MASON - paper

I have created a quick proposal which contains 3 parts. In the layout section I propose a layout of the paper, with bullet points for each section. Then I added the main points to be discussed. Here New suggestions for the web tool and PNA choice are welcome. The last section (draft) is just some sentences that I started to write, and can be ignored for now. For adding comments, please use the link for the google docs document:

https://docs.google.com/document/d/1g2JxVTaptuXcVfO9tcQGl5yK6x1FnasQHHLRaYk3YBY/edit?usp=sharing

Layout: **2**
- Introduction 2
- Methods 2
- Results 3
  - MASON web-tool for oligo design 3
  - PNA design for off-target identification 3
  - Global RNA-Seq to infer off-target mechanisms 3
- Discussion 3

Main points to be discussed: **4**
- MASON improvement / additional features 4
- PNA concentrations 4
- PNA choice 4

Draft: **7**
- Introduction 7
- Methods 7
- Results 7
- Discussion 7
- Literature 7

Layout:

1. Introduction

- General introduction on PNA
- Show that there is a need for tools like MASON
- Quickly summarize main points of paper

2. Methods

- How does MASON work?
- Details of conducted experiments:
- Choice of PNAs:
    - Choice of PNA, generated by MASON
        - PNA targeting a gene with many off-targets
        - PNA targeting a gene without many off-targets
        - acpP mutants with 2 mismatches ?
- RNASeq study:
    - [control (1) + PNAs (4-6)] * triplicate (3) = 15 - 21 samples
    - Protocol similar to the one in Linda's paper.
    - How to make sense of data? Are mismatches preventing PNA from altering mRNA level? If yes, at which positions?
- Software:
    - MASON webtool created using Python Flask
    - BBtools & R to analyse RNASeq data

3. Results

3.1. MASON web-tool for oligo design

- Why is there a need for this tool?
- Why is it better than currently available tools?
    - Easier user interface
    - Better off-target prediction
- Present MASON & describe the algorithm
- Describe possibilities to make use of MASON (web interface)
- Introduce how MASON was used here to design PNAs for off-target identification
- Figure of MASON output / web interface

3.2. PNA design for off-target identification

- Describe problem of off-targeting in PNA
- Show how/why we chose the PNAs using MASON
- Describe the chosen PNAs, the gene it targets and mismatches/off-targets in the genome
- Go through whole workflow of the experiment (Make a workflow figure)

3.3. Global RNA-Seq to infer off-target mechanisms

- Probably more than 1 subsection with the results of the RNAseq analysis
- Main goal: Identify off-target effects in RNASeq results
- Possible results:
    - PNA is highly specific and 1 bp mismatches in binding region prevents downregulation of respective gene
    - Effect of mismatches is region dependent (e.g. mismatches in the ends have lower mismatch-”penalty” than in the center region of the PNA)
    - Is PNA with a generally high amount of mismatches is less active on target mRNA?
    - Differences between PNA efficiency for different targets (infer patterns based on e.g. sequence, secondary structure, etc.)
    - Can we confirm (from Lindas paper) that off-targets in the coding sequence and not in the start region are not an issue?

- Volcano-plots with colors for genes with match in start region (and 1, 2 mismatches in different colors)
- Other figures about analysis of results
- Find patterns in off-targeting, e.g. correlation between off-target matching and PNA target regulation.

4. Discussion

- Compare MASON to other tool-boxes etc.
- Look into other studies which addressed off-targeting, as well as other papers with tools comparable to MASON.

Main points to be discussed:

MASON improvement / additional features

- Add self-complementarity warning to generated ASO
- Add an option which includes pathways (KEGG or GO).
    - For example, MASON would favour ASOs which have a mismatch in the start region of a gene in the same pathway as the target gene.

PNA concentrations

- Which concentrations should we use ?
- Probably it is best to just use a similar concentration which we used for Linda's paper (if we just use 10-mer PNAs it should be fine).

PNA choice

- As mentioned, we could probably test 4-6 PNAs per sequencing run, :

[control (1) + PNAs (4-6)] * triplicate (3) = 15 - 21 samples

We want to include:

- 1-3 PNA just targeting one gene with as few other matches in the genome as possible (PNA 1-4 in table 1)
- 1-3 PNA with a lot of matches (“worst case”) in the genome, so many possible off-targets (e.g.1 PNA 5-6 in table 1)
- Look into mutations of acpP PNA to investigate mismatch effects (see next page, table 2)
- Which PNA-length should we use? I now only included 10 mers, because by using length 10, we can be sure that the PNA actually enters the cell with the same PNA concentration.

Table 1: PNA targets which can be used for the RNASeq experiment. PNA1-4 have few matches in the genome of Salmonella. Here I show both full matches and ½ bp mismatches around other start regions. PNA4 , which has exactly 1 match (ompC) and no 0 mismatch off-targets in the whole genome, we have in stock. PNA5-6 have a high amount of 1 & 2 bp off-targets. PNA6 has only 1 target, whereas PNA5 has 5 fully complementary targets.

Nr.	PNA target sequence	Full start region matches	str	1 (2) bp mismatches in other start regions	Raw read counts in Popella et al. (untreated)	Uniprot name
PNA1 (few mismatches)	AGGTCC**ATG**G	rpl32	+	0 (7)	2000	50S ribosomal protein L32
PNA2 (few mismatches)	CCCACT**ATG**G	yjeQ	-	0 (9)	130	Small ribosomal subunit biogenesis GTPase RsgA
PNA3 (few mismatches)	C**ATG**TCCGAG	pflB	-	0 (15)	250	Formate C-acetyltransferase
PNA4 (few mismatches, in stock)	CCTAC**ATG**CG	ompC	-	1 (9)	1000	Outer membrane porin C
PNA5 (worst case, many possible off-targets)	GACGCC**ATG**A	ybeX	+	23 (215)	195	Putative transport protein
		ybjZ	+		105	Putative inner membrane protein
		aegA	+		27	Oxidoreductase FeS-binding subunit
		hslR	+		58	Heat shock protein 15
		yigC	+		309	Putative inner membrane protein
PNA6 (worst case, many possible off-targets)	GCAAC**ATG**GC	ddlB	+	22 (179)	150	D-alanine--D-alanine

Table 2: acpP mutants. We have PNAs for different double-mutations in the acpP gene. Barbara has prepared RNASeq samples for some of them (last column). She used very high concentrations of PNA compared to Svetlana and Kristina's MICs, so we're not sure whether it is useful to sequence those. Mutant 7 is particularly interesting, as it has full complementarity to kup, a non-essential gene involved in potassium transport.

PNA	Sequence	MIC (μM) Kristina	MIC (μM) (Svetlana)	RNASeq samples (μM) (Barbara)
wildtype	CTCATACTCT	0.6	1.25	-
mutant-1	GACATACTCT	2.5	5	28/56
mutant-2	CAGATACTCT	5	20	-
mutant-3	CTGTTACTCT	> 20	> 20	-
mutant-4	CTCTAACTCT	> 20	> 20	-
mutant-5	CTCAATCTCT	10	20	-
mutant-6	CTCATTGTCT	10	20	-
mutant-7	CTCATAGACT	5	20	54
mutant-8	CTCATACAGT	5	10	36
mutant-9	CTCATACTGA	2.5	5	12

Draft:

Introduction

Background
Most clinically available antibiotics target a large spectrum of microbes by cell wall removal or DNA-destruction (reviewed in (Baquero and Levin 2021)). Antibiotic treatment therefore eradicates a variety of bacteria which can disturb the natural balance of the microbiota of patients, often causing dysbiosis (reviewed in (Vangay et al. 2015)).
To prevent this detrimental side effect, new antibiotic treatments specifically targeting single pathogenic species would be a great step forward in antimicrobial research. As described in (Vogel 2020), sequence specific antisense drugs such as peptide nucleic acid (PNA) have promising attributes such as stability against nucleases. They can kill bacteria by binding to the ribosome binding site (RBS) of an essential gene and thereby evading ribosomal recognition (Nielsen papers).

- Specify issue of off-targeting
- Introducing MASON

Methods

Results

Discussion

Literature

(Dryselius et al. 2003) :
- designed PNA in *E. coli* against the whole mRNA region of beta lactamase and *acpP* genes.
- Found that there is an effect at the start codon and SD region.

(Good et al. 2001):
- Found one bp mismatch to be less efficient in 12 mer *lacZ* PNA (in *E. coli*).
- 9-12 mer PNAs are most effective

(Ghosal and Nielsen 2012):
- They use lots of different PNA combinations with targets around the start region of *acpP* and *ftsZ* in *P. aeruginosa*.
- They show that even PNAs with 2 mismatches in 11 mer PNA is sufficient to inhibit growth (higher MIC, but still possible to inhibit growth), however they did not include scrambled controls to see whether it is due to binding or cell penetration.