GREENLEAF LAB ssDNA SEQEUNCING SUMMARY

GEORGI K. MARINOV, ALEX TREVINO, ZOHAR SHIPONY

2019/12/23

1. TESTING KAS-SEQ

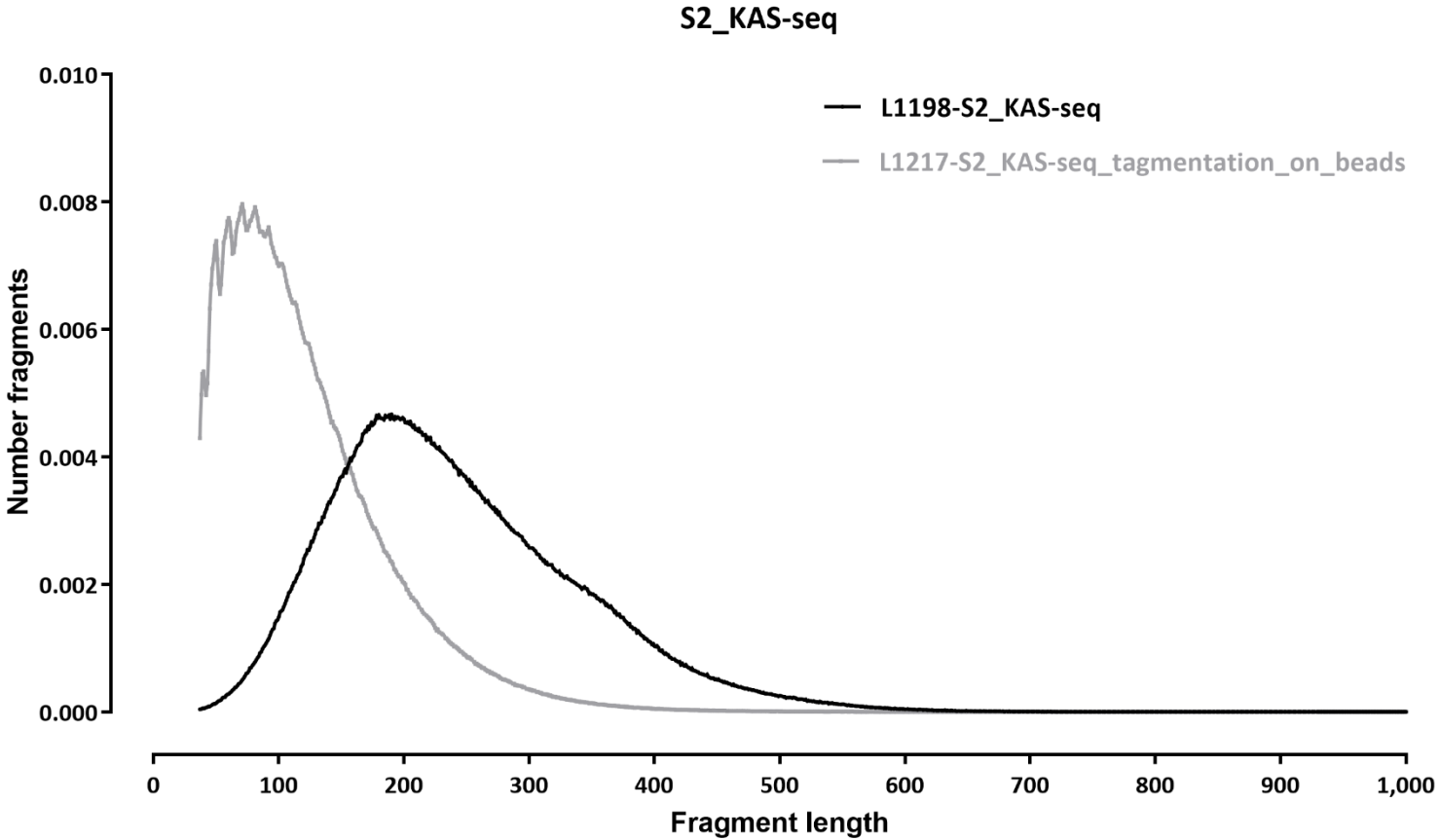
EXPERIMENT SUMMARY

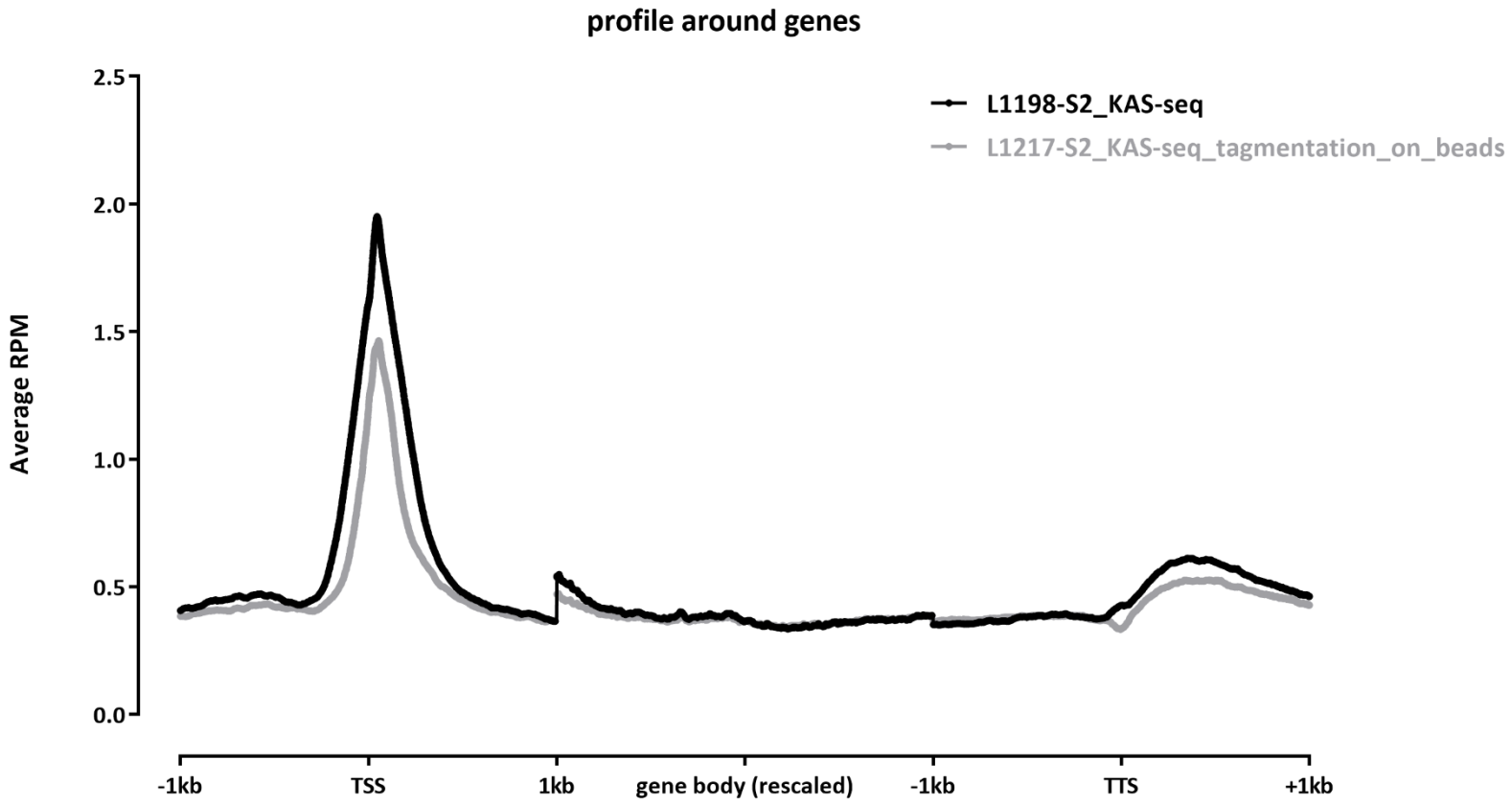
- We did our first test in *Drosophila* S2 cells (we needed to work with small genome, see further below why)

- We changed the protocol a little bit after the biotin capture step, adapting the procedure we have been using for Hi-C instead, which uses amplification directly on beads

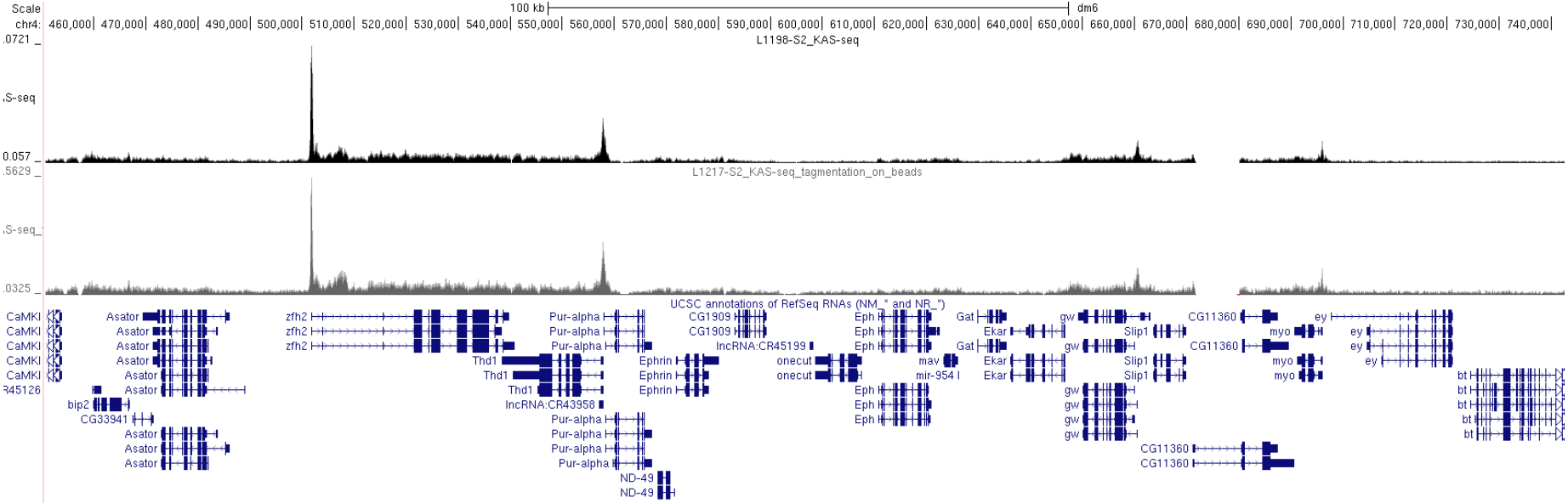
- We also tried tagmentation on beads as a potentially more efficient (and adaptable to high-throughput formats) library building strategy

FRAGMENT LENGTH DISTRIBUTION

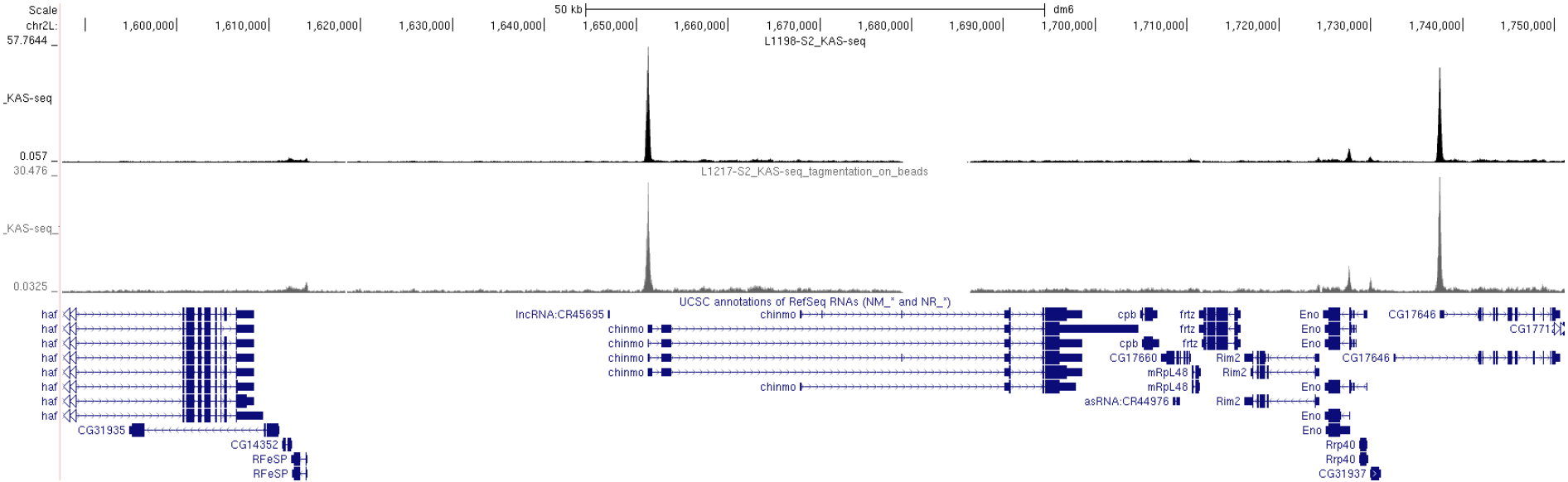



profile around genes
L1198-S2_KAS-seq
L1217-S2_KAS-seq_tagmentation_on_beads
Average RPM
2.5
2.0
1.5
1.0
0.5
0.0
-1kb
TSS
1kb
gene body (rescaled)
-1kb
TTS
+1kb

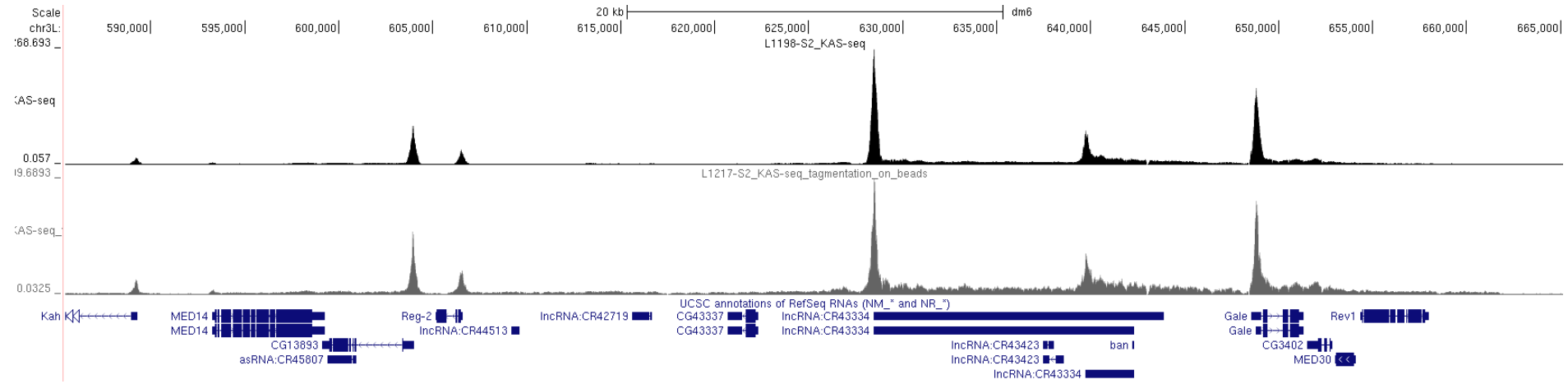
SNAPSHOTS


Scale
100 kb
dm6
chr4: 460,000 470,000 480,000 490,000 500,000 510,000 520,000 530,000 540,000 550,000 560,000 570,000 580,000 590,000 600,000 610,000 620,000 630,000 640,000 650,000 660,000 670,000 680,000 690,000 700,000 710,000 720,000 730,000 740,000
L1198-S2_KAS-seq
L1217-S2_KAS-seq_tagmentation_on_beads
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
CaMKI
Asator
zfh2
Pur-alpha
CG1909
Eph
Gat
gw
CG11360
ey
Thd1
Ephrin
lncRNA:CR45199
onecut
Ekar
Slip1
myo
bip2
CG33941
lncRNA:CR43958
mav
mir-954
bt
ND-49

SNAPSHOTS


Scale
chr2L:
1,600,000
1,610,000
1,620,000
1,630,000
1,640,000
1,650,000
1,660,000
1,670,000
1,680,000
1,690,000
1,700,000
1,710,000
1,720,000
1,730,000
1,740,000
1,750,000
50 kb
dm6
57.7644
L1198-S2_KAS-seq
_KAS-seq
0.057
30.476
L1217-S2_KAS-seq_tagmentation_on_beads
KAS-seq
0.0325
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
haf
haf
haf
haf
haf
haf
haf
haf
CG31935
CG14352
RFeSP
RFeSP
lncRNA:CR45695
chinmo
chinmo
chinmo
chinmo
chinmo
chinmo
cpb
cpb
CG17660
mRpL48
mRpL48
asRNA:CR44976
frtz
frtz
frtz
Rim2
Rim2
Rim2
Eno
Eno
Eno
Eno
Eno
Rrp40
Rrp40
CG31937
CG17646
CG17646
CG17714

SNAPSHOTS


Scale
chr3L:
20 kb
dm6
590,000
595,000
600,000
605,000
610,000
615,000
620,000
625,000
630,000
635,000
640,000
645,000
650,000
655,000
660,000
665,000
L1198-S2_KAS-seq
0.057
L1217-S2_KAS-seq_tagmentation_on_beads
0.0325
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
Kah
MED14
MED14
Reg-2
lncRNA:CR44513
lncRNA:CR42719
CG13893
asRNA:CR45807
CG43337
CG43337
lncRNA:CR43334
lncRNA:CR43334
lncRNA:CR43423
lncRNA:CR43423
lncRNA:CR43334
ban
Gale
Gale
Rev1
CG3402
MED30

PROTOCOL TEST SUMMARY

- KAS-seq works successfully in our hands

- Both ligation-based and tagmentation-based on-beads protocols work fine, though, at least in this test, the ligation resulted in a slightly better signal-to-noise ratio

2. SINGLE-MOLECULE READOUT (nanoKAS)

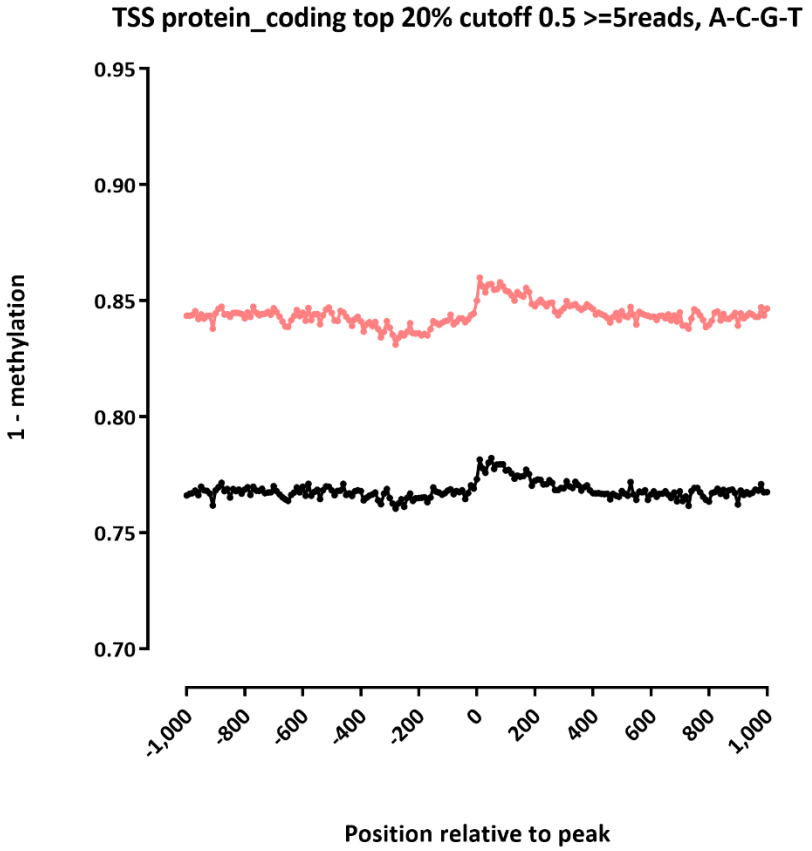
POSSIBILITIES AND ADVANTAGES

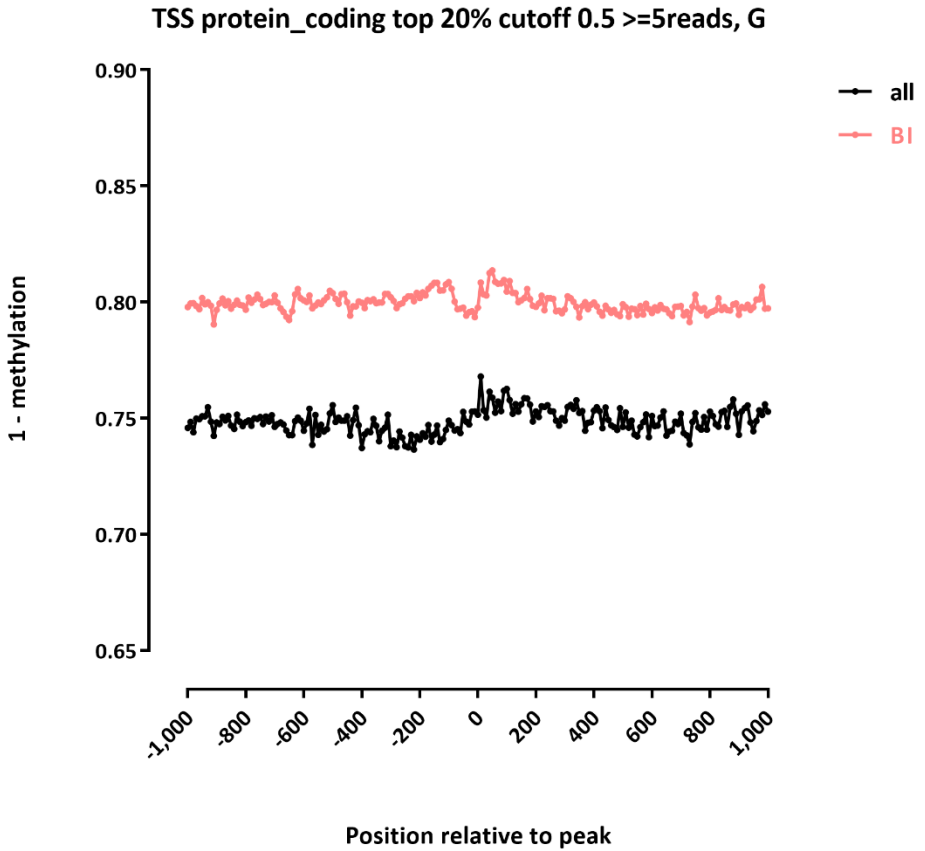
- Ketoxal is a bulky modification
- Should be detectable very well using nanopore sequencing
- Would provide base-pair readout of ssDNA
- Would provide linked long-range single-molecule readout of ssDNA
- We sequenced one MinION flowcell of HMW DNA from the same S2 experiment (prior to the click reaction)
- Note: this is why we tested in S2 cells – we were aiming at obtain sufficiently deep coverage to try some single-molecule analysis, thus a small genome being beneficial

RESULTS FROM THE FIRST 800K NANOPORE READS:

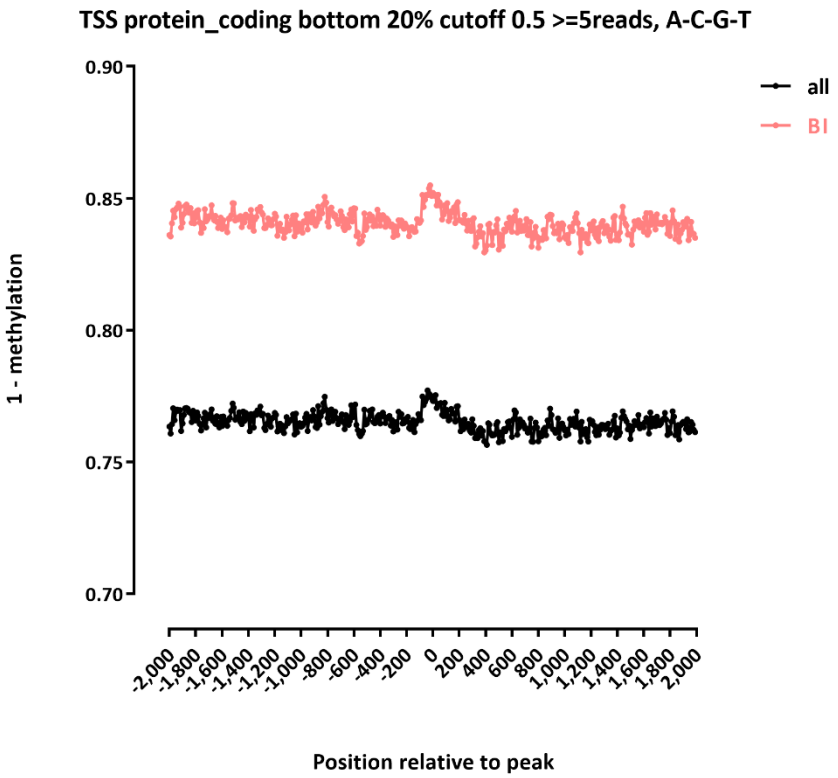
Species	Assay	#	Mapping	Base Contexts	Reads	Total Bases	Mean Read Length	Median Read Length
dm3	**ssDNA**	2019-11-06-S2-ssDNA-800k	**Tombo 1.5 denovo**	**generic-A-C-T-G**	533,199	975,589,084	1,830	806
dm3	**ssDNA**	2019-11-06-S2-ssDNA-800k	**Tombo 1.5 denovo**	**generic-G**	532,664	969,620,818	1,820	797

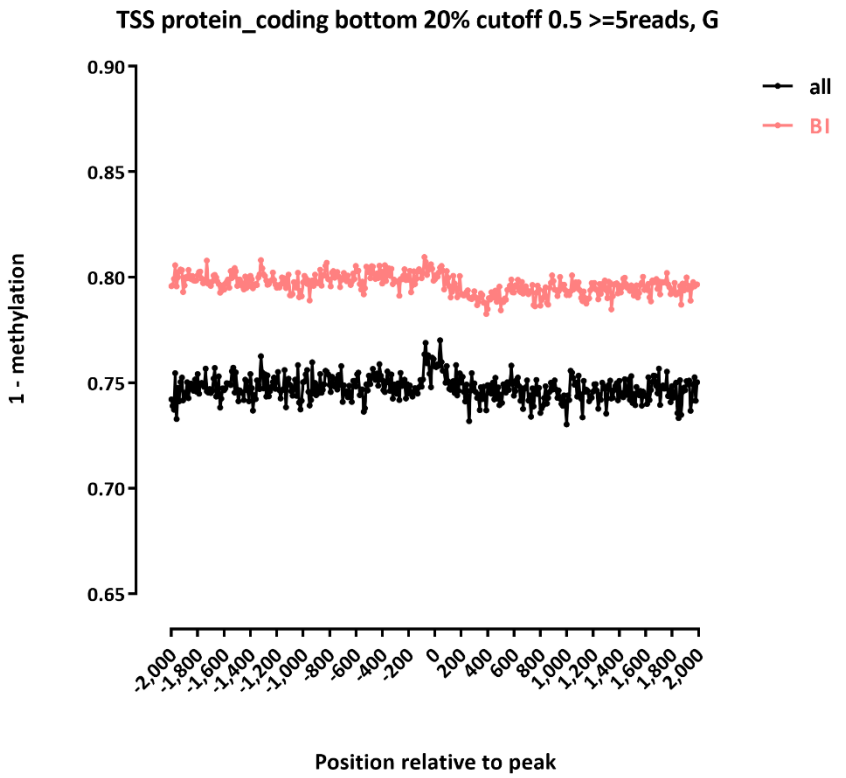
DETECTED MODIFICATION LEVELS AROUND HIGHLY EXPRESSED TSSs



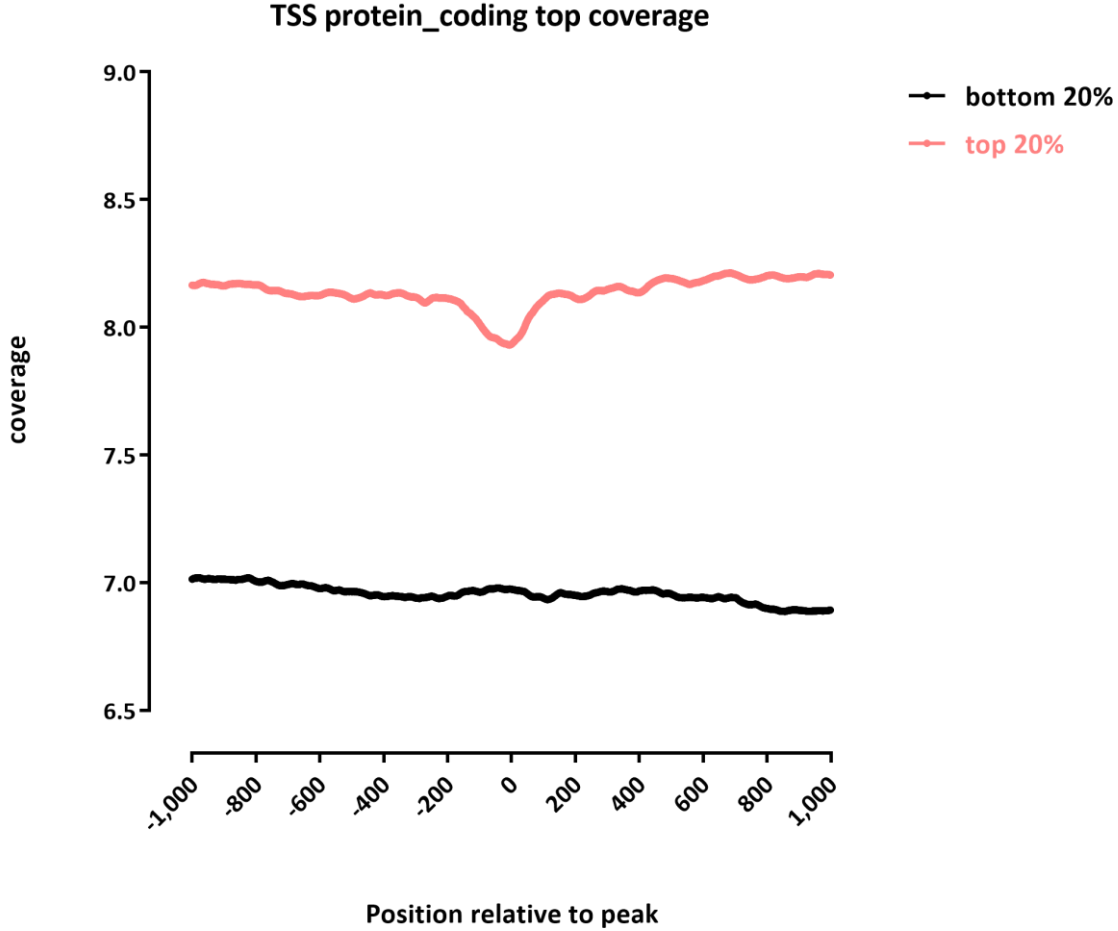


DETECTED MODIFICATION LEVELS AROUND SILENT TSSs





RAW READ COVERAGE



nanoKAS SUMMARY

- We were not able to obtain strong enrichment signal around TSS

- However, this is not for failure of the kethoxal modification to induce a current change through the pore

- The problem appears instead to be that kethoxal is causing the software to throw away modified reads once the modification is encountered (thus the dip in coverage around the TSS at highly expressed genes), which is because it is inducing very strong current changes

- We are in discussion with ONT about how to resolve these issues (we are having the same problem on several other projects right now). It is a purely algorithmic problem, not a physical one at the pore

3. EXPERIMENTS WITH CAS9

KAS-SEQ TEST IN PROKARYOTES

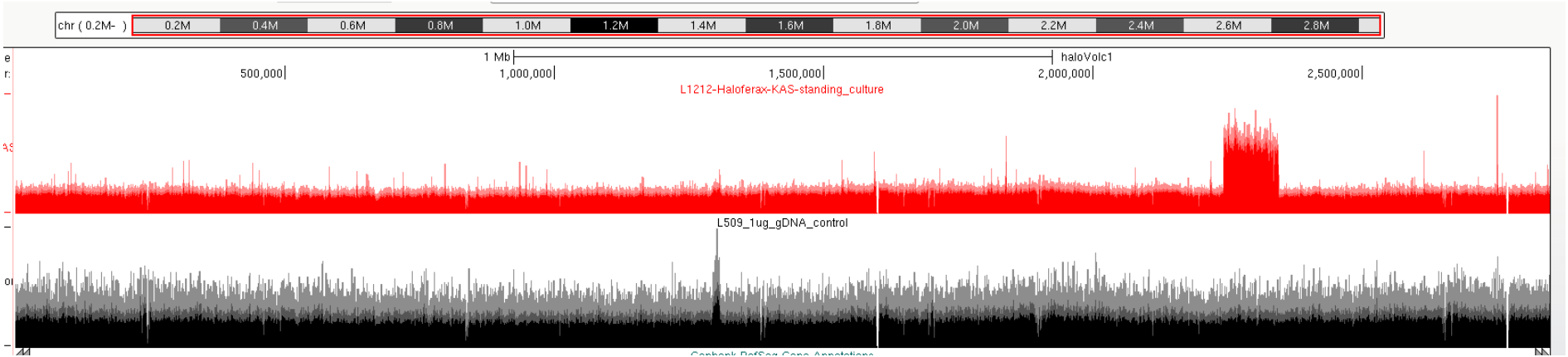
- In addition to S2 cells, we tested the protocol on a prokaryotic genome, using a *Haloferax volcanii* culture we had growing in the lab

- *Haloferax* is a species of halophilic archaea

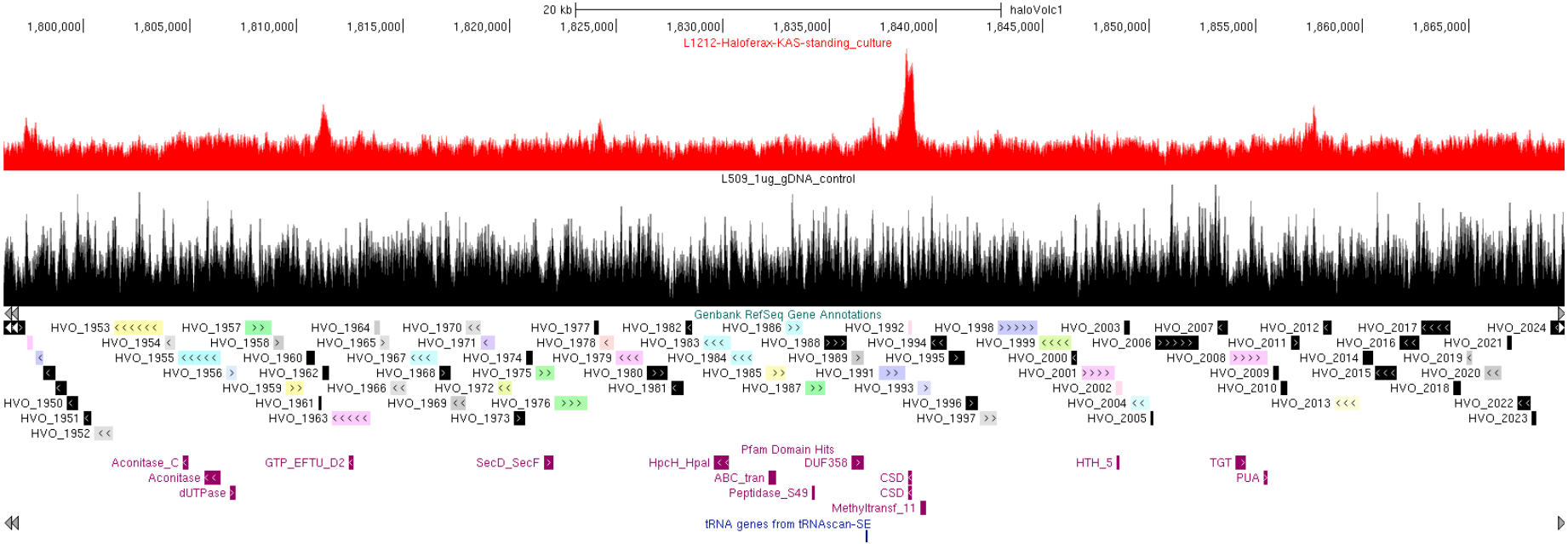
- It has a main chromosome and 5 plasmids

- The experiment was done on a dormant room-temperature culture (*Haloferax* normally grows at 45C)

MAIN CHROMOSOME

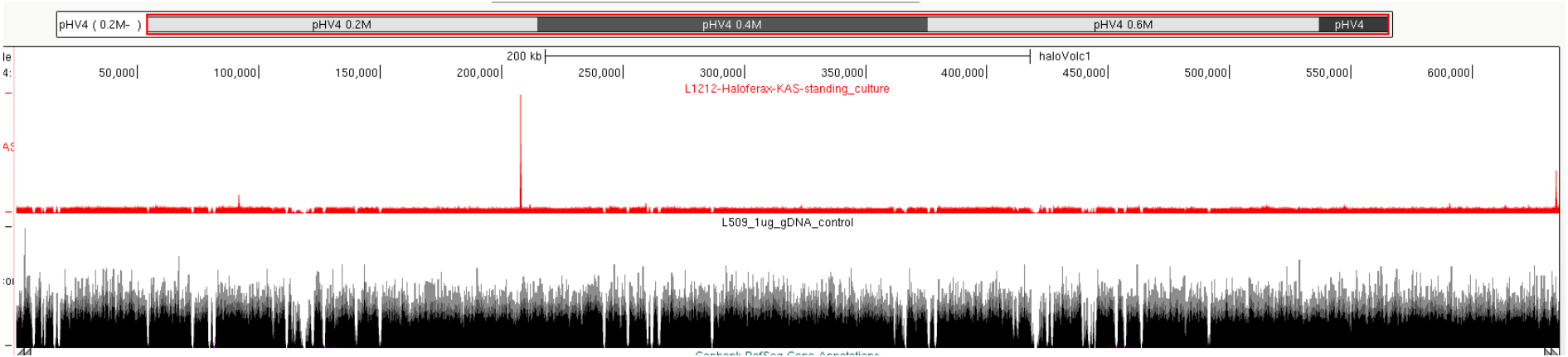


MAIN CHROMOSOME, ZOOM IN



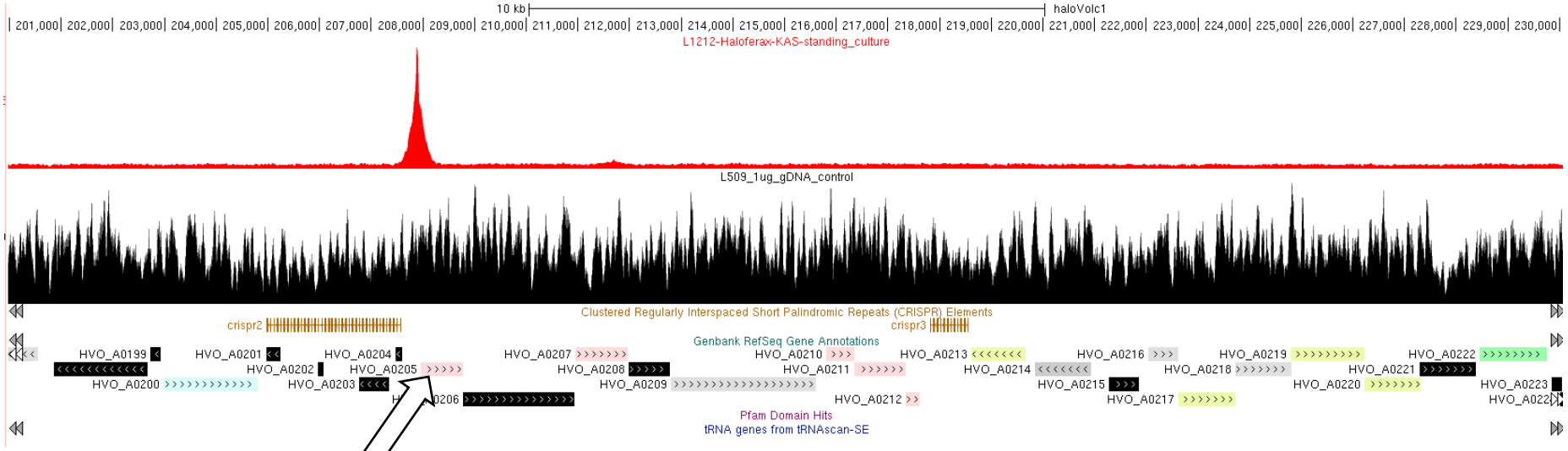
- We observe only modest enrichment at TSSs, consistent with the source material being a dormant culture

HOWEVER:



- <u>A huge peak is observed on one of the plasmids!</u>

ZOOM IN



this is Cas6

- There are 3 CRISPR arrays in *Haloferax*, the other one is on the main chromosome. Only this one shows a KAS peak

- We do not know right now why the CRISPR array beginning is so strongly enriched for ssDNA, but it made us think about CRISPRs (of coyrse, this clearly has implications about basic CRISPR biology that could be followed on at some point)

dCAS9/CAS9 TARGET MAPPING

- We do not know right now why the CRISPR array beginning is so strongly enriched for ssDNA, but it made us think about CRISPRs

- We have done quite a bit of work on studying Cas9 specificity in recent years, and there are many unresolved problems in the field

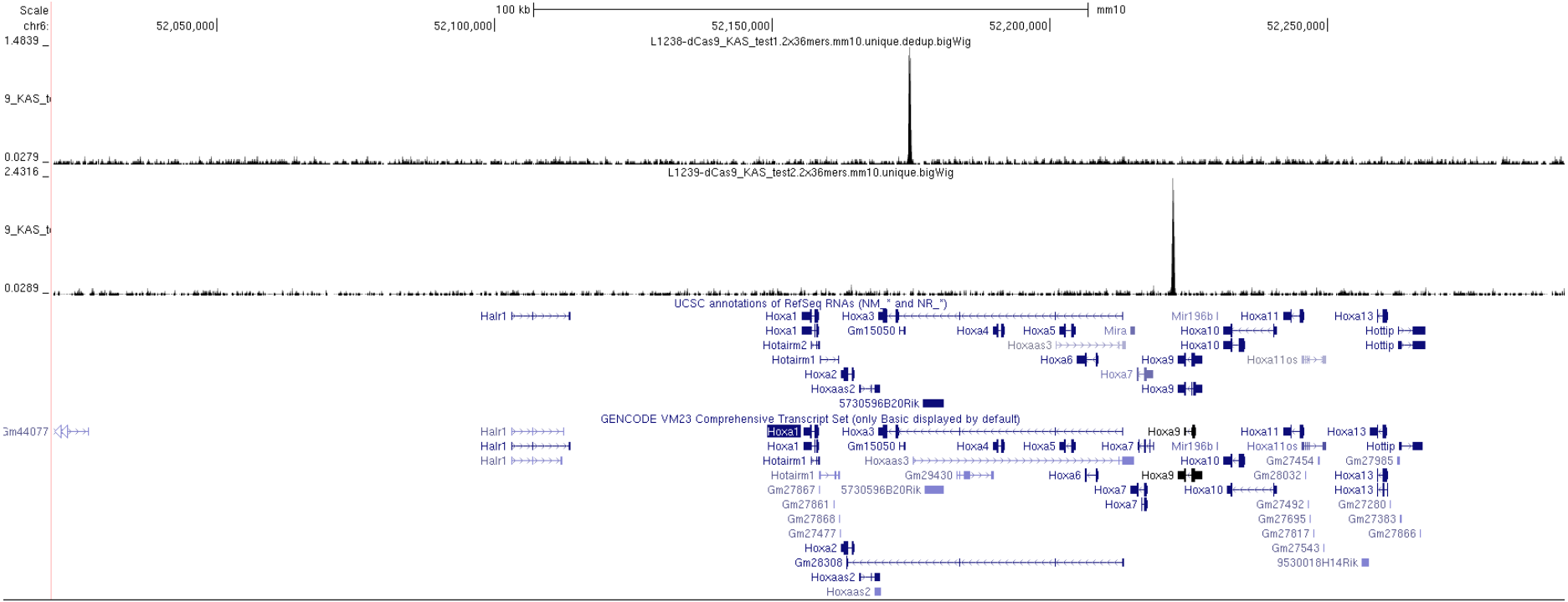
- Existing methods for mapping off-target sites are all very complicated and difficult to carry out

- In addition, there is no good method for mapping dCas9 off-target sites, as existing methods all rely on mapping actual cuts and use the catalyitically active Cas9 version. ChIP-seq is too noisy, plus it does not distinct actual strand invasion from other types of binding

- We reasoned kethoxal should be able to label sites of strand invasion

- We carried out a test using two sgRNAs targeting the mouse *HOXA* locus, by carrying out a ketoxal reaction on an *in vitro* sample of a dCas9 sgRNA RNP that was incubated with mouse gDNA

RESULTS


Scale
100 kb
mm10
chr6:
52,050,000
52,100,000
52,150,000
52,200,000
52,250,000
1.4839
L1238-dCas9_KAS_test1.2x36mers.mm10.unique.dedup.bigWig
0.0279
2.4316
L1239-dCas9_KAS_test2.2x36mers.mm10.unique.bigWig
0.0289
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
Halr1
Hoxa1
Hoxa3
Mir196b
Hoxa11
Hoxa13
Gm15050
Hoxa4
Hoxa5
Mir
Hottip
Hotairm2
Hoxaas3
Hotairm1
Hoxa6
Hoxa9
Hoxa11os
Hoxa2
Hoxa7
Hoxaas2
5730596B20Rik
GENCODE VM23 Comprehensive Transcript Set (only Basic displayed by default)
Gm44077
Gm29430
Gm27454
Gm27985
Gm28032
Gm27867
Gm27861
Gm27868
Gm27477
Gm28308
Gm27492
Gm27280
Gm27695
Gm27383
Gm27817
Gm27866
Gm27543
9530018H14Rik

dCAS9/CAS9 TARGET MAPPING SUMMARY

- Kethoxal seems to be an extremely promising way of mapping Cas9 targets (“CasKAS”)
- Easier and quicker than anything currently existing
- Mapping strand invasion rather than binding
- Could in principle be adapted to a high-throughput format in plates to profile numerous guides in parallel

4. STRAND-SPECIFIC KAS (ssKAS)

SUMMARY

- We have reasons to think that kethoxal labeling is not equal on both strands of the DNA, at least in some contexts
- Thus a strand-specific version of the protocol could reveal interesting new biology
- Nanopore sequencing will readily provide that information, once the readout issue is resolved, but there are Illumina-based approaches too, plus Illumina is much easier to work with in general
- We have not gotten around to sequencing either of them, but we are testing two approaches to convert denatured post-click DNA that has been captured into libraries
- The first one is an adaptation of the PBAT protocol for carrying out whole-genome bisulfite sequencing, which uses random priming to carry out a first/second strand synthesis off a template (thus preserving the strand specificity)
- The second one is to ligate adapters with random-sequence overhangs onto the 3’ end of the captured DNA on beats and carry out a first/second strand synthesis
- We will report on the success of these approaches after the break

5. SINGLE-CELL KAS (scKAS)

SUMMARY:

- We have debated a lot internally on how to approach the single-cell problem

- The challenge is that we need to simultaneously:
  - carry out the click reaction
  - enrich for modified DNA
  - index cells

- These three goals do not fit directly into any of the existing single-cell protocols

- We are testing two main strategies right now, on the success of which we will report after the break

APPROACH 1:

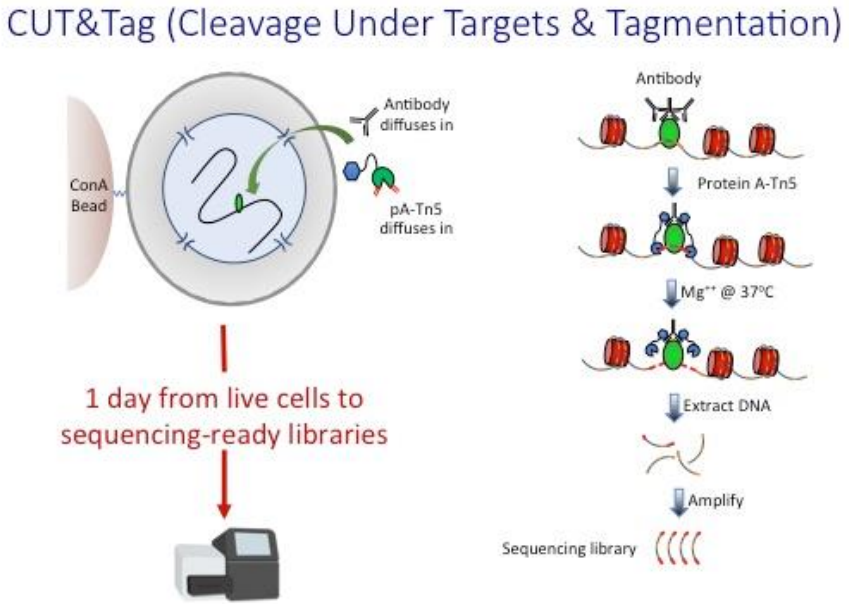
- This approach uses the 10X platform
- Kethoxal labeling is carried out in cells
- Cells are then fixed (we need to do this in order to “freeze” things in place and stop DNA repair reactions from eliminating the kethoxal)
- The click reaction is the carried out *in situ*
- Cells are then denatured in SDS (this approach has been used in multiple single-cell studies, and it preserves the cells as long as they are fixed)
- Then we feed the cells into the standard 10X scATAC protocol, where tagmentation and indexing happen in droplets (this is why we need to denature, in order to not convolve the KAS signal with the accessible chromatin signal)
- Then once GEMs are broken up, we do a biotin pull-down and library amplification

APPROACH 2:

- This approach adapts the CUT&Tag protocol for mapping protein-DNA interactions

- CUT&Tag uses a proteinA-Tn5 fusion to direct Tn5 to the sites where the protein of interest (labelled by a primary antibody) is located in the genome

- The advantage is that Tn5 directly adds primer landing sites to DNA fragments, which can then be amplified



APPROACH 2:

- This approach adapts the CUT&Tag protocol for mapping protein-DNA interactions
- Kethoxal labeling is carried out in cells
- Cells are then fixed (we need to do this in order to “freeze” things in place and stop DNA repair reactions from eliminating the kethoxal)
- The click reaction is then carried out *in situ*
- We may or may not have to denature in this case, in principle we think it could work without denaturing
- We are trying two strategies to bring Tn5 to kethoxal sites:
    - Use an anti-biotin antibody to bring the pA-Tn5 to ssDNA sites
    - Incubate with streptavidin, then use an anti-streptavidin antibody to bring pA-Tn5
- We expect the second strategy to work better
- The CUT&Tag approach also has the advantage of being scalable, as it can be directly adapted to a combinatorial indexing format, thus greatly increasing throughput