDKN2BAS or CDKN2B-AS1). Many SNPs located within or around this lncRNA have been associated in GWAS with a susceptibility to atherosclerotic vascular disease, coronary artery disease, stroke, myocardial infarction, aortic aneurysm, type 2 diabetes, and several types of cancers [160,161]. Exactly how these SNPs contribute to disease is not yet known. It is unclear whether they directly regulate the expression level of ANRIL or whether they disturb the binding site of transcription factor STAT1 (ANRIL’s repressor) in ANRIL’s enhancer (Figure 6B) [162]. SNPs may also modulate ANRIL transcripts by inducing exon skipping, resulting in shorter splice variants with reduced efficiency or non-functional isoforms (Figure 6C) [160].

Bioinformatics tools and high-throughput applications for studying lncRNAs

As lncRNAs are a relatively novel class of transcripts, there are not yet many bioinformatics tools to aid the
study of IncRNA function. To date, a couple of IncRNAs databases and catalogs are available (Table 1), such as Lincipedia [163], LncRNAdb [79], and the Human lincRNA catalog [73]. LncRNAdb contains a comprehensive collection of eukaryotic IncRNAs and relevant information, such as RNA sequence, structure, genomic location, expression profiles, subcellular localization, conservation and function [79]. The Human lincRNA catalog [73] focuses only on human lincRNAs and contains more than 8,000 lincRNAs defined by more than 30 properties. Expression data across 24 tissues and cell types is available in this database. Data from both databases can be visualized and analyzed using the University of California Santa Cruz (UCSC) genome browser [164]. In the same browser, it is also possible to annotate ncRNAs with the latest updated genome-wide data from large international consortia, such as ENCODE and Functional Annotation of the Mammalian Genome (FANTOM) [165–167].

Co-expression analysis is a promising strategy to predict the mechanism of action of IncRNAs of interest. This type of analysis predicts functions for molecules of unknown gene products, based on co-expression data and taking into account the known functions or mechanisms of the co-expressed genes ("guilt by association"). Two examples of such tools are Gene Network (www.genenetwork.nl/genenetwork) (Karjalainen and Franke et al., manuscript in preparation) or Gemma (http://www.chibi.ubc.ca/Gemma/home.html) [168]. However, the data most commonly used for these analyses comes from arrays. Although nearly all the arrays were designed to study the expression of protein-coding genes, several of them also carry probes for IncRNAs. Kumar et al. (2012) took advantage of this and used microarray data to investigate the association of SNPs with the expression levels of lincRNAs by using eQTL and a platform containing approximately 2,000 IncRNA probes [9]. They discovered 112 cis-regulated lincRNAs, of which 45% were replicated in an independent dataset. A remarkable 75% of these SNPs affected the expression of the lincRNA but did not affect the expression of neighboring protein-coding genes. Expression microarray platforms probing both protein-coding transcripts and non-coding RNAs will enable co-expression analyses of these RNA classes. The data can then be used to link IncRNAs to pathways, using existing data for the coding genes with which they are co-expressed.

In the near future, enough transcriptomic data is expected to be generated by next-generation sequencing to allow in depth co-expression analysis. The initial studies in the field were mostly focused on IncRNAs in the polyadenylated (polyA+) RNA fraction. More recently, a significant proportion of IncRNAs also present in the non polyadenylated
Genetic variation in the non-coding genome (polyA-) RNA fraction have been described [169]. For this reason, the latest focus is on both RNA fractions (polyA+, polyA-) as well as on different subcellular compartments (nuclear, cytoplasmic) [69,72]. It is expected that many more lncRNA transcripts will be identified [170].

Because lncRNA function is based on RNA structure rather than on RNA sequence, it is difficult to predict lncRNA targets or to pinpoint the functional domains in these molecules. However, a new tool has recently been described that can be used to predict regions within chromatin-associated non-coding RNAs that can insert themselves into DNA stretches, which are then likely to be able to form a triple helix structure [171].

The techniques used for the genome-wide identification of lncRNA targets and binding partners employ methods similar to the ones described above for finding miRNA targets. Chromatin Isolation by RNA Purification (ChIRP) [172] and Capture Hybridization Analysis of RNA Targets (CHART) [173] are based on cross-linking RNA in complex with DNA and/or protein, followed by capture techniques targeting the lncRNA in the complex. Subsequently, the protein component or DNA sequences in the complexes can be identified by mass spectrophotometry [172] or next-generation sequencing, respectively. Conversely, since lncRNAs have been reported to be involved in targeting chromatin-modifying complexes to specific DNA sequences, one can cross-link complexes and immunoprecipitate these (e.g. the PRC2 complex), and then analyze the RNA component by sequencing (RNA immunoprecipitation-sequencing (RIP-SEQ)) [174]. Finally, lncRNAs can be synthesized in vitro and hybridized with protein microarrays to identify lncRNA-protein interactions [99].

**Perspectives**

The rapid evolution of next-generation sequencing technologies and the expected drop in cost of assays means we will soon be able to sequence large numbers of genomes and transcriptomes. This will ultimately provide us with a nearly complete overview of disease-associated genetic variation. Transcriptome analysis will discover tens of thousands of new transcripts, many of which will be of a non-coding nature [170]. Most of the genetic variation associated with complex diseases is located within genomic regions that control transcription, rather than in protein-coding areas. As the ncRNAs are now emerging as important regulators of expression, it is becoming clear that deregulation of gene expression might be the lynch-pin in many disease mechanisms. Because of the above it is essential to understand the function of ncRNAs in health and disease.

What is already emerging and will complicate the study of the regulatory involvement of ncRNAs in human diseases even more, is that different classes of ncRNAs might interact. A few lncRNAs have
been reported to be precursors for different types of small ncRNAs, such as miRNAs or small nucleolar RNAs [69,175]. In addition, miRNAs and lncRNAs can regulate each other's expression. For example, Zhang et al. (2013) reported miR-21 which is able to negatively regulate the expression of the GAS5 lncRNA. On the other hand GAS5 was able to repress miR-21 [176]. LncRNAs can also act as microRNA ‘sponges’. The most striking example is the circular lncRNA CDR1 (cerebellar degeneration-related protein 1), which has been shown to contain approximately 70 binding sites for miR-7. By binding miR-7 molecules CDR1 prevents binding of this miRNA to other miR-7 targets [177,178].

The availability of genomic and transcriptomic data will also lead to an increase in the number of ncRNA eQTLs associated with disease [9]. Once the causal mutations have been connected to specific ncRNAs, the next step will be to identify the targets/interaction partners of these transcripts, which is essential to fully understanding the mechanism of disease. Because both miRNAs and lncRNAs may act on many targets, high-throughput methods will need to be designed and applied. Eventually, genome-editing techniques might be applied to pinpoint the effects of a single SNP or of small deletions/insertions by allowing alterations of wild-type alleles into risk alleles in eukaryotic cells [179].

The ‘omics’-revolution started with the description of the sequence of the human genome. Since then state-of-the-art high-throughput techniques and bioinformatic approaches have been developed, that were used to identify mutations that are associated with human diseases. It has now become apparent that a significant part of these mutations affect the expression or function of ncRNAs. To fully understand the function of these ncRNAs and how this function is affected in disease, biologists will need to investigate the function of individual ncRNAs. Because both miRNAs and lncRNAs exhibit cell type, and even cell developmental stage specific expression profiles, it is pivotal that this research will be performed in the relevant (disease associated) cell types.

Acknowledgements
The authors would like to thank Jackie Senior and Kate Mc Intyre for critically reading the manuscript. The authors of this review are paid by, or have obtained grants from the Netherlands Organization for Scientific Research (NWO-VICI 918.66.620; BH), the Coeliac Disease Consortium 2 (BSIK grant 03009; BH), the Capes Foundation, Ministry of Education of Brazil (BEX0891-10-0 to RA), the European Research Council Advanced Grant (ERC-671274; ZB & SW), and the Dutch Multiple Sclerosis Foundation (11-752 to SW).
Genetic variation in the non-coding genome

References


[19] G. Stefani, F.J. Slack, Small non-coding RNAs in animal development,
219–230.


[37] M.T. de la Morena, J.L. Eitson, I.M. Dozsmorov, S. Belkaya, A.R. Hoover,
Genetic variation in the non-coding genome


Genetic variation in the non-coding genome


[86] A.M. Khalil, M. Guttman, M. Huarte,


Genetic variation in the non-coding genome

651–669.


[121] T. Gutschner, M. Hammerle, S. Diederichs, MALAT1 -- a paradigm


[137] S. Chung, H. Nakagawa, M. Uemura,
Genetic variation in the non-coding genome


variation to a polycomb/trithorax
epigenetic switch in FSHD muscular


Genetic variation in the non-coding genome