“SAPIENZA” UNIVERSITY OF ROME
Department of Cellular Biotechnology and Haematology

Ph.D. program: “HUMAN BIOLOGY AND GENETICS”
XXVI cycle

RNAi machinery cooperates with SWI/SNF complexes in nucleosome positioning at Transcriptional Start Sites

Candidate:
Dr. Silvia Gioiosa

Scientific tutor:
Prof. Giuseppe Macino “Sapienza” University of Rome

Director of Doctoral Program:
Prof. Paola Londei “Sapienza” University of Rome

Board of examiners:
Prof. M.Eugenia Schininà, Roma, “Sapienza”
Prof. Antonio Antoccia, Roma, “Università di Roma3”
Prof. Graziano Pesole, Bari, “Università Aldo Moro”

Academic year 2012-2013

INDEX

ABSTRACT

In Eukaryotes, Argonaute (AGO) proteins have a well-established role in the cytoplasm in post-transcriptional regulation of gene expression in association with different classes of small non-coding RNAs (sRNAs).
In plants and yeast, it has been demonstrated that AGO proteins exert a role in the epigenetic regulation of chromatin modifications. Furthermore, AGO2 protein acts also in the nuclei of human cell lines and emerging literature reports that upon the transfection of sRNAs complementary to non-coding promoter transcripts, AGO2 is recruited on target promoters.
Previous results in our laboratory demonstrated that AGO2 and SWI/SNF have a physical interaction, which is independent of RNA or DNA, in human cell lines. As SWI/SNF is the major chromatin-remodelling complex in human, these data suggest that AGO2 might participate in the regulation of chromatin plasticity. In eukaryotes, the proper organization of chromatin is essential for the control of gene expression and is achieved through the concerted activity of histone modifications, DNA methylation and nucleosome positioning.
The focus of the present thesis has been the development of relevant bioinformatics pipelines for data processing, analysis and visualization, all aiming at dissection of the functional significance of the AGO2-SWI/SNF interaction.

Interestingly, this bioinformatics pipeline allowed me to identify a novel class of nuclear AGO2-bound sRNAs arising from genomic regions 150 nt around the Transcription Start Sites (TSS) bound by SWI/SNF (swiRNAs). Furthermore, swiRNAs present a Dicer-dependent processing and show an involvement in nucleosome occupancy at nucleosome +1. These data represent the first description of a molecular mechanism through which AGO2 is involved in nucleosome positioning in mammalian cells.

INTRODUCTION

The Argonaute protein family

Argonaute proteins constitute a highly conserved protein family whose number is much variable between species, ranging from one in S. pombe (Verdel et al., 2004) to 27 in *C.elegans* (Grishok et al., 2001).
In mammals eight Argonaute proteins have been identified (Sasaki et al., 2003).
These proteins are highly specialized binding modules that accommodate the small RNA component — such as microRNAs (miRNAs), short interfering RNAs (siRNAs) or PIWI-associated RNAs (piRNAs) — and coordinate downstream gene-silencing events by interacting with other protein factors.

Structural features of AGO proteins

Argonaute proteins are multi-domain proteins that contain an N-terminal domain, and a PAZ, middle (MID) and PIWI domain (Fig.I1a). The recent determination of the crystal structure of full-length human AGO2, consistent with previous studies of prokaryotic homologues, has revealed a bilobate architecture, with the MID and PIWI domains forming one

lobe, and the N-terminal and PAZ domains constituting the other (Fig.I1b) (Schirle NT, 2012).
The interaction between small RNAs and AGOs does not occur through sequence-specific contacts but through several contact points on the protein, mediated by distinct domains (Yan KS, 2003).

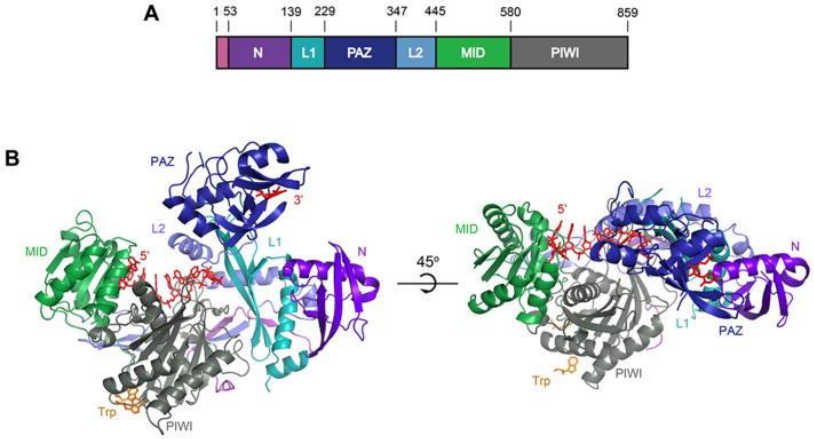


Fig. I1 A: Schematic representation of the Ago2 primary sequence. **B:** Front and top views of Ago2 with the N (purple), PAZ (navy), MID (green), PIWI (grey) domains and linkers L1 (teal) and L2 (blue). A generic guide RNA (red) can be traced for nucleotides 1–8 and 21. Tryptophan molecules (orange) bind to tandem hydrophobic pockets in the PIWI domain *(Schirle NT, 2012).*

In particular, the small RNA 5' and 3' termini are recognized by the PAZ and MID domains, respectively. The MID domain

contains a highly basic pocket, which specifically binds the characteristic 5' phosphate of sRNAs and anchors the sRNA onto Argonaute proteins (Ma JB, 2005). Once the sRNA has been loaded, Argonaute proteins silence target RNAs by slicing activity that resides in the PIWI domain. The PAZ domain contains a specific binding pocket that specifically recognizes the characteristic 2-nucleotide 3' overhang and the base-paired terminus of siRNA duplexes produced by Dicer processing (Lingel et al., 2003). A segment of the human PIWI domain has been shown to mediate protein-protein interaction between Argonaute proteins and Dicer, which may facilitate the incorporation of the siRNA into the effector RNA silencing complex (Doi et al., 2003).

The PIWI domain shows extensive homology to RNase H, an endoribonuclease that cleaves RNA–DNA hybrids (Parker JS, 2004). In vitro assays using RNA substrates complementary to exogenous siRNAs (Liu J, 2004) as well as endogenous miRNAs (Meister G, 2004) have identified AGO2 as the only member of the human AGO subfamily with endonuclease activity.

Sub-cellular localization of AGO proteins

AGO proteins have been implicated in both post-transcriptional and transcriptional gene-expression regulation. This dual level of regulation reflects the localization of AGO both in the cytoplasm and in the nucleus.

Previous studies have shown that AGO proteins are located in the nucleus of cells of yeast (Verdel et al., 2004), plants

(Zilberman et al., 2003), *Drosophila melanogaster* (Pal-Bahdra et al., 2004) and human cell lines (Robb GB, 2005).

The distinct sub-cellular localization of AGO proteins is mediated by accessory proteins and post-translational modifications, which coordinate AGO activity with the plethora of cellular signals. In the cytoplasm AGO proteins are enriched in distinct foci such as P-bodies and stress-granules. The former are cellular sites where mRNA turnover and storage occur (Eulalio A, 2007a), the latter are structures induced upon cellular stress and contain mRNAs stalled in the process of translation initiation (Leung AK, 2006). It has been reported that the AGO2 P-body localization is mediated by several accessory proteins such as GW182 (Eulalio A, 2007b), and influenced by post-translational modifications such as phosphorylation at serine-387 (Zeng Y, 2008).

Emerging evidence shows that the cytoplasmic and nuclear AGO2-functions are coordinated by a common set of proteins. Imp8, which is required for binding of AGO proteins to a variety of mRNA targets in the cytoplasm, modulates nuclear localization of AGO2 as well (Weinmann L, 2009). Similarly, TNRC6A, that interacts with AGOs and triggers translational repression and/or mRNA degradation in P-bodies, translocates AGO2 in and out the nucleus via its own recently discovered nuclear localization signals (NLS) and nuclear export signals (NES) (Nishi K, 2013).

Loading of AGO Proteins with small RNAs

Naturally produced dsRNAs may derive from several different sources. Molecules of dsRNA may emerge during viral infection and replication (Li et al., 2002;) or after transposition of mobile genetic elements (Ketting et al., 1999). Transcribed pseudogenes (Hirotsune et al., 2003), endogenous repetitive gene loci (Aravin et al., 2001) or microRNA genes (Bartel, 2004-a) represent other endogenous sources of dsRNAs.

Molecules of dsRNA are processed by Drosha and/or Dicer RNase III enzymes into small dsRNA of specific length and structure (Bernstein et al., 2001) which can enter into various gene silencing pathways that are collectively referred to as RNA silencing (Grishok et al., 2001).

Ago proteins most likely recognize the characteristic 2 nt 3' overhangs of small dsRNA by their PAZ domains.

Depending on the source of dsRNA, the products of Dicer processing are termed small interfering RNAs (siRNAs) or microRNAs (miRNAs).

siRNAs. After Dicer-mediated cleavage, one strand of the siRNA duplex (the guide strand) is loaded in the RISC-loading complex (Hammond et al., 2000), a ternary complex that consists of an AGO protein, Dicer and a dsRNA-binding protein (known as TRBP in humans). During loading, the non-guide (passenger) strand is cleaved and ejected.

miRNAs. These small RNAs are transcribed by polymerase II from endogenous miRNA genes resulting in a primary transcripts (pri-miRNAs) that contain ~ 65–70-nucleotide stem–loop structures. The hairpin structure is excised in the nucleus by the Drosha–DGCR8 complex to yield a precursor miRNA (pri-miRNAs). After its export to the cytoplasm, the

pri-miRNAs undergoes another endonucleolytic cleavage, which is catalysed by Dicer, generating a miRNA–miRNA* duplex of ~21–25 nucleotides (where miRNA is the guide strand and miRNA* is the passenger strand). For most miRNAs, only one strand accumulates as mature miRNA.
Such asymmetric loading is guided by the relative thermodynamic stability of the 5' ends of the small RNA duplex. The strand whose 5' end is less stably paired is preferentially incorporated into AGO complexes, whilst the passenger strand is subsequently degraded (Bartel DP, 2009).

Ago-Mediated Post-transcriptional Gene Silencing Pathways

Sequence specific gene silencing triggered by double-stranded RNA (dsRNA) is a fundamental gene regulatory mechanism present in almost all eukaryotes (Béclin et al., 2002; Denli and Hannon, 2003; Dykxhoorn et al.,2003; Finnegan and Matzke, 2003; Grewal and Moazed, 2003; Hannon, 2002; Plasterk, 2002). Gene silencing mediated by dsRNAs has been shown to act at the post-transcriptional level.
Depending not only on the bound small RNA but also on the specific mRNA that is targeted, Ago protein complexes mediate different posttranscriptional gene silencing mechanisms.
Post-transcriptional regulation. The mechanism by which a small RNA regulates its target mRNAs reflects both the specific AGO protein association and the extent of complementarity between the small RNA and the mRNA. In

general, a full complementarity to target mRNAs, typical of siRNAs, directs endonucleolytic cleavage (slicing) of the mRNA (Elbashir et al., 2001). After slicing, the cleaved target RNA is released, and the RISC is recycled for another round of slicing (Zamore PD, 2001). **(Fig. I2-b)**

miRNAs not only can guide RNA cleavage (Llave et al., 2002), but are predominantly considered to act as translational repressors on partially complementary, evolutionary conserved sequences in the 3' UTR of the target mRNAs (Aukerman and Sakai, 2003).

miRNAs generally interact with their mRNA targets through a limited base-pairing of only 2-7 nt (seed region) at the 5'end. With few exceptions, miRNAs-binding sites lie in the 3'UTR of target mRNA and are usually present in multiple copies. Typically, mRNA may have several putative targets for both the same and different miRNAs, and any particular miRNA may have hundreds of putative mRNA targets. Hence, given the substantial number of miRNAs, a big combinatorial network of miRNA-mediated post-transcriptional gene regulation exists (Selbach M, 2008).

Since their initial identification in 1993 (Lee RC. et al., 1993), there have been several efforts for the identification of miRNA targeted genes and computational target prediction remains one of the key means to analyse the role of miRNAs in biological processes. Therefore, many miRNA target prediction programs have been published, such as TargetScan (Lewis et al., 2005), Pictar (Lall et al., 2006) and Diana-microT (Maragkakis et al., 2009). Most of these programs are mainly based on sequence alignment of the miRNA seed region (nucleotides 2–7 from the

5′-end of the miRNA) to the 3′-UTR of candidate target genes leading to the identification of putative binding sites.

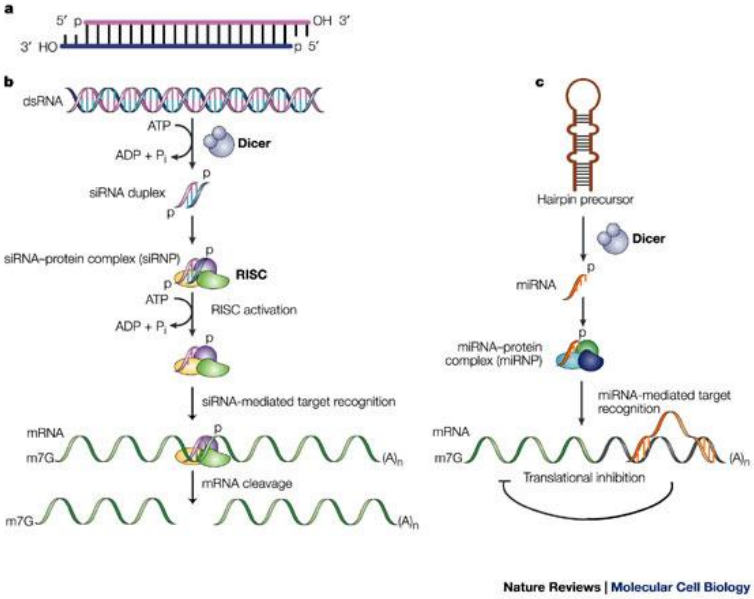


Fig I2: a. Short interfering (si)RNAs. Molecular hallmarks of an siRNA include 5′ phosphorylated ends, a 19-nucleotide (nt) duplexed region and 2-nt unpaired and unphosphorylated 3′ ends that are characteristic of RNase III cleavage products14. **b. The siRNA pathway.** Long double-stranded (ds)RNA is cleaved by Dicer, into siRNAs in an ATP-dependent reaction. siRNAs are then incorporated into the RNA-inducing silencing complex (RISC). Although the uptake of siRNAs by RISC is independent of ATP, the unwinding of the siRNA duplex requires ATP. Once unwound, the single-stranded antisense strand guides RISC to mRNA that has a

complementary sequence, which results in the endonucleolytic cleavage of the target mRNA. **c |The micro (mi)RNA pathway.** Although originally identified on the basis of its ability to process long dsRNA, Dicer can also cleave the ~70-nt hairpin miRNA precursor to produce ~22-nt miRNA. They target mRNA leading to translational repression.
(Dykxhoorn et al.,2003).

Their specificity is usually increased by exploiting the commonly observed evolutionary conservation of the binding sites or by using additional features such as structural accessibility (Kertesz et al., 2007), nucleotide composition (Grimson et al., 2007) as well as location of the binding sites within the 3′-UTR (Gaidatzis et al.,2007).

Degradation of mRNAs associated to AGO-miRNA complexes involves deadenylation (removal of the poly(A) tail) (Eulalio A, 2007c), and/or decapping (removal of the 7-methylguanosine cap structure) (Rehwinkel J, 2005). The exposure of mRNA termini leads to exonucleolytic digestion from the 5' and 3'ends by the action of the exosome and of the exonuclease XRN1, respectively (Parker R, 2004b). miRNAs might also silence their targets by sequestering mRNAs into P-bodies, which exclude the translation machinery, and releasing them upon specific signals allowing the re-enter into polyribosomes fraction for translation (Bhattacharyya SN, 2006).

Transcriptional regulation. In addition to their post-transcriptional role in the cytoplasm, AGO proteins also function in the nucleus either repressing or activating

transcription (Gagnon KT, 2012). In humans, the existence of nuclear RNAi pathways was first inferred from observations that nuclear miRNAs and siRNAs could cause cleavage of RNA targets, such as 7SK small nuclear RNA (Robb GB, 2005). In addition, endogenous human AGO2 was identified in the nucleus with highly specific antibodies (Rudel S, 2008). The first report of human transcriptional gene silencing (TGS) utilized exogenous siRNAs to silence expression of an integrated GFP reporter driven by the eukaryotic translation elongation factor alpha (EF1α) promoter (Morris KV, 2004). The silencing at transcriptional level was confirmed by nuclear run-on analysis, which is the gold standard for differentiating silencing effects mediated by TGS from those of the post-transcriptional pathways.

Subsequently, other reports described the silencing of various genes, including E cadherin, RASSF1, TGFb receptor II, progesterone receptor, major vault protein, androgen receptor, cyclooxygenase-2, CDH1, and c-myc (Ting AH, 2005; Castanotto D, 2005; Janowski BA, 2007; Kim JW, 2007; Green VA, 2011). Mechanistic details about TGS have begun to emerge after the observation that siRNA-induced silencing at the EF1α promoter was sensitive to the Pol II inhibitor α-amanitin (Weinberg MS, 2006). This observation led to the proposal of two models to explain the mechanism of TGS: (a) the siRNA binds to DNA, facilitated by the opening of the DNA duplex by the transcription machinery; (b) the siRNA binds to nascent promoter-associated RNA. Work by Han et al. favoured the latter by revealing a requirement for a promoter-associated transcript for small RNA-mediated gene silencing

(Han J, 2007). The antisense strand of the siRNA was shown to target an EF1α promoter-associated RNA variant with an extended 5'UTR. In addition, the association was inhibited by RNase A but resistant to RNase H treatment, implicating an RNA-RNA interaction in TGS.

Other studies have also demonstrated a requirement for promoter-associated transcripts for small RNA-mediated gene silencing. Some of these studies have shown that, similar to EF1α, silencing requires sense strand transcripts (Hawkins PG, 2009), while others have found that antisense transcription through the targeted region is necessary (Schwartz JC, 2008). The findings reported to date, support a model for TGS in which AGO2-small RNAs complexes are guided to their targets by complementary base-pairing with low copy, promoter-associated Pol II transcript.

AGO2 serves as a scaffold for the recruitment of chromatin-modifying complexes that favour heterochromatin formation and, therefore decreasing transcription of the targeted gene (Hall et al., 2002). More recently, small RNA-targeting of sequences beyond the mRNA 3'terminus of the PGR gene has been demonstrated to induce TGS (Yue X, 2010). Targeting genomic regions outside promoter sequence involves a looping mechanism bringing the 3' terminus and the promoter into close proximity to allow modulation of promoter activity. The report of transcriptional gene activation (TGA) showed increase expression of E- cadherin, p21, and VEGF on the transfection of promoter-targeting siRNAs into cultured cells (Li LC, 2006). The mechanism by which transcriptional activation is achieved shares common features with the

transcriptional silencing. The transcriptional opposite outcome depends on the specific region targeted by the small RNA and by the different set of modifying complexes recruited. Emerging evidence has revealed that the TGA/TGS in human cells can be mediated not only by exogenous small RNA but also by endogenous ones. Consistently, it has been reported that a substantial fraction of human miRNAs are present in the nucleus, with even an abundance greater than their cytoplasmic levels, and, in some cases, have shown to alter promoter activity (Kim DH, 2008; Place RF, 2008). It is the case of miR-373 and miR-320, in which the former induces and the latter represses transcription (Place RF, 2008) (Kim DH, 2008). Moreover, endo-siRNAs with a putative role in transcriptional regulation have been recently discovered also in mammals (Yang N, 2006). Many are the genomic sources of dsRNA triggering for endo-siRNAs: structured loci that pair intra-molecularly to produce long dsRNA, complementary overlapping transcripts and bidirectionally transcribed loci, protein-coding genes that associate with the cognate pseudogenes and from regions of pseudogenes that form inverted repeated structures (Ghildiyal M, 2009).

SWI/SNF chromatin re-modeller complexes

DNA storage and chromosome packaging are important biological processes common to all living organisms. In eukaryotes, the advent of nucleosome-based DNA organization not only promoted favourable packing ratios that increase

cellular DNA content but also facilitated the development of complex regulatory mechanisms. The basic unit of chromatin is the nucleosome, which consists of 146 base pairs of duplex DNA wrapped around a histone octamer comprised of two of each of the conventional histone proteins: H2A, H2B, H3 and H4. Highly related histone variants are also incorporated throughout the genome for regulatory purposes (Talbert PB, 2010). A fifth histone protein, H1, promotes higher order chromatin structures by encouraging condensation of neighbouring nucleosomes from "beads on a string" to the 30 nm fibre. As this resulting fibre accounts for only ~25-fold of a 5000- fold DNA-to-nucleus compaction ratio, several other mechanisms must contribute to higher order compaction and nuclear organization. Inherent to any packing solution is the need for reversibility, as DNA must remain accessible to cellular machinery as required. To accomplish this task, chromatin re-modelling proteins form crucial complexes that reposition nucleosomes. They utilize the energy of ATP to disrupt nucleosome-DNA contacts, move nucleosomes along DNA, and remove or insert histone octamers.

The best-studied family of chromatin remodeler is the evolutionarily conserved SWI/SNF family (Clapier CR, 2009). In mammals, these enzymes play an essential role in several aspects of embryonic development including pluripotency (Yan et al., 2008), cardiac development (Hang et al., 2010), dendritic morphogenesis (Kim JK et al., 2001) and self-renewal of neural stem cells (Kidder et al., 2009). In the adult, deletion or mutation of these proteins often leads to tumorigenesis as a consequence of a dysregulated cell cycle control (Weissman B,

2009). The SWI/SNF family of chromatin re-modellers spans eukaryotic lineages. Purified SWI/SNF complexes contain 10–12 polypeptides and have an apparent molecular mass of ~2 MDa in mammals (Wang et al., 1996). SWI/SNF complexes in mammalian cells contain one of two possible ATPases (Brg1 or Brm).

Human SWI/SNF subunits are often encoded by more than one gene, thus permitting combinatorial assembly and a diversity of related complexes, some of which are modulated during development and cell differentiation.

Canonical human SWI/SNF complexes contain a single ATPase, a "core" group of subunits consisting of Ini1 (integrase interactor 1), BAF155 (Brg1-associated factor), and BAF170, plus seven other accessory subunits, one of which is β-actin. Four of the accessory subunits are each encoded by a different gene family that has between two and four members, thus permitting 72 possible combinations among them when allowing for one protein from each family per SWI/SNF complex (Wu et al., 2009). In mammalian cells, the SWI/SNF family can be divided into BAF and PBAF subfamily (Fig. I3).

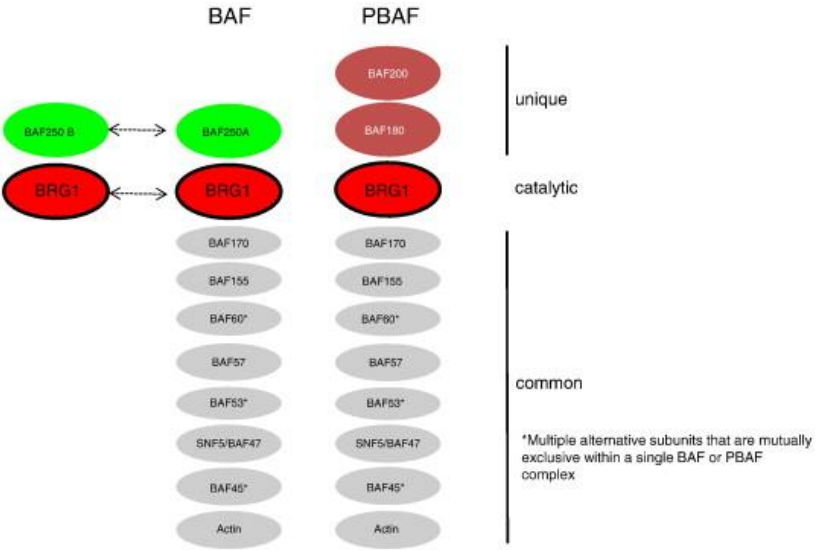


FIG I3: Mammalian SWI/SNF chromatin-remodeling complexes are subdivided into BAF and PBAF complexes. BAF complexes are defined by either BAF250A or BAF250B as the ARID-containing subunit (light green ovals), which are mutually exclusive (as denoted by dashed arrow with dual arrowheads), while PBAF complexes utilize BAF200 (blue oval) as the ARID-containing subunit. PBAF complexes also incorporate BAF180 (blue oval), which has 6 bromodomains that can bind to acetylated histones. BAF complexes are catalyzed by either BRG1 or BRM (red ovals), but no single complex contains both as denoted by dashed arrow with dual arrowheads. In contrast, PBAF is catalyzed exclusively by BRG1 (red oval). A number of other common subunits (BAFs for BRG1 or BRM associated factors) are also present (gray ovals from largest to smallest with numbers referring to protein size in KDa), and some of these subunits are represented by multiple

isoforms (as indicated by asterisks) that are encoded by separate genes although only 1 isoform is present in a single complex similar to BAF250A-BAF250B and BRG1-BRM. (Bevilacqua et al., 2013)

This subdivision is due to the exclusive presence of ARID1a-b only in BAF complexes and BAF180, BAF200, and BRD7 in PBAF complexes. These exclusive subunits provide functional specificity recruiting the complexes to specific loci and the association with specific proteins (Nie Z, 2000; Thompson M, 2009).

Both cellular and developmental contexts are highly important when considering the composition of SWI/SNF, as well as its eventual targets and actions. For example, specialized SWI/SNF complexes, such as those found in embryonic stem (ES) cells, include Brg1, BAF155, and BAF60A but exclude Brm, BAF170, and BAF60C (Kaeser et al., 2008; Ho et al., 2009). Notably, the majority of SWI/SNF subunits are capable of binding to DNA or chromatin. DNA-binding domains found among various SWI/SNF subunits include HMG, ARID (AT-rich interaction domain), SANT, and Krüppel domains (Wu et al., 2009). Binding appears to lack sequence specificity; however, recognition of certain DNA structures may be favoured, such as the minor groove of DNA and four-way helical junctions (Das et al., 2009).

Epigenetic modifications also contribute to SWI/SNF and chromatin associations. Bromodomains, found in both of the ATPases and in several other SWI/SNF subunits, are thought to recognize acetylated lysines in histones (Singh et al., 2007). BAF155 and BAF170 each contain a chromo-related domain,

which suggests a high affinity for methylated histones (Brehm et al., 2004). Indeed, the SWI/SNF complexes are defined polymorphic readers of epigenetic modifications (Wu JI, 2009). Moreover, their large size (about 12-fold bigger than a nucleosome) should enable reading of multiple histone modifications on adjacent nucleosomes providing significant affinity for both targeting and retention. Conversely, the indirect recruitment involves protein-protein interaction with histone modifying complexes (Zhang HS, 2000) and with transcriptional activators and repressors such as p53 (Lee D, 2002), cyclin E (Shanahan F, 1999), nuclear hormone receptors (Trotter KW, 2008). In the formation of productive SWI/SNF complexes, stochastic assembly and transient interactions among subunits may also play a significant role, as has been proposed for other large protein complexes, such as RNA polymerases and associated transcription factors (Dinant et al., 2006), spliceosomes (Rino et al., 2007), and DNA repair complexes (Luijsterburg et al., 2010).

Functions of SWI/SNF

One of primary functions of SWI/SNF complexes is to assist in gene regulation. The combination of ChIP with either DNA microarrays (ChIP-chip) or, more recently, sequencing (ChIP-seq) enables the genome-wide localization of chromatin-associated proteins. Several recent studies highlight the use of a genomic approach in elucidating regulatory regions associated with SWI/SNF localization patterns.

Euskircken et colleagues (2011) have recently identified the targets of four SWI/SNF components, Ini1 (SMARCB1), Brg1 (SMARCA4), BAF155 (SMARCC1) and BAF170 (SMARCC2), using ChIP-Seq experiments followed by bioinformatics data analysis with PeakSeq (Rozowski et al., 2009). Such genome-wide analysis of SWI/SNF binding sites, conducted in HeLa S3 cell lines, indicates that SWI/SNF complexes likely contribute to gene regulation through many different avenues. In fact, SWI/SNF complexes are reported to bind promoters, enhancers, CTCF sites and many regions occupied by Pol II (Euskirchen GM, 2011).

Furthermore, SWI/SNF complexes may facilitate looping interactions among these various elements. It has been shown in vitro that SWI/SNF can interact simultaneously with multiple DNA sites and generate loops between them (Bazett-Jones DP, 1999). DNA looping plays a significant role in gene regulation on a system-wide level by bringing linearly distant regions into close spatial proximity. Chromosomal looping interactions have been mapped at high resolution in vivo using 3C (chromosome conformation capture) and related methods. Examples of SWI/SNF-mediated higher order chromatin interactions are loops that can form in the β-globin locus control region (Kim S. et al., 2009, a), in the α-globin locus (Kim S. et al., 2009, b), throughout regions of the 200-kb T helper 2 (Th2) cytokine locus (Cai et al., 2006), and across the 150-kb CIITA locus (Ni et al., 2008).

Once targeted to specific promoters or enhancers, SWI/SNF complexes promote transition to and from the active and repressed chromatin state. These opposite effects are mediated

by interactions with different types of co-regulators. For example, SWI/SNF complexes interact with MYC, a transcription factor that regulates gene expression during cell cycle progression, apoptosis and differentiation. In particular, the core subunit SNF5 directly interacts with MYC and is capable of cooperating in the activation of MYC target genes in vitro (Cheng SW, 1999). By contrast, the ATPase subunits BRG1 binds to RB and facilitates the repression of RB target genes, including E2Fs and CCND1 (Trouche D, 1997). Transcriptional activation by SWI/SNF complexes is achieved by sliding or ejecting nucleosomes, allowing the transient exposure of binding sites for transcriptional activators and the stabilization of pre-initiation complex formation (Salma N, 2004). Additionally, SWI/SNF have been shown to be associated with regions downstream the promoter and to influence the RNA Pol II promoter escape (Soutoglou E, 2002) and transcription elongation (Corey LL, 2003).

On the contrary, SWI/SNF represses transcription inducing a more compact chromatin conformation and also allowing transcriptional repressors to bind to chromatin or facilitating other modifications of the chromatin such as deacetylation of histones (Sudarsanam and Winston; 2000). Besides its role in transcriptional regulation, recent findings illustrate that SWI/SNF act also post-transcriptionally regulating the production of alternative transcripts.

SWI/SNF interacts with components of the nascent pre-mRNA complex and with the transcription machinery (Cho H, 1998). These interactions cause a delay in the transcription elongation rate, which in turn affects splice site selection. Moreover,

SWI/SNF associates with several components of the spliceosome and with Sam 68, an ERK-activated enhancer of variant exon inclusion (Batsché E, 2006). The post-transcriptional regulation of SWI/SNF seems to be independent from its chromatin remodelling activity. In fact, mutation of the ATPase domain of hBrm does not impair its effect on splicing regulation (Batsché E, 2006). In order to better understand the very complex network of SWI/SNF cellular functions, proteomic analysis in human and mouse cell lines were exerted.

Immunoprecipitated proteins from non-cross-linked HeLa cells were identified by MS for six different SWI/SNF subunits: Brg1, Brm, Ini1, BAF155, BAF170, and BAF250A. Comprehensive analysis of SWI/SNF-interacting proteins revealed the presence of transcription factors, DNA repair proteins (ERCC5 and RAD50), DNA replication proteins (MCM2 and RPA1), and proteins important for chromosome integrity (NUF2, BUB1B, CENP-E, and PTTG1) (Euskirchen et al., 2011). In total, 158 SWI/SNF-interacting proteins have been described in HeLa cells across numerous studies from multiple investigators (Euskirchen et al., 2011), and ∼200 SWI/SNF-interacting proteins have been described in mouse ES cells, including histone- and DNA-modifying enzymes and regulators (e.g. HDAC1, JARID2, DNMT3b, and DNMT3L) (table2) (Ho et al., 2009) (Table 2).

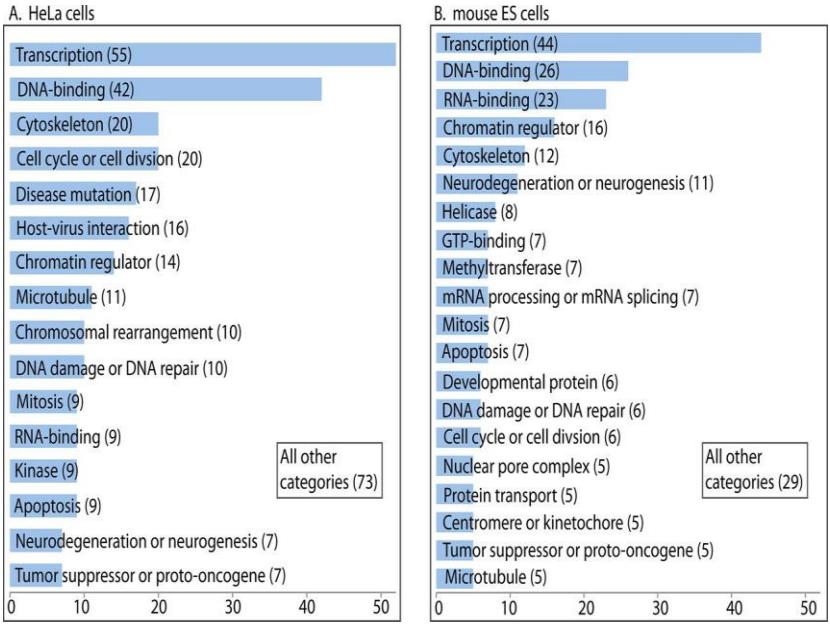


Table 2: Histograms showing the frequencies of UniProt keywords for proteins that co-purify with SWI/SNF factors. The keywords shown were retrieved from the UniProt Database for proteins that co-purify with a SWI/SNF factor for HeLa cells (*A*) and mouse ES cells (*B*;). Multiple (Euskirchen et al., 2012)

PREVIOUS RESULTS

In our laboratory AGO2-associated proteins have been investigated by SILAC/MS approach. As expected, well-known AGO2 interactors, such as TNRC6 complex (Pfaff et al.,2013) and HSP90 (Wu et al., 2011) were enriched in AGO2 immunoprecipitates (AGO2-IP) relative to mock IP (IgG-IP). Surprisingly, as shown in the Table P1, from our proteomic analysis emerged several nuclear proteins that were not known to be associated to AGO2. A Gene Ontology analysis revealed that, besides the cytoplasmic complexes such as RISC complex, AGO2 is also bound with high significance to many nuclear complexes implicated in mRNA processing (mRNA cleavage and polyadenylation) and in chromatin remodelling. Indeed, our results revealed for the first time the association of AGO2 to the main family of chromatin-remodelling complexes: the SWI/SNF family. Among the most highly enriched proteins in our AGO2-immunoprecipitations we found all the core proteins of SWI/SNF complexes (BAF155, BAF170, SNF5), both the ATPase subunits (BRG1 and BRM) and several accessory proteins including the subunits signature of BAF (ARID1a-b) and PBAF complexes (BAF180) (Table P1).

Proteins	Rank	Proteins	Rank	Proteins	Rank
BAF45D	**1**	FIP1	34	MYH9	67
SNF5	2	DSP	35	TUBGCP2	68
IGLC1	3	FAM83G	36	DDX3	69
BAF57	4	CDHF1	37	ACTA1	70
VPS35	5	VH-3 family (VH26)D/J	38	CLLD7	71
AGO2	6	PDHX	39	LSM12	72
TNRC6C	7	IGHV	40	KIAA0221	73
CSTF3	8	H2AFX	41	PSF	74
LAK1	9	CASP14	42	TBC1D5	75
BAF155	10	**BAF180**	**43**	KIAA0731	76
CSTF1	11	CPSF100	44	MOV10	77
CSTF2T	12	WDC146	45	NSEP1	78
ANX2	13	TNRC6A	46	RPS26	79
SMARCD2	14	**BAF170**	**47**	MYL6	80
ARID1A	**15**	AGO3	48	HNRNPK	81
KIAA0268	16	AGO1	49	P4HA	82
TNRC6B	17	CYPA	50	HNRNPM	83
CSNK1A1	18	**ARID2**	**51**	CTNNG	84
BCL7A	19	IGF2BP3	52	RPS24	85
SS18	20	C20orf99	53	CFL	86
BAF53A	**21**	PABPC4	54	HNRNPF	87
BAF60A	22	TFG	55	HNRNPA3	88
CLPX	23	PABP1	56	RPL1	89
KIAA0292	24	CDHF4	57	NCL	90
H3.3A	25	HGRG8	58	DDX5	91
ARID1B	26	CCR4-associated factor 1	59	DDX15	92
VPS26B	27	LMNB1	60	RPS6	93
BRG1	28	CSTF2	61	KIAA0264	94
CTAGE5	29	TUBG	62	MRLC2	95
H2BFD	30	MATR3	63	HSD14	96
BRM	**31**	HUR	64	HSP90AB1	97
DLAT	32	CPSF1	65	AUF1	98
H4/A	33	EDH17B4	66	DAP3	99

Table P1: SILAC/MS results. Top-list of AGO2-associated proteins ranked on fold enrichment over the control immunoprecipitation.

We further validated that AGO2 and SWI/SNF associate *in vivo* through immunoprecipitation and western blotting experiments on HeLa S3 cells (Fig. P2). Moreover, this association is independent of RNA molecules and involves a protein-protein interaction (Fig. P3). Our experiments demonstrated that only AGO2 and not AGO1 specifically interacts with SWI/SNF complexes (data not shown).

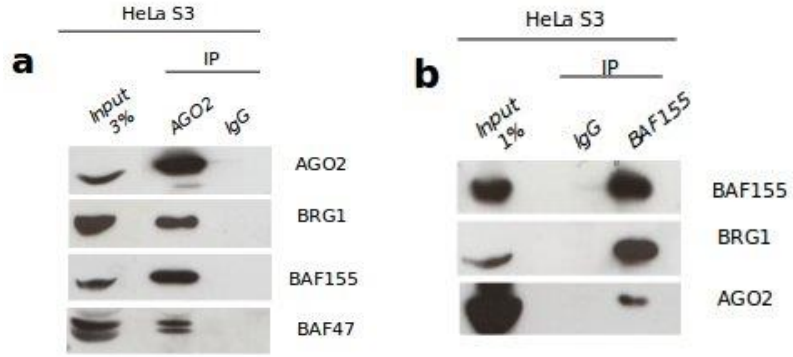
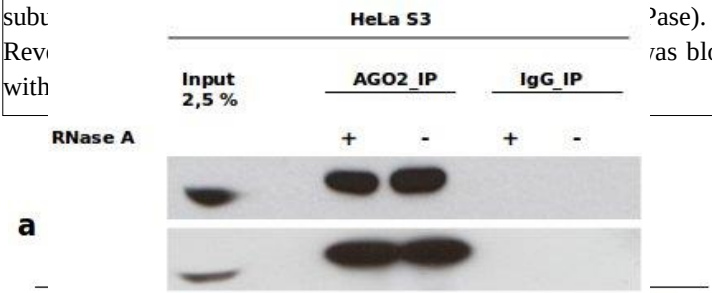
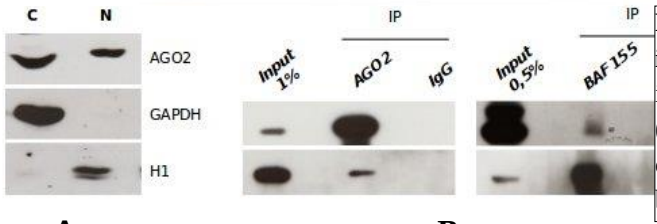


FIG P2: Human AGO2 interacts with SWI/SNF subunits BAF155 and BRG1. (A) HeLa S3 whole cell extracts were immunoprecipitated using anti-AGO2 or IgG (negative control) and analysed by western blotting using antibodies anti-BAF155 (core subu ase). **(B)** Rev as blotted with





A. **B.**

FIG P3: AGO independent. Total immunoprecipitated (AGO2_IP) or isotyp control. Input and in blot for the presence o

Fig P4: AGO2 and SWI/SNF complexes interact in the nucleus. (A) Control of cell fractionation by blotting for GAPDH and H1, cytosol and nuclear marker respectively. **(B)** AGO2-BAF155 co-immunoprecipitation from the nuclear fraction of HeLa S3 cells.

As shown in Fig. (P4) SWI/SNF-AGO2 interaction has been confirmed also in the nucleus, besides the validation performed on whole cell extracts (FIG P3). To this aim, a cell fractionation into cytosolic and nuclear fraction was performed, followed by AGO2 immunoprecipiation from the nuclear fraction only. The association with BAF155 was verified by western blot.

Furthermore, a chromatin fractionation experiment was performed (Fig. P5) to assess whether AGO2 is associated with chromatin. The AGO2 and BAF155 distribution profiles through the several resulting fractions were compared.

As expected, AGO2 was mainly detected in the S1 fraction, that includes the cytosolic and nucleoplasmic components. Interestingly, a substantial amount of AGO2, as well as BAF155, was revealed also in S2 fraction that specifically represents proteins associated to chromatin.

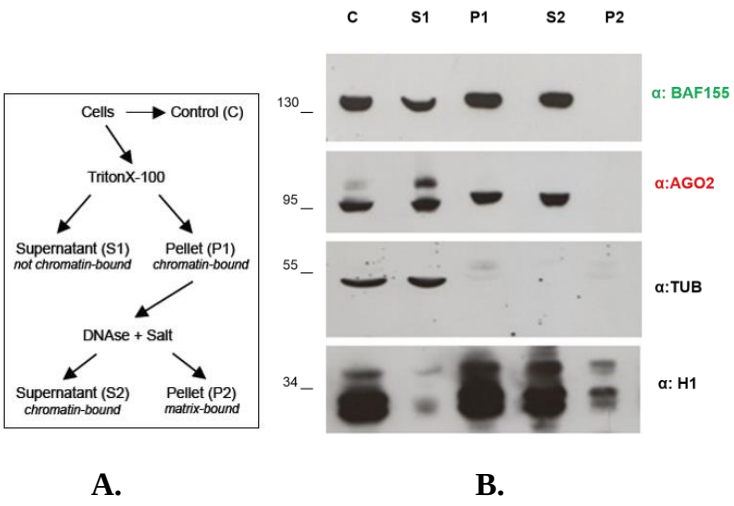


Fig P5. (A) Scheme of the procedure used to fractionate HeLa S3 cells (Cernilogar FM, 2011). Chromatin associated proteins should be found in fractions P1 (pellet) and S2 (supernatant). **(B) AGO2 is associated with chromatin in the nuclei of human cells**. Equal amounts of each resulting fraction was analyzed by western blotting for the presence of the indicated proteins. Tubulin (TUB) serves as a chromatin unbound marker, whereas H1 is a chromatin bound marker that is fully released only after high salt and DNase treatment.

In conclusion, these data suggested us that AGO2-SWI/SNF complexes might co-operate at chromatin level.

AIM OF THE THESIS

Studies over the past years have provided strong evidence regarding the multi-level regulation of gene expression operated by AGO proteins. In particular, human AGO2 protein, the only endonucleolytically active AGO protein, has been demonstrated to act either post-transcriptionally or at transcriptional level. In order to exert this multiplicity of roles, AGO2 needs to associate to different proteic and/or ribonucleic partners (sRNAs). As previously demonstrated in our laboratory, AGO2 interacts with SWI/SNF complexes in nuclei of human cell lines. Therefore, as SWI/SNF is the most important chromatin-remodelling complex, these findings could suggest that AGO2 has an active role in chromatin plasticity. In order to shed light on AGO2 possible roles in chromatin remodelling processes, I set up an extensive bioinformatics pipeline analysis aiming at the characterization of the small RNAs loaded onto nuclear AGO2 by high-throughput technologies. The outcome of my bioinformatics analysis indicates the existence of a novel, uncharacterised class of AGO2-bound sRNAs. Further analyses have then been conducted in order to investigate their biogenesis and molecular functions. The questions I will try to address regard the processing mechanisms they undergo, their possible involvement in mRNA expression level changes the correlation between this novel class of sRNAs and SWI/SNF, and least but not last, their possible involvement in nucleosome positioning. From this extensive bioinformatics analysis, I have

been able to identify new biological processes AGO2 is involved in.

RESULTS

Analysis of nuclear AGO2-associated small RNAs

The interaction of AGO2 with SWI/SNF complexes prompted us to hypothesize a concerted role of AGO2 and SWI/SNF in chromatin dynamics and possibly in gene expression regulation. Recently, it has been shown that in HeLaS3 cells SWI/SNF is recruited in the proximity of TSS of expressed genes (Euskirchen et al., 2011). The current model for gene expression regulation exerted by AGO2 indicates that AGO2 utilizes small RNAs (sRNAs) to recognize target molecules (Elbashir et al., 2001; Liu et al., 2004; Meister et al., 2004).
Moreover, there is evidence of the existence of a class of sRNAs arising from sense and antisense strand nearby TSS (TSSa RNA) (Seila et al., 2008; Valen et al., 2011). Starting from these considerations, I speculated whether AGO2 is loaded with TSSa RNAs in the nuclei of human cells.
Therefore, we performed an RNA Immunoprecipitation in order to profile by Next Generation Sequencing (NGS) the nuclear AGO2-bound sRNAs in HeLaS3. As control, a parallel immunoprecipitation using isotype matched IgG was performed. RNAs extracted from the total nuclear RNA sample (Input) and from the IPs samples were subjected to deep-

sequencing. Two independent biological replicates were performed.
Samples were sequenced on an Illumina HiSeq2000 machine. The sequencing experiment produced the following amount of data:
151.830.408 short reads (sequences 50 nucleotides in length) for the input sample;

- 149.821.287 short reads (sequences 50 nucleotides in length) for the AGO2-IP sample;
- 87.923.967 short reads (sequences 50 nucleotides in length) for the IgG-IP sample.

The term *read* refers to a short cDNA sequence, typically as output by a sequencing instrument. A corresponding string of quality values usually accompanies a read, where each value estimates the probability that the corresponding base was miscalled by the instrument software.
The analysis and biological interpretation of the huge amount of data obtained from this deep-sequencing experiment required me to set up a bioinformatics pipeline.
The first step of my bioinformatics analysis was the quality check through FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), a bioinformatics tool providing a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines.
After checking a good status of the experiment, I removed the adaptor sequences with Cutadapt (Martin, 2011). Indeed, when sequencing devices produce a list of sRNAs sequences, often the read length exceeds the length of the sRNA itself.

Depending on the device, this can result in sequenced reads that include sequences of the adaptor nucleic acid molecules employed in the cloning procedure at one or both ends of the read. The Adaptor Removal tool trims away these adaptor sequences making sRNAs data ready for analysis and processing by other tools.

At this point, I proceeded with the phase of the mapping, which means assessing which genomic regions the reads arise from.

The Bowtie (Langmead et al. 2009) bioinformatics tool enables the alignment of large sets of sequencing reads to a reference sequence, such as the human genome. Bowtie has many command-line options that allow the user to tune its behaviour according to the inputs and the desired result.

To get a complete view of the general origin of these sRNAs, I pre-filtered the data against highly expressed sRNA species that would otherwise obscure any other lowly expressed novel class of sRNAs. Therefore, the reads were iteratively aligned to different databases of known AGO2-associated classes of sRNA. After each alignment round, reads with no alignments were used in the next step. Alignments were performed requiring no mismatches in the first 18 nt of the reads (the "seed" region). Indeed, several authors have reported that this seed part of a read is expected to contain less miscalled bases due to the specifics of the NGS technologies (Nakamura et al,2011).

During the alignment round, reads were aligned to miRbase (v20.0) to remove all reads aligning to human miRNA precursors.

In a second step, unaligned reads were aligned against rRNA

and tRNA database to remove any reads aligning to tRNA and ribosomal RNA.

In a third step, unaligned reads were aligned against a manually cured Rfam database, a resource for RNA family information and multiple sequence alignments. Such manually cured Rfam database contains all sequences included in Rfam with the exception of those annotated as lincRNAs.

After these pre-filtering steps, unaligned reads were mapped against the human genome (hg19) index.

Some genomes, including the human genome, have substantial repetitive content (Batzer et al.,2002; Jurka et al., 2007; Britten et al. 2010; Hua-Van 2011; Kim P., et al., 2008), i.e. sub-sequences that appear multiple times throughout the genome. Repeats come in several forms (e.g. simple repeats, tandem repeats, segmental duplications, interspersed repeats), and arise via various biological processes (e.g. slipped strand mispairing or retro transposition). Repeats also affect alignments because reads originating from repetitive portions of the genome are difficult or impossible to unambiguously assign to a point of origin. Reads from repeats will tend to have many "valid" alignments, with no strong basis for preferring one over the others. Repetitive alignments in turn affect downstream analyses. The simplest way to deal with alignment ambiguity is to use Bowtie's −m option to filter out and/or annotate ambiguous evidence as such. With the −m 1 option, I specified a limit whereby alignments for reads that aligned to more than one genomic location were excluded from the output.

It is worth mentioning that no significant enrichment in Multiple Matching (MuM) reads was observed in AGO2

sample compared to Input and IgG, ruling out the possibility that nuclear AGO2 preferentially binds repetitive sequences. Therefore, only RNAs mapping to a single locus on hg19 genome assembly were further analysed, on the contrary Multiple Matching reads (MuM) were no longer investigated.
The reads which uniquely mapped to hg19 have been called "other sRNAs" because they do not belong to any known class of sRNA. A closer look at "other sRNAs" revealed that they mostly occur as clusters consisting of less than 50 molecules with slightly different 3' and 5' termini (Fig R1). A few clusters (0.86%) consisted of hundreds to thousands of identical sRNAs ("High copy" clusters). These "high-copy" clusters likely represent novel miRNAs or PCR artefacts. In order to avoid any possible bias, I focussed my analysis on "low copy" clusters (99.14 % of clusters) only.

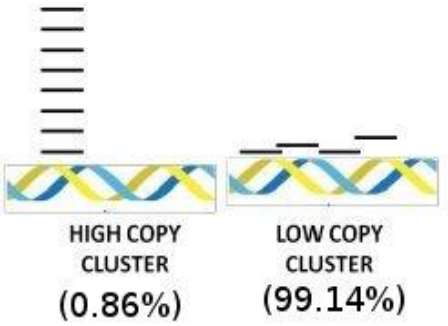


FIG R1: cluster division. "Other sRNAs" were divided in clusters (requiring same strandness) using mergeBed bioinformatics tool; most

clusters consisted of less than 50 sRNAs, while a few hundred clusters consisted of tens of thousands of identical reads. These high-copy clusters likely represent novel miRNAs or PCR artefacts. I therefore decided to focus my attention on sRNAs in low copy clusters (< 50 sRNAs) and removed reads lying into High-copy clusters (> 50 reads).

The percentages of different classes of non-coding sRNAs observed in the mock sample and in the AGO2-IP sample are depicted in Fig.R2.

As expected, the main class of sRNA in the AGO2 sample is represented by miRNAs. This specific miRNAs enrichment underlines the efficiency of our AGO2 immunoprecipitation, as miRNAs constitute the principal class of non-coding RNA associated to AGO proteins.

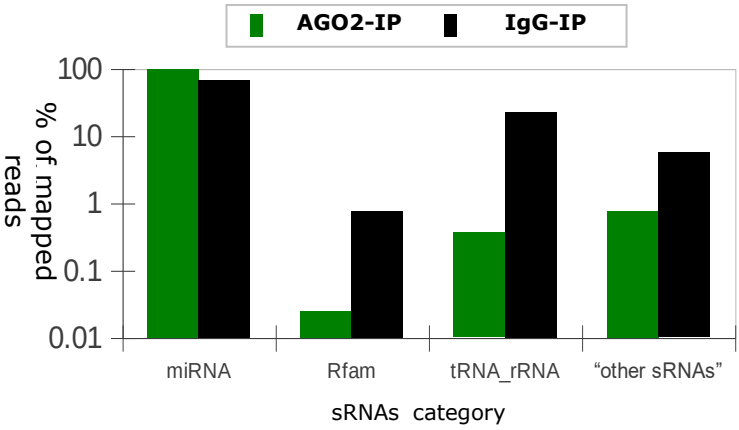


FIG R2: Genomic annotation of sRNA-Seq data. The histograms illustrate the sRNA classes identified in AGO2- or IgG-IP samples. The main RNA classes in the AGO2-IP sample are represented by miRNAs followed by "other sRNAs".

It is worth mentioning that, when ranking the four classes based on their abundance, the class of "other sRNAs" ranks second after miRNA class only in the AGO2-IP sample.
To evaluate whether "other RNAs" may have a functional role or were to be considered purely background noise, a size distribution analysis was performed (Fig. R3). The size range considered was from 18 to 50 nt, where 18 is the minimum length required for alignment during the mapping step and 50 is the sequencing length.

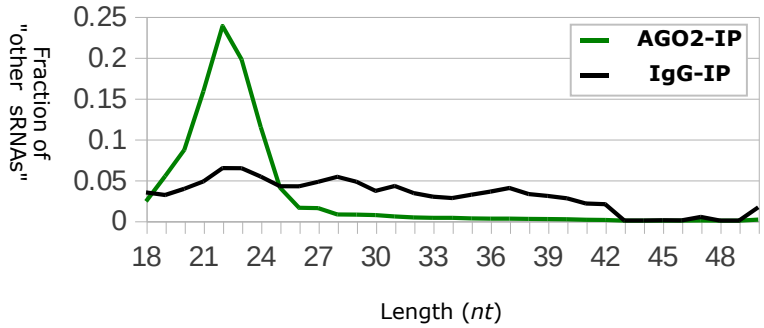


FIG R3: Size distribution of AGO2- and IgG-IP "other sRNAs". The fraction of "other sRNAs" in AGO2- or IgG-IP samples and corresponding length (*nt*) are plotted.

The “other RNAs” of the IgG-IPed sample showed a homogeneous distribution in the considered range. Intriguingly, the AGO2-IPed “other RNAs” displayed a size peak of 22-23 nt which represents the canonical length of small RNAs bound to AGO proteins. This result suggests that “other sRNAs” do not represent degradation products or aspecific materials.
In order to exclude the idea that these “other” RNAs could be putative miRNAs I performed a miRNA discovery analysis using miRanalyzer (Hackenberg et al., 2009). miRanalyzer is a web server tool very useful in the prediction of new microRNAs. This bioinformatics tool firstly clusters the input reads into putative mature microRNAs, then extracts candidate pre-microRNAs from the genome to select the energetically best candidate and finally applies five different Random Forest models (Breinman L., 2001) to calculate the probability that a given candidate is a microRNA (Hackenberg et al., 2011). MiRanalyzer did not report any candidate among AGO2 associated “other sRNAs”, thus corroborating the idea of a novel and uncharacterised class of non-coding sRNAs associated to AGO2.
In light of this finding, I proceeded to analyse how the newly identified sRNA class was distributed along the genome. Given the evidence of the existence of a class of short RNAs of heterogeneous length arising from sense and antisense strand nearby TSS (TSSa RNAs) (Seila et al., 2008; Valen et al., 2011), I looked for occurrence of TSSa RNAs among the class of “other sRNAs” identified in my samples.
Therefore, using a bioinformatics tool called intersectBed (from BedTools package: Quinlan and Hall, 2010) I could

intersect the genomic coordinates of my reads with the genomic coordinates of 1kb window around TSS of expressed genes. This step allowed me to compute the fraction of "other sRNAs" with at least a nt of overlap with a 1kb window around each TSS.

As shown in Fig. R4-a, I found a specific enrichment of sRNAs mapping on TSS in AGO2-IP relative to mock-IP (P-value < 2.2×10^{-16}, exact binomial test; the frequency of overlapping "other sRNAs" was used as null hypothesis).

As in our laboratory it was previously shown that AGO2 and SWI/SNF physically interact in the nuclei of HeLa S3 cell lines, I wanted to test whether there was a specific enrichment of "other sRNAs" lying within 1kb of SWI/SNF-bound regions. With this aim, the genomic coordinates of clusters of each sample (AGO2-IP and IgG-IP) were intersected with the genomic coordinates of SWI/SNF binding sites obtained through ChIP-seq experiments (Euskirchen et al., 2012) in HeLa cell lines and the number of sRNAs overlapping SWI/SNF binding sites was computed. Intriguingly, "other sRNAs" mapping to SWI/SNF binding sites are enriched in AGO2-IP compared to mock-IP (P-value < 2.2×10^{-16}, exact binomial test) (FigR4-b).

A. **B.**

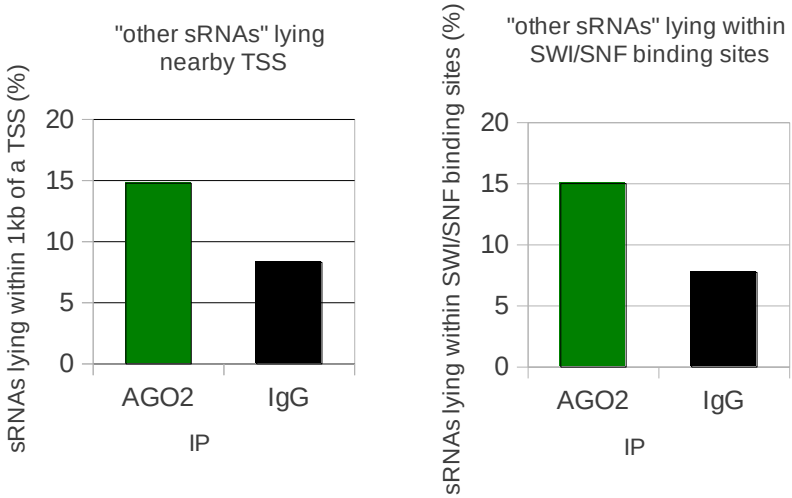
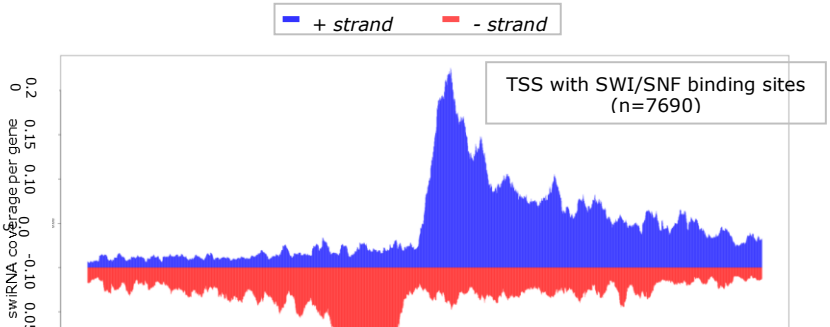


FIG. R4: a. Percentage of "other sRNAs" in AGO2- or IgG-IP samples mapping within 1 kb of TSS of expressed genes. Only "other sRNAs"

lying within clusters of < 50 molecules were considered. **b. Percentage of "other sRNAs" in AGO2- or IgG-IP samples mapping within SWI/SNF biding sites** (ENCODE ChIP-seq data). Only "other sRNAs" lying within clusters of < 50 molecules were considered.

Coverage analysis of TSSa RNAs on SWI/SNF-bound TSS (ENCODE ChIP-seq data) and all other TSS was then computed with coverageBed bioinformatics tools. using CoverageBed tool, I computed coverage of "other sRNAs" along a 1kb interval around each TSS, then processed the output in a custom perl script to obtain a cumulative coverage for all TSS of interst and finally plotted the results with R statistical environment (CRAN project: http://www.r-project.org/). The coverage was normalised to the number of the genes, thus getting the average sRNAs per gene. Fig. R5 highlights two very intriguing findings: firstly, the majority of AGO2-associated "other RNAs" are transcribed from SWI/SNF-bound TSS. For this reason, nuclear AGO2-associated TSSa RNAs have been named "swiRNAs". Secondly, Fig R5 clearly shows that swiRNAs arise from both sense and antisense strands. Even though previous works identified TSSa RNAs up to 100 nt in length transcribed from sense and antisense strand (Seila et al. 2008; Valen et al., 2011), the specific size distribution of AGO2-bound "other sRNAs" I observe indicates the existence of a novel and uncharacterized class of AGO2-bound sRNAs.



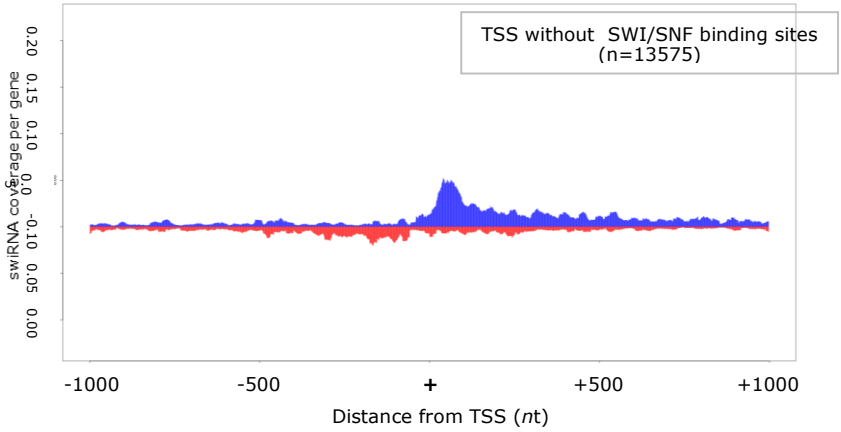


Fig R5: a. Average per gene coverage of AGO2-associated "other sRNAs" around SWI/SNF bound TSS (ENCODE ChIP-seq data).

Average per gene coverage is defined as the sum of sRNAs mapping to a given position (relative to TSS) divided by number of TSS analysed. Only "other sRNAs" lying within clusters of < 50 molecules were considered. **b. Average per gene coverage of AGO2-associated "other sRNAs" around TSS without SWI/SNF binding sites** (ENCODE ChIP-seq data). Average per gene coverage is defined as the sum of sRNAs mapping to a given position (relative to TSS) divided by number of TSS analysed. Only "other sRNAs" lying within clusters of < 50 molecules were considered.

swiRNAs are processed in a Dicer-dependent manner

In humans, Dicer processes both long dsRNAs and pri-miRNAs into siRNA duplexes and miRNA duplexes, respectively (Zhang et al., 2002; Provost et al., 2002).
I reasoned that complementary long TSSa RNA pairs might form a dsRNA molecule suitable for DICER processing, pinpointing long TSSa RNAs as swiRNAs precursors.
Thus, in order to investigate this hypothesis, I looked at the 3' termini of pairs of swiRNAs lying on opposite strands of the same locus. The idea is that I can compute the distance of the 3' termini and plot it in function of the frequencies at which I observe these distances.
My results could lead to two different scenarios: in the first case I could observe a heterogeneous distribution of distances, because of random chance. In the second case, I could observe a sharp peak in correspondence of the most distance between

the 3’ termini of pairs of swiRNAs lying on opposite strands of the same locus, which would be suggestive of an enzymatic processing. As depicted in fig. R6, the final plot indicates a clear enrichment for the 2 nt 3’ overhang, which is a specific hallmark of DICER processing.

Starting from this strong suggestion, I decided to shed more light on Dicer involvement in swiRNAs processing.

Therefore, in order to study the effects of loss of DICER on swiRNAs processing, we performed an RNA Immunoprecipitation and profiled by Next Generation Sequencing (NGS) the sRNAs specifically bound to nuclear AGO2 in wild type HCT116 human colon cancer cell lines and DICER^{ex5} HCT116 cell lines. In HCT 116 DICER^{ex5} cell lines, *DICER* gene exon 5, which encodes for the helicase domain, was deleted via homologous recombination (Cummins et al., 2006).

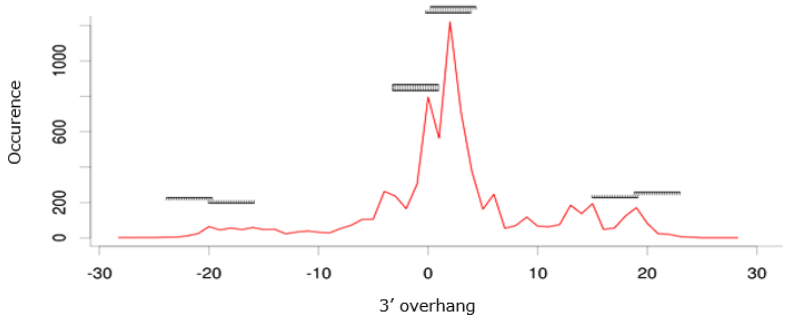


FIG. R6: swiRNAs display a 2 nt 3' overhang typical of DICER-processed sRNAs. The length (nt) of 3' overhang for pairs of complementary swiRNAs was plotted. Negative values represent a 5' overhang. Most pairs display a 2 nt 3' overhang, suggestive of DICER processing.

It is worth mentioning that we have demonstrated that the interaction of AGO2 and SWI/SNF subsists also in the nuclei of HCT116 and DICER^{ex5} HCT116 cells, as previously shown for HeLa cell lines (Data not shown). As control, a parallel immunoprecipitation using isotype matched IgG was performed. RNAs extracted from the total nuclear RNA sample (Input) and from the IPs samples were subjected to deep-sequencing.
The sequencing experiment produced the following amount of data:

- 176.154.736 short reads in HCT116 wt input sample.
- 161.631.990 short reads in HCT116 wt RIP-Ago2 sample.
- 159.612.886 short reads in HCT116 wt RIP IgG wt sample.
- 145.659.010 short reads in HCT116 DICER^{EX5} input sample.
- 165.007.012 short reads in HCT116 DICER^{EX5} RIP-Ago2 sample.
- 146.507.557 short reads in HCT116 DICER^{EX5} RIP-IgG sample.

In order to pre-process the reads and to map them to the human genome I used the same bioinformatics pipeline I set-up for the AGO2-associated sRNAs observed in HeLa cell lines.
As shown in the figure R7, the presence of “other sRNAs” is confirmed in HCT116 cell lines as well as in HCT116 DICER^{EX5} sub-clone. Furthermore, also the “other sRNAs” size distribution profile is very similar to the one observed in HeLa cell lines. Intriguingly, we can observe a reduction of “other sRNAs” mapping within 1kb of TSS of expressed genes in absence of Dicer (FIG. R7-f).

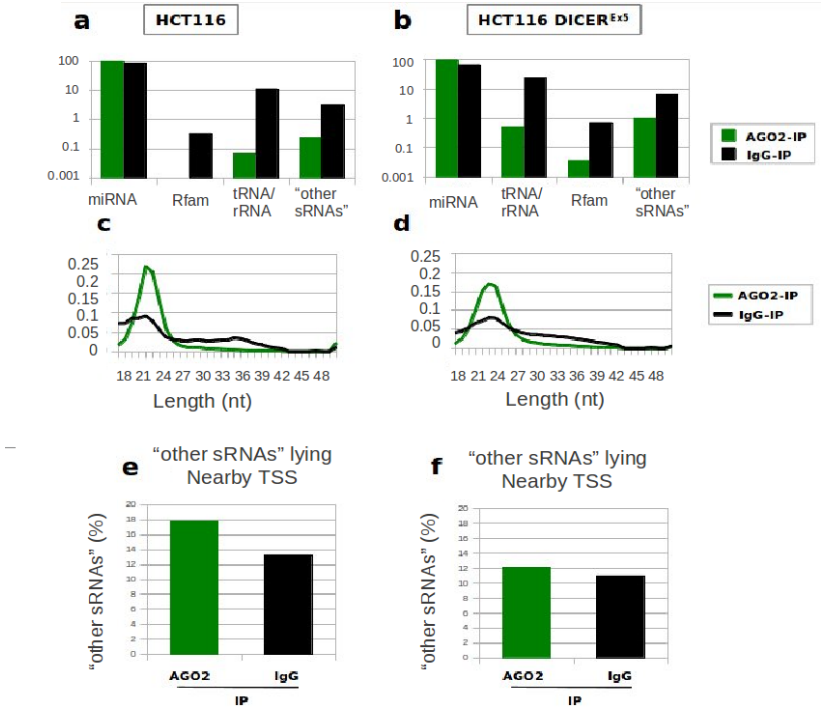


FIG R7: a-b. Genomic annotation of sRNA-Seq data in HCT116 and HCT116 DICER^{Ex5}. The histograms illustrate the sRNA classes identified in AGO2- or IgG-IP samples. **c-d.** Size distribution of AGO2- and IgG-IP "other sRNAs" . Number of reads of "other sRNAs" in AGO2- or IgG-IP samples and corresponding length (*nt*) are plotted. **e-f.** Percentage of "other sRNAs" in AGO2- or IgG-IP samples mapping within 1 kb of TSS of expressed genes (TSS were defined by mRNA-seq). Only "other sRNAs" lying within clusters of < 50 molecules were considered.

Therefore, in order to test whether the absence of Dicer could affect swiRNAs processing, I computed the coverage of swiRNAs around TSS of expressed genes both in HCT116 and DICER^{ex5} cell lines.
Coverage at each position represents the number of swiRNAs mapping at the indicated distance from a TSS, normalized by the total number of "other sRNAs".
Accordingly, I observed a drastic reduction of swiRNAs in HCT116 DICER^{EX5} relative to parental HCT116, which confirms that swiRNAs are processed by DICER (Fig R8).
As a control, I checked for AGO2-bound sRNAs mapping on Transcription Termination Sites (TTS) (Valen et al., 2011) and observed that they are not affected by DICER depletion (Fig. R9).
Based on to the specific size, the DICER mediated processing and the AGO2 association, I can conclude that swiRNAs are a novel class of sRNAs distinct from any other class of sRNAs mapping near TSS previously described (Seila et al., 2008; Valen et al., 2011; Taft et al., 2009).

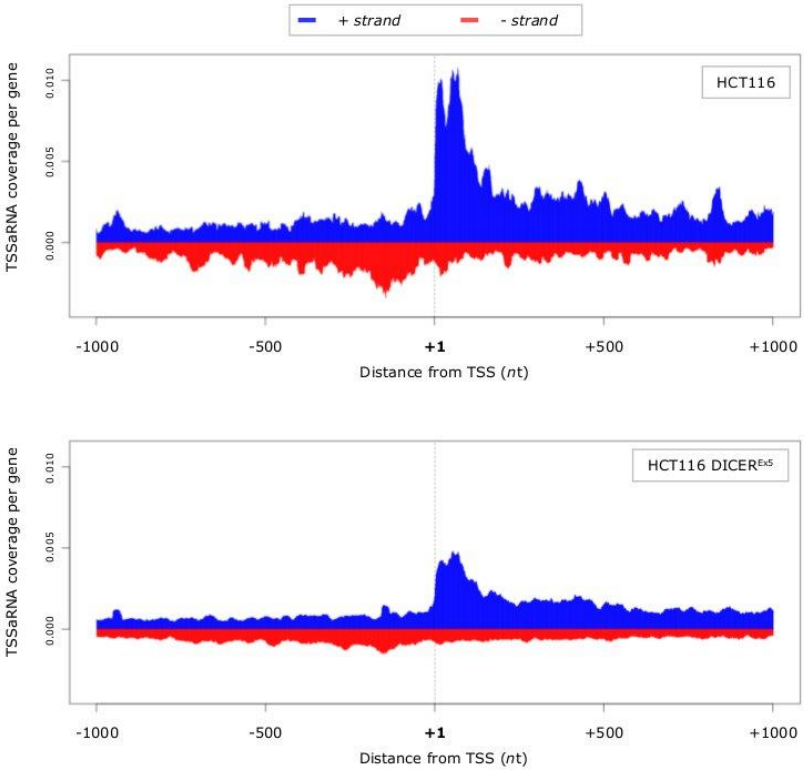


FIG. R8: Dicer depletion affects swiRNAs processing. Coverage of swiRNAs in parental HCT116 cell line (upper panel) and in HCT116 DICER^{Ex5} cell line (lower panel) (blue: sense strand; red: antisense strand). Coverage at each position represents the number of swiRNAs mapping at the indicated distance from a TSS, normalized by the total number of "other

sRNAs". Only "other sRNAs" lying within clusters of < 50 molecules were considered.

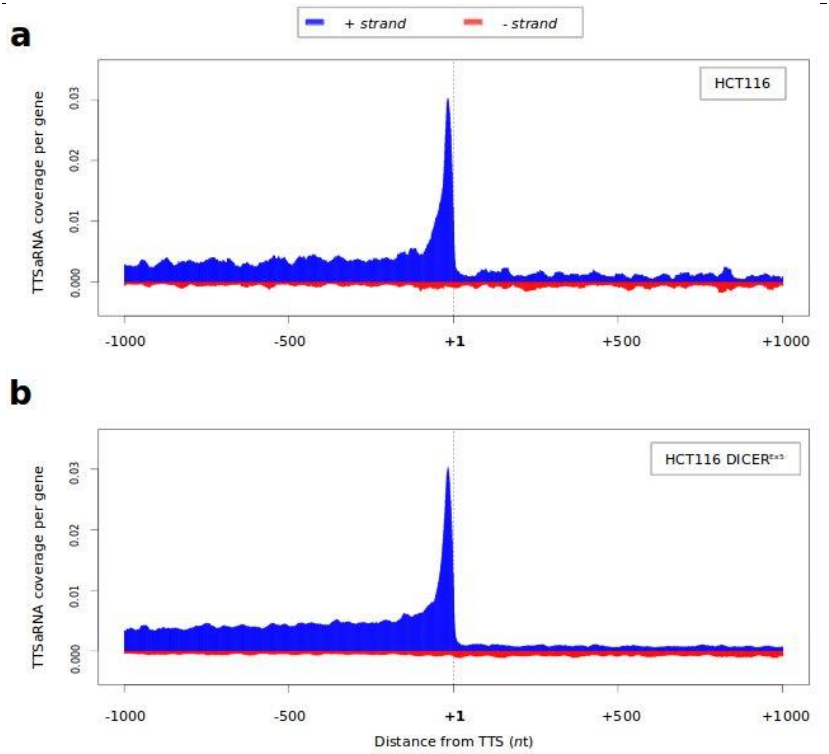


FIG. R9:AGO2-associated Transcription Termination Site associated (TTSa) RNA are not processed in a DICER-dependent manner. Coverage of swiRNAs in parental HCT116 cell line (panel a.) and in HCT116 DICER^{Ex5} cell line (panel b.) (blue: sense strand; red: antisense strand). Coverage at each position represents the number of "other" RNAs

mapping at the indicated distance from a TTS, normalized by the total number of "other sRNAs". Only "other sRNAs" lying within clusters of < 50 molecules were considered.

Analysis of mRNA expression level changes upon Ago2 knock-down

The identification of a new macromolecular complex including AGO2, core components of SWI/SNF and swiRNAs suggests the involvement of AGO2 in SWI/SNF-mediated nucleosome positioning. As nucleosome positioning affects DNA accessibility and mRNA transcription levels, I wanted to test whether any transcriptomic changes could be induced upon AGO2 depletion. Therefore, we knocked down AGO2 in HeLaS3 cells and looked at mRNA expression profiles in control and AGO2 Knock-Down HeLaS3 cells through NGS experiments conducted in two independent biological replicas. As shown in Fig. R10, down-regulation of AGO2 protein was verified by western blot analysis.

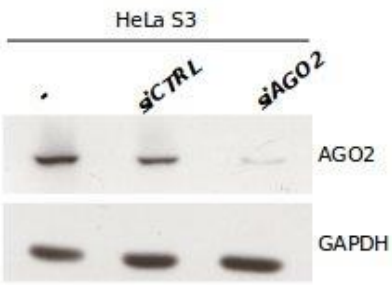


Fig. R10: Down-regulation of AGO2 protein in HeLaS3 cells. Cells were transfected with siCTRL or siAGO2. Down-regulation of AGO2 protein was verified by western blot. GAPDH was used as loading control.

Then, total mRNA was extracted in the three samples (untreated, siCtrl and siAgo2) and sequenced on an Illumina HiSeq2000 machine in paired-end mode. Paired-end sequencing represents a simple modification to the standard single-read DNA library preparation and facilitates reading both the forward and reverse template strands of each cluster. In addition to sequence information, both reads contain long range positional information, allowing for highly precise alignment of reads. The paired-end run set up for my experiment produced 2 × 100 bp reads.
The amount of my sequencing data is as follows (sum of two biological replicates):

- 37.403.780 paired-end reads input sample.
- 37.059.506 paired-end reads siCtrl sample.
- 44.727.903 paired-end reads siAgo2 sample.

Therefore, I performed the routine quality check with Fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
If a sequencer is unable to make a base call with sufficient confidence then it will normally substitute an N rather than a conventional base call. It's not unusual to see a very low proportion of N appearing in a sequence, especially close to the end of a sequence. Indeed I noticed that also my reads presented an increase of the N strarting from 87th position on. The presence of the N during the mapping step can strongly condition the percentage of reads that correctly align to genome, because the more mismatches I have during the mapping the less percentages of reads I can align to the reference genome. In order to preserve my depth of

sequencing, I decided to trim my reads from the 87th position on with the bioinformatics tool Trimmomatic (Lohse et al., 2012; http://www.usadellab.org/cms/?page=trimmomatic).

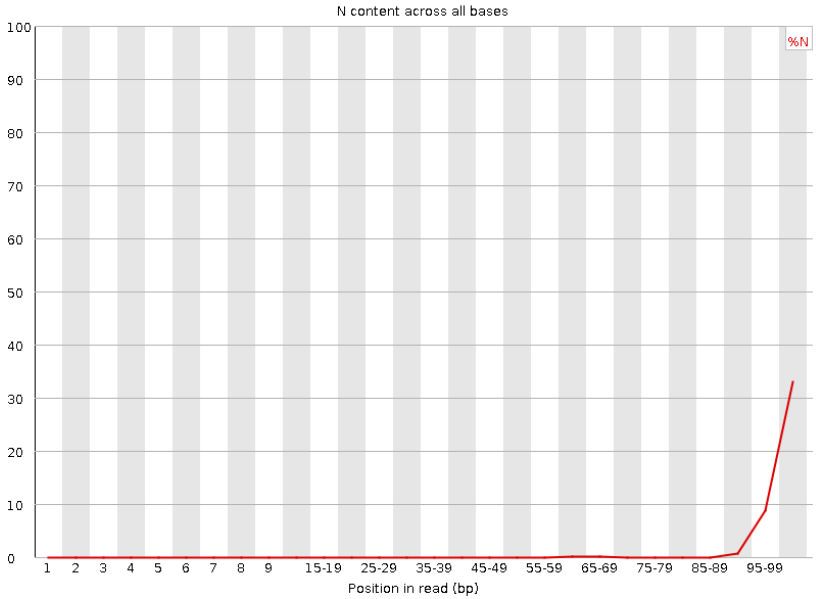


Fig R11: N % across all bases. This module plots out the percentage of base calls at each position for which an N was called. Here it is represented the Input sample. The other samples show a very similar N% distribution.

Afterwords, reads were mapped with the bioinformatics tool TopHat 2.0.9 (Kim D. et al., 2013). TopHat is a fast splice junction mapper for long RNA-Seq reads. By first mapping RNA-Seq reads to the human genome, TopHat identifies

potential exons, since many RNA-Seq reads will contiguously align to the genome. Using this initial mapping information, TopHat builds a database of possible splice junctions and then maps the reads against these junctions to confirm them. Short read sequencing machines can currently produce reads 100bp or longer but many exons are shorter than this so they would be missed in the initial mapping. TopHat solves this problem mainly by splitting all input reads into smaller segments which are then mapped independently. The segment alignments are put back together in a final step of the program to produce the end-to-end read alignments. TopHat generates its database of possible splice junctions from two sources of evidence. The first and strongest source of evidence for a splice junction is when two segments from the same read (for reads of at least 45bp) are mapped at a certain distance on the same genomic sequence or when an internal segment fails to map - again suggesting that such reads are spanning multiple exons. With this approach, "GT-AG", "GC-AG" and "AT-AC" introns will be found ab initio. The second source is pairings of "coverage islands", which are distinct regions of piled up reads in the initial mapping. Neighbouring islands are often spliced together in the transcriptome, so TopHat looks for ways to join these with an intron (http://tophat.cbcb.umd.edu/manual.shtml). The mapping rate for my NGS data was very high, indeed 100% of my reads very properly aligned on the human genome. After the reads were mapped on the genome, it was necessary a step of transcripts assembly. Cufflinks (v 2.1.1) (Trapnell et al., 2013) constructs a parsimonious set of transcripts that "explains" the reads

observed in an RNA-Seq experiment. I therefore ended up with a list of genes expressed in HeLaS3 cells under my growth conditions. Differential gene expression analysis was exerted in parallel with two different softwares: Cuffdiff (Trapnell et al., 2013; http://cufflinks.cbcb.umd.edu/howitworks.html#difftest) and DESeq, an R package released by BioConductor (Anders and Huber, 2010). Accordingly, the results obtained from these two algorithms have shown to be highly comparable. In detail, DESeq steps as follows. As a first processing step, I needed to estimate the effective library size. This step is sometimes also called normalisation, even though there is no relation to normality or a normal distribution. The effective library size information is called the size factors vector, since DESeq package only needs to know the relative library sizes. If the counts of non-differentially expressed genes in one sample are, on average, twice as high as in another (e.g. because the library was sequenced twice as deeply), the size factor for the first sample should be twice that of the other sample. Having estimated the dispersion for each gene, it is straight-forward to look for differentially expressed genes. To contrast two conditions, e.g., to see whether there is differential expression between conditions “siCtrl” and “siAGO2”, I have performed a binomial test in order to compute p-values associated to the log2 fold changes computed. As shown in Fig. R12 (differentially expressed genes are coloured in red) mRNA expression profile analysis did not highlight significant changes in the expression level of mRNAs transcribed from Ago2 knock-down sample compared to the control one.

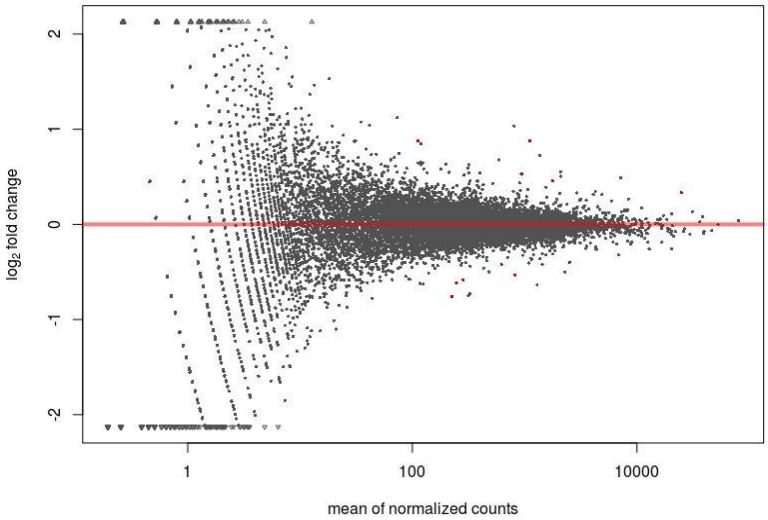


Fig. R12: The function *plotMA*, from the R package DESeq, plots the log2 fold changes against the mean normalised counts, colouring in red those genes that are significant at 10% FDR.

It is also instructive to look at the histogram of p-values (Figure R13). The enrichment of low p-values would stem from the differentially expressed genes, while those not differentially expressed are spread uniformly over the range from zero to one (except for the p values from genes with very low counts, which take discrete values and so give rise to high

counts for some bins at the right). As shown in the figure there is no enrichment for low p-values, which means that no significant difference is observed in the transcriptome upon Ago2 knock-down. This prompted me to think that Ago2 and SWI/SNF interaction might exert some other kind of role in the cell.

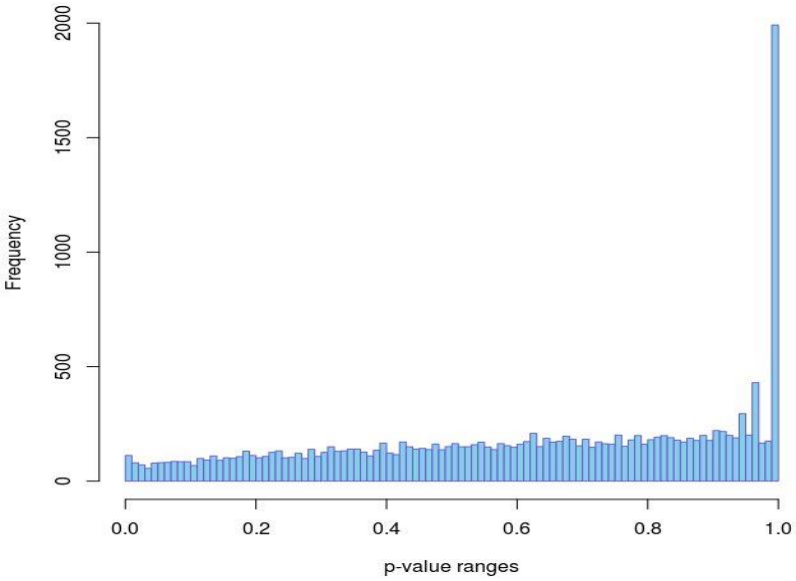


Fig R13: Histogram of p-values resulted from the call to nbinomtest() function from DESeq R package. The lack of enrichment for low p-values in this plot indicates that there are no significant differences in the transcriptional levels of the two compared samples.

Analysis of nucleosome occupancy changes upon Ago2 knock-down

Therefore, the next step of my PhD work was to assess whether nucleosome positioning underwent any changes upon AGO2 Knock-Down.
The locations and occupancies of nucleosomes can be assayed through the use of enzymatic digestion with micrococcal nuclease (MNase), an endo-exo nuclease that preferentially digests naked DNA and the DNA in linkers between nucleosomes, thus enriching for nucleosome-associated DNA. To determine nucleosome organization genome-wide, DNA fragments, recovered following MNase digestion, were sequenced using high-throughput sequencing technologies (MNase-seq) both in control and AGO2 Knock-down cells.
Paired End sequencing of MNase-digested fragments gave rise to the following amount of data:

- 592.534.546 paired-end reads in Ago2 Knock-down cells.
- 711.051.028 paired-end reads in control cells.

Sequence quality was assessed using FastQc v 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
Paired End reads were then aligned to the human genome using Bowtie v 0.12.7 (Langmead et al., 2009) allowing at the most 1 mismatch in the first 30 nts of the read.

Paired-end sequencing facilitates reading both the forward and reverse template strands of each cluster. For a pair to be "properly paired" it needs to have both reads mapped to the same sequence within a given distance. In my analysis, only those reads flagged by Bowtie as “Properly paired” reads were further analysed, in order to detect with the maximum precision the position of each nucleosome.

After the mapping step, I checked for the size distribution of the fragments.

As expected, I observed a major peak at 147 and a minor peak at 128 nt. Indeed, MNase digestion is known to produce fragments of variable size; however shorter fragments (< 100 nt) are likely to represent the DNA footprint of transcription factors, polymerases and other DNA binding proteins. The DNA footprint of a nucleosome is expected to be about 146 nt long. However, it is known that at low efficiency MNase may occasionally nick the 146 nt long nucleosomal DNA at about 10 nt (one DNA helix turn) into the nucleosome at each end, giving rise to a small satellite population of 127 nt fragments which do represent nucleosomes. DNA fragments longer than 200 nt are unlikely to represent bona fide nucleosome footprints. We therefore selected Paired End reads with a size between 100 and 200 nt.

As shown in Fig.R14, the size distribution of DNA fragments is strikingly similar in the two samples, ruling out the possibility of a bias due to unequal digestion and thus allowing a proper comparison between the two experimental conditions.

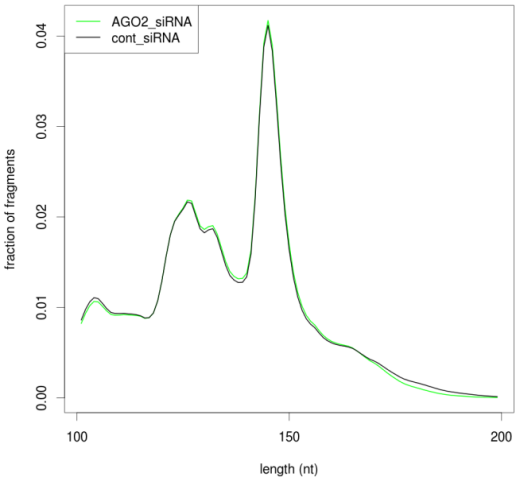


Fig.R14: Distribution of fragment length is remarkably similar in siCTRL- or siAGO2-treated HeLaS3 cells. This indicates a comparable digestion efficiency for the two samples. Over-digestion of a sample compared to the other would have resulted in a higher 128 nt peak and a lower 147 nt peak. Y-axis: fraction of fragments. X-axis: fragment length (green line: AGO2 Knock-Down HeLaS3 cells; black line: ctrl siRNA HeLaS3 cells).

After the mapping step, I used a bioinformatics tool, Danpos (Chen et al., 2013), in order to carry on comparative analysis of nucleosome physical organization at single-nucleotide resolution.

The physical organization of nucleosomes can be described as an array of nucleosome units across the genome.

In different cells, the exact positions of the nucleosomes within each unit may deviate more or less while centring around a most preferred position. This deviation of nucleosome positions within each unit in a cell population is referred to as fuzziness. Thus, each nucleosome can be described by a most preferred position (hereafter referred to as nucleosome position) and its fuzziness, along with an occupancy value referring to the frequency with which the unit is occupied in a cell population (Kaplan et al. 2010; Pugh 2010).

Therefore, starting from reads that have been mapped to the reference genome, I have executed a preliminary data processing with Danpos. Firstly, the program removes clonal reads with identical sequences resulting from possible over-amplification during sample preparation. The clonal reads can be determined based on their extremely high coverage relative to the mean coverage across the genome using a Poisson *P*-value cut-off. Then, nucleosome occupancy was calculated as the count of adjusted reads covering each base pair in the genome. After the pre-processing step, Danpos used a quantile normalization option in order to make the occupancy levels comparable in the two samples. Therefore, a Poisson test was applied to the nucleosome differential signals computed, in order to evaluate the statistical significance of the results. Then, a peak calling was performed on the single-nucleotide-resolution differential signal to identify differential peaks, thus classifying them into three categories including nucleosome position shifts, fuzziness changes, and occupancy changes. Unfortunately, I did not observe any statistically significant changes between Ago2 knock-down and control sample for the

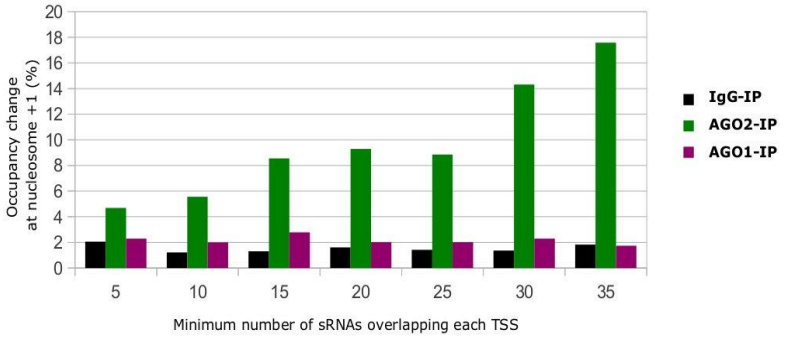
required *P*-value cut-off of 1×10^{-5} (or a false discovery rate <0.01). The same results were obtained using another nucleosome calling algorithm, nucleR (Flores and Orozco, 2011), in defining nucleosome position, occupancy, and fuzziness changes. I hypothesized that this weak statistical significance could be suggestive of the fact that most of the nucleosomes did not undergo differential positioning upon Ago2 knock-down or that each change, taken individually, was not significant. Indeed, we have to keep in mind that FDR represents the adjusted p-value computed for multiple testing, that is for all the nucleosomes I find all over the genome. This method, on one hand reduces the false positive rate observation but, on the other hand, could mask actually significant signals arising from a little subclass of nucleosomes which, on the contrary, undergo differential occupancy changes.

Therefore, I decided to change my strategy and focused my attention only on the average nucleosome occupancy profiles around all TSS of expressed genes in the two samples. Since AGO2 is associated with swiRNAs, mapping nearby TSS and SWI/SNF has been reported to bind TSS of expressed genes (Tolstorukov et al., 2013), I used the TSS coordinates of the 21.265 genes expressed in HeLaS3 derived on my previous RNA-seq analysis of mRNAs under our growth conditions (see above paragraph). Nucleosome Occupancy was computed using coverageBed (from BedTools package, Quinlan et al., 2010) and a custom perl script to determine cumulative occupancy over multiple loci.

Occupancy was normalized by the total number of sequenced nucleotides in each library taking into account Properly Paired

read pairs with a length between 100 and 200. Consistently with previous results (Schones et al.,2008; Jiang et al., 2009; Hartley et al., 2009), I observed in both samples a nucleosome-free region located immediately upstream the TSS, flanked by two well-positioned nucleosomes, referred to as -1 and +1 nucleosome. By restricting my analysis only on regions corresponding to nucleosome +1 relative to TSS, I found a mild (1.98%) but highly significant (P-value = 1.5×10^{-15}, paired t-test; see experimental procedures for the computing of the p-value) decrease in the average occupancy at nucleosome +1 in AGO2 Knock-Down cells (data not shown). Afterwards, I restricted my analysis only on TSS overlapped by swiRNAs.
Therefore, I have created several groups of genes selected based on the minimum number of swiRNAs mapping within ± 150 nt of TSS and repeated the computation for each group. Nucleosome +1 occupancy was defined as the sum of the coverage at each nucleotide position in the interval between nt +100 and nt +300 relative to TSS for each group of genes.
A paired t-test was performed to compute P-value of the difference observed between average coverage at nucleosome +1 in AGO2 knock-down and Ctrl siRNA cells.
For each group of n genes (whose TSS were overlapped by at least m swiRNAs) 10000 random permutations were performed to estimate FDR. In each permutation, n random genes were chosen (among the 21265 genes expressed in HeLaS3 cells) and the P-value was computed comparing occupancy at nucleosome +1 with the same procedure outlined above for the real case. All real cases with a P-value < 0.01 were tested, and FDR was always estimated to be < 0.01.

Intriguingly, AGO2-dependent reduction in nucleosome occupancy positively correlated with the number of swiRNAs mapping within ± 150 nt from TSS (Fig R15-a). On the contrary, there was no correlation between nucleosome +1 occupancy changes and IgG-IP nor AGO1-IP sRNAs overlapping each TSS. These data are in agreement with the fact that AGO1 does not interact with SWI/SNF (see preliminary data) and indicate that the observed phenotype is not due to a differential recovery of nucleosomes mapping near the more expressed genes (Fig. R15-b).



a.

b.

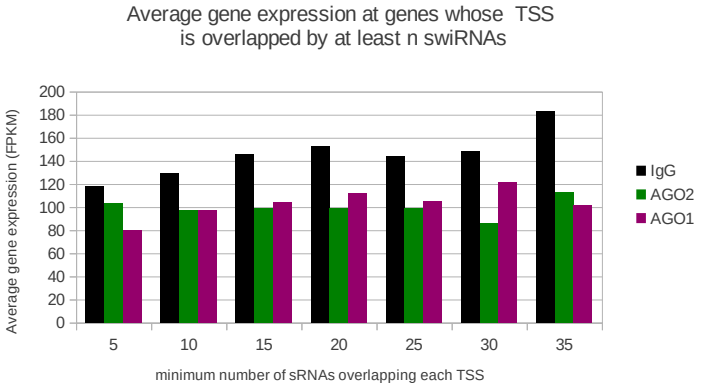


Fig.R15: a. The occupancy at the nucleosome +1 is reduced in AGO2 Knock Down cells and depends on swiRNAs mapping between -150 /+150 nt relative to TSS. Chromatin from siCtrl- or siAGO2-treated HeLaS3 cells was digested by MNase and recovered DNA fragments were sequenced. Bars height represents percent reduction of nucleosome occupancy (siAGO2 vs siCTRL) at TSS (+/- 150 nt) overlapped by at least the indicated number of swiRNAs (green), IgG-IP "other sRNAs" (black) and AGO1-associated "other sRNAs" (purple). Differences depicted with green bars (AGO2 IPed sRNAs) are highly significant (P value < 0.01, paired t-test). **b. The number of AGO2-associated swiRNAs overlapping each TSS does not correlate with gene expression level.** Bar height represents average gene expression in untreated HeLaS3 cells at TSS (+/- 150 nt) overlapped by at least the indicated number of swiRNAs (green), IgG IP "other sRNAs" (black) and AGO1 associated "other sRNAs" (purple). Expression values in FPKM summarized at gene level and obtained from Cuffdiff software were used to compute average expression

values reported in this bar-plot. sRNAs coverage was computed using TSS coordinates restricted to the 21265 loci expressed in HeLaS3 cells under our growth conditions. No further filtering based on the level of expression of mRNAs was applied.

As shown in Fig R16, occupancy at nucleosome +1 downstream of TSS for the subclass of genes with at least 30 swiRNAs lying nearby their TSS was strongly (14%, paired t-test P-value = 0.0001687, FDR < 0.01) affected by AGO2 depletion. This phenotype resembles the one observed in murine cells where core components of SWI/SNF have been genetically ablated (Tolstorukov et al., 2013).

Normalized nucleosome occupancy

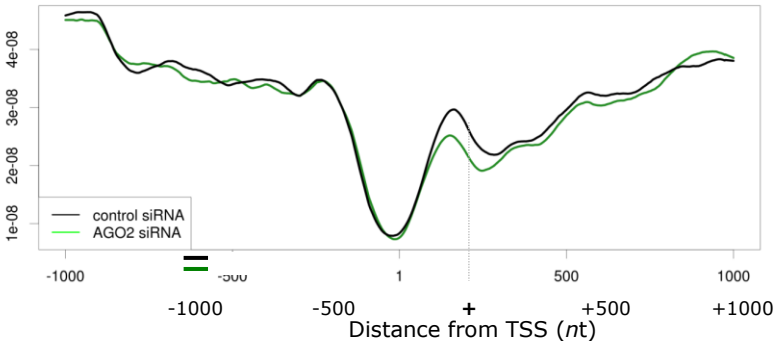


Fig.R16. Nucleosome occupancy profile for siCTRL and siAGO2 cells was plotted for TSS of HeLaS3 cells with at least 30 swiRNAs mapping between nt -150 / +150 relative to TSS (siCTRL, black line; siAGO2, green line).

Overall, these results strongly support the idea that the newly identified AGO2-associated sRNAs (swiRNAs) are key players in nucleosome +1 positioning.

EXPERIMENTAL PROCEDURES

Cell culture and transfection. HeLa S3, Jurkat and HEK293T cells were grown in DMEM medium supplemented with 10% (v/v) fetal bovine serum, 2mM L-Glutamine and Penicillin-Streptomycin. HCT 116 WT and DICER^{Ex5} cells[1] were grown in McCoy's 5A medium supplemented with 10% (v/v) fetal bovine serum, 2mM L-Glutamine and Penicillin-Streptomycin. Transfections were done with 10 nM siRNAs (siAGO2: a pool

containing the following siRNAs GCAGGACAAAGAUGUAUAA[dT][dT][2] and CGUCCGUGAAUUUGGAAUCAU[dT][dT] (Sigma); siCTRL: AGCUUCAUAAGGCGCAUGC[dT][dT]) for 4 days using INTERFERin® as transfecting agent according to the manufacturer's instructions (Polyplus Transfection).

Western blot. Western blot analyses were carried out according to standard procedures using the following antibodies: anti-AGO2 (11A9, Ascenion)[4], anti-GAPDH (14C10, Cell Signaling technology), anti-Histone Antibody, H1 + core proteins (F152.C25.WJJ, Millipore), anti-β-TubulinI (SAP.4G5, SIGMA), goat-anti mouse and anti-rabbit IgG-HRP conjugated (Bio-Rad), anti-rat IgG-HRP conjugated (Jackson).

Preparation of total, nuclear and cytosolic cell extract. For total cell extract, cells were washed twice with cold PBS, scraped off the culture plates with PBS and then lysed in appropriates volumes (20×10^6 cells/ml) of IP-buffer (150 mM KCl; 25 mM NaCl; 2 mM EDTA; 0.5% NP40; 0.5 mM DTT; protease inhibitor (Sigma)) for 20 min on ice. Lysate was clarified by centrifugation at 16,000 g for 10 min at 4°C. Cytosolic and nuclear extracts were prepared as following: cells were resuspended, firstly, in two volumes of ice-cold buffer I (0.3 M sucrose in 60 mM KCl; 15 mM NaCl; 5 mM MgCl2; 0.1 mM EGTA; 15 mM Tris-HCl pH 7.5; 0.5 mM DTT; protease inhibitors), then added other two volumes of ice-cold Buffer II (0.3 M sucrose in 60 mM KCl; 15 mM NaCl; 5 mM MgCl2; 0.1 mM EGTA; 15 mM Tris-HCl pH 7.5; 0.5

mM DTT, 0.4% NP-40, protease inhibitors) in order to obtain 25×10^6 cells per ml, and placed on ice for 10 min. Afterwards, cell lysate was layered on 24 ml of a sucrose cushion (1.2 M sucrose in 60 mM KCl; 15 mM NaCl; 5 mM MgCl2; 0.1 mM EGTA; 15 mM Tris-HCl pH 7.5; 0.5 mM DTT) and centrifuged in a pre-chilled swing-out rotor at 10,000 g for 20 minutes at 4°C. The upper phase containing the cytoplasmic fraction was collect and the pellet containing cell nuclei was lysed in IP-buffer (20×10^6 cells/ml) and clarified. Fractionation efficiency was evaluated by western blotting analysis. H1 and GAPDH proteins were used as controls for the nuclear and cytosolic fraction, respectively.

Immunoprecipitation (IP) and co-immunoprecipitation. Antibodies (anti-AGO1 (4B8, Ascenion); anti-AGO2 (11A9, Ascenion) and isotype-matched IgG (as mock IP) were coupled to Protein-G-sepharose beads (Sigma) for 2 hr at 4°C and washed once in IP buffer and twice in IP-wash buffer (50 mM Tris-HCl, pH 7.5; 300 mM NaCl; 5 mM MgCl2; and 0.05% Nonidet P-40). Whole-cell lysate (40×10^6 - 80×10^6 cells) or nuclear fraction (60×10^6 - 160×10^6 cells) were incubated overnight at 4°C under constant rotation with antibodies pre-coupled beads. An aliquot of total extracts was taken out as Input. IP samples were washed once with IP-buffer for 5 min and three times with IP-wash buffer. Proteins were eluted from beads by boiling for 5 min in SDS-PAGE sample buffer (Sigma). Immunoprecipitation efficiency and co-immunoprecipitating proteins were analyzed by western blotting.

Nuclear RNA-Immunoprecipitation (RIP). Nuclei of HeLa S3, Jurkat, HCT116 WT and DICER^{Ex5} cell lines were lysed in IP-buffer supplemented with RNasin (Promega). Nuclear lysates were clarified and pre-cleared in the presence of Protein G-sepharose beads for 2 hr 4°C under constant rotation and then filtered through a 0.45 μm filter. Lysate aliquots were taken out as Input for RNA and protein isolation. Antibodies and isotype-matched IgG (mock IP) were coupled to Protein G-sepharose beads for 2 h at 4°C in IP-buffer containing 1 mg/ml heparin (Sigma). The pre-cleared lysate corresponding to 150×10^6 nuclei for AGO2-RIP or 350×10^6 cells for AGO1-RIP were incubated with antibody-coupled beads overnight at 4°C under constant rotation. IP samples were washed once with IP-buffer for 5 min at 4°C, and three times with IP-wash buffer at 4°C. An aliquot was taken out for western blot. For RNA isolation, Input and IP samples, were DNAseI- and Proteinase K (Roche)-treated. Total RNA and co-precipitated RNA were extracted by phenol:chloroform:isopropyl alcohol and precipitated in ethanol.

Isolation of nucleosomal DNA by Micrococcal Nuclease (MNase) digestion. Digestion of chromatin from untreated, siCTRL- or siAGO2-treated HeLa S3 cells (2×10^6) was performed with 50U of MNase (New England Biolabs) in 300uL of Permeabilization buffer (15mM Tris-HCl pH 7.4, 300mM sucrose, 60 mM KCl, 15 mM NaCl, 4mM $CaCl_2$, 0.5mM EGTA, 0.2% NP-40, 0.5mM β-mercaptoethanol) for 20 min at 37°C. The reaction was stopped by addition of 300uL of Stop Buffer (50mM Tris-HCl pH 8, 20mM EDTA, 1% SDS)

for 2 min on ice. RNA was degraded with 75 µg RNAse A for 1 h at 37°C and cellular proteins digested with 30 µg Proteinase K for 1h at 55°C. Nucleosomal DNA was purified by phenol:chloroform extraction and precipitated in ethanol. DNA size was verified by separation on 2% agarose gel and by Agilent 2100 Bioanalyser (Agilent Technologies). DNA was mostly digested as 1 n nucleosomal DNA.

Library construction and sequencing. Library construction and sequencing was performed by the Institute of Applied Genomics (IGA) Technology Services (Italy). Samples were sequenced on an Illumina HiSeq2000.

mRNA Sequencing Analysis. Reads were aligned using TopHat 2.0.9 and transcripts were assembled with cufflinks (v 2.1.1) with the following options:

-g iGenome_hg19_genes.gtf -M mask_hg19.gtf -b hg19.fa

where: iGenome_hg19_genes.gtf contains annotation of known transcripts (iGenome UCSC hg19). mask_hg19.gtf contains genomic coordinates of annotated rRNAs and tRNAshg19.fa is the fasta file containing sequence of the human genome hg19.transcripts were merged using cuffcompare (2.1.1) using the -R option, to exclude from final report any transcript which was not overlapped by any fragment in our RNA-seq dataset. I therefore ended up with a list of genes expressed in HeLaS3 cells under our growth conditions. All analyses conducted using TSS coordinates (sRNAs coverage, nucleosome occupancy) were restricted to

the 21265 loci expressed in HeLaS3 cells under our growth conditions. No further filtering based on the level of expression of mRNAs was applied. Expression values in FPKM summarized at gene level were used to compute average expression values reported in Supplementary Figure 10.

- DESeq analysis:

```
##count table creation:

samtools sort -n accepted_hits.bam
accepted_hits.sorted
samtools view accepted_hits.sorted.bam | grep -v
"chrM" > dup_rem.sam
rm accepted_hits.sorted.bam
htseq-count -m intersection-nonempty dup_rem.sam
bowtie0.12.7/indexes/transcriptome_data/iGenome_
hg19_genes.gff > sample1.counts.tsv
rm dup_rem.sam
R
library("DESeq")
data <- data.frame(c("input.r1", "siCtrl.r1",
"siAgo.r1" ,"input.r2", "siCtrl.r2",
"siAgo.r2"),
c("sample1.counts.tsv","sample2.counts.tsv",
"sample3.counts.tsv"
,"sample4.counts.tsv","sample5.counts.tsv",
"sample6.counts.tsv"), c("untreated", "siCtrl",
"siAgo", "untreated", "siCtrl", "siAgo" ))
colnames(data) <- c("sample", "name",
"condition")
```

```
cds <-
newCountDataSetFromHTSeqCount(sampleTable=data,
directory=".")
cds <- estimateSizeFactors( cds )
sizeFactors(cds)
norm.cds <- counts( cds, normalized=TRUE )
cds = estimateDispersions( cds )
plotDispEsts(cds)
siAgo.vs.siCtrl <- nbinomTest(cds, "siAgo",
"siCtrl")
plotMA(siAgo.vs.siCtrl)
hist(siAgo.vs.siCtrl$pval, breaks=100,
col="skyblue", border="slateblue", main="")
siAgo.vs.input <- nbinomTest(cds, "siAgo",
"untreated")
plotMA(siAgo.vs.input)
siAgo.vs.siCtrl.Sig =
siAgo.vs.siCtrl[(siAgo.vs.siCtrl$padj < 0.1), ]
sum(siAgo.vs.siCtrl$padj<0.1, na.rm=T)
#[1] 13
#sum(siAgo.vs.input$padj < 0.1, na.rm=T)
[1] 2
```

Small RNA Sequencing analysis.

- Reads quality was checked using FastqC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

```
fastqc Sample$i.fastq.gz
```

- Sequences were quality trimmed (H.Li: https://github.com/lh3/seqtk) and adapters were removed using cutadapt v 1.0 (Martin, M. , 2011) .

```
seqtk trimfq <(gunzip -c Sample$i.fastq.gz)
| cutadapt -m18 -a TGGAATTCTCGGG - | gzip >
$TDIR/trimmed_adapt_rem.gz
```

- Sequences were iteratively aligned to different databases to identify RNAs which did not belong to any known AGO2 associated class of RNA. After each alignment round reads with no alignments were used in the next step. Alignements were performed using Bowtie 0.12.7 (Langmead et al., 2009)with the following options: -n 0 -l 18 (requiring no mismatches in the first 18 nt of the reads). RNAs mapping to a single locus on hg19 at this step were called “other sRNAs” and further analyzed. “other sRNAs” alignements were converted into BED format for both AGO2 sample and IgG control (AGO2_sRNAs.BED and IgG_sRNAs.BED) and further analysed using the bedtools package as follows: “other sRNAs” were divided in clusters (requiring same strandness) using mergeBed program (mergeBed -s); most clusters consisted of less than 50 sRNAs, while a few hundred clusters consisted of tens of thounsands of identical reads. These high-copy clusters likely represent novel miRNAs or PCR artifacts. We therefore decided to focus our attention on sRNAs in low copy clusters (< 50 sRNAs) and removed from AGO2_sRNAs.BED and IgG_sRNAs.BED reads lying into High-copy clusters (> 50 reads).

```
bowtie -p 4 -t -n 0 -l 18 -q --al
$TDIR/aligned_to_miRNA --un
$TDIR/unaligned_to_miRNA
indexes_for_sRNA_analysis/mirbase_20 <(zcat
$TDIR/trimmed_adapt_rem.gz) >
$TDIR/alignements_to_miRNA 2>
$TDIR/logs/miR_bowtie_log
tar -pczf $TDIR/alignements.to.miRNA.tar.gz
$TDIR/alignements_to_miRNA
rm $TDIR/alignements_to_miRNA
tar -pczf $TDIR/aligned_to_miRNA.tar.gz
$TDIR/aligned_to_miRNA
rm $TDIR/aligned_to_miRNA
bowtie -p 4 -t -n 0 -l 18 -q --al
$TDIR/aligned_to_tRNA_rRNA --un
$TDIR/unaligned_to_tRNA_rRNA
indexes_for_sRNA_analysis/hsa_rRNA_tRNA
$TDIR/unaligned_to_miRNA >
$TDIR/alignements_to_tRNA_rRNA 2>
$TDIR/logs/tRNA_rRNA_bowtie_log
tar -pczf $TDIR/unaligned_to_miRNA.tar.gz
$TDIR/unaligned_to_miRNA
rm $TDIR/unaligned_to_miRNA
tar -pczf $TDIR/aligned_to_tRNA_rRNA.tar.gz
$TDIR/aligned_to_tRNA_rRNA
rm $TDIR/aligned_to_tRNA_rRNA
tar -pczf $TDIR/alignements_to_tRNA_rRNA.tar.gz
$TDIR/alignements_to_tRNA_rRNA
rm $TDIR/alignements_to_tRNA_rRNA
bowtie -p 4 -t -n 0 -l 18 -q --al
$TDIR/aligned_to_Rfam_manually_cured --un
$TDIR/unaligned_to_Rfam
indexes_for_sRNA_analysis/Rfam_hsa_manually_cure
```

```
d $TDIR/unaligned_to_tRNA_rRNA >
$TDIR/alignements_to_Rfam_manually_cured 2>
$TDIR/logs/Rfam_manually_cured_bowtie_log
tar -pczf $TDIR/unaligned_to_tRNA_rRNA.tar.gz
$TDIR/unaligned_to_tRNA_rRNA
rm $TDIR/unaligned_to_tRNA_rRNA
bowtie -p 4 -t -n 0 -l 18 -q -m 1 --al
$TDIR/aligned_to_hg19 --max $TDIR/Mum_to_hg19 --
un $TDIR/unaligned_to_hg19
~/miei_eseguibili/bowtie-0.12.7/indexes/hg19
$TDIR/unaligned_to_Rfam > $TDIR/sRNAs 2>
$TDIR/logs/hg19_bowtie_log
tar -pczf $TDIR/unaligned_to_Rfam.tar.gz
$TDIR/unaligned_to_Rfam
rm $TDIR/unaligned_to_Rfam
tar -pczf $TDIR/aligned_to_hg19.tar.gz
$TDIR/aligned_to_hg19
rm $TDIR/aligned_to_hg19
tar -pczf $TDIR/unaligned_to_hg19.tar.gz
$TDIR/unaligned_to_hg19
rm $TDIR/unaligned_to_hg19
awk 'BEGIN {FS= "\t"; OFS="\t"} {print $3, $4,
$4+length($5), $1, 1,$2}' $TDIR/sRNAs >
$TDIR/sRNAs.bed
awk '{print $3-$2}' $TDIR/sRNAs.bed | sort -n |
uniq -c | sed 's/^[ ]*//' > $TDIR/sRNA_length
mergeBed -s -i $TDIR/sRNAs.bed -n >
$TDIR/clusters.bed
awk -v n=$4 -v OFS='\t' '{if ($4 <= 50) print
$1,$2,$3,"name",$4,$5 }' $TDIR/clusters.bed >
$TDIR/LC_clusters.bed
```

```
awk -v n=$4 -v OFS='\t' '{if ($4 > 51) print
$1,$2,$3,"name",$4,$5 }' $TDIR/clusters.bed >
$TDIR/HC_clusters.bed
intersectBed -a $TDIR/sRNAs.bed -b
$TDIR/LC_clusters.bed -s -u -wa >
$TDIR/sRNAs_in_LC.bed
intersectBed -a $TDIR/sRNAs.bed -b
$TDIR/HC_clusters.bed -s -u -wa >
$TDIR/sRNAs_in_HC.bed
awk '{print $3-$2}' $TDIR/sRNAs_in_LC.bed | sort
-n | uniq -c | sed 's/^[ ]*//' >
$TDIR/sRNA_in_LC_length
awk '{print $3-$2}' $TDIR/sRNAs_in_HC.bed | sort
-n | uniq -c | sed 's/^[ ]*//' >
$TDIR/sRNA_in_HC_length
done
```

- Length distribution computing:

```
samtools view *.bam | cut -f10 | perl -e
'my @c=(0) x 51; while ( $_ =<STDIN> )
{chomp; $c[length ($_)-1]=$c[length ($_)-
1]+1;} $OUT=join("\n",@c); print
"$OUT\n";' | sed = | sed 'N;s/\n/\t/g' >
length.distr.sample*
```

- To compute the overlap between sRNAs and other genomic features we calculated the fraction of “other sRNAs” with at least a nt overlap with the feature of interest (e.g. CpG islands) using intersectBed tool. Exact binomial test was used to compute P-values,

using the frequency of “other sRNAs” overlapping a given feature as the expected frequency (null hypothesis). coverageBed tool was used with the -s and -S options to compute coverage of sense “other sRNAs” and antisense “other sRNAs” around each TSS. The TSS coordinates were obtained by analysis of HeLaS3 mRNA-seq (see above). For analysis of HCT116 and HCT116 DICER ^{Ex5} cells TSS and TTS of human genes were retrieved from ENSEMBL human gene annotation (version 69)[17]. Occupancy profiles around each TSS were piled up and normalized by the total number of “other sRNAs” to obtain the cumulative profile (TTS).

```
coverageBed -s -a sRNAs_in_LC.bed -b
1000_nt_around_ENSG_TSS.bed -d | cut -f 6,7,8
| perl parse_and_join_coverage.pl >
coverage.LC.Sample$.sense.1000nt.arnd_tss

coverageBed -S -a sRNAs_in_LC.bed -b
1000_nt_around_ENSG_TSS.bed -d | cut -f 6,7,8
| perl parse_and_join_coverage.pl >
coverage.LC.Sample$.antisense.1000nt.arnd_tss
```

(where parse_and_join_coverage.pl is a custom perl script).

- Computing of the norm.value:

```
wc -l sRNAs_in_LC.bed
```

- Graphics production:

```
R
Sample$.sense <-
read.table("coverage.LC.Sample$.sense.1000nt.arn
d_tss")/norm.value
Sample$.antisenso <-
read.table("coverage.LC.Sample$.antisense.1000nt
.arnd_tss")/norm.value
X11(width=16,height=8)
#par(mar=c(8,8,8,8))
par(cex = 1.5, mar =c(5,5,5,2))
#png (filename="TSSrna_density(Sample$).png" ,
width= 2000 , height= 1000)
plot(Sample$.senso[,1], col="blue",
type="h",xaxt="n", xlab="Distance from TSS
(nt)", ylab="Average TSSRNA\n coverage per
gene", main="Coverage of TSSa RNAs associated
with AGO2 in Sample$", ylim=c(-0.004,0.011))
abline(v=1000, lty=2, lwd=0.5)
lines(Sample$.antisenso[,1], col="red",
type="h")
axis(1, at = c(0,500,1000,1500,2000), labels =
c("-1000","-500",1,500,1000), las=1);
legend(0,0.009, legend = c("TSSRNA (sense)",
"TSSRNA (antisense)"), col = c("blue","red"),
lty = 2, lwd = 10);
savePlot(filename =
"Immagini/TSSrna_density(Sample$).tiff", type =
c("tiff"), device = dev.cur())


dev.off()
```

- Nuclear AGO1 associated sRNAs were analysed using the same methodology outlined for AGO2.

MNAse Sequencing analysis

592534546 reads were sequenced for AGO2 Knock Down sample. 711051028 reads were sequenced for control sample.

Sequence quality was assessed using FastQc v 0.10.1. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

Paired End reads were aligned using bowtie v 0.12.

```
bowtie -S -p 4 -3 10 -5 1 -n 1 -l 25 -m 1 -q --
nomaqround      bowtie-0.12.7/indexes/hg19      -1
Sample$_R1_001.fastq    -Sample$_R2_001.fastq    2>
bowtie_L006.log | samtools view -Sb -f 0x2 - >
Sample$.bam

samtools    view    -f    0x02    Sample$.bam    >
Prop.Paired.Sample$.bam
```

##Length distribution computing:

```
cat    <(samtools    view    Prop.Paired.Sample$.bam)
<(samtools view Prop.Paired.Sample$.bam | awk -
F'\t' 'function abs(x){return ((x < 0.0) ? -x : x)}
{if (abs($9) < 162 && abs($9) > 115 ) print $0}') |
samtools view -Sb - | bamToBed -i - | sed
'N;s/\n/\t/' | cut -f 1,2,9 | slopBed -b 1 -g
../human.hg19.genome -i - | awk 'NF >0 { print ($3-
$2) }' | sort | uniq -c> length.distr.Sample$
```

- I therefore selected Paired End reads with a size between 100 and 200 nt.

After such selection I obtained:

206561235 fragments for AGO2 Knock down sample

240239066 fragments for Control sample

#Danpos:

```
for i in {1..22..1}
do
out1="chr$i.1.bam"
chr="chr$i"

cat <(samtools view -H Sample$_PropPair.bam)
<(samtools view Sample$_PropPair.bam) | fgrep -w
"$chr" | samtools view -Sb - >
Sample$/danpos/$out1
python danpos-2.1.2/danpos.py
Sample$/danpos/:Sample$i/danpos/ -o
samples.comparing$chr -p 1
rm Sample$/danpos/$out1
done

chrx="chrX"
$out1 ="chrx.bam"
cat <(samtools view -H Sample$_PropPair.bam)
<(samtools view Sample$_PropPair.bam) | fgrep -w
"$chrx" | samtools view -Sb - >
Sample$/danpos/$out1
```

```
python danpos-2.1.2/danpos.py
Sample$/danpos/:Sample$i/danpos/ -o
samples.comparing$chrx -p 1
rm Sample$/danpos/$out1
done
```

#NucleR:

```
#alignments must be split by chromosome before
#using nucleR:
#!/bin/bash

for i in {1..22..1}
do
out1="chr$i.bed"
chr="chr$i"

grep -w "$chr" <(gunzip -c
Sample$.Prop_Paired.100_200_sorted.bed.gz) | cut
-f 1,2,3,6  > Alignments_split_by_chr/$out1

done

chrx="chrX"
$out1 ="chrx.bed"

grep -w "$chrX" <(gunzip -c
Sample$.Prop_Paired.100_200_sorted.bed.gz) | cut
-f 1,2,3,6  > Alignments_split_by_chr/$out1
done

R
```

```
library("rtracklayer")
chr$.sample$ <- import.bed("chr$.bed",
asRangedData=T)
library("nucleR")
reads_trim = processReads(chr$.sample$,
type="paired", fragmentLen=200, trim=40)
cover_trim = coverage.rpm(reads_trim)
rm(reads_trim)
gc()
htseq_raw = as.vector(cover_trim[[1]])
htseq_fft = filterFFT(htseq_raw,
pcKeepComp=0.02)
rm(htseq_raw)
gc()
 peaks = peakDetection(htseq_fft,
threshold="25%", score= T, width=140)
rm(htseq_fft)
gc()
nuc_calls = ranges(peaks[peaks$score >
0.1,])[[1]]
export(nuc_calls,"bed.files/sample$.nuc_calls.ch
r$.bed",format="bed")
red_calls = reduce(nuc_calls)
red_class = RangedData(red_calls,
isFuzzy=width(red_calls) > 140)
sum( red_class$isFuzzy=="TRUE")
write.table(as.data.frame(red_class),
"bed.files/Sample$.fuziness.chr$.bed", sep="\t",
row.names=F, quote=F)
detach(package:rtracklayer)
export.wig(cover_trim,"wig.files/sample$.chr$")
q()
```

- Nucleosome Occupancy at the genomic loci of interest was computed using coverageBed with default options and a custom perl script to determine cumulative occupancy over multiple loci. Occupancy was normalized by the total number of sequenced nucleotides in each library taking into account Properly Paired read pairs with a length between 100 and 200. HeLaS3 TSS coordinates were based on RNA-seq analysis of mRNAs (see below). Occupancy profile at TSS overlapped by at least 30 swiRNAs was smoothed using supsmu function in R statistical environment with option span = 50. Nucleosome +1 occupancy was defined as the sum of the coverage at each nucleotide position in the interval between nt +100 and nt +300 relative to TSS. For each group of genes that were selected based on the minimum number of swiRNAs mapping within ± 150 nt of TSS. A paired t-test was performed to compute P-value of the difference observed between average coverage at nucleosome +1 in AGO2 knock down and Ctrl siRNA cells. For each group of n genes (whose TSS were overlapped by at least m swiRNAs) 10000 random permutations were performed to estimate FDR. In each permutation n random genes were chosen (among the 21265 genes expressed in HeLaS3 cells) and the P-value was computed comparing occupancy at nucleosome +1 with the same procedure outlined above for the real case. All real cases with a P-value < 0.01 were tested, and FDR was always estimated to be < 0.01.

DISCUSSION

The main goal of the present thesis has been to set up an extensive bioinformatics analysis in order to shed light on the

functional roles of AGO2-SWI/SNF interaction in nuclei of human cell lines. Indeed, we have previously demonstrated in our laboratory that nuclear AGO2 interacts with SWI/SNF complexes in nuclei of human cell lines (HCT116; HeLa S3; HEK293T; Jurkat cell lines) and their interaction is independent of DNA or RNA.

The results here reported highlight both the bioinformatics pipeline, set up for the analysis of the huge amount of data produced from Next-Generation Sequencing experiments, as well as the biological relevance of the obtained outputs. Notably, I have identified a novel class of endogenous, AGO2-associated and Dicer-dependent small RNAs (sRNAs) in nuclei of HeLa S3 cell lines. The bioinformatics recursive mapping to several classes of already annotated sRNAs (eg. miRNAs, rRNAs/tRNAs, snoRNAs), allowed me to exclude from my analysis any previously described class of sRNAs, searching for novel sRNAs. This data investigation brought to my attention low-copy abundance transcripts that, moreover, do not show the biological characteristics necessary to be considered putative microRNAs, as highlighted by the output of my analysis with miRanalyzer.

Previous reports (Euskirchen et al., 2011) have characterized a genome-wide map of SWI/SNF binding sites in HeLa S3 cell lines. The use of bioinformatics tools allowed me to intersect these publicly released data sets (ENCODE ChIP-seq data) and the genomic coordinates of the filtered sRNAs present in my data sets. Such analyses allowed me to figure out that this novel class of AGO2-bound sRNAs arise from SWI/SNF-bound TSS, therefore here referred at as “swiRNAs”.

Moreover, the Dicer processing I have demonstrated, allows to hypothesize that complementary long TSSa RNA pairs (Valen et al., 2011; Seila et al., 2008) might be swiRNAs precursors.
Several experimental evidence indicates that AGO proteins in association with sRNAs can regulate nuclear processes in yeast and plants (Grewal et al., 2007; Matzke et al., 2009), suggesting that RNAi might similarly operate in animals. Our identification of a new macromolecular complex including AGO2, core components of SWI/SNF and swiRNAs suggests the involvement of AGO2 in novel and uncharacterized functions in mammals. It has been recently shown that genetic ablation of key subunits of SWI/SNF specifically affects nucleosome occupancy at TSS (Tolstorukov et al.,2013), suggesting that SWI/SNF function is of great importance for the proper positioning of nucleosomes around TSS. My data highlight that AGO2 depletion in HeLa S3 cells affects nucleosome occupancy as well. Importantly the canonical tools used to compute nucleosome occupancy changes, such as Danpos and NucleR, failed to identify differences between AGO2 knock down cell lines and control. This is because these algorithms perform statistical tests on each single nucleosome present all over the genome. On the contrary, my bioinformatics pipeline focused only on a subclass of nucleosomes: the ones around TSS. Thus, computing the average nucleosome occupancy profiles around all TSS, I managed to highlight actual differences upon AGO2 knock down. In order to get very precise results, I used the TSS coordinates of the 21.265 genes expressed in HeLaS3 derived on my previous RNA-seq analysis of mRNAs under the same

growth conditions. Interestingly, the overall nucleosome occupancy changes, even if statistically significant, were not much appreciable when considering all TSS coordinates. On the contrary, by restricting my analysis only on TSS bound by swiRNAs and I observed an important reduction of nucleosome +1 occupancy. Finally, the more swiRNAs map around the considered TSS, the bigger was the change in nucleosome +1 occupancy in AGO2 knock- down HeLaS3 cells. Our data also highlight that larger numbers of AGO1 or IgG associated sRNAs mapping on TSS did not correspond to stronger reduction of nucleosome +1 occupancy, suggesting that this effect is mediated by swiRNAs and not by other sRNAs.

Taken together, these data have brought to the formulation of the biological model presented in FIG. D1. We propose that swiRNAs mediate recruitment of nuclear AGO2 and SWI/SNF onto targeted TSS to ensure proper nucleosome +1 positioning. Further experiments with AGO2 knock-out cell lines will be necessary in order to elucidate whether the small, although statistically highly significant, occupancy reduction observed in AGO2 knock-down may be due to the fact that residual AGO2 protein in AGO2 knock-down cells may still recruit, although to a lesser extent, SWI/SNF onto targeted TSS.

mRNA expression profile analysis, which on the contrary does not reveal any significant change in gene expression upon Ago2 knock-down, needs to be more deeply investigated. Indeed, the nucleosome +1 differential positioning in AGO2 knock-down cells could alter in the canonical splice patterns (Ameyar-Zazoua M. et al., 2012). Moreover, an alternative TSS usage for the same gene needs to be investigated as well.

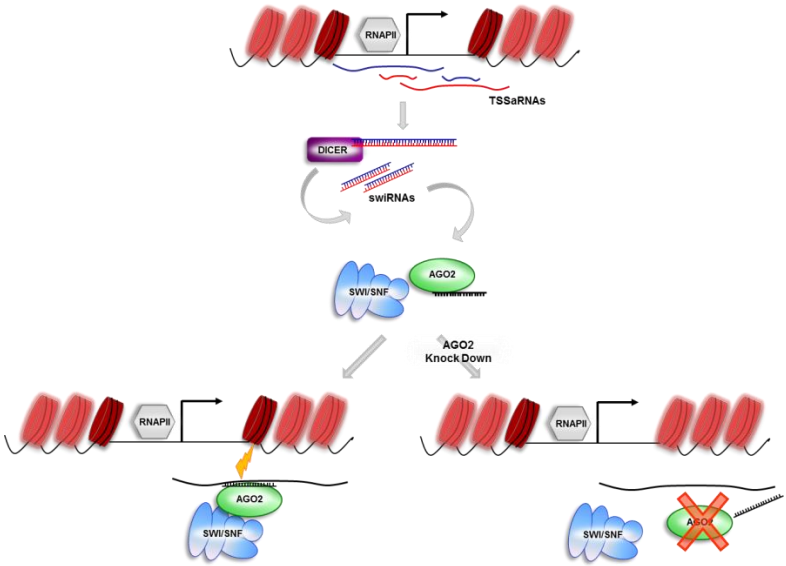


Fig D1: A schematic model depicting the role of AGO2-SWI/SNF-swiRNA complex in chromatin remodelling in mammalian cells. Divergent transcription at active promoters gives rise to sense (blue) and antisense (red) TSSa RNAs. Complementary TSSa RNAs might form dsRNA, which are subsequently processed by DICER to produce swiRNAs. In the nucleus AGO2 loaded with swiRNAs recruits SWI/SNF complex on target TSSs through complementarity between swiRNA and nascent promoter transcripts. The AGO2-swiRNA-SWI/SNF complex maintains the typical nucleosome occupancy signature at TSS, which is characterized by a nucleosome free region spanning the TSS, flanked by two well-positioned nucleosome (+1 and –1 nucleosome; dark red). Following depletion of AGO2, the occupancy of nucleosome +1 decreases (light red).

REFERENCES

- Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC, Mathieu J, Hamiche A, Ait-Si-Ali S, Muchardt C, Batsché E, Harel-Bellan A. Sep 9. Argonaute proteins couple chromatin silencing to alternative splicing. Nat Struct Mol Biol. 2012 Oct;19(10):998-1004. doi: 10.1038/nsmb.2373. Epub 2012
- Anders S. and Huber W. « Differential expression analysis for sequence count data". Genome Biology, 2010: 11,R106.
- Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol. 2001 Jul 10;11(13):1017-27.
- Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. The Plant Cell 15: 2730–2741. doi: 10.1105/tpc.016238
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan 23;116(2):281-97.
- Bartel DP. "MicroRNAs: target recognition and regulatory functions." Cell, 2009: 136, 215–233.
- Batsché E, Yaniv M, Muchardt C. "The human SWI/SNF subunit Brm is a regulator of alternative splicing." Nat Struct Mol Biol, 2006: 13(1):22-9.
- Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. Nature Rev. Genet. 3, 370–379 (2002).
- Bazett-Jones D. P., Côté J., Landel C. C., Peterson C. L., Workman J. L. (1999) The SWI/SNF complex creates loop domains in DNA

and polynucleosome arrays and can disrupt DNA-histone contacts within these domains. Mol. Cell. Biol. 19, 1470–1478.

- Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. Stress-induced reversal of microRNA repression and mRNA P-body localization in human cells. Cold Spring Harb Symp Quant Biol. 2006;71:513-21.
- Béclin C, Boutet S, Waterhouse P, Vaucheret H. (2002) A branched pathway for transgene-induced RNA silencing in plants. Curr Biol. 2002 Apr 16;12(8):684-8.
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ.n E, Caudy AA, Hammond SM, Hannon GJ. (2002) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409(6818):363-6.
- Bernstein E, Kim SY, Carmell MA, et al. Dicer is essential for mouse development. Nat Genet 2003; 35: 215–7.
- *Bevilacqua A, et al, SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease,*Cardiovasc Pathol (2013), http://dx.doi.org/10.1016/j.carpath.2013.09.003
- Brehm A., Tufteland K. R., Aasland R., Becker P. B. (2004) The many colors of chromodomains. BioEssays 26, 133–140.
- Breiman L « Random forests". Machine learning 2001;45:5-32.
- *Britten, R. J. Transposable element insertions have strongly affected human evolution. Proc. Natl Acad. Sci. USA 107, 19945–19948 (2010).*
- Burge, S. W. et al. Rfam 11.0: 10 years of RNA families. Nucl. Acids Res. gks1005 (2012).
- Cai S., Lee C. C., Kohwi-Shigematsu T. (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat. Genet. 38, 1278–1288
- Castanotto D, Tommasi S, Li M, Li H, Yanow S, Pfeifer GP, Rossi JJ. "Short hairpin RNA-directed promoter in HeLa cells." Mol Ther, 2005: 12:179–183.
- Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, et al. (2011) Chromatin-associated RNA interference

components contribute to transcriptional regulation in Drosophila. Nature480: 391–395.

- Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, Dent S, He X, Li W. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res. 2013 Feb;23(2):341-51. doi: 10.1101/gr.142067.112. Epub 2012 Nov 28.
- Cheng SW, Davies KP, Yung E, Beltran RJ, Yu J, Kalpana GV. "c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function." Nat Genet, 1999: 22(1):102-5.
- Cho H, Orphanides G, Sun X, Yang XJ, Ogryzko V, Lees E, Nakatani Y, Reinberg D. "A human RNA polymerase II complex containing factors that modify chromatin structure." Mol Cell Biol, 1998: 18(9):5355-63.
- Clapier CR, Cairns BR. "The biology of chromatin remodeling complexes." Annu Rev Biochem, 2009: 78:273- 304.
- Corey LL, Weirich CS, Benjamin IJ, Kingston RE. "Localized recruitment of a chromatin-remodeling activity by an activator in vivo drives transcriptional elongation." Genes Dev, 2003: 17(11):1392-401.
- Cummins, J. M. et al. The colorectal microRNAome. PNAS 103, 3687–3692 (2006).
- Das S., Cano J., Kalpana G. V. (2009) Multimerization and DNA-binding properties of INI1/hSNF5 and its functional significance. J. Biol. Chem. 284,19903–19914.
- Denli AM, Hannon GJ. RNAi: an ever-growing puzzle. Trends Biochem Sci. 2003 Apr;28(4):196-201. Review.
- Dinant C., Luijsterburg M. S., Höfer T., von Bornstaedt G., Vermeulen W., Houtsmuller A. B., van Driel R. (2009) Assembly of multiprotein complexes that control genome function. J. Cell Biol. 185, 21–26.
- Doi N, S Zenno, R Ueda, H Ohki-Hamazaki, K Ui-Tei, K Saigo.Short-interfering-RNA-mediated gene silencing in

mammalian cells requires Dicer and eIF2C translation initiation factors. Curr. Biol., 13 (2003), pp. 41–46

- Dykxhoorn DM, Novina CD, Sharp PA. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol. 2003 Jun;4(6):457-67. Review.
- Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 2001;15(2):188-200
- Eulalio A, Behm-Ansmant I, Izaurralde E. "P bodies: at the crossroads of post-transcriptional pathways. ." Nat. Rev Mol Cell Biol, 2007a: 8: 9–22 (a).
- Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. "P-body formation is a consequence, not the cause, of RNA-mediated gene silencing." Mol Cell Biol, 2007b: 27(11):3970-81 (b).
- Eulalio A, Rehwinkel J, Stricker M, Huntzinger E, Yang SF, Doerks T, Dorner S, Bork P, Boutros M, Izaurralde E. "Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing." Genes Dev, 2007c: 21(20):2558-70 (c).
- Euskirchen G. M., Auerbach R. Snyder M. « SWI/SNF Chromatin-remodeling Factors: Multiscale Analyses and Diverse Functions. « The Journal of Bioch. Chemistry, 2012 : 287(37): 30897–30905.
- Euskirchen, G. M. et al. Diverse Roles and Interactions of the SWI/SNF Chromatin Remodeling Complex Revealed Using Global Approaches. PLoS Genet 7, e1002008 (2011).
- Flores O, Orozco M. nucleR: a package for non-parametric nucleosome positioning. Bioinformatics. 2011 Aug 1;27(15):2149-50. doi: 10.1093/bioinformatics/btr345. Epub 2011 Jun 7.
- Finnegan EJ, Matzke MA. The small RNA world. J Cell Sci. 2003 Dec 1;116(Pt 23):4689-93.
- Gagnon KT, Corey DR. Argonaute and the nuclear RNAs: new pathways for RNA-mediated control of gene expression. Nucleic Acid Ther. 2012 Feb;22(1):3-16. doi: 10.1089/nat.2011.0330. Epub 2012 Jan 27.

- Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics. 2007 Mar 1;8:69.
- Ghildiyal M, Zamore PD. "Small silencing RNAs: an expanding universe." Nat Rev Genet, 2009: 10(2):94-108.
- Green VA, Weinberg MS. "Small RNAinduced transcriptional gene regulation in mammals mechanisms,therapeutic applications, and scope within the genome." Prog Mol Biol Transl Sci, 2011: 102: 11– 46.
- Grewal S. I. S. and Moazed D. Heterochromatin and Epigenetic Control of Gene Expression. 8 August 2003: Vol. 301 no. 5634 pp. 798-802.
- *Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007 Jul 6;27(1):91-105.*
- Grishok A., Pasquinelli A.E., Conte D., Li N., Parrish S., Ha I., Baillie D.L., Fire A., Ruvkun G., and Mello C.C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 106:23.
- Hackenberg M, Sturm M, Langenberger D, Falcon-Perez JM, Aransay AM "miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments". 2009;37:W68-W76.
- Hall I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. 2002. Establishment and maintenance of a heterochromatin domain. Science 297: 2232–2237.

- Hammond S.M. , E. Bernstein, D. Beach, G. J. Hannon, 404, 293 (2000).
- Han J, Kim D, Morris KV. "Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells." Proc Natl Acad Sci U S A, 2007: 104(30):12422-7.

- Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, Zhou B, et al. Chromatin regulation by Brg1 underlies heart muscle development and disease. 2010;466:62-67.
- Hannon GJ. « RNA interference. » Nature (2002). Jul 11;418(6894):244-51.
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 2009;10:R32.
- Hartley, P. D. & Madhani, H. D. Mechanisms that Specify Promoter Nucleosome Location and Identity. Cell 137, 445–458 (2009).
- Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. "Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells." Nucleic Acids Res, 2009: 37(9):2984-95.
- Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature. 2003 May 1;423(6935):91-6.
- Ho L., Ronan J.,L., Wu J., Staahl B.T.,Chen L., Kuo A., Lessard J., Nesvizhskii A. I., Ranish J., Crabtree G. R.(2009) An embryonic stem cell chromatin-remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. U.S.A. 106, 5181–5186.
- Höck J, Meister G. "The Argonaute protein family." Genome Biol, 2008: 9(2): 210.
- Hu, J. et al. Promoter-associated small double-stranded RNA interacts with heterogeneous nuclear ribonucleoprotein A2/B1 to induce transcriptional activation. Biochem. J. 447, 407–416 (2012).
- Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filee, J. & Capy, P. The struggle for life of the genome's selfish architects. Biol. Direct 6, 19 (2011).

- Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR. "Activating gene expression in mammalian cells with promoter-targeted duplex RNAs." Nat Chem Biol, 2007: 3:166–173.
- Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10, 161–172 (2009).
- Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. Annu. Rev. Genomics Hum. Genet. 8, 241–259 (2007).
- Kaeser M. D., Aslanian A., Dong M. Q. ,Yates J. R. 3rd., Emerson B. M.(2008)BRD7, a novel PBAF-specific SWI/SNF subunit, is required for target gene activation and repression in embryonic stem cells. J. Biol. Chem. 283,32254–32263.
- Kaplan N, Hughes TR, Lieb JD, Widom J, Segal E 2010. Contribution of histone sequence preferences to nucleosome organization: Proposed definitions and methodology. Genome Biol11: 140.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007 Oct;39(10):1278-84. Epub 2007 Sep 23.
- Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell. 1999 Oct 15;99(2):133-41.
- Kidder BL, Palmer S, Knott JG. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. 2009;27:317-328.
- Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14, R36 (2013).
- Kim, P. M. et al. Analysis of copy number variants and segmental duplications in the human genome: evidence for a change in the

process of formation in recent evolutionary history. Genome Res. 18, 1865–1874 (2008).

- Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011 Aug 11;12(8):R72. doi: 10.1186/gb-2011-12-8-r72.
- Kim DH, Villeneuve LM, Morris KV, Rossi JJ. "Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells." Nat Struct Mol Biol, 13(9):793-7 (2006).
- Kim JK, Huh SO, Choi H, Lee KS, Shin D, Lee C, Nam JS, et al. Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. 2001;21:7787-7795.
- Kim JW, Zhang YH, Zern MA, ROSSI JJ,Wu J. "Short hairpin RNA causes the methylation of transforming growth factor-beta receptor II promoter and silencing of the target gene in rat hepatic stellate cells." Biochem Biophys Res Commun, 2007: 359: 292–297.
- Kim S. I., Bultman S. J., Kiefer C. M., Dean A., Bresnick E. H. (2009) BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. Proc. Natl. Acad. Sci. U.S.A. 106, 2259–2264 (a).
- Kim S. I., Bresnick E. H., Bultman S. J. (2009) BRG1 directly regulates nucleosome structure and chromatin looping of the α-globin locus to activate transcription. Nucleic Acids Res. 37, 6019–6027 (b).
- Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods. 2009;6:291–295.
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39, D152–D157 (2010).
- Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y. L. L., Dewey, C. N. N., et al. (2006). A genome-wide map of conserved microRNA

targets in C. elegans. Curr. Biol. 16, 460–471. doi: 10.1016/j.cub.2006.01.050

- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25 (2009).
- Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843-854.
- Lee D, Kim JW, Seo T, Hwang SG, Choi EJ, Choe J. "SWI/SNF complex interacts with tumor suppressor p53 and is necessary for the activation of p53-mediated transcription." J Biol Chem, 2002: 277(25):22330-7.
- Leung A.K.L., Sharp P.A. (2013). Quantifying Argonaute proteins in and out of GW/P-bodies: Implications in microRNA activities. 768: 165–182.
- *Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20. doi: 10.1016/j.cell.2004.12.035*
- Li H, Li WX, Ding SW. Induction and suppression of RNA silencing by an animal virus. Science. 2002 May 17;296(5571):1319-21.
- Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature 426: 465–469. doi: 10.1038/nature02123
- Liu J, Carmell MA, Rivas FV. et al. Argonaute2 is the catalytic engine of mammalian RNAi. Science. 2004;305(5689):1437-41.
- Llave, C; Xie, ZX; Kasschau, KD; et al. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA . SCIENCE Volume: 297 Issue: 5589 Pages: 2053-2056 Published: SEP 20 2002
- Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. RobiNA: a user-friendly, integrated software solution

for RNA-Seq-based transcriptomics. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W622-7.

- Luijsterburg M. S., von Bornstaedt G., Gourdin A. M., Politi A. Z., Moné M. J., Warmerdam D. O., Goedhart J., Vermeulen W., van Driel R., Höfer T. (2010) Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. J. Cell Biol. 189, 445–463.
- Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ. "Structural basis for 5'-endspecific recognition of guide RNA by the A. fulgidus Piwi protein." Nature, 2005: 434:666-670.
- Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M, Kourtis K, Koziris N, Dalamagas T, Hatzigeorgiou AG. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. 2011;39:W145-W148.
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, pp. 10–12 (2011).
- Matzke M., T. Kanno, L. Daxinger, B. Huettel, and A. J. Matzke, "RNA-mediated chromatin-based silencing in plants," Current Opinion in Cell Biology, vol. 21, no. 3, pp. 367–376, 2009.
- Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G,Tuschl T. "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell, 2004: 15:185-197.
- Meister, G. Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. 14, 447–459 (2013).
- Morris KV, Chan SW, Jacobsen SE, Looney DJ. "Small interfering RNA-induced transcriptional gene silencing in human cells." Science, 2004: 305(5688):1289-92.
- Morris, K. V., Santoso, S., Turner, A.-M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet. 4, e1000258 (2008).

- Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, et al. (2011) Sequence-specific error profile of illumina sequencers. Nucleic Acids Research 39: e90.
- Ni Z., Abou El Hassan M., Xu Z., Yu T., Bremner R. (2008) The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. Nat. Immunol. 9, 785–793.
- Nie Z, Xue Y, Yang D, Zhou S, Deroo BJ, Archer TK, Wang W. "A Specificity and Targeting Subunit of a Human SWI/SNF Family-Related Chromatin-Remodeling Complex." Mol Cell Biol, 2000: 20(23):8879-88.
- Nishi K, Nishi A, Nagasawa T, Ui-Tei K. "Human TNRC6A is an Argonaute-navigator protein for microRNA- mediated gene silencing in the nucleus." RNA, 2013: 19(1):17-35.
- Pal-Bhadra, M. et al. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science 303, 669–672 (2004).
- Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature 434: 663–666 doi: 10.1038/nature03462.
- Parker R, Song H. The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol. 2004 Feb;11(2):121-7.
- Pfaff, J. et al. Structural features of Argonaute-GW182 protein interactions. Proc. Natl. Acad. Sci. U.S.A. 110, E3770–3779 (2013).
- Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci U S A. 2008 Feb 5;105(5):1608-13. doi: 10.1073/pnas.0707594105. Epub 2008 Jan 28.
- Plasterk RH. RNA silencing: the genome's immune system. Science. 2002 May 17;296(5571):1263-5.
- Provost, P. et al. Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J.21, 5864–5874 (2002).

- Pugh BF 2010. A preoccupied position on nucleosomes. Nat Struct Mol Biol 17: 923.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. 26:841–842. doi:10.1093/bioinformatics/btq033.
- Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E. "A crucial role for GW182 and the DCP1:DCP2decapping complex in miRNA-mediated gene silencing." RNA, 2005: 11(11):1640-7.
- Rino J., Carvalho T., Braga J., Desterro J. M., Lührmann R., Carmo-Fonseca M. (2007) A stochastic view of spliceosome assembly and recycling in the nucleus. PLoS Comput. Biol. 3, 2019–2031.
- Robb GB, Brown KM, Khurana J, Rana TM. "Specific and potent RNAi in the nucleus of human cells ." Nat Struct Mol Biol, 2005: 12(2):133-7.
- Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, et al. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol 27: 66–75. doi:10.1038/nbt.1518.
- Rüdel S, Flatley A, Weinmann L, Kremmer E, Meister G. "A multifunctional human Argonaute2-specific monoclonal antibody." RNA, 2008: 14(6):1244-53.
- Salma N, Xiao H, Mueller E, Imbalzano AN. "Temporal recruitment of transcription factors and SWI/SNF chromatin-remodeling enzymes during adipogenic induction of the peroxisome proliferator-activated receptor gamma nuclear hormone receptor." Mol Cell Biol, 2004: 24(11):4651-63.
- Sasaki T, Shiohama A, Minoshima S, Shimizu N. “Identification of eight members of the Argonaute family in the human genome small star, filled.” Genomics, 2003:Sep;82(3):323-30.
- Seila, A. C. et al. Divergent Transcription from Active Promoters. Science 322, 1849–1851 (2008).
- Schirle NT, MacRae IJ. "The Crystal Structure of Human Argonaute2." Science, 2012: 336, 1037.

- Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R,Rajewsky N. "Widespread changes in protein synthesis." Nature, 2008: 455: 58–63.
- Shanahan F, Seghezzi W, Parry D, Mahony D, Lees E. "Cyclin E associates with BAF155 and BRG1, components of the mammalian SWI-SNF complex, and alters the ability of BRG1 to induce growth arrest." Mol Cell Biol, 1999: 19(2):1460-9.
- Schones, D. E. et al. Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell 132, 887–898 (2008).
- Schwartz JC, Younger ST, Nguyen NB, Hardy DB, Monia BP, Corey DR, Janowski BA. "Antisense transcripts are targets for activating small RNAs." Nat Struct Mol Biol, 2008: 15(8):842-8.
- Singh M., Popowicz G. M., Krajewski M., Holak T. A. (2007) Structural ramification for acetyl-lysine recognition by the bromodomain of human BRG1 protein, a central ATPase of the SWI/SNF remodeling complex. ChemBioChem 8, 1308–1316.
- Stein A, Takasuka TE, Collings CK. Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? Nucleic Acids Res. 2010;38:709–719.
- Soutoglou E, Talianidis I. "Coordination of PIC assembly and chromatin remodeling during differentiation- induced gene activation." Science, 2002: 295(5561):1901-4.
- Sudarsanam P., Winston F. (2000). The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet. 16(8):345-51.
- Taft, R. J. et al. Tiny RNAs associated with transcription start sites in animals. Nat Genet 41, 572–578 (2009).
- Talbert PB, Henikoff S. "Histone variants--ancient wrap artists of the epigenome." Nat Rev Mol Cell Biol, 2010: 11(4):264-75.
- Thompson M. "Polybromo-1: The chromatin targeting subunit of the PBAF complex." Biochimie, 2009: 309e319.
- Ting AH, Schuebel KE, Herman JG, Baylin SB. "Short double-stranded RNA induces transcriptional gene silencing in human

cancer cells in the absence of DNA methylation." Nat. Genet, 2005: 37: 906–910.

- Tolstorukov, M. Y. et al. Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. Proc. Natl. Acad. Sci. U.S.A. 110, 10165–10170 (2013).
- Trapnell, C., Williams, Pertea, Mortazavi, Kwan, van Baren, Salzberg, Wold & Lior Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotech 28, 511-515 (2010).
- Trapnell C, Hendrickson D,Sauvageau S, Goff L, Rinn JL, Pachter L, Differential analysis of gene regulation at transcript resolution with RNA-seq .Nature Biotechnology doi:10.1038/nbt.2450.
- Trotter KW, Archer TK. "The BRG1 transcriptional coregulator." Nucl Recept Signal, 2008: 6:e004.
- Trouche D, Le Chalony C, Muchardt C, Yaniv M, Kouzarides T. "RB and hbrm cooperate to repress the activation functions of E2F1." Proc Natl Acad Sci U S A, 1997: 94(21):11268-73.
- Valen, E. et al. Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. Nat. Struct. Mol. Biol. 18, 1075–1082 (2011).
- Vasudevan S, Tong Y, Steitz JA. "Switching from repression to activation: microRNAs can up-regulate translation." Science, 2007: 318(5858):1931-4.
- Verdel, A. et al. RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303, 672–676 (2004).
- Wang W., Côté J., Xue Y., Zhou S.,Khavari P. A., Biggar S. R., Muchardt C.,Kalpana G. V., Goff S. P., Yaniv M., Workman J. L., Crabtree G. R. (1996) Purification and biochemical heterogeneity of the mammalian SWI/SNF complex. EMBO J. 15,5370–5382.
- Weinberg MS, Villeneuve LM, Ehsani A, Amarzguioui M, Aagaard L, Chen ZX, Riggs AD, Rossi JJ, Morris KV. "The antisense strand of small interfering RNAs directs histone

methylation and transcriptional gene silencing in human cells." RNA, 2006: 12(2):256-62.

- Weinmann L, Höck J, Ivacevic T, Ohrt T, Mütze J, Schwille P, Kremmer E, Benes V, Urlaub H, Meister G. "Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs." Cell, 2009 : 136(3):496-507.
- Weissman B, Knudsen KE. Hijacking the chromatin remodeling machinery: impact of SWI/SNF perturbations in cancer. Cancer Res. 2009 Nov 1;69(21):8223-30. doi: 10.1158/0008-5472.CAN-09-2166. Epub 2009 Oct 20. Review.

- Wu J. I., Lessard J., Crabtree G. R. (2009) Understanding the words of chromatin regulation. Cell 136, 200–206.
- Wu, C. et al. Hypoxia potentiates microRNA-mediated gene silencing through posttranslational modification of Argonaute2. Mol. Cell. Biol. 31, 4760–4774 (2011).
- Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou MM. "Structure and conserved RNA binding of the PAZ domain." Nature, 2003: 426:468-474.
- Yan Z, Wang Z, Sharova L, Sharov AA, Ling C, Piao Y, Aiba K, et al. BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. 2008;26:1155-1165.
- *Yang N, Kazazian HH Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol. 2006 Sep;13(9):763-71. Epub 2006 Aug 27.*
- Yue X, Schwartz JC, Chu Y, Younger ST, Gagnon KT, Elbashir S, Janowski BA, Corey DR. Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. Nat Chem Biol. 2010 Aug;6(8):621-9. doi: 10.1038/nchembio.400. Epub 2010 Jun 27.

- Zamore P.D., 2001.Thirty-three years later, a glimpse at the ribonuclease III active site. Mol. Cell, 8 (2001), pp. 1158–1160.
- Zeng Y, Sankala H, Zhang X, Graves PR. "Phosphorylation of Argonaute 2 at serine-387 facilitates its localization to processing bodies." Biochem J, 2008: 413(3):429-36.
- Zhang HS, Gavin M, Dahiya A, Postigo AA, Ma D, Luo RX, Harbour JW, Dean DC. "Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF." Cell, 2000: 101(1):79-89.
- Zhang, H., Kolb, F.A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. 21, 5875–5885 (2002)
- Zilberman, D. , Cao, X. & Jacobsen, S.E. Argonaute4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299, 716–719 (2003).

AKNOWLEDGEMENTS

A kind thanks to Prof. Giuseppe Macino and to all of my colleagues who have accompanied me during these three years: Emanuela Cipolletta, Ilaria Laudadio, Claudia Carissimi, Lorena Verduci, Roberta Bottini, Gianluca Azzalin and Katia Broccoletti.

A big Thanks to Dr. Teresa Colombo for accompanying me into the "bioinformatics world" with her love for science and her great charge of enthusiasm.

A big Thanks to Dr. Valerio Fulci for his professionalism and for teaching me many bioinformatics tips during my last year of PhD.

Many many thanks to Samantha Cannazzaro, Giulia Pierini and Alessio Renna for all the coffees and laughs.

The greatest thanks goes to Michele, to my family (cats included) and friend