iLocks: a novel tool for RNA assays with improved specificity

Tomasz Krzywkowski

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Friday 24 November 2017 at 10.00 in Högbomsalen, Geovetenskapens hus, Svante Arrhenius väg 12.

Abstract

The Central Dogma of molecular biology describes a framework for how genetic information is transferred in cells, placing RNA as a messenger between DNA and translated proteins. During the last years, interest in RNA research has grown tremendously due to the increasing understanding and recognition of the importance of RNA in regulation of gene expression, biochemical catalysis, and genome integrity surveillance. Most importantly, RNA content, unlike DNA, changes constantly, fine-tuning the cellular response to match the environmental conditions. There is a clear potential for RNA biomarkers, reflecting both the natural and pathological conditions *in vivo*.

Various methods have been developed to study RNA, of which the most common tools and techniques are described in this thesis. Since many of these gold standard methods are based on detecting RNA derivative (cDNA), there is a wide scope for efficient alternative tools directly targeting RNA. In Paper I, the spatiotemporal expression of human adenovirus-5 mRNA in epithelial and blood cells infected with the virus has been studied. For this, padlock probes and rolling circle amplification (RCA) were used to visualize, quantify and analyse both viral and host cell cDNAs in different infection scenarios, at single cell level. In Paper II, direct RNA detection fidelity has been evaluated using padlock probes. A novel type of probe (iLock) that is activated on RNA via invasive cleavage mechanism, prior to RCA was developed in this approach. Using iLocks, a substantial improvement of direct RNA sensing fidelity has been observed. In Paper III, RNA modifications were introduced in otherwise DNA iLock probes to enhance the probes' efficiency on miRNAs. Using chimeric iLock probes, multiplexed differentiation of conserved miRNA family members were performed with next- generation sequencing-by-ligation readout. Efficient replication of chimeric probes used in Paper III implies that the Phi29 DNA polymerase readily accepts RNA-containing circles as amplification substrates. In Paper IV, real-time RCA monitoring for measurement of amplification rates and analysis of amplification patterns of various RNA-containing circles was achieved. Moreover, the RCA products were sequenced as a proof for the reverse-transcriptase activity of the Phi29 DNA polymerase.

This thesis effectively contributes to a better understanding of mechanisms influencing RNA detection with, but not limited to, padlock probes. It expands the available RNA analyses toolkit with novel strategies and solutions, which can be potentially adapted for RNA-focused research, in general and molecular diagnostics, in particular.

Keywords: *RNA, miRNA, non-coding RNA, padlock probes, rolling circle amplification, invader, single cell, in situ, adenovirus, virology, diagnostics.*

Stockholm 2017
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-147734

ISBN 978-91-7797-043-9
ISBN 978-91-7797-044-6

Department of Biochemistry and Biophysics

Stockholm University, 106 91 Stockholm

ILOCKS: A NOVEL TOOL FOR RNA ASSAYS WITH IMPROVED SPECIFICITY

Tomasz Krzywkowski

Stockholm
University

iLocks: a novel tool for RNA assays with improved specificity

Tomasz Krzywkowski

ISBN print 978-91-7797-043-9
ISBN PDF 978-91-7797-044-6

Printer: Universitetsservice US-AB, Stockholm 2017
Distributor: Department of Biochemistry and Biophysics, Stockholm University

To Magda, my wife, and
science, our mistress

List of Publications

This thesis is based on the following publications:

I **Krzywkowski T***, Ciftci S*, Assadian F, Nilsson M, Punga T. Simultaneous Single-Cell In Situ Analysis of Human Adenovirus Type 5 DNA and mRNA Expression Patterns in Lytic and Persistent Infection. *Journal of Virology* **91**, 1-17 (2017).

II **Krzywkowski T**, Nilsson M. Fidelity of RNA templated end-joining by Chlorella virus DNA ligase and a novel iLock assay with improved direct RNA detection accuracy. *Nucleic Acids Research* **1**, 1-9 (2017).

III **Krzywkowski T**, Kühnemund M, Nilsson M. Detection of miRNAs using chimeric DNA/RNA iLock probes utilizing novel activity of PBCV-1 DNA ligase: RNA-templated ligation of ssRNA. *Final stage manuscript.*

IV **Krzywkowski T***, Kühnemund M*, Di Wu, Nilsson M. Reverse transcriptase activity of Phi29 DNA polymerase. *Submitted manuscript.*

* equal contribution

Related work by the author

Publications

I El-Heliebi* A, Hille C*, Laxman N*, Svedlund J*, Haudum C, Ercan E, Kroneis T, Chen S, Smolle M, Rossmann C, **Krzywkowski T**, Ahlford A, Darai E, G von Amsberg, Alsdorf W, König F, Löhr M, Kruijff I de, Riethdorf S, Gorges T M, Pantel K, Bauernhofer T, Nilsson M, Sedlmayr P. *In situ* detection and quantification of AR-V7, AR-FL, PSA and KRAS point mutations in circulating tumor cells. *Submitted manuscript.*

II Geny S, Moreno P M D*, **Krzywkowski T***, Andersen N K, Isse A J, El-Madani A M, Lou I C, Pabon V, Gissberg O, Anderson B A, Zaghloul E M, Zain R, Hrdlicka P J, Jorgensen P T, Nilsson M, Lundin K E, Pedersen E B, Wengel J, Smith E. Next generation bis-locked nucleic acids with stacking linker and 2'-glycalamino-LNA show enhanced invasion into duplex DNA. *Nucleic Acids Research*, **44**, 2007-2019 (2016).

III Clausson C-M*, Arngården L*, Ishaq O, Klaesson A, Kühnemund M, Grannas K, Koos B, Qian X, Ranefall P, **Krzywkowski T**, Brismar H, Nilsson M, Wählby C, Söderberg O. Compaction of rolling circle amplification products increases signal integrity and signal-to-noise ratio. *Scientific Reports* **5**, 1-10 (2015).

* equal contribution

Book chapters

I Punga T, Ciftci S, Nilsson M. **Krzywkowski. T.** Simultaneous detection of virus DNA and mRNA in individual cells. *Current Protocols in Microbiology*. (2018). *In preparation*

II **Krzywkowski T**, Nilsson M. Padlock probes to detect single nucleotide polymorphisms. *RNA Detection: Methods and Protocols, Methods in Molecular Biology*. Imre Gaspar (ed.). (02.2018)

III **Krzywkowski T**, Hauling T, Nilsson M. In situ single molecule RNA genotyping using padlock probes and rolling circle amplification. *Genotyping, Methods in Molecular Biology* **1492**. Stefan White (ed.). (2017)

IV Hauling T, Ke R, **Krzywkowski T**, Nilsson M. In Situ Detection of Individual mRNA Molecules with Padlock Probes and Target-Primed Rolling-Circle Amplification in Fixed Mouse Brain Tissues. *In Situ Hybridization Methods, Neuromethods* **99**. Giselbert Hauptmann (ed.). (2015)

Contents

Abbreviations

A	adenine
cDNA	complementary DNA
C	cytosine
ds	double-stranded
FISH	fluorescent in situ hybridization
G	guanine
HAdV	human adenovirus
HGP	Human Genome Project
LAMP	loop-mediated isothermal amplification
mRNA	messenger RNA
miRNA	micro RNA
NA	nucleic acid
NASBA	nucleic acid sequence-based amplification
nt	nucleotide
OLA	oligonucleotide ligation assay
PLP	padlock probe
PBCV-1	Paramecium bursaria Chlorella virus 1
PCR	polymerase chain reaction
ncRNA	non-coding RNA
qPCR	quantitative PCR
rRNA	ribosomal RNA
RNA-Seq	RNA sequencing
RCA	rolling circle amplification
SNP	single nucleotide polymorphism
ss	single-stranded
ENCODE	The Encyclopedia of DNA Elements
T	thymine
UTR	untranslated region
U	uracil

Introduction

Scientific progress has reached unbelievable momentum and continues to surprise me. Until the late 19th century, bloodletting was a standard cure and preventive measure against agents destabilizing fine harmony of our "humours". Solving the structure of DNA, hacking the genetic code that encrypts molecular RNA blueprints, as well as revealing DNA's and RNA's link to structures and functions of proteins, marked a turning point in the history of research. We've only just begun to submerge into molecular level of cells, governing life and death.

A laboratory in which I have conducted my doctoral research operates on such molecular level. The years I have devoted to science educated me and refined my understanding of principles underlying disease diagnostics and personalised medicine. In the first chapter of this work, I describe important facts and notions that constitute a fundamental part of my research and aim to provide a good ground for understanding of this thesis' content. Additionally, I explain how recent discoveries have shifted scientific focus to RNA, which for half a decade remained obscured and was considered a mere messenger between DNA and proteins.

The majority of my work was devoted to the development of innovative RNA detection techniques. I do not claim this thesis to be a comprehensive summary of all nucleic acids sensing methods, however, I present the most standardized tools—in the RNA field—used in world spread laboratories.

In the third part of this thesis, I describe and discuss my scientific publications that constitute this thesis. Following with concluding remarks, I shortly acknowledge all friends and colleagues who supported me along this journey.

Stockholm, July 2017

Central dogma: a new look

"Where ignorance is bliss, 'tis folly to be wise". These famous words of Thomas Gray can be aptly used to encapsulate current state of biology; governed by very few rigid principles, growing in complexity with every discovery. Unlike in mathematics, exceptions can be found to every rule if one is willing to look hard enough.

The Central Dogma is an influential and immortal point of reference in modern biology. Presented originally by Francis Crick in 1958 (1) it illustrates the flow of genetic information from DNA—the central information repository—to protein—the effector building blocks. The interpretation of the dogma was initially restricted to the linear flow of information from DNA to RNA and from RNA to protein (Figure 1).

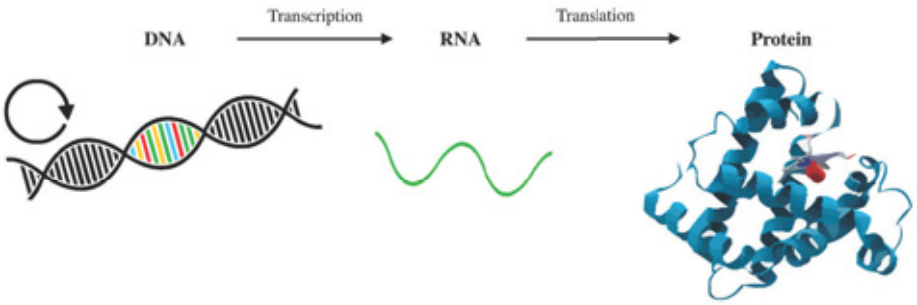

***Figure 1**. The Central Dogma of molecular biology, according to James Watson in 1965 (2). Arrows adjoining intermediates of the dogma—DNA, RNA and protein—represent unidirectional flow of information, meaning that RNA molecules are made from DNA. Consequently, proteins are assembled from the information coded in the template RNA. A circular arrow above DNA represents replication.*

After the Dogma was published, it has been frequently challenged. Relevant discoveries forced scientists to refine this seemingly one-way channel of information. In 1970, RNA-dependent DNA polymerase has been discovered in tobacco mosaic virus (TMV), questioning the role of DNA as an exclusive storage of hereditary information (3, 4). Barbara McClintock's work on retrotransposons (5), mobile elements that can "escape" canonical pathway and reverse-transcribe back into DNA from their own RNA templates, was finally recognized in the late 1960s–1970s and crowned with the Nobel prize in 1983. The discovery of RNA editing in the early 1990s initiated a great pursuit after epigenetic factors, which drive protein structure changes (and thus phenotype) without altering gene sequence on a DNA level (6). Interestingly, coding RNA of a certain protein in *Trypanosoma brucei* becomes altered to such extent where accurate prediction of the protein primary structure from the sequence of the corresponding gene is no longer possible (7).

In the refined Dogma, constraining point is placed in the last step of information transfer, translation. To date, no life forms are known to exist that would harbour reverse-translation: rewriting RNA molecules (or DNA) from protein templates. While inheritance of phenotypical traits mediated epigenetically by proteins was reported in yeast and animals (8, 9), hardcoding DNA from a primary structure of a protein is impossible due to genetic code degeneracy, and would be possibly deleterious.

Shifted paradigm: growing interest in the RNA research

The "disruptive" discoveries listed above, tremendously shaken the view on the role and function of RNA in the Central Dogma. This thesis would not be completed without mentioning 13 years of the iterative and strenuous sequencing of the human genome, under the banner of the greatest explorative endeavour in the history of biology: Human Genome Project (HGP). The primary objective of this multinational effort was to plot a detailed DNA map, permitting all researchers to study human inherited diseases and providing a template for the future genome assemblies (10, 11). One of the project outcomes was the realization, that out of 2,851,330,913 nucleotides constituting our DNA, only 1.2% is covered by protein coding genes (Figure 2) that undergo transcription.

***Figure 2**. Simplified pie-charts representing proportion between protein coding and non-coding segments of the human genome, according to HGP (left) and ENCODE project (right).*

At the time, evidence for the existence of protein non-coding RNA (ncRNA) transcripts (excluding ribosomal RNAs, transfer RNAs or small RNAs) was available, however, it was unclear whether such ncRNA molecules bear any biological functions or rather constitute cellular "noise".

Just before HGP was finalized, the Encyclopedia of DNA Elements (ENCODE) project had been launched. ENCODE objective was to provide a more informative image of functional elements encoded in the human genome, with the use of high-throughput methods (12). Approximately 1% of the genome was selected, half from previously well-annotated segments and

half was sampled randomly. Not surprisingly, an abundance of novel ncRNAs, previously thought to be transcriptionally muted, was identified. Remarkably, as much as 74% of DNA was found to be transcribed to RNA molecules, as confirmed by at least two independent technologies (Figure 2).

Reflecting on examples I have just presented, one can deduce that RNA, as a molecule, is more than just a transient messenger in the Dogma and its central position, is far from coincidental. Though the number of protein-coding genes in human DNA lies between 20 000–25 000 (11), a number of proteins is much bigger (13). RNA is evolutionary older than DNA (13) and it used to be responsible for carrying genetic information. Today, information is stored in DNA but it still needs to be transcribed to RNA and most of the proteome complexity is dictated by processes that take place after transcription. Apart of RNA's passive, yet undeniably essential role as a platform for ribosomes as well as mRNA's codon reader, research of the past two decades has linked RNA to the multitude of active functions, previously reserved to proteins:

- Catalysis of chemical reactions: a flag example is the peptidyl transferase function of a ribosome, responsible for peptide bond formation in a growing protein chain, is fully executed by ribosomal RNA, rRNA (14, 15). Interestingly, Harry Noller from Ludwika Zimniak's lab demonstrated that rRNA (natively complexed with ribosomal proteins) retained peptidyl transferase activity, even after vigorous protein extraction procedures (16)
- Allosteric regulation of mRNA processing: RNAs, known as "riboswitches", were demonstrated to control gene expression in response to temperature changes or binding of selected metabolites (17)
- Signalling and immune surveillance: the microbial equivalent of the mammalian adaptive immune system, conferring cell resistance to external genetic objects like plasmids or phages, is encoded in prokaryotic DNA. Clustered regularly interspaced short palindromic repeats (CRISPR) form an archive of barcodes of potential foreign invaders. RNA transcribed from CRISPR can recognize foreign nucleic acids (NA) and stimulate defence mechanisms against bacteriophages (18, 19)
- Silencing of gene expression by regulatory components targeting mRNAs (I will expand this particular class of RNAs later in this thesis).

We have begun to appreciate the new image of RNA and numerous functions it plays in a cell. It is now obvious that RNA is no longer just an ordered series of nucleotides, sequence is just one component of its complex entity. Catalysis of enzymatic reactions as well as regulation of many physiological processes imposes reconsideration of the RNA as molecule with proportional, if not more essential, function to proteins. Interest in RNA is growing (Figure 3).

Classes and roles of RNAs in human biology

Numerous types of RNAs have been characterised thus far; however, exceptions in physicochemical properties of RNAs within the groups make accurate classification not trivial. It is important to mention that some of the RNA classes presented below are even further sub-classified. For the simplicity of this paragraph, most essential types of RNA are briefly described and those absent in human were excluded.

RNAs can be broadly divided into protein coding and ncRNAs, however, such distinction is a great oversimplification. Instead, RNAs can be classified into 3 categories according to their cellular function. RNAs that constitute the biggest proportion of total RNA mass in a cell, belong to the first group and are involved in protein synthesis. This group includes mRNAs, rRNAs and transfer RNAs (Table 1).

Table 1. *Group I: RNAs involved in translation*

Class	Abbreviation	Function	% RNA mass	Reference
messenger	mRNA	Genetic information carrier	3–5	
ribosomal	rRNA	Associates with set of proteins to form a ribosome	80–85	(14, 15)
transfer	tRNA	Translates triplet codons in mRNA to aminoacids	10–13	(20)

Common examples of ncRNAs in group II—involved in post-transcriptional modification of RNA—is ribonuclease P, small nuclear- and small nucleolar RNAs (Table 2).

Table 2. *Group II: RNAs involved in post-transcriptional modification of RNA*

Class	Abbreviation	Function	Reference
small nuclear	snRNA	mRNA splicing	(21)(22)
small nucleolar	snoRNA	rRNA processing and ribosome assembly	(23)(24)
ribonuclease P	RNaseP	tRNA processing	(25, 26)

snRNAs are on average 150 nucleotides (nt) long and mainly responsible for maturation of primary mRNA transcripts. Removing introns (splicing), non-essential segments of mRNA coding message, is a major step in post-transcriptional mRNA editing. snoRNAs, present in human cells in the form of small nucleolar ribonucleoprotein particles (snoRNP), transiently associate with pre-processed rRNAs and assist in their maturation. RNaseP is a ribozyme, a remarkable example of protein-like RNA with catalytic activity, which I mentioned in the previous paragraph. The function of this ribozyme is to truncate immature tRNA molecules (26). To date, RNaseP is the second ribozyme, after ribosome, capable of substrate turnover. For its discovery, Sidney Altman was awarded the Nobel Prize in Chemistry in 1989.

During last 15 years, ncRNAs have been in the centre of scientific attention and we have witnessed discoveries of plentiful tiny regulatory RNAs, classified in group III (Figure 3).

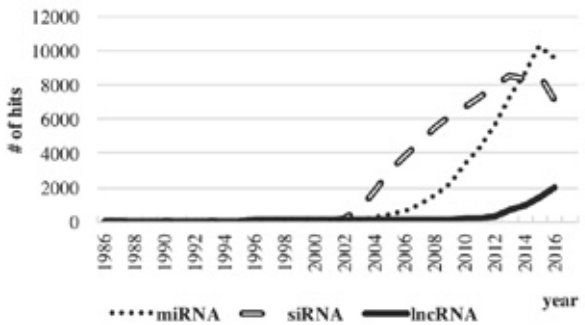

Figure 3. *Number of search hits per year for selected classes of ncRNAs between 1986 and 2016. Since 2014, yet another class of long non-coding RNAs (lncRNAs) is emerging. Based on https://www.ncbi.nlm.nih.gov/pubmed/*

The most fundamental properties of these RNAs are: their short length (between 20 and 30 nt), ability to associate with numbers of effector proteins and their profound silencing effect on target genes. Micro RNAs (miRNA), small interfering RNAs (siRNA), piwi-interacting RNAs (piRNA), short hairpin RNAs (shRNA), and antisense RNAs (aRNA) are representative candidates in group III (Table 3).

Table 3. *Group III: regulatory RNAs*

Class	Abbreviation	Function	Reference
micro	miRNA	Gene expression regulation	(27–29)
small interfering	siRNA	Gene expression regulation	(30)
piwi-interacting	piRNA	Silencing of transposones	(31, 32)
short hairpin	shRNA	Gene expression regulation	(33)
antisense	aRNA	Gene expression regulation	(34, 35)

In 1993, *Caenorhabditis elegans' lin-4* RNA, the first short ncRNA was discovered in Victor Ambros' lab (27). Rosalind Lee and co-workers found multiple short *lin-4* RNA transcripts, containing complementary sequence to *C. elegans lin-14* mRNA 3'-UTR (untranslated region). Authors suggested that the *lin-4* transcript regulates *lin-14* translation via an antisense RNA–miRNA interaction. Seven years later, a second regulatory RNA transcript *let-7* was suggested to control developmental transitions in *C. elegans* (36). Discovery of exceptional *let-7* conservancy across wide range of animal species (28) taken RNA field by storm and started scientific revolution (Figure 3). Under a standard term of microRNAs (miRNA), coined by Ruvkun in 2001 (29), thousands on new miRNAs have been identified in animals and plants and number is growing every year. 2588 unique miRNAs identified in human (at a point of writing this thesis) have been demonstrated to regulate gene expression by antisense mRNA binding, which results in RNA decay (37, 38) or alternatively, in translation inhibition by repression of ribosome attachment (39) and elongation (40, 41). Interestingly, next to their gene silencing function, miRNAs can also induce transcript levels (42) and effectively regulate all fundamental cellular processes (43).

Just when miRNA research was germinating, Andrew Fire in Craig Mello's lab has observed that introduction of double stranded RNA (dsRNA) into *C. elegans* gonads interfered with the expression of endogenous mRNA transcript (30). The discovery of dsRNAs (later called small interfering RNAs (siRNAs)), capable of RNA silencing via a mechanism prevalent from protozoa to vertebrates started a revolutionary era of RNA interference

(RNAi) (44). siRNAs are considered as exogenously administered dsRNAs, while miRNAs as endogenous and single-stranded. These early distinctions became less significant as our understanding of miRNA biology has increased and short, endogenous dsRNA were discovered (45, 46). Post-transcriptional RNAi can be moderated by either miRNA or siRNA; these two are the key participants in gene expression regulation. RNAi using single stranded (ss), antisense RNA (aRNA) has been reported in bacteria (34, 35) and together with synthetic RNA hairpins—shRNA (33)—they expand a repertoire of oligo-tools used for RNAi in eukaryotic cells. The ability to moderate genes expression made RNAi oligonucleotides a promising candidates for the treatment of genetic disorders in human (33, 47).

As our understanding of complicated cellular circuitry grows, and scientific literature with it, various lately described RNAs receive unique acronyms such as piwi-interacting RNAs (piRNA). These exceptional, short RNAs associate with Piwi proteins (essential in the RNAi pathway) and are abundantly expressed in germline cells. Similarly to CRISPR, mammalian piRNA loci occur in clusters and map to repetitive sequences (32), including fragmented transposons. piRNAs play the role of genome guardians, protecting cells from the destructive effect of mobile, genetic elements (31).

The last group of RNAs, which also deserves a short introduction, encompasses long non-coding RNAs (lncRNA). Even though this RNA class remained unknown for a long time, lncRNAs are receiving progressively more attention (Figure 3). Arbitrarily defined as longer than short RNAs from group III, lncRNAs are involved in regulating numerous biological processes including loci imprinting, allosteric regulation of enzyme activity and chromosome structure (48, 49). As the overall complexity of the organism correlates better with the number of lncRNAs than mRNAs (50), the rank of lncRNA will most likely continue to grow in parallel with our understating of their role in human biology. The function of the most lncRNAs is still unknown, yet certain lncRNA received coding-gene-like status for their defined involvement in fundamental biological processes like: X chromosome inactivation (*XIST*), telomere elongation (*TERC*) or chromatin rearrangements (*HOTAIR*) (51).

Finally, it should be emphasized that this chapter encompasses a larger fraction of RNA classes, selected by their general share in the landscape of biological processes. As interest in RNA field is accelerating, it is very likely that new classes will soon be defined or the balance of importance between RNA classes presented in this chapter will change.

Implications of RNA sequence variations and abundance on transcriptome function

Hermann Joseph Muller began his career at the times when genetics was not so precisely defined and all biological hypotheses and conclusions were based on direct observations. In 1927, through series of experiments, Muller demonstrated that exposure of fruit flies to various doses of X-ray radiation causes mutations (52). 50 years later, after the DNA structure was resolved and understood, the first heritable mutations were precisely annotated in the human genome (53, 54). Till today, the number of described human gene lesions responsible for inherited diseases exceeds 140 000 for nearly 6 000 protein coding genes (Human Gene Mutation Database, HGMD; http://www.hgmd.org; August 2017) and the number continues to grow. Variation in the DNA sequence of a functional gene will be propagated into mRNA molecule during transcription and finally into protein. While the genetic code is degenerated and some frequently occurring mutations will code exactly the same aminoacids (a phenomenal example of evolutionary genetics safety mechanism), certain changes will be deleterious and will result in a phenotypic change. A study from 2007 showed that 53% of mis-sense mutations—that result in aminoacid substitution—are associated with a moderate proteins' phenotype alteration and 20% with proteins' loss of function (55). Although the majority of disease-associated DNA variants are single nucleotide polymorphisms (SNP), a significant number of mutations is attributed to nucleotide insertions or deletions (56). Functional (or phenotypical) gene variations are primarily localised within gene coding regions (60%), though DNA changes occurring outside coding segments can still affect gene expression, mRNA splicing or binding of proteins associated with RNA maturation (56).

In contrast to the sequence variants residing within gene coding segments, posing obvious consequences for the functioning of the encoded protein, we begin to understand implications of mutations in non-protein-coding segments. Genome-wide association studies and high-throughput sequencing projects like 1000 Genomes project (http://www.internationalgenome.org) dramatically increased our understanding of the mutation-disease relationship. Indeed, sequence alterations within gene promoter or enhancer can affect gene transcription, post-transcriptional regulation and mRNA half-life (57).

Gene expression can also be affected if mutations are localized directly within gene-regulatory elements like miRNAs or miRNAs' binding site in the mRNA 3'-UTR (58). miRNAs and mRNAs are part of the mutually regulated network. As miRNA genes are often localized in loci subjected to frequent genetic damage, alternated expression of an oncogene-silencing miRNA will promote oncogene expression (59, 60). As a consequence, not only variations in sequence but also miRNA abundance became a hallmark of several human diseases. Amplification of a locus coding for a miRNA that silences expression of a tumour suppressor can promote cancer growth (60). Indeed, mutations in miRNAs have been associated with cancer (61–63) as well as with schizophrenia (64), osteoporosis (65), autoimmune diseases (66) and more.

A great example of a miRNA-mRNA network, where mutations affecting miRNA abundance are reflected on the level of matched mRNAs is *KRAS* and *let-7* miRNA. Approximately 20–25% of tumours harbour activating mutations in KRAS gene, stimulating proliferative, downstream pathways and conferring a survival advantage for tumours (67). *let-7* miRNA, the first miRNA ever discovered, was also identified as one of the first tumour suppressor miRNAs downregulated in cancer (68). Mutations in *KRAS* 3'-UTR, targeted by *let-7,* as well as *let-7* expression found to be a prognostic biomarker in cancer (69, 70).

Analysis of RNA

Role of RNAs as biomarkers

As we have learned from the simplified picture of The Central Dogma, DNA, RNA, and proteins are organized in a way that supports information storage, readout and execution. Characterisation of individual components as biomarkers became available after a) we defined specific links between biological conditions and properties of biomolecules b) learned how to accurately differentiate that properties from a biological noise. Diagnostic methods based on detection of proteins are ubiquitous in hospitals, giving clinicians rapid answers regarding patients' state (blood types, infections or activity of liver enzymes). Analysis of nucleic acids (NA) on the other hand, have a slightly broader line of applications. In comparison to proteins or RNA, DNA is profoundly stable (71) and unique for every person. Thus, defining DNA fingerprint is broadly used in forensics or parenthood tests (72). High-resolution DNA sequence analyses allow us to put any organism, including human, in the context of time (evolution) as well as space (geographical migrations). Automated platforms reduced the cost of DNA sequencing to the point where personal sequencing becomes a regular clinical a practice (73). It is estimated that by 2018, up to 70% of lung cancer patients and 60% of colorectal cancer patients will have some level of tumour DNA sequenced (73). Such knowledge is extremely valuable, as its interpretation helps to establish the personalised line of treatment, that could significantly slow down disease progression (74).

In contrary to DNA, RNA level changes under different physiological and pathological conditions. Thus, there is a great potential for RNA to become accurate predictive factor of a person's condition. By measuring the expression of genes and regulatory molecules, various immune- or chemotherapy targets can be potentially identified in cancer patients. Indeed, there are few examples of established cancer diagnostic tests based on mRNA expression profiling, including OncotypeDX (75), Pathwork TOO (76) and Prosigna (76), some of which are approved by Food and Drug Administration. Recently, there is an enormous interest in the extracellular, RNA containing vesicles as well as in circulating RNA molecules as potential biomarkers for non-invasive diagnostic applications. In fact, circulating RNA molecules

(including mRNAs and lncRNAs) were used to track tissue changes in pregnant woman and fetus, sampled from mother's blood (77). The same study also showed that neuron-specific RNAs are differentially expressed in individuals with Alzheimer's disease and this difference is reflected in blood circulating RNAs (77). Finally, a tumour type—as well as the tissue of origin—can be accurately determined based on miRNA expression signatures (78). miRNAs' clinical potential is reflected by the number of ongoing clinical trials investigating their roles as biomarkers under various conditions (https://clinicaltrials.gov/).

Examples above visualise how much information can be recovered from the transcriptome of an individual person. Even though genome remains unchanged throughout life (in a healthy cell), genes expression changes in response to the environment and this constant fluctuation can be used in diagnostics.

Challenges in RNA detection

Detection of a biomarker, directly in cells or biological fluids, without prior isolation or amplification is a desired feature of a diagnostic assay, however, technically challenging. The polymerase chain reaction (PCR) is used to exponentially amplify few copies of sample DNA to concentrations facilitating detection (79). Most NA sensing assays are based on DNA PCR, as it provides required robustness, detection sensitivity, and accuracy. The difficulty in exponential RNA amplification is that enzymes carrying out similar activity are not present in nature. Thermostable DNA polymerases were evolutionary engineered to readily amplify (replicate) DNA fragments and withstand thermal cycling, features exploited in PCR. DNA dependent RNA polymerase on the other hand (most commonly used bacteriophage T7 RNA polymerase), requires a double stranded DNA promoter sequence to initiate RNA synthesis (limiting applicability) and cannot be cycled due to enzyme thermolability (80). Thermostable RNA polymerases are available, but having optimal activity at 55-65°C they do not tolerate high NA melting temperatures, typical for PCR. T7 RNA polymerase is used as *in vitro* and *in vivo* RNA transcription tool to generate various RNA products (tRNAs, ribozymes) and express proteins *in vivo* (80). Polymerases able to replicate RNA from RNA are exclusively reserved to RNA viruses and are not used as tools for research (81, 82). Curiously, a unique RNA dependent RNA replicase has been discovered in Qβ bacteriophage. Qβ replicase is able to copy very particular type of structured RNA. Even though RNA detection with Qβ replicase is highly sensitive, the requirement of specialised templates is its' major limitation (83). To increase the sensitivity of RNA sensing assays,

templates are typically reverse-transcribed to complementary DNA (cDNA) and subsequently detected using the variety of means (84, 85).

Single stranded RNAs have a tendency to form self-folding structures, where some segments are base paired while others not. Intramolecular stability of RNA, a consequence of the presence of 2' OH group in the RNA pentose, allows for the formation of tertiary structures, conferring important properties. For example, correct folding is required for tRNA, ribozyme and rRNA function (14) or for miRNA biogenesis (86). Longer RNA molecules (mRNAs or lncRNAs) will show higher tendency to form secondary structures that can hinder accessibility of certain RNA fragments in targeted detection methods (87–89). To overcome this limitation, many mRNA sequencing methods use polythymidylate (poly(T)) primes in cDNA synthesis step, since mRNA's 3' poly(A) tail is unlikely to participate in secondary structures (90).

Lastly, a big challenge in RNA sensing is its instability. RNA hydrolysis occurs when a phosphodiester bond linking successive nucleotides in RNA brakes. Compared to DNA, RNA is more susceptible to hydrolysis, due to the presence of 2' hydroxyl group in the RNA's pentose. In alkaline conditions, 2' OH becomes readily deprotonated and can catalyse nucleophilic attack on the adjacent phosphorus causing RNA cleavage (Figure 4).

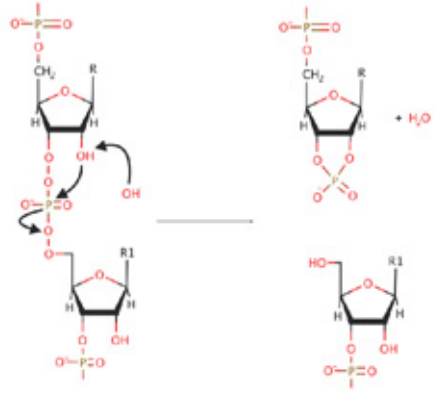

Figure 4*. RNA hydrolysis under alkaline conditions. Deprotonated 2' OH group in the RNA pentose acts as a nucleophile towards phosphorus in adjacent phosphodiester bond (arrow). As a result, the upstream RNA chain forms 2'3'-cyclic monophosphate derivative and downstream chain initiates at 5'-OH.*

DNA lacking 2' OH group is stable under such conditions (13). RNA hydrolysis can occur spontaneously and is further induced at a higher temperature, by temperature cycling, sample processing and prolonged or non-optimal storage (91, 92). RNA can also be rapidly digested by RNase enzymes, ubiquitous in cells, tissues and resistant to standard enzyme inactivation methods (93, 94). RNA fragility forces careful isolation and handling

procedures and rapid sample freezing or RNA isolation effectively protects RNA from the undesired hydrolysis (93).

Principles governing nucleic acid detection

Detection of NA may seem like an easy task due to one fundamental and simple property of RNA and DNA: base pairing. NA polymer comprises four nucleotide bases: adenine (A), guanine (G), cytosine (C) and thymine (T) or uracil (U) in RNA. Each base is composed of three subunits: a nitrogenous base, pentose (ribose in RNA and deoxyribose in DNA) and a phosphate group. Thanks to the chemical properties of these DNA and RNA subunits, nucleotides can assemble via sterically precise hydrogen bonds into double stranded helices. Formation of bonds between corresponding bases is called base pairing, or more generally, hybridisation. Adenine forms a base pair with thymine and guanine with cytosine. DNA/RNA hybridisation via base pairing is one of the two key principles governing detection of NA. Genomic DNA exists as the double stranded helix and requires a lot of energy, in a form of heat, to break hydrogen bonds between the associated DNA strands. With each base removed from the DNA end, however, this breaking (melting) point will slightly decrease. By careful manipulation of a complementary probe length, a melting point of a probe/target duplex can be established to favour hybridisation of the desired duplex solely. Complementary NA probes, designed to base pair with a particular target fragment and labelled to facilitate visualisation, are successfully used in various DNA/RNA detection methods (described briefly later).

DNA replication and repair are meticulously controlled to ensure genomic continuity. The second principle that drives accurate detection of NA is utilization of inherent fidelity of DNA reactive enzymes. Enzyme-assisted methods proved to be more robust than passive, complementary probe base pairing (95) as distinct types of bonds have to form between enzyme and substrate to catalyse the reaction. Methods that acquired a status of universal techniques are based on one or combination of the following reactions: ligation, polymerisation and NA lysis. The ultimate NA sensing assay often combines two of aforementioned principles to minimize detection of undesired substrates. For example, linear oligonucleotides can be designed to form a circle-like duplex with DNA or RNA. Short, terminal target-complementary arms promote base pairing with a correctly matching substrate, while ligase, supplied in the reaction, seals a nick between hybridised arms. These probes, called “padlock probes” (96) are described later.

Methods to detect RNA

Various methods were developed to provide whole RNA or specific (targeted) RNA profiling *in vitro* and *in situ*. Below, I briefly describe techniques that received gold standard status and are used worldwide.

Hybridisation-based methods

The main advantage of methods that are based solely on hybridisation is single molecule sensitivity, provided that the event of hybridisation is efficiently detected. Complementary detection probes (decorators) are often conjugated with reporter chemicals, fluorescent or non-fluorescent. Fluorophore-conjugated probes allow for direct visualisation while non-fluorescents molecules can be targeted by secondary detection reagents yielding stronger detection response. Fluorescent hybridisation is a popular technique with a range of sub-techniques for *in situ* or *in vitro* applications. In traditional, fluorescent RNA *in situ* hybridization (FISH), mRNA molecule is "covered" with multiple fluorescently-labeled decorators that generate signal *in situ,* observed under a fluorescence microscope (97). To increase mRNA detection throughput, multiple mRNA-complementary DNA probes can be imprinted on a glass in microarrays format for massively parallel mRNA hybridization. Sample mRNAs are labelled prior the hybridization step and signal strength is proportional to RNA abundance allowing for quantitative mRNA expression measurements (98). Multiple mRNAs can also be detected simultaneously using RNA barcoding techniques *in situ* (99) and *in vitro* (100). The disadvantage of these methods is low throughput, dictated by the low brightness of hybridized probes that demands use of high-magnification objectives. To circumvent this limitation, the number of sites in mRNA molecule, to which decorators can hybridize, can be increased by formation of mRNA-dependent DNA scaffolds. Decorators are designed to target scaffolds instead, leading to a local concentration of fluorophores (101, 102), signal enhancement and thus, faster imaging (Figure 5).

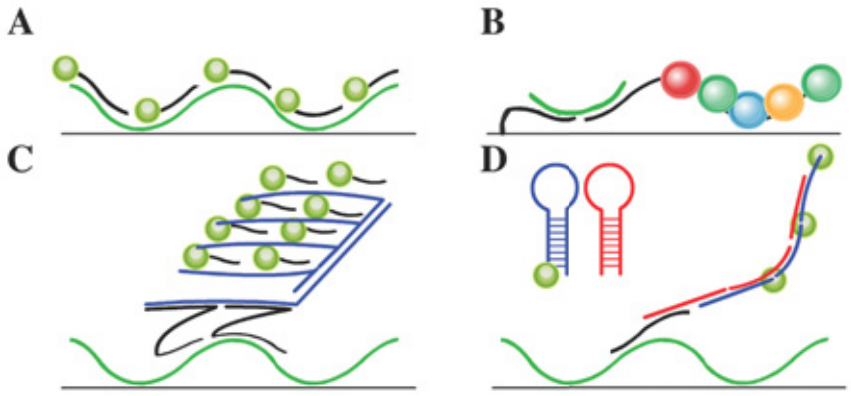

***Figure 5**. Different hybridisation-based RNA detection strategies. A: in FISH, series decorator probes (black), conjugated with a fluorophore (green ball) bind to the mRNA (green). B: In nCounter mRNA barcoding* (99)*, mRNA stabilizes a barcoded reporter on the solid surface. Order of fluorophores in the reporter identifies the transcript. C: in branched DNA (bDNA) FISH* (101) *signal is multiplied when decorators target a DNA scaffold (blue) made of DNA initiators (black). D: in hybridization chain reaction (HCR)* (102)*, presence of mRNA molecule and initiator hairpin (here shown in opened form, black) catalyzes chain opening of secondary hairpins (blue, red) one of which is fluorophore-conjugated.*

An additional benefit of hybridization-based signal amplification presented in the Figure 5C and D is targeting relatively shorter RNA fragments. This allows for more flexible target selection and for avoiding those likely to form intermolecular, secondary structures. bDNA FISH was successfully applied to detect mRNA splice variants, lncRNA (103) and can be potentially applied for snoRNAs, pre-miRNAs, and circular RNAs. To increase the affinity of decorators to mRNAs, various DNA analogs were synthesized. In a locked nucleic acid (LNA) 2'-O-4'-C methylene bridge "locks" the furanose ring of ribose sugar in C3' endo conformation (104). This gives LNAs a superior binding property in comparison to DNA. In artificial peptide nucleic acid (PNA), a sugar-phosphate backbone is replaced with N-(2-aminoethyl)-glycine units linked by peptide bonds (105–107). Neutral charge of PNA backbone increases its affinity to negative DNA/RNA. Increased affinity of nucleotide analogs was exploited for detecting short RNAs. *In situ* detection of miRNAs is facilitated using LNA-modified oligonucleotides with high efficiency (108–111). Like in traditional FISH, the number of miRNA targets that can be detected simultaneously is limited by optical limitations of microscope used in the experiment. The second disadvantage of hybridization-based methods is low sequence detection resolution. Since single base variations may not destabilize hybridizing probes completely, detection of single base variations remains a challenge. For more accurate NA sensing, enzyme-assisted methods have proven to be superior.

Enzyme-assisted methods

Due to limitations in RNA detection discussed earlier, very few methods can be robustly used for direct sensing. In this chapter I, will introduce the most prevalent methods based on RNA–DNA reverse transcription and highlight those, that can be used for direct RNA analyses.

Polymerase-based methods

PCR is a technique that revolutionised the field of molecular diagnostics. Simplicity, amplification fidelity, low price, and robustness made it not only a vital component of sample preparation steps in various techniques, but also a powerful diagnostic assay as such. In the most basic form of PCR, the temperature cycling regulates melting of dsDNA, annealing of primers, exponential amplification performed by thermostable polymerase (79), and the final products are visualised electrophoretically. If the desired outcome is the detection of mRNA, then reverse transcriptase (RT) is added prior to PCR to generate a cDNA copy of mRNA. In quantitative (qPCR), amplicon build-up can also be visualised in real-time, using intercalating (112, 113) or cleavable (114) fluorescent molecules and fluorescence detector. Such approach allows for multiplexed assessment of differences in basal RNA concentration and is one of the most prevalent techniques for targeted mRNA detection worldwide. By using short, hairpin-like RT primers, miRNAs can be targeted in qPCR (115) (Figure 6).

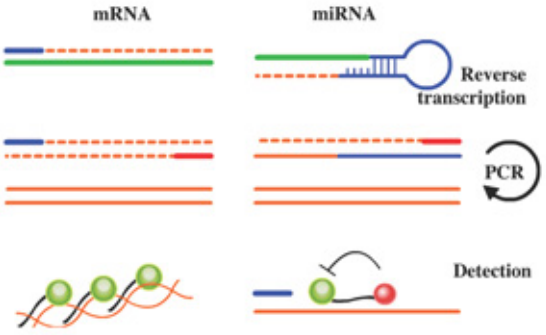

***Figure 6**. Detection of mRNA and miRNA with PCR. To detect mRNAs (left, section), RNA of interest (green) is reverse-transcribed to DNA using RT primer (blue). cDNA (orange) is copied with another primer (red) and amplification is repeated several times. For miRNAs, stem-loop RT primer is used in RT (right section). To detect amplified products, DNA intercalators (bottom left) or cleavable probes (bottom right) can be used. In the latter, fluorophore (green) is quenched (red), blocking background fluorescence. During amplification, primer is extended and polymerase cleaves a fluorophore.*

High temperature cycling, used in PCR, makes traditional amplification inappropriate for NA detection *in situ, in vivo* or in heat-sensitive applications. For this reason, a myriad of isothermal amplification techniques has emerged and were applied for DNA and RNA sensing. Loop-mediated isothermal amplification (LAMP) uses multiple sets of primers and a polymerase with high displacement and polymerisation activity (116). At constant temperature, series of consecutive strand displacements and replications result in exponential DNA amplification that can be detected using various downstream methods. For mRNA detection, RT can be added (117) or alternatively DNA polymerase with RT activity, can be applied (118). Unlike methods described above, nucleic acid sequence-based amplification (NASBA) and signal mediated amplification of RNA technology (SMART) utilize T4 RNA polymerase to generate RNA in presence of DNA or mRNA target. In NASBA, mRNA is reverse-transcribed and replicated to form dsDNA harbouring a T7 RNA polymerase promoter sequence. mRNAs transcribed from the promoter are again turned to dsDNA from and transcription continuous exponentially (119). In SMART, RNA amplification is initiated by formation of three-way-junction between mRNA/DNA and DNA probes. Extension of one probe creates a dsDNA containing a T4 RNA polymerase promoter, initiating transcription (120).

Sister strands in dsDNA can also be displaced using DNA untangling enzymes. In helicase-dependent amplification (HDA) (121) as well as in recombinase polymerase amplification (RPA) (122), DNA is enzymatically displaced to bring primers to the desired location. Amplified DNA is used as a substrate in another cycle and reaction proceeds exponentially.

Another method, that exploits a thermostable polymerase, but not its polymerase activity, is an Invader assay. 5'->3' exonucleolytic activity of eubacterial polymerases or flap endonucleases (FEN) was shown to be well orchestrated if proper structures between probe and target are formed (Figure 7) (123). This rigorous sequence and structure regime was exploited to create highly sensitive and specific assay for *in vitro* NA sensing, including SNPs in DNA (124), mRNA (125) and miRNAs (126).

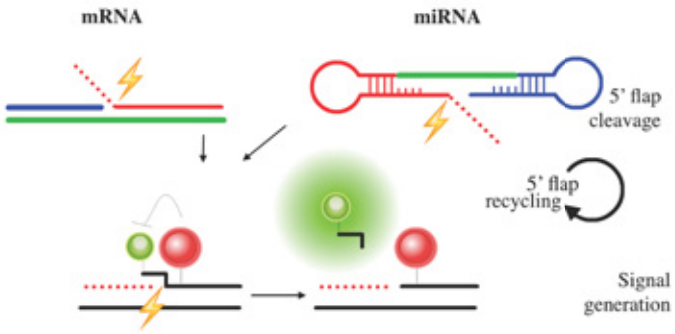

***Figure** 7. Scheme of the invader method for mRNA and miRNA detection. Detection probes (blue and red) bind the RNA target. To increase the stability of the construct, miRNA-specific probes form stem-loop hairpins and are enriched with additional chemical modifications* (126). *3' probe (blue) displaces a non-complementary 5' arm (flap), which is cleaved by a polymerase or endonuclease (marked as spark). Released 5' flaps participate in a secondary invasion cycle, displacing oligonucleotide with quenched fluorophore (green). After displacement, fluorophore is cleaved and fluorescence increases. Assay is performed at high temperature, facilitating turnover of both probes and cleaved 5' flaps, increasing assay sensitivity.*

Ligase-based methods

Another type of enzymes that are adapted for NA sensing are DNA ligases. The discrimination between different DNA strands, including SNPs was first reported in oligonucleotide ligation assay (OLA) (127). In OLA, two oligonucleotides are hybridising adjacently on the DNA target. If perfectly matched, probes are joined and a product can be detected using different techniques. For example, OLA was proposed in the format of a multiplexed, microtiter colorimetric plate (128) or in an assay where products are distinguished by different electrophoretic mobility, detected in the DNA-sequencers (129). Since RNA-templated DNA oligonucleotide ligation activity was reported for T4 DNA ligase (130), OLA was adapted for SNP detection on RNA (131).

In ligase chain reaction (LCR), ligation of annealed probes with a thermostable enzyme follows DNA denaturation. During temperature cycling, analogously to PCR, ligated products can serve as templates for new reaction cycles leading to exponential amplification of DNA (129, 132). Variants of LCR assays have been proposed, in which probes are directly ligated on mRNA (133) or miRNA (134) with T4 RNA ligase 2, providing templates for following amplification cycles.

In the 1990s, a bacteriophage Phi29 DNA polymerase was shown to accept circular templates as substrates for linear DNA amplification (132, 135). This process, known as rolling circle amplification (RCA), can be combined with padlock probes ligation for sensitive detection of NA (136). In the most basic concept, padlock probes (PLP) are linear oligonucleotides, where 3' and 5' termini are complementary to the target sequence and adjoined with a target non-complementary DNA linker (Figure 8,(96)). In the presence of the DNA/RNA analyte, PLP arms hybridise in a juxtaposed fashion forming a DNA circle which can be sealed into a continuous molecule. Due to the helical nature of dsDNA, hybridised PLP becomes catenated to the target. Ligation event is target dependent and highly specific, allowing for sensitive detection of SNPs (137, 138). Circularised PLPs can be used as templates for RCA, during which circles are replicated into long, tandem copies of the original probe template (Figure 8).

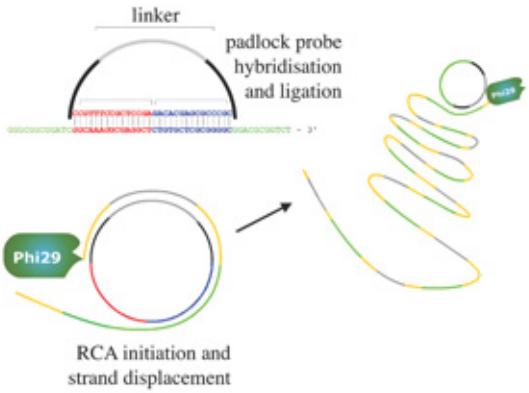

***Figure 8**. Nucleic acid detection with padlock probes and RCA.*

PLP technology can be integrated into many assay formats. Rolling circle products (RCPs) can be imaged and quantified *in vitro* by hybridising fluorophore-conjugated decorators or amplification can be monitored real-time with complementary molecular beacons (139) or intercalating dyes (140). Alternative methods to detect RCPs were demonstrated, such as decoration of RCPs with magnetic nanoparticles (141) or measurement of electrical conductivity of metal ion-coated DNA threads (142). Finally, since RCP comprises series of identical repeated sequences, enzymatic fragmentation yields monomers that can be characterised electrophoretically (143), participate in secondary ligation and RCA (144) or separated and detected on a paper strip (145). With various readout methods listed above, PLPs were used in the detection of Epstein-Barr virus (143), influenza (144), porcine circovirus and tuberculosis (141). Using conventional epi-fluorescent microscope, chromosomal and mitochondrial DNA (137, 148) as well as cDNA (149) and mRNA (150) can be visualised in cultured cells or tissues *in situ*.

Targets can be detected on sub-cellular level, allowing for quantitative and spatial analysis of expression. Finally, another advantage of PLP technology is the excellent scalability with a minimal loss of performance (141) or specificity (151). 55 000 probes targeting 10 000 human exons were amplified and sequenced *in vitro* (151). Using barcoded PLPs, as many as 31 cDNAs were targeted *in situ*. Individual DNA barcode in each RCP was decoded sequentially using sequencing-by-ligation chemistry (SBL) (152).

RNA-Sequencing

High-throughput sequencing revolutionised biology and medicine during the last two decades. Next Generation Sequencing (NGS) delivers the unprecedented amount of RNA expression data in biological samples. Importantly, sequence of each RNA molecule is reassembled during the process, facilitating discovery of SNPs, splice variants or novel genes. RNA-Seq comprises multiple steps, some of which were described in previous chapters. Even though direct sequencing of RNA has been demonstrated (153), vast majority of RNA-Seq experiments are conducted on DNA sequencing machines, circumventing RNA-associated limitations. Depending on the sequencing platform used or class of RNA to be analysed, preparation of cDNA library varies. Generally, mRNA transcripts are often enriched based on poly(A) tail presence (90), while shorter RNAs can be size-selected (154, 155). RNAs are reverse-transcribed and adapter sequences are added to cDNA ends for clonal amplification and sequencing (156). Interestingly, RCA was also used for clonal library amplification for whole genome sequencing by hybridisation (157) and whole genome SBL (156). The accuracy of the sequencing process is dictated by the fidelity of DNA polymerases (sequencing-by-synthesis) or ligases (sequencing-by-ligation) used in the reaction (159).

Present investigations

Paper I. Simultaneous single-cell *in situ* analysis of human adenovirus type 5 DNA and mRNA expression patterns in lytic and persistent infection

Background

Human adenovirus 5 (HAdV-5) is one of more than 60 species of adenovirus described so far (160). HAdV-5 efficiently infects epithelial cells, releasing viral progeny from lysed cells within few days after the infection (161). Additionally, HAdv-5 have been shown to establish long-term infections in lymphocytes residing in tonsil tissue (162). Infection leads to virus DNA proliferation as well as expression of mRNAs associated with virus replication cycle. Due to population averaging methods used for most DNA/RNA expression assessments, detailed cellular data was unavailable. Interestingly, HAdV-5 in long-term infections can be transcriptionally silent, thus single-cell studies would be superior and more informative to visualise cell-to-cell heterogeneity and to study gene expression and the progress of the infection.

Summary

In the paper I, we have developed a novel method allowing for simultaneous visualisation of HAdV-5 genomic DNA and spliced mRNAs on the subcellular level. After the desired mRNA targets were reverse transcribed to cDNA, viral DNA was rendered single-stranded by a combined endo- and exonucleolytic digestion. Padlock probes targeting viral RNAs, associated with early and late infection stages, were used in parallel with probes targeting the HAdV-5 genome. No cross-reactivity between RNA- and genome-specific padlock probes was observed due to the careful selection of mRNA targets on the spliced junctions. We have studied spatiotemporal changes in HAdV-5 infected HeLa cells and described highly heterogeneous populations of cells sampled at every time point studied. Interestingly, we have observed clear mRNA splice variants expression patterns, dependent on the virus replication rate. To study the long-term infection, we have infected human B-lymphocyte cell line (BJAB) with HAdV-5. 6 days post-infection cells were sampled and the protocol was applied. Cells were clustered using

non-supervised methods into 3 clusters, resembling expression patterns in cells during the lytic infection in HeLa cells. Interestingly, out of ~5 300 cells sampled, we identified 12 cells showing considerably lower mRNA expression but high HAdV-5 DNA content.

Discussion

An alternative method to visualise HAdV DNA and mRNA *in situ* has been proposed, but additional RNase or DNase digestion steps were required to separate genome-derived signal from RNA due to probe cross-reactivity (163). Our novel method allows for simultaneous, multiplexed detection of HAdV-5 ds/ssDNA and mRNAs with great specificity and no signal overlap. The method provided quantitative, single-cell data on viral replication and RNA expression during the lytic infection in HeLa cells. Though our data were concordant with traditional, whole-cell population methods, we were able to highlight discrete cell subpopulations with unique expression characteristics. For example, cells with considerably high virus amount but little viral-derived RNAs were detected. Those cells would have been missed if the experiment was performed using PCR, on cellular isolates. Finally, we have delivered evidence that overexpression of pVII, histone-like protein involved in HAdV-5 DNA packing, hinders accessibility to viral DNA, limiting sensitivity of hybridisation-based methods *in situ*. We hypothesise that this explains why detection of the virus at low copy number (initiation of infection) was not possible with our method. Technique is suitable for molecular characterisation of viral infection, not restricted to HAdV-5. As no DNase, RNase or proteinase procedures are required, cellular morphology is preserved and specimen could be additionally stained for markers of host immune response (164).

Paper II. Fidelity of RNA templated end-joining by Chlorella virus DNA ligase and a novel iLock assay with improved direct RNA detection accuracy

Background

Most of the enzymatic methods for NA sensing are based on DNA. If RNA detection is desired, cDNA copy of the original molecule is typically formed as a substrate for high-fidelity enzymes. Similarly, in high-resolution ligation-based methods, RNA molecules are converted prior detection (149, 165). However, reverse-transcription is prone to bias and was shown to introduce artefacts that can interfere with accurate expression assessment

(153). Ligase-based assays using RNA as substrate/ligation template could provide a low-cost and bias-free alternative. In recent years, efficient RNA-templated DNA end joining was demonstrated for Paramecium bursaria Chlorella virus 1 DNA ligase (PBCV-1)(166), but little was known with respect to ligation fidelity. In Paper II, we systematically characterised fidelity of the PBCV-1 DNA ligase on RNA, mRNA and miRNA. Since ligation activity on RNA was not dependent on accurate probe-target base pairing, we proposed RNA sensing approach based on novel circularizable probes, iLocks.

Summary

We have evaluated RNA splinted PBCV-1 DNA ligase DNA nick sealing fidelity on 3' and 5' probe termini. Due to high mismatch tolerance, PBCV-1 DNA ligase could not be used for single nucleotide variants detection or discrimination of highly similar miRNA sequences. To circumvent this drawback, we adapted invasive cleavage step prior to ligation (167). Hybridisation of an iLock (padlock probe with a 5' flap) is verified by the polymerase and non-complementary extension is removed only upon correct base pairing. Following the cleavage, probe arms are ligated by the PBCV-1 DNA ligase. We observed substantial improvement (up to nearly 100% for almost all tested bases) of ligation specificity on RNA with iLock probes. We tested our method to accurately differentiate single base variants in mRNA and closely related *let-7* miRNA family members. Finally, we showed that invasive cleavage-ligation combination can be integrated into traditional assay formats for detecting the low copy numbers of RNA.

Discussion

Considering intrinsic polymerisation biases and inefficiency of the reverse-transcription step, common in most RNA detection assays, direct RNA detection is desired for accurate RNA profiling. However, traditional enzymes, commonly used in nucleic acid (NA) sensing do not perform well on RNA. PBCV-1 DNA ligase has proved to be highly error-prone when DNA oligonucleotides are ligated on RNA. We have integrated a principle of the invader assay into a padlock probe, to make use of thermostable *Taq* DNA polymerase's ability to recognize and act on sequence-specific probe/target structures (167). Specificity of the RNA sensing in the iLock assay is dictated by two independent sequence recognition events. During the first event, *Taq* DNA polymerase reads the DNA/RNA heteroduplex and cleaves the protruding 5' extension only if iLock probe is matching the target. 5' extension itself as well as lack of the 5' phosphate group, effectively prevents probes from non-templated ligation. During the second event, following the extension cleavage, ligase seals the nick, thus overall method specificity is medi-

ated by fidelities of both enzymes combined. Considerable improvement of RNA sensing specificity as well capability of integration in alternative assay formats makes our method a valuable addition to a NA sensing toolkit. By using thermostable ligases, cleavage and ligation reaction could be potentially performed in a single step, thus shortening the protocol time. We have also observed higher efficiency of iLock probe activation and ligation for longer RNA targets when compared to miRNAs. This suggests that method can be further optimised to provide sensitive and specific sensing of short RNA targets.

Paper III. Detection of miRNAs using chimeric DNA/RNA iLock probes utilizing novel activity of PBCV-1 DNA ligase: RNA-templated ligation of ssRNA

Background

As explained in several paragraphs throughout this thesis, miRNAs, despite their short size, are extremely important regulatory RNAs that are commonly used as biomarkers in pathological conditions. Their detection is technically challenging and often requires specialized equipment or expensive reagents. In the paper II, we have presented and systematically studied a novel RNA detection assay using DNA iLock probes, undergoing intermolecular trimming in response to perfect base paring with a target (168). We have demonstrated specific detection of different RNA classes, including mRNA and miRNA. Though assay showed excellent, single-base resolution specificity, the overall end-point efficiency for short RNA targets was lower comparing to longer RNAs. We hypothesized that miRNA detection sensitivity could be increased if efficiency of catalytic reactions is improved or when probe/target binding increased. Higher ligation activities on RNA were observed for T4 RNA ligase 2 (T4Rnl2) provided that terminal base(s) of the 3'-OH acceptor substrate were RNA (169). Motivated by this principle, we decided to investigate, if ligation of chimeric substrates is possible for PBCV-1 ligase. As mentioned earlier, prior ligation, iLock probes need to become enzymatically cleaved (activated). Invasive cleavage on RNA has only been observed for handful of enzymes, and has been relatively less studied on RNA. Because displaced and cleaved 5' extension is natively RNA (replication of lagging strands in *Taq* is initiated from RNA primer) we hypothesized that endonucleolytic activity of *Taq* DNA polymerase could also be improved for RNA containing iLock probes.

Summary

We characterized, previously unreported, RNA template-dependent 3'-RNA/5'-DNA hybrid nick sealing efficiency and fidelity for PBCV-1 DNA ligase and T4Rnl2 (such activity was known for T4Rnl2, but not properly characterized) for detecting RNA single nucleotide polymorphisms and miRNAs. We have observed considerable ligation efficiency improvement for RNA containing padlock probes on short RNA targets. Ligation activity was structure- but not sequence-specific for PBCV-1 DNA ligase. Higher enzymatic efficiency, despite poor fidelity, motivated us to investigate if miRNA detection using iLock probe RNA detection assay could also be improved. We have studied various RNA substitutions in iLock probes, how they affect probe activation and ligation. Our experiments highlighted two particular positions in the iLock probe that enhance probe activation and ligation on miRNA if substituted with RNAs (single substitution at the 3' end and a single substitution within a 5' flap). Using chimeric iLock probes, we performed multiplexed differentiation of conserved *let-7* miRNA family members with next-generation sequencing-by-ligation readout.

Discussion

This study introduces multiple novelties. To begin with, RNA-templated chimeric nick sealing activity of PBCV-1 DNA ligase has not been shown in the literature. Our experiments delivered evidence that the ligase readily ligates RNA/DNA nicks, provided that RNA modification is restricted to the 3' terminus. Considerable ligation inhibition (when RNA was localised to the 5' end) combined with high mismatch ligation tolerance, suggested that ligase activity is not sequence-specific, but rather dependent on a B-form helix conformation at the substrate 5' end. Another novelty was observation that amplification of RNA enriched padlock probes is in fact possible. This suggested that Phi29 DNA polymerase displays reverse transcriptase activity. Phi29 RT is more broadly studied in paper IV. In this work, we have investigated how RNA substitutions affect functionality of iLock probes, superior over traditional padlock probes for specific RNA sensing. We have defined two concrete nucleotide positions in the iLock probe that significantly increased probe efficiency on miRNAs. Our experiments showed that chimeric iLock probes can be used for multiplexed *in vitro* miRNA profiling using sequencing-by-ligation readout. Having in mind that padlock probes are commonly used for nucleic acid visualisation *in situ,* we believe that iLock assay can be optimised for multiplexed detection of miRNAs in cultured cells or tissue specimen.

Paper IV. Reverse transcriptase activity of Phi29 DNA polymerase

Background

To increase interaction strength between hybridising oligonucleotides, various nucleotide modifications or nucleotide analogues are used (126, 170). Elevated catalytic activities have been observed for certain ligases provided that terminal 3'-OH base(s) are RNA (169). In the paper III, we have introduced RNA substitutions in otherwise DNA iLock probes, which resulted in substantially increased miRNA detection efficiency using the iLock RNA detection assay. Compatibility of the RNA/DNA chimeric probes with Phi29 DNA polymerase, implied that the polymerase accepts RNA-containing circular substrates as rolling circle amplification (RCA) templates. At certain reaction conditions, reverse transcriptase activity can be enabled for *Taq* or *Bst* DNA polymerase I Large Fragment (171). For Phi29 DNA polymerase however, no such activity has been reported, and only limited literature exists describing the effect of nucleotide substitutions on the rate of RCA (172). In paper IV, we study the mechanism through which Phi29 DNA polymerase amplifies RNA-containing circular templates. We compare the effect of multiple RNA substations on RCA rate, amplicons size and fluorescence intensity.

Summary

Despite the common notion that the Phi29 DNA polymerase cannot reverse-transcribe RNA, we have observed that the enzyme readily accepts RNA-containing circular substrates for RCA. This observation motivated us to study the effect of various types and number of RNA substitutions on the rate of Phi29 DNA polymerase replication. By using SybrGold labelling and real-time PCR fluorescence build-up, we were able to monitor amplification rate of individual RNA-containing circles simultaneously, throughout the whole amplification process. We have observed that substrates with single RNA substitutions were amplified with similar rates as non-chimeric DNA circles. For 2 or 3 consecutive RNA substitutions, RNA pyrimidines were tolerated during RCA, but not purines. Inhibited RCA, observed for circles containing longer RNA stretches, was partially recovered by interspacing RNA bases with DNA. To identify the mechanism by which the Phi29 DNA polymerase replicates chimeric substrates, circles containing single and double, consecutive RNA substitutions were amplified and RCA products monomerised and sequenced using NGS system. We demonstrate that Phi29 DNA polymerase incorporated expected dNTPs for the corresponding RNA base with very high accuracy.

Discussion

In this work, we demonstrated that Phi29 DNA polymerase copies RNA containing circles via reverse transcription. According to our experiments, single RNA substitutions exerted no effect on the rate of RCA. Fidelity of RCA for circles containing a single RNA base was indifferent from those of pure DNA. Incorporation error rate for RNA positions was based on the analysis of full-length sequenced RCA monomers, extracted from the sequencing dataset. Interestingly, substrates with 2 and 3 consecutive RNA bases demonstrated both higher incorporation error as well as inhibited replication rate as measured in the real-time RCA experiments. Since for circles containing consecutive purines only a small fraction of sequencing reads was full length, this could explain their poor performance as substrates in RCA. Discovery of reverse transcriptase activity of Phi29 DNA polymerase extends range of applications of Phi29-mediated assays. This include usage of ligatable RNA/DNA chimeric oligonucleotides, exhibiting higher ligation efficiency on RNA targets (padlock probes or molecular inversion probes), as exemplified in Paper III. Increased ligation efficiency of chimeric probes combined with RCA offers an attractive alternative for direct RNA sensing assays.

Concluding remarks

Since the inception, the Central Dogma has been frequently revised but still provides a backbone for the flow of information in cells. As the history of molecular biology has been enriched with numerous findings, the picture of the Dogma has broadened and become more complex (Figure 9).

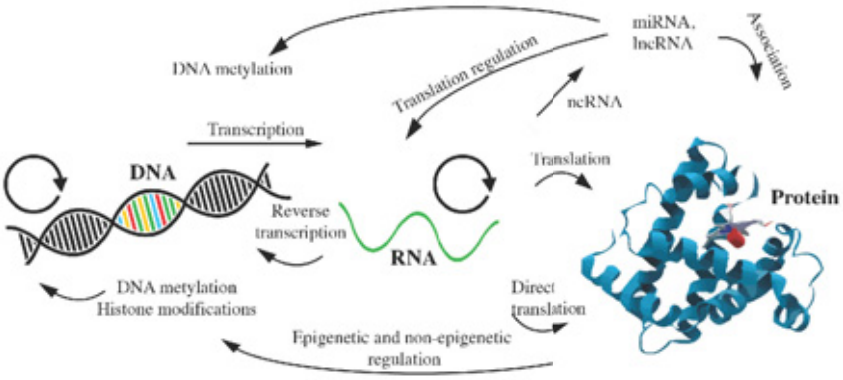

***Figure 9**. Revisited picture of the Central Dogma. Some of the RNA players described in this thesis interact in a more complicated manner than initially presumed. Circular arrows represent DNA/RNA replication*

Importantly, we slowly begin to realize the significance of RNA in regulating gene expression, chromosome architecture or cellular metabolism, among other roles. As technologies advance and costs to perform RNA sensing experiments decline, we can expect that tremendous amount of RNA data will emerge, outpacing the rate of its analysis. As a consequence, new associations between RNA states or levels and diseases will be discovered. Toolkit for RNA sensing continues to expand, and aim of this thesis is to contribute to this expansion. Padlock probes, used as a common RNA sensor in all my experiments, are very advantageous as they allow for highly specific nucleic acid recognition and facilitate efficient detection. Their convenient and simple integration into various readout formats allows for RNA detection *in vitro* as well as *in situ*, providing spatial information of RNA expression in single cells.

In **Paper I**, we have proposed a novel experimental procedure that allows for concordant visualisation of mRNAs and episomal DNA in adenovirus-5

infected human cells. Due to the fact that mRNA sequence is conserved in viral DNA, differentiation the two is challenging. Despite this hurdle, no off-target signal was generated from probe cross-hybridisation. We have applied this technique to study spatial and temporal changes in the virus spread as well as expression of RNAs derived from the virus and host cells. Importantly, we have showed that the method is applicable in persistent adenovirus infections, more relevant from the clinical point of view. By using computational cell grouping, we have clustered cells according to the infection stage and we were able to define rare cells, possibly marking an onset of persistency in the adenovirus infection cycle. The method presented in Paper I was a combination of procedures allowing for detection of mRNA and double stranded DNA, individually. mRNA detection procedure is relatively standardized and robust. However, it includes RNA to DNA conversion which I described in the previous chapters, that limits overall RNA detection sensitivity (149). In **Paper II**, I have aimed to develop a novel tool that would allow for a direct detection of RNA. Such an approach has a potential to liberate RNA sensing from reverse-transcription bias and thus facilitate more precise RNA quantitation. I have found that RNA-dependent ligases exhibit high tolerance for sealing padlock probes, even if base-paired unspecifically with RNA. To circumvent this limitation, I have exploited the intrinsic fidelity of eubacterial polymerase in structure-dependent probe/RNA target recognition. When the polymerase was used prior ligation, probes were trimmed in the desired position. Only then probes could have been ligated. In Paper II, I have characterised mechanisms governing removal of the extension and its effect on ligation specificity. This novel type of probes, named iLocks, became a primary interest of my other projects. iLocks showed an excellent RNA sensing fidelity on synthetic RNA targets, mRNAs and miRNAs compared to traditional padlock probes. However, because of the short size of miRNAs and possibly, the instability of hybridising probes, iLock assay for miRNAs showed a limited efficiency. In **Paper III** we have tested if substitutions of certain DNA nucleotides in padlock and iLock probes with RNA can stimulate catalytic activities of ligases and the eubacterial polymerase. In line with the hypothesis that RNA nucleotides could influence base-pairing or enzyme activities, we have observed improved detection of miRNAs using chimeric probes. Moreover, we have identified two particular positions in the iLock probe sequence that, when substituted with RNA, increased the rate of probe activation. Using the modified design, we proposed a novel iLock miRNA detection assay, allowing for multiplexed detection of miRNAs *in vitro*. It is worth to point out that alternative methods exist that allow for miRNA detection *in vitro,* some of which I have covered in the previous chapters. Those methods are very robust but suffer from a low multiplexibility, high costs, demand for advanced equipment. Unprecedented multiplexing capabilities of padlock probes com-

bined with their universal applicability may establish their position as great tools for miRNA visualisation *in situ*.

Lastly, successful application of chimeric probes in Paper III is attributed to the ability of the Phi29 DNA polymerase to replicate RNA-containing circles. Such activity, when polymerase can use RNA as a template for DNA synthesis, is called reverse transcription and was not known to exist in Phi29 DNA polymerase. We reported this activity in **Paper IV** and study mechanisms how various types, numbers and patterns of RNA substitutions affect Phi29 amplification rate. We delivered evidence that Phi29 polymerase accurately copies RNA bases within a circular substrate via reverse transcription, exhibiting preference towards RNA pyrimidines.

This thesis has contributed to the understanding of mechanisms relevant for RNA detection with padlock probes. I have characterised mechanisms of action for RNA reactive enzymes, new probe constructs as well as effects of nucleotide analogues substitution in padlock probes on their activity on RNA. Considering significant potential of RNA as biomarker and that direct detection of RNA is not prevalent, I hope this work will stimulate and guide further research in this fascinating area.

Populärvetenskapliga sammanfattningen på svenska

Människans kropp består av 30–70 biljoner celler. Alla celler bär identiskt genetiskt material (DNA) men de kan ha olika roller i kroppen, från celler i en näthinna till celler i en tånagel. Denna funktionella mångfald styrs av olika faktorer genom cellens liv och återspeglas på olika nivåer. Medan proteiner ofta anses vara byggstenar som bildar cellens kropp, har vi börjat uppfatta mekanismer som dikterar när, var och vilka proteiner som uttrycks i celler. Även om antalet proteinkodande gener i vårt DNA ligger mellan 20 000–25 000 är antalet proteiner mycket större. För att skapa ett funktionellt protein läses en blåkopia från dess gen-DNA och sparas noggrant i en RNA molekyl. RNA är mer reaktivt och evolutionärt äldre än DNA och det brukade förmodligen vara ansvarigt för lagring av genetisk information. Idag lagras vår genetiska information i DNA men det behöver fortfarande läsas av till en RNA molekyl för att skapa ett protein och den största cellmångfalden dikteras av processer som äger rum efter att RNA har skapats. Vi identifierde tidigt den centrala rollen av RNA som bärare av proteinmeddelanden och som en aktiv komponent i proteinbildningen. Forskning under de senaste två decennierna har kopplat RNA till många andra aktiva funktioner som tidigare var reserverade för proteiner, inklusive katalys av kemiska reaktioner, reglering av läsning av proteinmeddelanden eller skydd mot främmande DNA. RNA-nivån förändras under olika fysiologiska och patologiska förhållanden, och det finns en stor potential i RNA att vara en bättre markör att förutsäga en persons tillstånd.

I det här arbetet tillämpade jag "hänglåsprober" (padlock probes) för att detektera RNA. Traditionellt detekteras RNA indirekt med hjälp av hänglåsprober, vilket kan påverka resultatet. Således finns det stort utrymme för förbättring eftersom direkt RNA-detektion inte bara kan bestämma RNA-mängden i celler bättre, utan även förkorta analystiden. Denna avhandling presenterar alternativa metoder för att detektera RNA och exempel från mitt arbete visar hur RNA kan visualiseras *in situ*–inom celler–och *in vitro*–utanför celler. Dessutom har jag utvecklat en uppdaterad version av en hänglåsprobe, kallad **iLock**-probe. Även om iLocks liknar hänglåsprober, fungerar de på RNA direkt och måste aktiveras på RNA för att tillåta detektering. Jag visar hur iLocker kan användas för att upptäcka olika typer av RNA, vilket kan vara viktigt för klinisk tillämpning och forskningsapplikationer.

Acknowledgement

It was a long journey but feels like it was just a brief moment. Work presented in this thesis as well as my degree is an effect of a combined exertion of many people, not only scientists. To begin with, I wish to thank my research mentor who was navigating me through many, many, many! scientific challenges. Professor **Mats Nilsson** is a person of uncountable ideas and endless optimism. Stimulating environment you create brings science and fiction together, and working on this edge was the best part of my PhD. Thank you for giving me free hand to express all my ideas and never giving up on "you know which one project". We will get there! ☺

I wish to thank my opponent, Docent **Weng-Onn Lui** and my PhD committee, Professor **Ann-Christine Syvänen,** Professor **Pia Ädelroth** and Professor **Joakim Lundberg** for accepting to participate in my defence.

I wish to leave a line of appreciation for a **Professor Victor Ambros**. Thank you for allowing me doing research at your lab as well as everything you have taught me. You are an amazing person and exceptionally inspirational scientist. I spent an amazing time in your lab and I will never forget how kind you were to me during harsh early weeks.

Next, I would like to highlight my gratitude towards many collaborators, Docent **Mark Friedländer**, Professor **Edvard Smith** and many other people who helped me on various stages of various projects. I am very grateful to Docent **Tanel Punga** from the Department of Medical Biochemistry and Microbiology at Uppsala University. Thank you for your professional approach to our project and always believing in accomplishing our goals. My special thanks are extended to the staff of **Olink** and **Qlinea** companies for the partnership across variety occasions.

I am particularly grateful to **Jessica Östlind** for being my MolDia collaborator across so many experiments and projects. Thank you for kind support in my Swedish assignments, never saying no when I needed you help in anything and for being a good neighbour in the office for such long time.

I wish to extend my thanks to the MolDia lab. You are all an exceptional group of people. I learned a lot from many of you and you made me a better

person. Thank you, **Eva**, for your smiles, laughter, Metaxas, cat-sitting, craziness, all our uneasy discussions and of course, your Sisyphean endeavours to keep the lab in a good shape. I was sceptical. I was wrong ☺ Good job! A very big bow to **Xiaoyan**–our lab cyborg. Where should I start? I wish we have worked more together. I cannot understand how one can squeeze so much ambition, diligence and coolness in such tiny (relatively) person ☺ Thank you for bringing all this positive energy and fikas to the lab and spreading it around. I am particularly grateful to **Malte,** for keeping high-level discussions during meetings, pushing good ideas, everything we have accomplished together and lastly, for being a great and welcoming flatmate for my friends and family in Uppsala ☺ Special thanks to **Elin** for being a fantastic mRNA comrade and many hours of biochemical philosophy. **Di**, it was always a great pleasure discussing science with you and I hope there will be many to come! I wish to thank **Pavan** for his inspirational vision how to use our technology in a practical and meaningful way. Thank you, **Thomas**, for setting up our imaging system that at the end helped us getting all our data. Did you go to Singapore at the end☺? I would like to leave a line of gratitude to my former co-supervisor, **Marco**. Thanks for teaching me our techniques. I tried to look up to your high academic standard throughout all my years as PhD student. At the same time, thank you **Anja** for your assistance when I shifted towards *in vitro* research. I wish to acknowledge **Amal** for being my close friend and all warm moments we spent together. It was a great honour to work with **Annika Ahlford**. Thank you, Annika, for listening to me, being there for me, for all good advices, smiles and kindness.

Great THANK YOU to the rest of present and former members of the MolDia gang, **Ivan, Sibel, Navya, Markus, Felix, Carina, Chenglin Chika**, **Anna Engström**, **Tagrid**, **Rongqin**, **David**, **Lotte** and **Tim**. You are great guys and thanks for making my PhD time so special. I also wish to acknowledge all students in the lab, both present and former, **Anne, Susana, Nikola**, **Tanvir**, **Didrik**, **Yonatan** and **Sharif** for complementing the lab with your curiosity-driven mind-set, ambitious perspectives, passion and being always ready for lab-unrelated activities!

My special thanks are extended to **Nara**. I admire your professionalisms and limitless knowledge. Thank you, my friend, for helping me with proof-reading my manuscripts and watching my back. I will never forget that.

My sincere thanks to other people in the Gamma and Alpha house who made this place special every day. Thanks, **Akis**, **Alex**, **Sergey**, **Lara**, **Cecilia**, **Anatoly** and **Ida Hafstrand**.

I also wish to acknowledge my friends across the ocean, members of Victor Ambros' lab. Thank you so much **Sungwook**, **Oscar**, **Charles**,

Alejandro, **Isana** and **Katie** for your hospitality, for introducing me to the worm world and all discussions we had at the lunch table. **Orkan**, special thanks to you my dear friend. Hikes, games, beers, car lifts, jokes… but especially for the pool access☺ Şerefe! **Cat**, hahahha. My crazy faluche-wearing friend. Thank you for turning my stay into unforgettable experience. It was some ride! Thank you for all the moments and stories ☺

I am very grateful to whole DBB staff, especially Professor **Stefan Nordlund**, Professor **Pia Älderoth**, Professor **Peter Brzezinski**, **Pia Harryson, Agneta Norén** and **Maria Sallander** for my PhD management, assistance, teaching coordination and for the support with daily administration.

Last but not least, it gives me honest pleasure to acknowledge my friends and family. Bro-hug to **Jakub Lewicki** and fist bump to **Jakub Rybiński** for my stress management Thursdays ☺ and being the best gaming buddies one can ever imagine. Keep the K/D high! I am lucky to be your friend. My wife, **Magda**…☺ I dedicate this thesis to you.

Finally, I wish to thank my family for supporting me from the very beginning till today. Dziękuje mamo i tato za wspieranie moich decyzji. To dzięki Wam jestem tu gdzie jestem i osiągnąłem to, co osiągnąłem.

References

1. Crick,F. (1958) On Protein Synthesis. *Symp. Soc. Exp. Biol.*, **12**, 138–163.
2. Watson,J.D. (1965) Molecular Biology of the Gene. Benjamin,W.A. (ed).
3. Temin,H.M. and Mizutani,S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226**, 1211–1213.
4. Baltimore,D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226**, 1209–1211.
5. McClintock,B. (1953) Induction of instability at selected loci in maize. *Genetics*, **38**, 579–99.
6. Cattanco,R. (1991) Different types of messenger RNA editing. *Annu. Rev. Genet.*, **25**, 71–88.
7. Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) The MURF3 gene of T. brucei contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. *Cell*, **62**, 901–911.
8. Reed B. Wickner, Herman K. Edskes,F.S. and T.N. (2007) Prions of fungi: inherited structures and biological roles. *Nat. Rev. Microbiol.*, **5**, 611–618.
9. Wickner,R.B. (2012) Discovering protein-based inheritance through yeast genetics. *J. Biol. Chem.*, **287**, 14432–14442.
10. Consortium,I.H.G.S. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
11. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.
12. Consortium,T.E.P. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
13. Elliott,D. and Ladomery,M. (2011) Molecular Biology of RNA Oxford University Press.
14. Nissen,P., Hansen,J., Ban,N., Moore,P.B. and Steitz,T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–30.
15. Moore,P.B. and Steitz,T.A. (2003) After the ribosome structures: how does peptidyl transferase work? *RNA*, **9**, 155–159.
16. Noller,H.F., Hoffarth,V. and Zimniak,L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, **256**, 1416–1419.
17. Breaker,R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.*, **4**, 1–15.

18. Jansen,R., Van Embden,J.D.A., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.
19. Sander,J.D. and Joung,J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347–55.
20. Schmeing,T.M. and Ramakrishnan,V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–1242.
21. Thore,S., Mayer,C., Sauter,C., Weeks,S. and Suck,D. (2003) Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA: Common features of RNA binding in archaea and eukarya. *J. Biol. Chem.*, **278**, 1239–1247.
22. Guthrie,C. and Patterson,B. (1988) Spliceosomal snRNAs. *Annu. Rev. Genet.*, **22**, 387–419.
23. Mole,L.D.B. and Sabatier,P. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.
24. Tollervey,D. and Kiss,T. (1997) Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.*, **9**, 337–342.
25. Pannucci,J.A., Haas,E.S., Hall,T.A., Harris,J.K. and Brown,J.W. (1999) RNase P RNAs from some Archaea are catalytically active. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 7803–8.
26. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
27. Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–54.
28. Pasquinelli, a E., Reinhart,B.J., Slack,F., Martindale,M.Q., Kuroda,M.I., Maller,B., Hayward,D.C., Ball,E.E., Degnan,B., Müller,P., *et al.* (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, **408**, 86–89.
29. Ruvkun,G. (2001) Glimpses of a tiny RNA world. *Science*, **294**, 797–9.
30. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806–811.
31. Brennecke,J., Aravin,A.A., Stark,A., Dus,M., Kellis,M., Sachidanandam,R. and Hannon,G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, **128**, 1089–1103.
32. Choudhuri,S. (2010) Small noncoding RNAs: Biogenesis, function, and emerging significance in toxicology. *J. Biochem. Mol. Toxicol.*, **24**, 195–216.
33. Burnett,J. and Rossi,J. (2012) RNA-Based Therapeutics: Current Progress and Future Prospects. *Chem. Biol.*, **19**, 60–71.
34. Brantl,S. (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.*, **10**, 102–109.

35. Brantl,S. (2002) Antisense-RNA regulation and RNA interference. *Biochim Biophys Acta*, **1575**, 15–25.
36. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli, a E., Bettinger,J.C., Rougvie, a E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–906.
37. Behm-ansmant,I., Rehwinkel,J., Doerks,T., Stark,A., Bork,P. and Izaurralde,E. (2006) mRNA degradation by miRNAs and GW182 requires both CCR4 : NOT deadenylase and DCP1 : DCP2 decapping complexes. *Genes Dev.*, **20**, 1885–1898.
38. Wu,L., Fan,J. and Belasco,J.G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 4034–4039.
39. Kiriakidou,M., Tan,G.S., Lamprinaki,S., De Planell-Saguer,M., Nelson,P.T. and Mourelatos,Z. (2007) An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell*, **129**, 1141–1151.
40. Petersen,C.P., Bordeleau,M.-E., Pelletier,J. and Sharp,P. a (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell*, **21**, 533–542.
41. Olsen,P.H. and Ambros,V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.*, **216**, 671–680.
42. Place,R.F., Li,L.-C., Pookot,D., Noonan,E.J. and Dahiya,R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 1608–1613.
43. Filipowicz,W., Bhattacharyya,S.N. and Sonenberg,N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
44. Ullu,E., Tschudi,C. and Chakraborty,T. (2004) RNA interference in protozoan parasites. *Cell. Microbiol.*, **6**, 509–519.
45. Okamura,K. and Lai,E.C. (2008) Endogenous small interfering RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **9**, 673–678.
46. Piatek,M.J. and Werner,A. (2014) Endogenous siRNAs: regulators of internal affairs. *Biochem. Soc. Trans.*, **42**, 1174–1179.
47. Kole,R., Krainer,A.R. and Altman,S. (2012) RNA therapeutics: Beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.*, **11**, 125–140.
48. Quinn,J.J. and Chang,H.Y. (2015) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.*, **17**, 47–62.
49. Flynn,R.A. and Chang,H.Y. (2014) Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell*, **14**, 752–761.
50. Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F., *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.

51. Wang,K.C. and Chang,H.Y. (2011) Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell*, **43**, 904–914.
52. Muller, Hermann,J. (1927) Artificial transmutation of the gene. *Science*, **66**, 84–88.
53. Chang,J.C. and Kan,Y.W. (1979) Beta 0 thalassemia, a nonsense mutation in man. *Proc. Natl. Acad. Sci. U. S. A.*, **76**, 2886–2889.
54. Orkin,S.H., Alter,B.P., Altay,C., Mahoney,M.J., Lazarus,H., Hobbins,J.C. and Nathan,D.G. (1977) Application of endonuclease mapping to the analysis and prenatal diagnosis of thalassemias caused by globin-gene deletion. *N. Engl. J. Med.*, **299**, 166–172.
55. Kryukov,G. V, Pennacchio,L. a and Sunyaev,S.R. (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.*, **80**, 727–739.
56. Cooper,D.N., Chen,J.M., Ball,E. V., Howells,K., Mort,M., Phillips,A.D., Chuzhanova,N., Krawczak,M., Kehrer-Sawatzki,H. and Stenson,P.D. (2010) Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.*, **31**, 631–655.
57. Li,G., Pan,T., Guo,D. and Li,L.-C. (2014) Regulatory variants and disease: the E-cadherin -160C/A SNP as an example. *Mol. Biol. Int.*, **2014**, 1–9.
58. Sethupathy,P. and Collins,F.S. (2008) MicroRNA target site polymorphisms and human disease. *Trends Genet.*, **24**, 489–497.
59. Laganà,A., Russo,F., Sismeiro,C., Giugno,R., Pulvirenti,A. and Ferro,A. (2010) Variability in the incidence of miRNAs and genes in fragile sites and the role of repeats and CpG islands in the distribution of genetic material. *PLoS One*, **5**, 1–8.
60. Calin,G.A., Sevignani,C., Dumitru,C.D., Hyslop,T., Noch,E., Yendamuri,S., Shimizu,M., Rattan,S., Bullrich,F., Negrini,M., *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 2999–3004.
61. Shen,J., Ambrosone,C.B. and Zhao,H. (2009) Novel genetic variants in microRNA genes and familial breast cancer. *Int J Cancer*, **124**, 1178–1182.
62. Shen,J., DiCioccio,R., Odunsi,K., Lele,S.B. and Zhao,H. (2010) Novel genetic variants in miR-191 gene and familial ovarian cancer. *BMC Cancer*, **10**, 47.
63. Mullany,L.E., Herrick,J.S., Wolff,R.K. and Slattery,M.L. (2017) Single nucleotide polymorphisms eithin microRNAs, microRNA targets, and microRNA biogenesis genes and their impact on colorectal cancer survival. *Genes. Chromosomes Cancer*, **56**, 285–295.
64. Sun,G., Yan,J., Noltner,K., Feng,J., Li,H., Sarkis,D.A., Sommer,S.S. and Rossi,J.J. (2009) SNPs in human miRNA genes affect biogenesis and function. *RNA*, **15**, 1640–1651.
65. Li,H., Xie,H., Liu,W., Hu,R., Huang,B., Tan,Y.F., Liao,E.Y., Xu,K., Sheng,Z.F., Zhou,H. De, *et al.* (2009) A novel microRNA targeting HDAC5 regulates osteoblast differentiation in mice and contributes to

primary osteoporosis in humans. *J. Clin. Invest.*, **119**, 3666–3677.
66. Cai,T., Li,J., An,X., Yan,N., Li,D., Jiang,Y., Wang,W., Shi,L., Qin,Q., Song,R., *et al.* (2017) Polymorphisms in MIR499A and MIR125A gene are associated with autoimmune thyroid diseases. *Mol. Cell. Endocrinol.*, **440**, 106–115.
67. Eser,S., Schnieke,A., Schneider,G. and Saur,D. (2014) Oncogenic KRAS signalling in pancreatic cancer. *Br. J. Cancer*, **111**, 817–822.
68. Xia,Y., Zhu,Y., Zhou,X. and Chen,Y. (2014) Low expression of let-7 predicts poor prognosis in patients with multiple cancers: A meta-analysis. *Tumor Biol.*, **35**, 5143–5148.
69. Kjersem,J.B., Ikdahl,T., Guren,T., Skovlund,E., Sorbye,H., Hamfjord,J., Pfeiffer,P., Glimelius,B., Kersten,C., Solvang,H., *et al.* (2012) Let-7 miRNA-binding site polymorphism in the KRAS 3'UTR; colorectal cancer screening population prevalence and influence on clinical outcome in patients with metastatic colorectal cancer treated with 5-fluorouracil and oxaliplatin +/- cetuximab. *BMC Cancer*, **12**, 534.
70. Kim,M., Chen,X., Chin,L.J., Paranjape,T., Speed,W.C., Kidd,K.K., Zhao,H., Weidhaas,J.B. and Slack,F.J. (2014) Extensive sequence variation in the 3' untranslated region of the KRAS gene in lung and ovarian cancer cases. *Cell Cycle*, **13**, 1030–1040.
71. Allentoft,M.E., Collins,M., Harker,D., Haile,J., Oskam,C.L., Hale,M.L., Campos,P.F., Samaniego,J.A., Gilbert,M.T.P., Willerslev,E., *et al.* (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.*, **279**, 4724–4733.
72. Butler,J.M. (2011) Forensic DNA testing. *Cold Spring Harb. Protoc.*, **6**, 1438–1450.
73. Pedersen,G. and Kanigan,T. (2016) Clinical RNA sequencing in oncology: where are we? *Per. Med.*, **13**, 209–213.
74. Kris,M.G., Johnson,B.E., Berry,L.D., Kwiatkowski,D.J., Iafrate, a J., Wistuba,I.I., Varella-Garcia,M., Franklin,W. a, Aronson,S.L., Su,P.-F., *et al.* (2014) Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *Jama*, **311**, 1998–2006.
75. Paik,S., Shak,S., Tang,G., Kim,C., Baker,J., Cronin,M., Baehner,F.L., Walker,M.G., Watson,D., Park,T., *et al.* (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.*, **351**, 2817–2826.
76. Dumur,C.I., Lyons-Weiler,M., Sciulli,C., Garrett,C.T., Schrijver,I., Holley,T.K., Rodriguez-Paris,J., Pollack,J.R., Zehnder,J.L., Price,M., *et al.* (2008) Interlaboratory performance of a microarray-based gene expression test to determine tissue of origin in poorly differentiated and undifferentiated cancers. *J. Mol. Diagn.*, **10**, 67–77.
77. Koh,W., Pan,W., Gawad,C., Fan,H.C., Kerchner,G.A., Wyss-coray,T., Blumenfeld,Y.J., El-sayed,Y.Y. and Quake,S.R. (2014) Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 7361–7366.
78. Volinia,S., Calin,G.A., Liu,C., Ambs,S., Cimmino,A., Petrocca,F., Visone,R., Iorio,M., Roldo,C., Ferracin,M., *et al.* (2006) A microRNA

expression signature of human solid tumors defines cancer gene targets. *PNAS*, **103**, 2257–2261.
79. Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G. and Erlich,H. (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.*, **51**, 263–273.
80. Tabor,S. (2001) Expression Using the T7 RNA Polymerase/Promoter System. In *Current protocols in molecular biology*.Vol. Chapter 16, p. 16.2.1–16.2.11.
81. Lai,M.M.C. (2005) RNA replication without RNA-dependent RNA polymerase: surprises from Hepatitis delta virus. *J. Virol.*, **79**, 7951–7958.
82. Ahlquist,P. (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science*, **296**, 1270–1273.
83. Cahill,P., Foster,K. and Mahan,D.E. (1991) Polymerase chain reaction and Qβ replicase amplification. *Clin. Chem.*, **37**, 1482–1485.
84. Romero,A., Blanco-Urgoiti,B. and Ponz,F. (1997) Amplification and cloning of a long RNA virus genome using immunocapture-long RT-PCR. *J. Virol. Methods*, **66**, 159–163.
85. Tellier,R., Bukh,J., Emerson,S.U. and Purcell,R.H. (1996) Amplification of the full-length hepatitis A virus genome by long reverse transcription-PCR and transcription of infectious RNA directly from the amplicon. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 4370–4373.
86. Kim,V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.*, **6**, 376–385.
87. Behrens,S. and Ru,C. (2003) In situ accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. *Microbiology*, **69**, 1748–1758.
88. Mir,K.U. and Southern,E.M. (1999) Determining the influence of structure on hybridization using oligonucleotide arrays. *Nat. Biotechnol.*, **17**, 788–792.
89. Howell,W.M., Grundberg,I., Faryna,M., Landegren,U. and Nilsson,M. (2010) Glycosylases and AP-cleaving enzymes as a general tool for probe-directed cleavage of ssDNA targets. *Nucleic Acids Res.*, **38**, e99.
90. Picelli,S., Björklund,Å.K., Faridani,O.R., Sagasser,S., Winberg,G. and Sandberg,R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
91. Groelz,D., Sobin,L., Branton,P., Compton,C., Wyrich,R. and Rainen,L. (2013) Non-formalin fixative versus formalin-fixed tissue: A comparison of histology and RNA quality. *Exp. Mol. Pathol.*, **94**, 188–194.
92. Hatzis,C., Sun,H., Yao,H., Hubbard,R.E., Meric-Bernstam,F., Babiera,G. V., Wu,Y., Pusztai,L. and Symmans,W.F. (2010) Effects of tissue handling on rna integrity and microarray measurements from resected breast cancers. *J. Natl. Cancer Inst.*, **103**, 1871–1883.
93. Medeiros,F., Rigl,C.T., Anderson,G.G., Becker,S.H. and Halling,K.C. (2007) Tissue handling for genome-wide expression analysis: a review of the issues, evidence, and opportunities. *Arch. Pathol. Lab. Med.*,

131, 1805–1816.
94. Miyamoto,T., Okano,S. and Kasai,N. (2009) Irreversible thermoinactivation of ribonuclease-A by soft-hydrothermal processing. *Biotechnol. Prog.*, **25**, 1678–1685.
95. Pastinen,T., Kurg,A., Metspalu,A., Peltonen,L. and Syvänen,A.C. (1997) Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.*, **7**, 606–614.
96. Nilsson,M., Malmgren,H., Samiotaki,M., Kwiatkowski,M., Chowdhary,B.P. and Landegren,U. (1994) Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science*, **265**, 2085–2088.
97. Levsky,J.M. (2003) Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.*, **116**, 2833–2838.
98. Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
99. Lubeck,E. and Cai,L. (2012) Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods*, **9**, 743–748.
100. Geiss,G.K., Bumgarner,R.E., Birditt,B., Dahl,T., Dowidar,N., Dunaway,D.L., Fell,H.P., Ferree,S., George,R.D., Grogan,T., *et al.* (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.*, **26**, 317–325.
101. Player,A.N., Shen,L.P., Kenny,D., Antao,V.P. and Kolberg,J.A. (2001) Single-copy gene detection using branched DNA (bDNA) in situ hybridization. *J. Histochem. Cytochem.*, **49**, 603–612.
102. Choi,H.M.T., Beck,V. a. and Pierce,N. a. (2014) Next-generation in situ hybridization chain reaction: Higher gain, lower cost, greater durability. *ACS Nano*, **8**, 4284–4294.
103. Xing,Z., Lin,A., Li,C., Liang,K., Wang,S., Liu,Y., Park,P.K., Qin,L., Wei,Y., Hawke,D.H., *et al.* (2014) LncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell*, **159**, 1110–1125.
104. Petersen,M. and Wengel,J. (2003) LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol.*, **21**, 74–81.
105. Selvaraju,S.B., Kapoor,R. and Yadav,J.S. (2008) Peptide nucleic acid-fluorescence in situ hybridization (PNA-FISH) assay for specific detection of Mycobacterium immunogenum and DNA-FISH assay for analysis of pseudomonads in metalworking fluids and sputum. *Mol. Cell. Probes*, **22**, 273–280.
106. Stender,H. (2003) PNA FISH: an intelligent stain for rapid diagnosis of infectious diseases. *Expert.Rev.Mol.Diagn.*, **3**, 649–655.
107. Yaroslavsky,A.I. and Smolina,I. V. (2013) Fluorescence imaging of single-copy DNA sequences within the human genome using PNA-directed padlock probe assembly. *Chem. Biol.*, **20**, 445–453.
108. Sylwestrak,E.L., Rajasethupathy,P., Wright,M.A., Jaffe,A. and Deisseroth,K. (2016) Multiplexed intact-tissue transcriptional analysis

at cellular resolution. *Cell*, **164**, 792–804.
109. Kloosterman,W.P., Wienholds,E., Bruijn,E. De, Kauppinen,S. and Plasterk,R.H.A. (2006) In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat. Methods*, **3**, 2005–2007.
110. Silahtaroglu,A.N., Nolting,D., Dyrskjøt,L., Berezikov,E., Møller,M., Tommerup,N. and Kauppinen,S. (2007) Detection of microRNAs in frozen tissue sections by fluorescence in situ hybridization using locked nucleic acid probes and tyramide signal amplification. *Nat. Protoc.*, **2**, 2520–2528.
111. Renwick,N., Cekan,P., Masry,P.A., Mcgeary,S.E., Miller,J.B., Hafner,M., Li,Z., Mihailovic,A., Morozov,P., Brown,M., *et al.* (2013) Multicolor microRNA FISH effectively differentiates tumor types. *J. Clin. Invest.*, **123**, 2694–2702.
112. Gibson,U.E.M., Heid,C.A. and Williams,P.M. (1996) A Novel Method for Real Time Quantitative RT-PCR. *Genome Res.*, **6**, 995–1001.
113. Heid,C., Stevens,J., Livak,K. and Williams,P. (1996) Real time quantitative PCR. *Genome Res.*, **6**, 986–994.
114. Holland,P.M., Abramson,R.D., Watson,R. and Gelfand,D.H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 7276–7280.
115. Chen,C., Ridzon,D. a, Broomer,A.J., Zhou,Z., Lee,D.H., Nguyen,J.T., Barbisin,M., Xu,N.L., Mahuvakar,V.R., Andersen,M.R., *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*, **33**, e179.
116. Notomi,T., Okayama,H., Masubuchi,H., Yonekawa,T., Watanabe,K., Amino,N. and Hase,T. (2000) Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.*, **28**, e63.
117. Lee,M.-S., Lin,Y.-C., Lai,G.-H., Lai,S.-Y., Chen,H.-J. and Wang,M.-Y. (2011) One-step reverse-transcription loop-mediated isothermal amplification for detection of infectious bursal disease virus. *Can. J. Vet. Res.*, **75**, 122–127.
118. Chander,Y., Koelbl,J., Puckett,J., Moser,M.J., Klingele,A.J., Liles,M.R., Carrias,A., Mead,D.A. and Schoenfeld,T.W. (2014) A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.*, **5**, 1–11.
119. Guatelli,J.C., Whitfield,K.M., Kwoh,D.Y., Barringer,K.J., Richman,D.D. and Gingeras,T.R. (1990) Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 1874–1878.
120. Wharam,S.D., Marsh,P., Lloyd,J.S., Ray,T.D., Mock,G.A., Assenberg,R., McPhee,J.E., Brown,P., Weston,A. and Cardy,D.L.N. (2001) Specific detection of DNA and RNA targets using a novel isothermal nucleic acid amplification assay based on the formation of a three-way junction structure. *Nucleic Acids Res.*, **29**, e54.
121. Vincent,M., Xu,Y. and Kong,H. (2004) Helicase-dependent isothermal

DNA amplification. *EMBO Rep.*, **5**, 795–800.
122. Piepenburg,O., Williams,C.H., Stemple,D.L. and Armes,N.A. (2006) DNA detection using recombination proteins. *PLoS Biol.*, **4**, 1115–1121.
123. Lyamichev,V., Brow,M. a and Dahlberg,J.E. (1993) Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778–783.
124. Lyamichev,V., Mast,A.L., Hall,J.G., Prudent,J.R., Kaiser,M.W., Takova,T., Kwiatkowski,R.W., Sander,T.J., Arruda,M. De, Arco,D.A., *et al.* (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.*, **17**, 292–296.
125. Eis,P.S., Olson,M.C., Takova,T., Curtis,M.L., Olson,S.M., Vener,T.I., Ip,H.S., Vedvik,K.L., Bartholomay,C.T., Allawi,H.T., *et al.* (2001) An invasive cleavage assay for direct quantitation of specific RNAs. *Nat. Biotechnol.*, **19**, 673–676.
126. Allawi,H.T., Dahlberg,J.E., Olson,S., Allawi,H.T., Dahlberg,J.E., Olson,S., Lund,E. and Olson,M. (2004) Quantitation of microRNAs using a modified Invader assay. *RNA*, **10**, 1153–1161.
127. Landegren, Kaiser, Sanders and Hood (1988) A Ligase-Mediated Gene Detection Technique. *Science*, **241**, 1077–1080.
128. Nickerson,D.A., Kaiser,R., Lappin,S., Stewartt,J., Hood,L. and Landegrent,U. (1990) Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay (DNA amplification/gene detection/genome mapping). *Genetics*, **87**, 8923–8927.
129. Grossman,P.D., Bloch,W., Brinson,E., Chang,C.C., Eggerding,F.A., Fung,S., Lovannisci,D.A., Woo,S. and Winn-deen,E.S. (1994) High-density multiplex detection of nucleic acid sequences: Oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res.*, **22**, 4527–4534.
130. Fareed,G.C., Wilt,E.M., Richardson,C. and Richardsons,C.C. (1971) Enzymatic breakage and joining of deoxyribonucleic acid, VIII. Hybrids of ribo- and deoxyribonucleotide homopolymers as substrates for polynucleotide ligase of bacteriophage T4. *J. Biol. Chem.*, **246**, 925–932.
131. Nilsson,M., Antson,D.O., Barbany,G. and Landegren,U. (2001) RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.*, **29**, 578–581.
132. Barany,F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 189–193.
133. Hu,Y., Jia,H., Wang,Y., Cheng,Y. and Li,Z. (2014) Sensitive quantification of messenger RNA with a real-time ligase chain reaction by using a ribonucleotide-modified DNA probe. *Chem. Commun.*, **50**, 13093–13095.
134. Zhang,P., Zhang,J., Wang,C., Liu,C., Wang,H. and Li,Z. (2014) Highly sensitive and specific multiplexed MicroRNA quantification using size-

coded ligation chain reaction. *Anal. Chem.*, **86**, 1076–1082.
135. Liu,D., Daubendiek,S.L., Zillman,M.A., Ryan,K. and Kool,E.T. (1996) Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J. Am. Chem. Soc.*, **118**, 1587–1594.
136. Banér,J., Nilsson,M., Mendel-Hartvig,M. and Landegren,U. (1998) Signal amplification of padlock probes by rolling circle replication. *Nucleic acids res.*, **26**, 5073–5078.
137. Nilsson,M., Krejci,K., Koch,J., Kwiatkowski,M., Gustavsson,P. and Landegren,U. (1997) Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in juman chromosomes 13 and 21. *Nat. Genet.*, **16**, 252–255.
138. Antson,D.O., Isaksson,A., Landegren,U. and Nilsson,M. (2000) PCR-generated padlock probes detect single nucleotide variation in genomic DNA. *Nucleic Acids Res.*, **28**, e58.
139. Nilsson,M., Gullberg,M., Dahl,F., Szuhai,K. and Raap,A.K. (2002) Real-time monitoring of rolling-circle amplification using a modified molecular beacon design. *Nucleic Acids Res.*, **30**, e66.
140. Eriksson,R., Jobs,M., Ekstrand,C., Ullberg,M., Herrmann,B., Landegren,U., Nilsson,M. and Blomberg,J. (2009) Multiplex and quantifiable detection of nucleic acid from pathogenic fungi using padlock probes, generic real time PCR and specific suspension array readout. *J. Microbiol. Methods*, **78**, 195–202.
141. Engström,A., Zardán Gómez de la Torre,T., Strømme,M., Nilsson,M. and Herthnek,D. (2013) Detection of Rifampicin Resistance in Mycobacterium tuberculosis by Padlock Probes and Magnetic Nanobead-Based Readout. *PLoS One*, **8**, e62015.
142. Russell,C., Welch,K., Jarvius,J., Cai,Y., Brucas,R., Nikolajeff,F. and Al,R.E.T. (2014) Gold nanowire based electrical DNA detection using rolling circle amplification. *ACS Nano*, **8**, 1147–1153.
143. Zhang,D.Y., Zhang,W., Li,X. and Konomi,Y. (2001) Detection of rare DNA targets by isothermal ramification amplification. *Gene*, **274**, 209–216.
144. Gullberg,M., Mendel-hartvig,M., Landegren,U., Nilsson,M., Dahl,F. and Bane,J. (2004) Circle-to-circle amplification for precise and sensitive DNA analysis. *PNAS*, **101**, 4548–4553.
145. Pavankumar,A.R., Engström,A., Liu,J., Herthnek,D. and Nilsson,M. (2016) Proficient detection of multi-drug-resistant Mycobacterium tuberculosis by padlock probes and lateral flow nucleic acid biosensors. *Anal. Chem.*, **88**, 4277–4284.
146. Dou,D., Hernández-Neuta,I., Wang,H., Östbye,H., Qian,X., Thiele,S., Resa-Infante,P., Kouassi,N.M., Sender,V., Hentrich,K., *et al.* (2017) Analysis of IAV Replication and Co-infection Dynamics by a Versatile RNA Viral Genome Labeling Method. *Cell Rep.*, **20**, 251–263.
147. Hamidi,S.V., Ghourchian,H. and Tavoosidana,G. (2015) Real-time detection of H5N1 influenza virus through hyperbranched rolling circle amplification. *Analyst*, **140**, 1502–1509.

148. Larsson,C., Koch,J., Nygren,A., Janssen,G., Raap,A.K. and Landegren,U. (2004) In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat. Methods Methods*, **1**, 227–232.
149. Larsson,C., Grundberg,I. and Söderberg,O. (2010) In situ detection and genotyping of individual mrna molecules. *Nat. Methods*, **7**, 395–397.
150. Lagunavicius,A., Merkiene,E., Kiveryte,Z., Savaneviciute,A., Zimbaite-Ruskuliene,V., Radzvilavicius,T. and Janulaitis,A. (2009) Novel application of Phi29 DNA polymerase: RNA detection and analysis in vitro and in situ by target RNA-primed RCA. *RNA*, **15**, 765–771.
151. Porreca,G.J., Zhang,K., Li,J.B., Xie,B., Austin,D., Vassallo,S.L., LeProust,E.M., Peck,B.J., Emig,C.J., Dahl,F., *et al.* (2007) Multiplex amplification of large sets of human exons. *Nat. Methods*, **4**, 931–936.
152. Ke,R., Mignardi,M., Pacureanu,A., Svedlund,J., Botling,J., Wählby,C. and Nilsson,M. (2013) In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*, **10**, 857–60.
153. Ozsolak,F., Platt,A.R., Jones,D.R., Reifenberger,J.G., Sass,L.E., McInerney,P., Thompson,J.F., Bowers,J., Jarosz,M. and Milos,P.M. (2009) Direct RNA sequencing. *Nature*, **461**, 814–818.
154. Lu,C., Meyers,B.C. and Green,P.J. (2007) Construction of small RNA cDNA libraries for deep sequencing. *Methods*, **43**, 110–117.
155. Hafner,M., Landgraf,P., Ludwig,J., Rice,A., Ojo,T., Lin,C., Holoch,D., Lim,C. and Tuschl,T. (2008) Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods*, **44**, 3–12.
156. Hrdlickova,R., Toloue,M. and Tian,B. (2017) RNA-Seq methods for transcriptome analysis. *WIREs RNA*, **8**, e1364.
157. Pihlak,A., Baurén,G., Hersoug,E., Lönnerberg,P., Metsis,A. and Linnarsson,S. (2008) Rapid genome sequencing with short universal tiling probes. *Nat. Biotechnol.*, **26**, 676–684.
158. Drmanac,R., Sparks,A.B., Callow,M.J., Halpern,A.L., Burns,N.L., Kermani,B.G., Carnevali,P., Nazarenko,I., Nilsen,G.B., Yeung,G., *et al.* (2009) Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science*, **1469**, 78–81.
159. Mardis,E.R. (2008) Next-generation DNA sequencing methods. *Annu. Genomics Hum.Genet*, **9**, 387–402.
160. Lion,T. (2014) Adenovirus infections in immunocompetent and immunocompromised patients. *Clin. Microbiol. Rev.*, **27**, 441–462.
161. Tollefson,A.E., Ryerse,J.S., Scaria,A., Hermiston,T.W. and Wold,W.S. (1996) The E3-11.6-kDa adenovirus death protein (ADP) is required for efficient cell death: characterization of cells infected with adp mutants. *Virology*, **220**, 152–162.
162. Assadian,F., Sandström,K., Bondeson,K., Laurell,G., Lidian,A., Svensson,C., Akusjärvi,G., Bergqvist,A. and Punga,T. (2016) Distribution and molecular characterization of human adenovirus and Epstein-Barr virus infections in tonsillar lymphocytes isolated from

patients diagnosed with tonsillar diseases. *PLoS One*, **11**, 1–19.
163. Pombo,A., Ferreira,J., Bridge,E. and Carmo-Fonseca,M. (1994) Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells. *EMBO J.*, **13**, 5075–5085.
164. Weibrecht,I., Lundin,E., Kiflemariam,S., Mignardi,M., Grundberg,I., Larsson,C., Koos,B., Nilsson,M. and Söderberg,O. (2013) In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. *Nat. Protoc.*, **8**, 355–372.
165. Mezger,A., Ohrmalm,C., Herthnek,D., Blomberg,J. and Nilsson,M. (2014) Detection of rotavirus using padlock probes and rolling circle amplification. *PLoS One*, **9**, e111874.
166. Lohman,G.J.S., Zhang,Y., Zhelkovsky,A.M., Cantor,E.J. and Evans,T.C. (2014) Efficient DNA ligation in DNA-RNA hybrid helices by Chlorella virus DNA ligase. *Nucleic Acids Res.*, **42**, 1831–1844.
167. Ma,W., Kaiser,M.W., Schaefer,J.J., Hatim,T., Takova,T., Neri,B.P., Lyamichev,V.I., Ma,W., Kaiser,M.W., Lyamicheva,N., *et al.* (2000) RNA Template-dependent 5′ Nuclease Activity of Thermus aquaticus and Thermus thermophilus DNA Polymerases. *J. Biol. Chem.*, **275**, 23693–24700.
168. Krzywkowski,T. and Nilsson,M. (2017) Fidelity of RNA templated end-joining by chlorella virus DNA ligase and a novel iLock assay with improved direct RNA detection accuracy. *Nucleic Acids Res.*, **1**, gkx708.
169. Nandakumar,J. and Shuman,S. (2004) How an RNA Ligase Discriminates RNA versus DNA Damage. *Molcular Cell*, **16**, 211–221.
170. Braasch,D.A. and Corey,D.R. (2001) Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem. Biol.*, **8**, 1–7.
171. Tse,W.T. and Forget,B.G. (1990) Reverse transcription and direct amplification of cellular RNA transcripts by Taq polymerase. *Gene.*, **88**, 293–296.
172. Tang,S., Wei,H., Hu,T., Jiang,J., Chang,J., Guan,Y. and Zhao,G. (2016) Suppression of rolling circle amplification by nucleotide analogs in circular template for three DNA polymerases. *Biosci. Biotechnol. Biochem.*, **80**, 1555–1561.