RNA-DNA hybrids

Form, Flexibility and Function

Brian Luke
IMB Mainz
Johannes Gutenberg Universität, Mainz
RNA-DNA hybrids

1. Types of RNA-DNA hybrids, General
2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)
3. R-loops, their formation, detection regulation and distribution
4. Functional R-loops
5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H
6. R-loops as regulators of DNA methylation
Many types of RNA-DNA hybrids exist on genomic DNA


-also CRISPR/Cas9 makes RNA-DNA hybrids
RNA-DNA hybrids – the building blocks of RNA and DNA

Ribonucleotides

Deoxyribonucleotides
RNA-DNA hybrids – ss polymerized RNA and DNA molecules

- DNA is relatively stable compared to RNA

- The 2’OH of ribonucleotides can hydrolyze the sugar backbone by nucleophilic attack on the phosphate bond
The presence of 2′OH makes RNA susceptible to hydrolysis

- Hydrolysis of RNA occurs and this is particularly frequent in alkaline conditions (high pH)
RNA-DNA hybrids – ribonucleotides in the context of dsDNA

- Ribonucleotides can be inserted into the DNA backbone and basepair with deoxyribonucleotides

- This is considered an RNA-DNA hybrid molecule

- The effect is that the stability of the DNA is now compromised
Many types of RNA-DNA hybrids exist on genomic DNA


Brian Luke
b.luke@imb-mainz.de
1. Types of RNA-DNA hybrids, General

2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)

3. R-loops, their formation, detection regulation and distribution

4. Functional R-loops

5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H

6. R-loops as regulators of DNA methylation
rNTPs are frequently incorporated into DNA by the replicating polymerases

- DNA polymerases accidentaly incorporate rNTPs into DNA which are then called rNMPs (ribonucleoside monophosphate)

- Why does this happen?

- DNA polymerases have a tyrosine steric gate that recognizes the 2‘OH of rNTPs and prevents their entry into the catalytic site

- However the gate is not perfect rNTPs are used

- This is in part due to the high concentration of rNTPs compared to dNTPs

- Depending on the base, rNTPs are between 30 and 200-fold more concentrated than dNTPs in the cell
rNTPs incorporation is frequent

- In yeast approximately 10,000 rNMPs are incorporated into dsDNA per S phase
- this means approximately one rNMP per 6500 bases of DNA
- This makes ribonucleotides the most frequently occurring of all types of DNA damage
- Numbers are similar in human cells

So what is the problem with rNMP insertions?
rNMPs can lead to genome instability

- rNMPs are inherently prone to hydrolysis which would leave ss nicks in the DNA

- Upon DNA replication ss nicks are converted into ds breaks

- Moreover, the presence of ribonucleotides leads replication stress and polymerase stalling

- Most importantly, rNMPs are acted on by Top1 which can lead to mutagenesis (see later)
rNMP hybrids may lead to other alterations as well

- RNA-DNA hybrids form A form DNA instead of the usual B-form.

- This likely also prevents nucleosome assembly and may affect local epigenetic marks

So how do we get rid of rNMPs that have been inserted into the genome?
RER: RNase H2-initiated, faithful ribonucleotide excision repair

RNase H2

RNase H2

Fen1/Exo1

DNA Ligase I
Ribonucleotide excision repair (RER)

Mutations in RNase H2 result in a neurological syndrome called Aicardi Goutières Syndrome (AGS)

- neurodegeneration with severe ataxia
- auto immune
- high levels of genome instability

-RNase H2 is also mutated in metastatic castration-resistant prostate cancer and in CLL

There is an RER „back-up“ mechanism that is responsible for a lot of the problems when RNase H2 is missing
Topoisomerase 1 (Topo1) as a Ribonucleotide excision repair (RER) backup

Bottom line: RNase H2 repairs rNMPs in an error–free manner

When RNase H2 doesn’t work

Top1 steps in, but makes mistakes and leads to more mutations
We can measure the mutations that Top1 makes

Could these Top1 mutations increase with age?
Somatic mutation rates scale with lifespan across mammals

Equation:

\[ k = 3,206.4 \]

95% CI: 2,683.9–3,728.9

FVE = 0.82

ELB range: 1,828.1–5,378.7

Vanessa Pires
vanpires@uni-mainz.de
We collected old cells ...
....and measured Top1 mutations at rNMPs

Top1 mutations increase with age
-Why.....do rNMPs increase with age, do Top1 levels increase with age?
do RNase H2 levels decrease......these are ongoing questions that are lab
is trying to answer
RNA-DNA hybrids

1. Types of RNA-DNA hybrids, General
2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)
3. R-loops, their formation, detection regulation and distribution
4. Functional R-loops
5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H
6. R-loops as regulators of DNA methylation
Many types of RNA-DNA hybrids exist on genomic DNA

R-Loops refer to RNA that is base-paired to DNA in a Watson-Crick manner resulting in one strand of DNA being displaced.

Brian Luke
b.luke@imb-mainz.de
Where do we find R-loops?

How are R-loops formed?

There are likely many ways that R-loops form and there is no clear consensus on this.

1. Prevailing view is that R-loop formation is transcription coupled and that the transcript basepairs with the DNA template behind the transcription machinery. Here there is **negative supercoiling** and more chance for helix melting.

2. That R-loops can form *in trans*. There is evidence that **R-loops can become coated with Rad51 and do strand invasion**. Plasmid expressed RNA can form R-loops on chromosomes at homologous sequences. Therefore R-loops can form *in trans*.....do they?
How are R-loops regulated?

1. Prevent them from forming

2. Remove them once they have formed
How are R-loops prevented from forming?

**Prevent them from forming**

a. **Transcriptional repressors** – suppressing transcription especially at repetative regions prevents an RNA and hence an R-loop

b. **Elongation factors** – transcription stalling can also facilitate R-loop formation – therefore progressive elongation is critical

c. **RNA binding** – processing factors that promote RNA maturation, splicing and export prevent the RNA from re-annealing

Brian Luke
b.luke@imb-mainz.de
How are R-loops regulated?

Remove them once they have formed

a. **Helicases**—some helicases have preference for unwinding RNA-DNA hybrids *in vitro*, and may also do so *in vivo* e.g. BLM, SETX, AQR, DHX9 and a host of other factors that have been identified as R-loop interactors

a. **Nucleases**—The RNase H enzymes (RNase H1 and RNase H2) are the primary enzymes responsible for R-loop cleavage
How are R-loops regulated?

RNases H1 and H2 can hydrolyze the RNA moiety of R-loops

RNase H2 accounts for most of the RNase H activity in the cell
Why do R-loops need to be removed?

The stabilisation of R-loops leads to increased DNA damage and genome instability.

Example of yeast mutants accumulating R-loops
Costantino et al, 2018, Mol Cell
Why do R-loops need to be removed?

Kotsantis et al, 2016, Nat Comm
Why do R-loops need to be removed?

R-loops collide with replication

- Causes replication fork stalling and then processing of DNA
- May not be the RNA but rather the transcription machinery itself (i.e. a crash with the polymerase)
- Also the displaced strand is more vulnerable
- In general it is thought that head-on collisions are more detrimental than co-directional collisions
- It has also been suggested that the RNA-DNA hybrid per se is not the problem but rather the local compaction of the chromatin

Hamperl et al, 2017, Cell
Why do R-loops need to be removed?

R-loops collide with replication

- R-loops lead to local increases in H3S10 phosphorylation and hence chromatin compaction
- This prevents proper replication and repair
- Mutants that prevent compaction can rescue the phenotypes despite the fact that the R-loops persist

Castellano-pozo et al, 2013, Mol Cell
### Why do R-loops need to be removed – disease associated with R-loops

<table>
<thead>
<tr>
<th>Disease</th>
<th>R-loop Factor/Locus</th>
<th>Proposed Mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast/Ovarian</td>
<td>Estrogen</td>
<td>Estrogen-induced R-loops cause DNA damage and genome instability.</td>
<td>Stoik et al, 2016</td>
</tr>
<tr>
<td></td>
<td>BRCA1</td>
<td>BRCA1 interacts with SETX and suppresses R-loops and DNA breaks at gene terminators.</td>
<td>Hatchi et al, 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNAPII pausing contributes to BRCA1-associated R-loop accumulation and breast cancer development.</td>
<td>Zhang et al, 2017</td>
</tr>
<tr>
<td></td>
<td>BRCA2</td>
<td>BRCA2 is sequestered in cells expressing heterochromatin-associated non-coding RNAs, leading to genome instability.</td>
<td>Zhu et al, 2018</td>
</tr>
<tr>
<td>Ewing's sarcoma</td>
<td>EWS-FLI, BRCA1</td>
<td>R-loop depletion causes transcriptional stress, resulting in functional depletion of BRCA1 and subsequent DNA damage.</td>
<td>Gorthi et al, 2018</td>
</tr>
<tr>
<td>Myelodysplastic syndromes (MDS)</td>
<td>SRSF2, U2AF1</td>
<td>R-loops induced by splicing factor mutations cause replication stress and impair bone cell function.</td>
<td>Chen et al, 2018</td>
</tr>
<tr>
<td>Multiple myeloma and Burkitt's lymphoma</td>
<td>TRD3-TOP3B</td>
<td>TRD3-TOP3B complex relieves negative supercoiling and reduces R-loop levels at c-MYC and Igh to suppress chromosomal translocations.</td>
<td>Yang et al, 2014</td>
</tr>
<tr>
<td>Alternative lengthening of telomeres (ALT)-dependent cancers</td>
<td>Telomeric repeat-containing RNA (TERRA)</td>
<td>TERRA R-loops are upregulated in cancer cells and promote homologous recombination to preserve telomeres by the ALT pathway.</td>
<td>Arora et al, 2014</td>
</tr>
<tr>
<td>Fanconi anemia (FA)</td>
<td>FANCM, FANCD2</td>
<td>FA factor deficiency leads to increased R-loop levels, exacerbating TRC and causing genome instability.</td>
<td>Schwab et al, 2015; Garcia-Rubio et al, 2015</td>
</tr>
<tr>
<td>AOA2</td>
<td>SETX</td>
<td>SETX resolves R-loops in neuronal cells; R-loops are elevated in neural progenitor cells from AOA2 patients with SETX mutations.</td>
<td>Bouchard et al, 2015</td>
</tr>
</tbody>
</table>

Crossley et al, 2019, Mol Cell
How are R-loops detected?

All R-loop detection to date is based on two different methods...

1. The S9.6 monoclonal antibody.

- This antibody was developed to increase the hybridization of RNA to DNA for microarray experiments, was later determined to bind to RNA-DNA hybrid in a structure specific and sequence independent manner.

- Epitope is a hybrid of at least 6 base-pairs.

- Easy to handle.

- Specificity issues, the antibody recognizes dsRNA and other structured RNAs....many controls required.

Used for DRIP (DNA-RNA IP), to pull down the hybrid and analyse by qPCR or NGS.

- Can be used for IF as well as for Southwestern dot blotting.

Brian Luke
b.luke@imb-mainz.de

www.imb.de
How are R-loops detected?

All R-loop detection to date is based on two different methods

2. Catalytic dead RNase H1

- This reagent can recognize an R-loop but does not degrade it and it then gets „stuck“ on the R-loop

- an antibody against RNase H1 can then pull down the complex

- less non-specific binding compared to S9.6
- reagent has to be made
- get stabilization of R-loops

Used for R-ChIP

Antibody to Rnh1 pulls down associated sequences which are subsequently quantified by either qPCR or sequencing
All methods are variations of S9.6 and cat dead RNase H1 precipitations

<table>
<thead>
<tr>
<th>Method Name</th>
<th>Fragmentation Method</th>
<th>Detection Method</th>
<th>Molecule Sequenced</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Primary Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRIP-seq</td>
<td>Restriction digest</td>
<td>S9.6</td>
<td>dsDNA</td>
<td>Robust signal, widely adopted, easy to set up</td>
<td>Low resolution, no strand specificity, not in situ</td>
<td>Ginno et al., 2012</td>
</tr>
<tr>
<td>DRIVE-seq</td>
<td>Restriction digest</td>
<td>Catalytically inactive RNase H</td>
<td>dsDNA</td>
<td>Provides independent verification of some DRIP-seq results</td>
<td>Low enrichment, low resolution, no strand specificity, reagent not commercially available, not in situ</td>
<td>Ginno et al., 2012</td>
</tr>
<tr>
<td>S9.6-ChIP-seq</td>
<td>Sonication after cross-linking</td>
<td>S9.6</td>
<td>dsDNA</td>
<td>May overcome bias and resolution issues in DRIP-seq</td>
<td>Not strand specific, cross-linking could affect results</td>
<td>El Hage et al., 2014</td>
</tr>
<tr>
<td>S1-DRIP-seq</td>
<td>Sonication</td>
<td>S9.6</td>
<td>dsDNA</td>
<td>Higher resolution than DRIP-seq</td>
<td>Not strand specific, not in situ</td>
<td>Wahba et al., 2016</td>
</tr>
<tr>
<td>DRIPc-seq</td>
<td>Restriction digest</td>
<td>S9.6</td>
<td>RNA</td>
<td>Strand specific, high resolution</td>
<td>Not in situ, requires lengthier sample preparation, S9.6 may recognize dsRNA</td>
<td>Sanz et al., 2016</td>
</tr>
<tr>
<td>RDIP-seq</td>
<td>Sonication</td>
<td>S9.6</td>
<td>RNA</td>
<td>Strand specific, high resolution</td>
<td>Not in situ, lengthier preparation, S9.6 recognizes dsRNA</td>
<td>Nadel et al., 2015</td>
</tr>
<tr>
<td>ssDRIP-seq</td>
<td>Sonication</td>
<td>S9.6</td>
<td>ssDNA</td>
<td>Strand specific, easy compared to other strand-specific techniques</td>
<td>Not in situ, low resolution</td>
<td>Xu et al., 2017</td>
</tr>
<tr>
<td>Bis-DRIP-seq</td>
<td>Restriction digest</td>
<td>S9.6</td>
<td>dsDNA with bisulfite conversions</td>
<td>Strand specific, provides additional control to ensure S9.6 signal arises from an R-loop in situ</td>
<td>Requires many replicates</td>
<td>Dumalie and Jaffrey, 2017</td>
</tr>
<tr>
<td>R-ChIP-seq</td>
<td>Sonication</td>
<td>Catalytically inactive RNase H</td>
<td>ssDNA</td>
<td>Strand specific, in situ capture</td>
<td>Cell line must be engineered to express catalytically inactive RNase H construct, inactive RNase H may alter hybrid dynamics</td>
<td>Chen et al., 2017</td>
</tr>
</tbody>
</table>

Crossley et al, 2019, Mol Cell
RNA-DNA hybrids

1. Types of RNA-DNA hybrids, General
2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)
3. R-loops, their formation, detection regulation and distribution
4. Functional R-loops
5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H
6. R-loops as regulators of DNA methylation
Are R-loops all bad? – could be used to find a needle in the haystack

- Replication stress
- DNA damage
- Genome instability
- disease

- Ideal way to find a specific sequence
- Gene regulation
- Damaged DNA
R-loops as transcriptional regulators

- R-loops are frequently found in **promoter regions**
- Specifically promoter regions with **CpG islands**
- Frequently associated with a strong **GC skew**
- R-loops can exert different activities at different promoters
- Frequently associated with decreased DNA methylation

**A few examples**

- RASSF1A is a tumor suppressor
- Expression of antisense RNA (ANRASSF1) forms an R-loop which recruits the polycomb repressive complex 2
- This suppresses transcription and increased cell proliferation
R-loops as transcriptional regulators

- an anti-sense R-loop formed in the Gata3 promoter leads to the recruitment of MLL and promotes transcription
- this RNA can also be expressed in trans and will activate Gata3 by the same mechanism
- therefore regulatory R-loops can act in trans....but do they?
R-loops promote DNA repair

- at a break the 3` end is recognized by RNAPII
- the RNA forms a hybrid which gets recognized by BRCA1 and eventually BRCA2
- the hybrid is then degraded by RNase H2 and then Rad51 is loaded
- this then drives HR
- there are other theories regarding how hybrids work to repair DNA

D’Alessandro et al, 2018, Nat. Comm
Why are hybrids at Double strand breaks? – 2 predominant models

1. DSB as a promoter

Marnef and Legube, 2021
RNA-DNA hybrids

1. Types of RNA-DNA hybrids, General
2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)
3. R-loops, their formation, detection regulation and distribution
4. Functional R-loops
5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H
6. R-loops as regulators of DNA methylation
The double-edge sword of telomere shortening

Checkpoint arrest replicative senescence (tumor suppressive)

Important in long-lived organisms

Cell division / age

Important in long-lived organisms
Telomere length must be balanced to prevent disease.

How can the telomere shortening process be regulated to ensure a non-pathogenic balance is achieved?
Recombination ensures that premature senescence is avoided

Population doublings (telomerase negative cells)

Why does only the short telomere get extended and not the others?

Le et al, 1999
Telomeres are transcribed into TERRA molecules

1. Why are telomeres transcribed?

2. Do the R-loops have any physiological relevance?
TERRA-less telomeres do not recombine

![Graph showing TERRA expression and telomere length](image)

*PD* Balk and Maicher et al, 2013, *Nat Struc Mol Biol*
R-loop levels can influence rates of senescence

Balk and Maicher et al, 2013, *Nat Struc Mol Biol*

Graf et al, 2017, *Cell*
R-loops may promote elongation of the shortest telomeres

Population doublings

Are TERRA and R-loops really getting made at the short telomeres?
TERRA and R-loops accumulate at the critically short telomeres

TERRA at short telomeres?

premature senescence

HDR

Why is TERRA at short telomeres?
TERRA and R-loops are preferentially removed at long telomeres

Graf et al, 2017, Cell

So why does this drive recombination?
R-loops promote elongation of the shortest telomeres
RNA-DNA hybrids

1. Types of RNA-DNA hybrids, General
2. Ribonucleotide incorporation and ribonucleotide excision repair (RER)
3. R-loops, their formation, detection regulation and distribution
4. Functional R-loops
5. R-loops at telomeres and the timing of RNA-DNA hybrid removal by RNase H
6. R-loops as regulators of DNA methylation
R-loops affect the local methylation state via 2 means

- R-loops at CGIs prevent the association of DNA methyltransferases resulting in hypomethylation of promoters (Ginno et al, 2012, Grunseich et al, 2018)

- The GADD45 proteins binds directly to R-loops and recruits the TET DNA demethylases
- Approximately 4% of TET1 binding sites at CGI promoters may be R-loop dependent
- This has been characterized extensively at the TCF21 locus whereby an anti-sense RNA (TARID) forms an R-loop
RNA-DNA Hybrid are also critical for centromere function