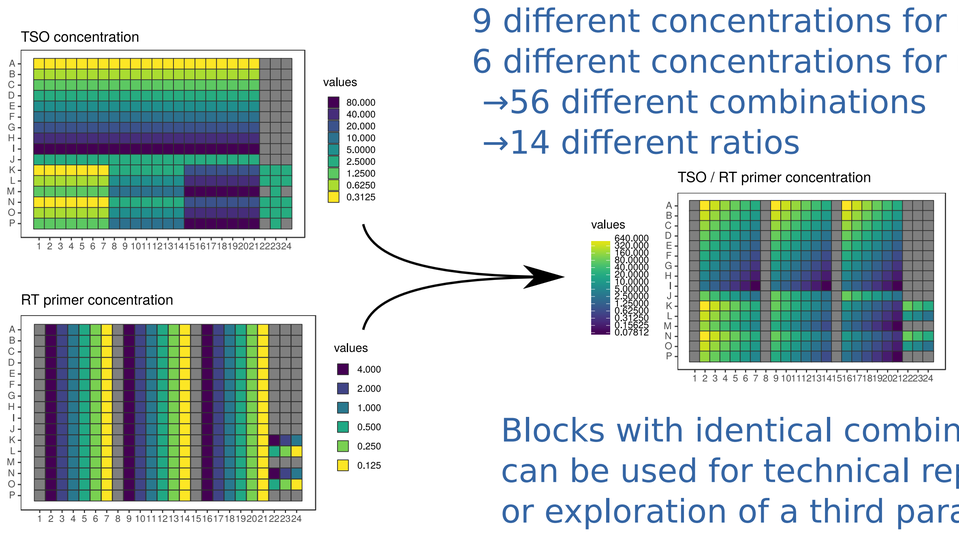
Machine-driven parameter-space exploration of biochemical reactions

Charles Plessy[1,2,3,4], Stéphane Poulain[1,2], Sachi Kato[1,2], Ophélie Arnaud[2], Piero Carninci[1,2]

1: RIKEN Center for Integrative Medical Sciences, Division of Genomic Medicine
2: RIKEN Center for Life Science Technologies, Division of Genomic Technologies
3: Present address: Okinawa Institute of Science and Technology Graduate University, Genomics and Regulatory Systems Unit.
4: For correspondence: charles.plessy@oist.jp

Acknowledgements: *we would like to thank Iris Chen, Hiro Ishida, Carl Jarman and Mark Lasinski from Labcyte, and Hisakazu Komine from Kiko-Tech for making these experiments possible.*
Funding: *MEXT grant to RIKEN CLST, JSPS Grand-in-Aid for Scientific Research S*

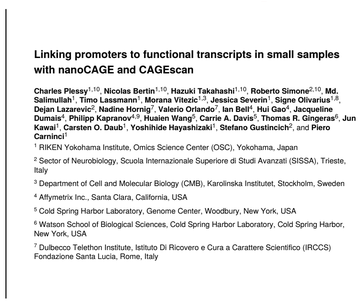
Nanoliter-scale reactions allow for testing hundreds of parameter combinations



9 different concentrations for reagent 1
6 different concentrations for reagent 2
→56 different combinations
→14 different ratios

Blocks with identical combinations can be used for technical replication or exploration of a third parameter

Proof of principle: optimisation of a reverse transcription reaction used in transcriptome sequencing



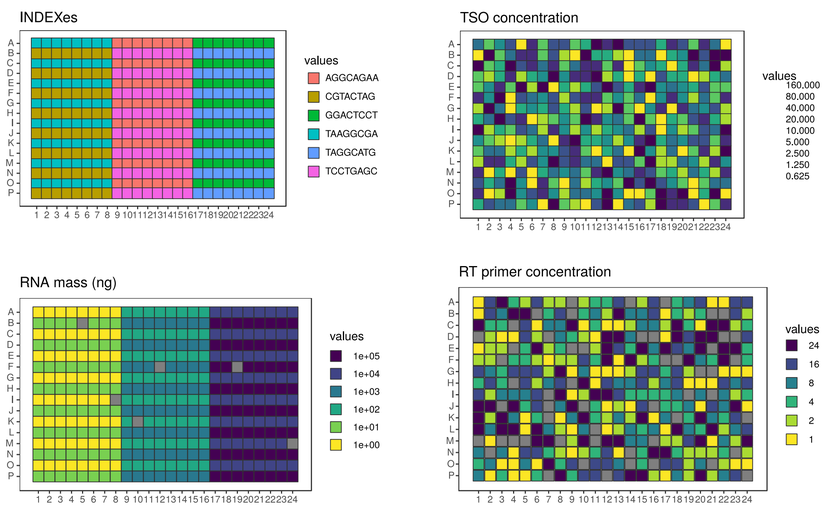
Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan

Charles Plessy[1,10], **Nicolas Bertin**[1,10], **Hazuki Takahashi**[1,10], **Roberto Simone**[2,10], **Md. Salimullah**[1], **Timo Lassmann**[3], **Morana Vitezic**[1,3], **Jessica Severin**[1], **Signe Olivarius**[1,8], **Dejan Lazarevic**[2], **Nadine Hornig**[7], **Valerio Orlando**[7], **Ian Bell**[4], **Hui Gao**[4], **Jacqueline Dumais**[4], **Philipp Kapranov**[4,9], **Huaien Wang**[5], **Carrie A. Davis**[5], **Thomas R. Gingeras**[6], **Jun Kawai**[1], **Carsten O. Daub**[1], **Yoshihide Hayashizaki**[1], **Stefano Gustincich**[2], and **Piero Carninci**[1]

[1] RIKEN Yokohama Institute, Omics Science Center (OSC), Yokohama, Japan
[2] Sector of Neurobiology, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[3] Department of Cell and Molecular Biology (CMB), Karolinska Institutet, Stockholm, Sweden
[4] Affymetrix Inc., Santa Clara, California, USA
[5] Cold Spring Harbor Laboratory, Genome Center, Woodbury, New York, USA
[6] Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA
[7] Dulbecco Telethon Institute, Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS) Fondazione Santa Lucia, Rome, Italy

As a proof of principle, we explored the reaction parameters critical to the template-switching reaction that takes place in the reverse-transcription step of the nanoCAGE method (for quantitative sequencing of RNA start sequences). In addition to the reverse-transcription primer, these reactions use an additional "*template-switching*" oligonucleotide to introduce linker sequences at the cDNA end. Optimal concentrations for both oligonucleotides were assessed at RNA amounts varying from 10 pg (single-cell scale) to 100 ng (bulk sample scale). nanoCAGE libraries were sequenced at low cost on a MiSeq sequencer.

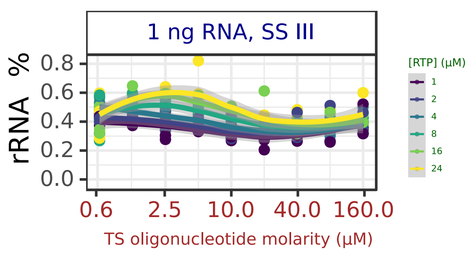
Randomisation removes spatial biases. Controls coordinates identify replicates



We used a high-performance, contact-less, nanoliter-scale liquid transfer platform, the *Labcyte Echo 525*, to assemble the reactions after randomising their coordinates. Volume transfer sheets were produced by a R script simulating the machine. Open-sourcing possible upon request.

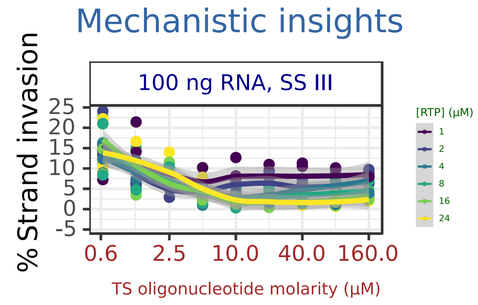
Examples of findings

Non-linear patterns



The high dynamic range and extensive replication are needed for observing complex relations

Mechanistic insights

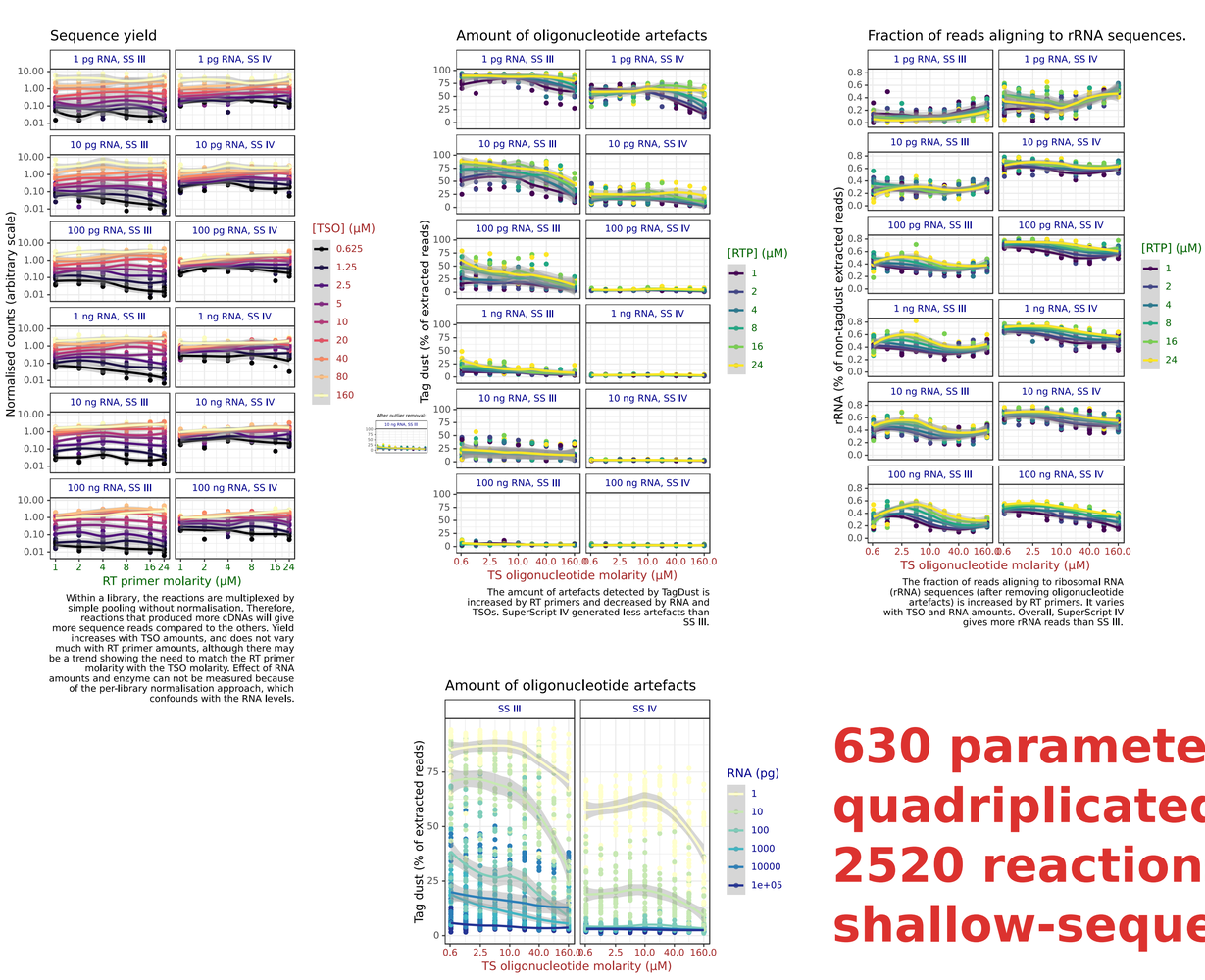


Systematic exploration of the parameter space is essential for finding counter-intuitive results. Here, we show that the amount of artefacts generated by the TS oligonucleotide decrease when its concentration is increased.

Perspectives

→ Broadly applicable as long as the readout cost per reaction is low (qPCR, fluorometry, multiplexed quantitative sequencing)
→ More complex strategies (random parameter values, ….)
→ Further automatisation of experiment design and execution (AI).

Our results in details:

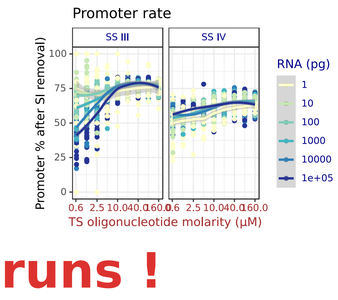


Within a library, the reactions are multiplexed by simple pooling without normalisation. Therefore, reactions that produced more cDNAs will give more sequence reads compared to the others. Yield increases with TSO amounts, and does not vary much with RT primer amounts, although there may be a trend showing the need to match the RT primer molarity with the TSO molarity. Effect of RNA amounts and enzyme can not be measured because of the per-library normalisation approach, which confounds with the RNA levels.

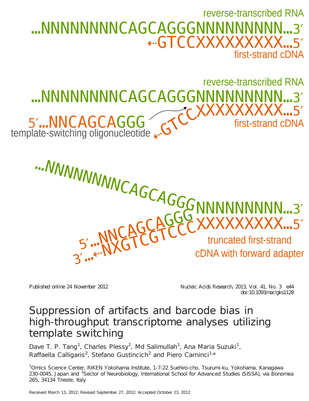
The amount of artefacts detected by TagDust is increased by RT primers and decreased by RNA and TSOs. SuperScript IV generated less artefacts than SS III.

The fraction of reads aligning to ribosomal RNA (rRNA) sequences (after removing oligonucleotide artefacts) is increased by RT primers. It varies with TSO and RNA amounts. Overall, SuperScript IV gives more rRNA reads than SS III.

Promoter rates increases with TSO and RT primer molarity. It is higher with SuperScript III than with SS IV.



630 parameter combinations quadriplicated in a total of 2520 reactions (plus controls), shallow-sequenced in 2 MiSeq runs !



Published online 24 November 2012 Nucleic Acids Research, 2013, Vol. 41, No. 3 e44 doi:10.1093/nar/gks1128

Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching

Dave T. F. Tang[1], Charles Plessy[1], Md Salimullah[1], Ana Maria Suzuki[1], Raffaella Calligaris[2], Stefano Gustincich[2] and Piero Carninci[1,*]

[1]Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan and [2]Sector of Neurobiology, International School for Advanced Studies (SISSA), via Bonomea 265, 34134 Trieste, Italy

Received March 13, 2012; Revised September 27, 2012; Accepted October 23, 2012