SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

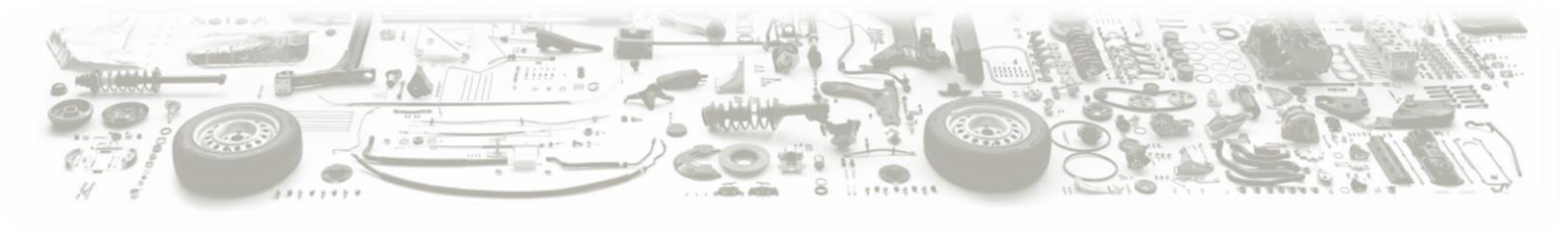
S1.2

scM& T-seq: A method for parallel bisulfite and RNA-sequencing in single cells

Stephen Clark[1], Christof Angermueller[2], Heather Lee[1], Iain Macaulay[3], Tim Hu[4], Felix Krueger[1], Sebastien Smallwood[1], Chris Ponting[4], Thierry Voet[5], Gavin Kelsey[1], Oliver Stegle[2], Wolf Reik[1]

[1]*Babraham Institute, Cambridge, United Kingdom*
[2]*European Bioinformatics Institute, Cambridge, United Kingdom*
[3]*Wellcome Trust Genome Campus, Cambridge, United Kingdom*
[4]*University of Oxford, Oxford, United Kingdom*
[5]*KU Leuven, Leuven, Belgium*

Our current understanding of how DNA methylation is associated with gene transcription is based on measurements from bulk cell samples which obscure inter-cellular differences and prevent analyses of rare cell types. Thus, parallel sequencing of the methylome and transcriptome in single cells will make important contributions to understanding both normal development and disease. Our group recently demonstrated the feasibility of genome-wide bisulfite sequencing to profile methylation at almost 50% of all CpG loci from a single cell (scBS-seq). We have since optimised our protocol in order to substantially increase throughput using an automated liquid handling robot and to allow compatibility with G& T-seq, a recently published method for the separation of full-length cDNA and genomic DNA from the same single cell. Using our updated method, scM& T-seq (methylome and transcriptome), we have profiled whole genome DNA methylation and gene expression in mouse embryonic stem cells with equivalent levels of accuracy of either technique performed in isolation. By analysing epigenomic and transcriptomic data from the same cells we identify novel correlations underlying possible regulatory mechanisms that were previously masked in bulk cell analyses thereby demonstrating the utility of multi-dimensional single cell sequencing.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

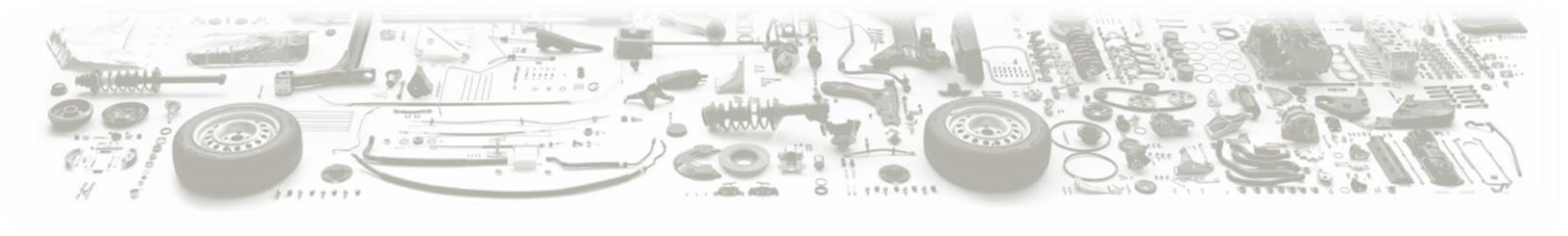
S1.5

Simultaneous transcriptomic and lineage analysis within a single cell

Rahul Satija[1,2], Christian Mayer[2], Rachel Bandler[2], Ashley Powers[1], Gord Fishell[2]

[1]*New York Genome Center, New York City, NY, United States*
[2]*NYU Medical Center, New York, NY, United States*

New technologies for single cell genomics offer transformative potential for the unbiased discovery and molecular characterization of cell types, states, and developmental transitions. However, despite the importance of the lineage relationships that connect these populations, these hierarchies cannot be accurately inferred from gene expression information alone. For example, while high throughput single cell profiling has revealed multiple new classes of cortical interneurons in the mouse brain, little is known about the phylogenetic relationships between these groups or the progenitor classes that give rise to each class. To address this, we present a novel approach to simultaneously measure the transcriptome and lineage of a single cell. We combine genetic lineage tracing approaches, where progenitor cells are retrovirally-labeled with a unique genomic barcode that is maintained in its progeny, with high-throughput single cell genomics, enabling us to sequence both the cell's transcriptome as well as its lineage barcode. We first validate our approach on clonally derived populations of HEK293 cells, and demonstrate how attempts to amplify the barcode from either the genome and transcriptome return the same lineage tags. We apply this approach in an initial study of both embryonic and adult murine interneurons, and believe that these methods will enable similar studies across distinct biological disciplines. Our protocol is compatible with 3' end tagging strategies (i.e Drop-Seq), which will enable joint studies of lineage and gene expression to be extended to tens of thousands of single cells.



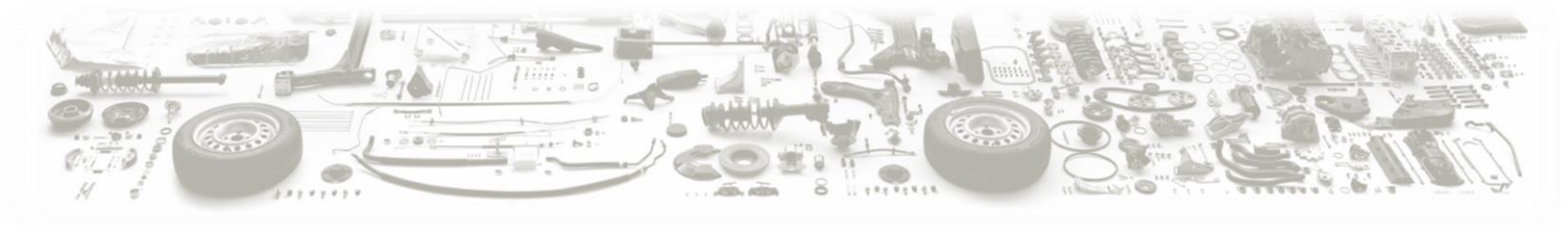
S2.4

Spatial organization shapes the turnover of a bacterial transcriptome

Jeffrey Moffitt, Xiaowei Zhuang

Harvard University, Cambridge, MA, United States

Single molecule fluorescence in situ hybridization (smFISH) has been an instrumental tool in the study of both the spatial distribution and abundance of RNAs within cells and tissues, yet its application to systems-level questions has been limited by the relatively small number of mRNA species that can be stained simultaneously in single cells. Here we extend smFISH to the transcriptome scale by developing a method to create diverse but precisely defined sets of smFISH probes that can be used to simultaneously stain tens to thousands of mRNA species and to image these RNAs with nanometer-scale resolution. Using this approach, we mapped the spatial organization of the E. coli transcriptome by grouping and staining all mRNAs that share common properties, such as the genomic location of the encoding gene, or the location or function of the encoded protein. We find that the E. coli transcriptome is organized with mRNAs that encode inner-membrane proteins primarily localized at the membrane whereas all other mRNAs are scattered throughout the cell. This membrane enrichment is controlled by the translation of signal peptides recognized by the signal recognition particle. Using time-resolved RNA-seq, we measured the lifetimes of all E. coli mRNAs and found that membrane-localized mRNAs are degraded significantly faster than mRNAs that are distributed throughout the cytosol. This destabilization arises because core members of the RNA degradation machinery are also membrane enriched. Our work reveals that the spatial organization of the transcriptome can shape the post-transcriptional fate of bacterial mRNAs. The method described here can be used to probe the spatial organization of the transcriptome and its physiological consequences in many different cell types.



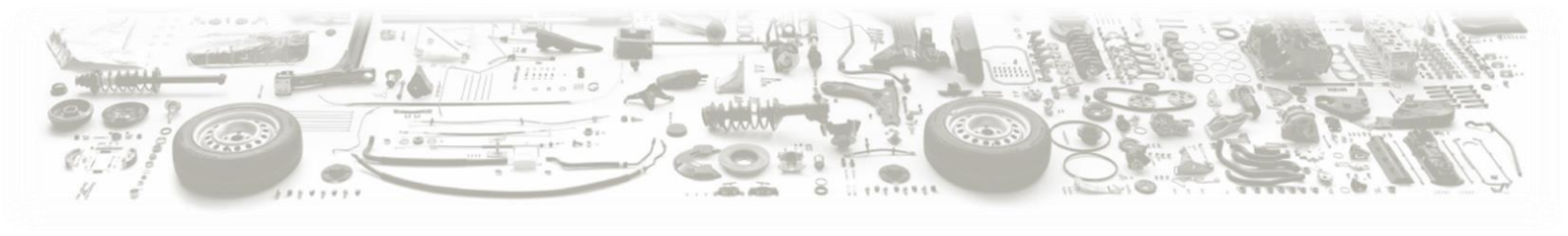
S2.5

In situ RNA profiling by FISH SCALYS (Sequential Coding anALYSis)

Long Cai

Caltech, Pasadena, CA, United States

I will report on our new results using FISH SCALYS to profiled gene expression in a diverse range of tissues. In FISH SCALYS, mRNAs in cells are barcoded by sequential rounds of hybridization, imaging, and probe stripping. The number of barcodes available with this approach scales exponentially as F^N, where F is the number of distinct fluorophores and N is the number of hybridization rounds. In particular, I will discuss our work in gene profiling in brain slices as well as in developing embryos.



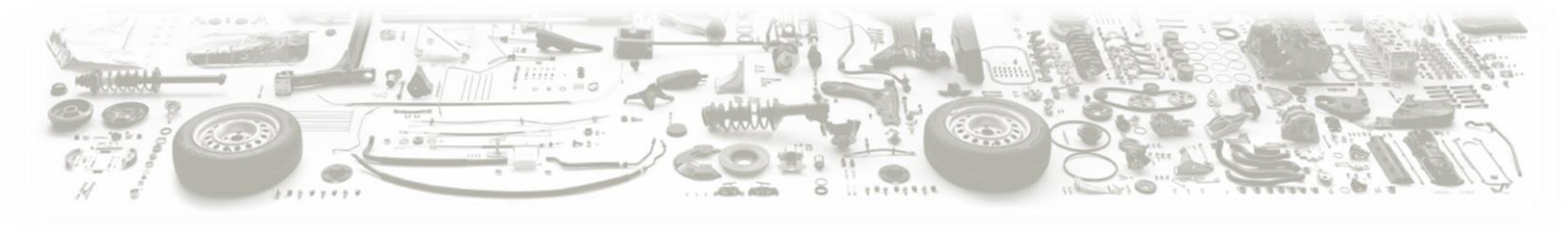
S3.3

Single-cell genomics allows detection of rare cells in primary lung cancer with genomic alterations present in most metastatic tumor cells

Diana Spierings, Ali Saber, Nancy Halsema, Hilda van den Bos, David Porubsky, Aaron Taudt, Maria Colome-Tatche, Klaas Kok, Anke van den Berg, Harry Groen, Peter Lansdorp

University Medical Center Groningen, Groningen, Netherlands

Small-cell lung cancer (SCLC) is characterized by aggressive growth and high metastatic potential. Genomic heterogeneity within a tumor has been thought to be relevant to cancer resistance towards chemotherapy. Unfortunately, classic genomic profiling by comparative genomic hybridization (CGH) or whole-exome sequencing (WES) of bulk material can only identify global genomic alterations of a cell population.
In the present study, we used single-cell whole genome sequencing (WGS) to assess copy number variations (CNVs) in a primary SCLC tumor and its liver metastasis. Single nuclei isolated from frozen tissue sections were flow sorted and WGS libraries were prepared without upfront whole genome amplification to avoid amplification biases. Subsequently 192 barcoded single-cell WGS libraries (96 per sample) were sequenced on a single lane of the Illumina platform resulting in sufficient coverage to determine CNVs.
Hierarchical clustering of single-cell CNV profiles showed clear differences in the copy number of chromosomes 11, 16, 18 and 22 between the primary tumor and metastasis in agreement with previous bulk WES data of the same material. Further examination of the single-cell CNV profiles revealed more heterogeneity within the primary tumor than was observed in the metastatic tumor. Interestingly, within the heterogeneous population of the primary tumor, two cells displayed a CNV pattern similar to that observed in most metastasis tumor cells. Rare cells cannot be detected when assessing CNVs using CGH or WES of bulk material. In conclusion, our single-cell WGS method is a powerful technique to assess intra-tumor heterogeneity providing insight into tumor cell evolution and metastasis.



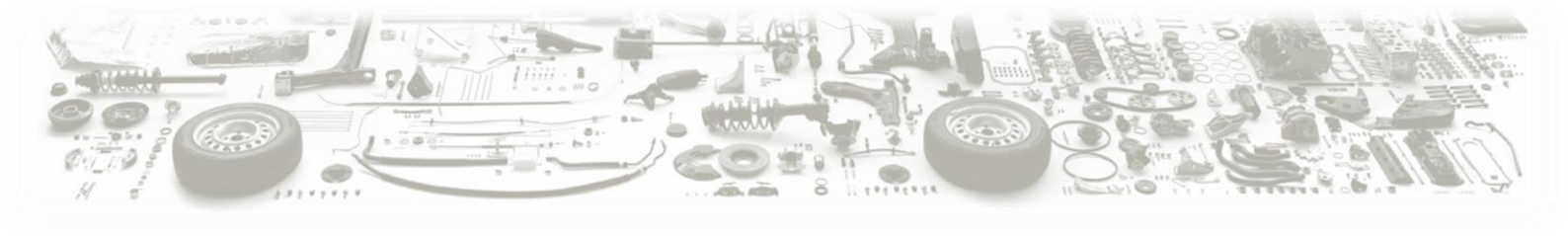
S3.5

Linking transcriptional and genetic intratumoral heterogeneity

Peter Kharchenko

Harvard Medical School, Boston, MA, United States

The ability to directly assess cell heterogeneity is particularly critical in the context of cancer therapy, where presence of phenotypically distinct subclonal populations fuels relapse and resistance to treatment. The transcriptional heterogeneity within such tumors and its impact on disease progression is poorly understood. Furthermore, the relationship between better-studied genetically-distinct subclonal populations and transcriptional heterogeneity cannot be easily assessed with the current bulk analysis or single-cell techniques.
To investigate these questions we have developed a computational approach to infer subclonal hierarchy based on the single-cell RNA-seq measurements. The sequence information available in the RNA-seq reads is used to assess the probability of single nucleotide variants and structural rearrangements in a given cell, taking into account biases from monoallelic expression inherent to single-cell measurements. Bayesian phylogenetic inference is then used to reconstruct the ensemble of likely genetically subclonal architectures within a given sample, and isolating statistically confident subclonal populations.
We apply the developed approach to investigate the relationship between genetic and transcriptional heterogeneity within tumors of multiple myeloma patients, demonstrating significant transcriptional differences among subclones, including persistent transcriptional features distinguishing metastasis-associated clones in the context of a pre-metastatic primary tumor.



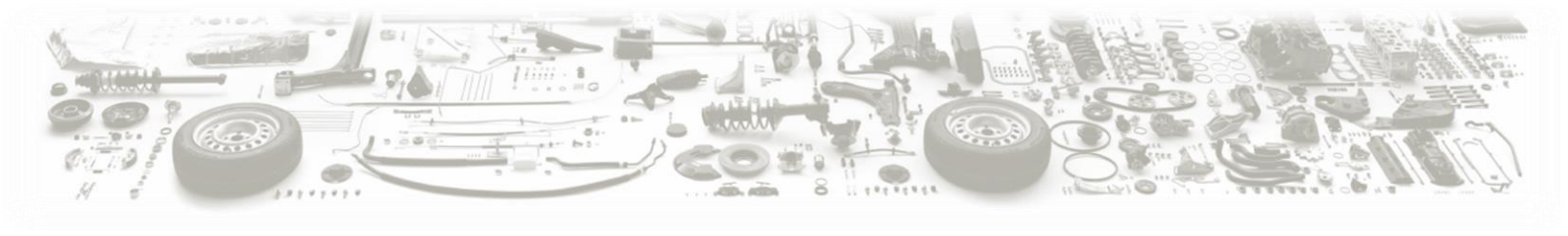
S4.4

Visualizing structural variation in single cells to explore genomic heterogeneity in human populations

Ashley Sanders[1], Mark Hills[1], Ester Falconer[1], David Porubsky[2], Victor Guryev[2], Peter Landorp[2]

[1]*BC Cancer Agency, Vancouver, BC, Canada*
[2]*ERIBA, Groningen, Netherlands*

Studies of genome heterogeneity and plasticity aim to resolve how genomic features underlie phenotypes and disease susceptibilities. Identifying genomic features that differ between individuals and cells can help uncover the functional variants that drive specific biological outcomes. For this, single cell studies are paramount, as it becomes increasingly clear that the contribution of rare but functional cellular subpopulations is important for disease prognosis, management and progression. Until now, studying these associations has been challenged by our inability to map structural variants accurately and comprehensively. To overcome this, we employed the template strand sequencing method, Strand-seq, to preserve the structure of individual homologues and visualize genomic variants in single cells. We used this method to rapidly discover, map, and genotype human polymorphisms with unprecedented resolution. This allowed us to explore the distribution and frequency of structural variation in a heterogeneous cell population, identify several polymorphic domains in complex regions of the genome, and locate rare alleles in the reference assembly. We then extended this analysis to comprehensively map the complete set of inversions in an individual's genome and define their unique invertome. We predict characterizing the inversion profiles of patients will have important implications for personalized medicine. Finally, we generated a non-redundant, global reference of structural variants in the human genome and better characterized their architectural features. Taken together, we describe a powerful new framework to study structural variation and genomic heterogeneity in single cell samples, whether from individuals for population studies, or tissue t--ypes for biomarker discovery.?



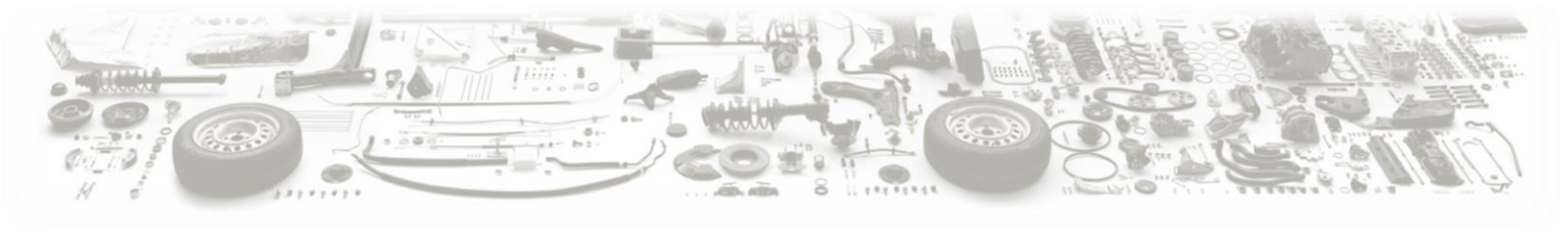
S5.1

Counting your favorite transcripts in single cells

Liran Valadarsky, Hadas Keren-Shaul, Ido Amit

Weizmann Institute of Science, Rehovot, Israel

Recent years have seen a revolution in genome-wide single-cell transcription technologies, improving sensitivity, throughput and accuracy. In contrast, development of similar technology for meso-scale measurement of a targeted set of tens to hundreds of markers is greatly lacking. We develop a sensitive, high-throughput, sequencing-based method (STAR-seq: Single-cell TARgeted amplification and sequencing) for the quantitation of a selected set of genes from single cells. We assayed dozens of genes in hundreds of cells at an average of a few hundred reads/cell and report a yield equivalent to currently existing single-cell technologies. STAR-seq opens the way for sensitive and accurate measurement of selected RNA transcripts in single cells for both research and diagnostics.



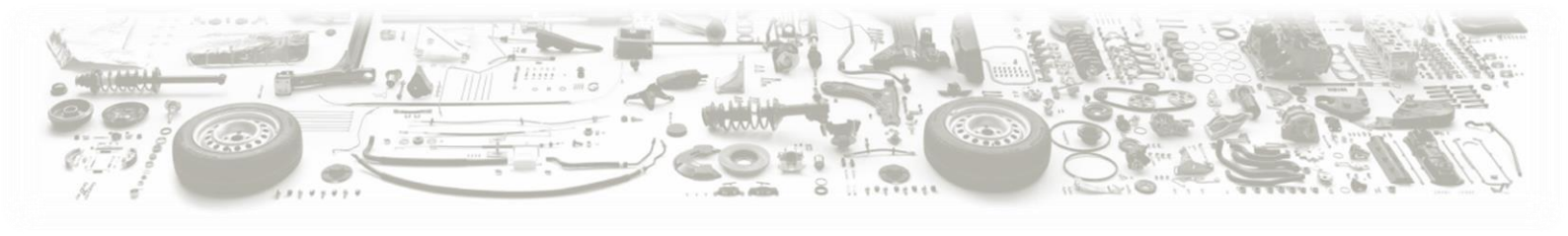
S5.2

Whole-organ lineage tracing of the human pancreas

Martin Enge, Hatice Efsun Arda, Geoff Stanley, Seung Kim, Stephen Quake

Stanford University, Palo Alto, CA, United States

Major outstanding questions concerning organ development, such as the nature of cell state transitions on a whole transcriptome level and the contribution of different sets of mesenchymal cells to organ development can be addressed using single-cell RNA-seq methods. We use a combination of unbiased cell collection with enrichment of rare cell populations to gain a comprehensive understanding of pancreas development. Towards this goal, we have collected full-length transcriptomic data from almost 10000 cells from pancreatic rudiments taken from diverse stages of fetal gestation, juvenile (proliferative) and adult (non-proliferative). Sequencing full length transcripts at high sequencing depth on a large number of cells allows us to capture even subtle cell state differences such as alternative splicing events along a complex series of cell state transitions.



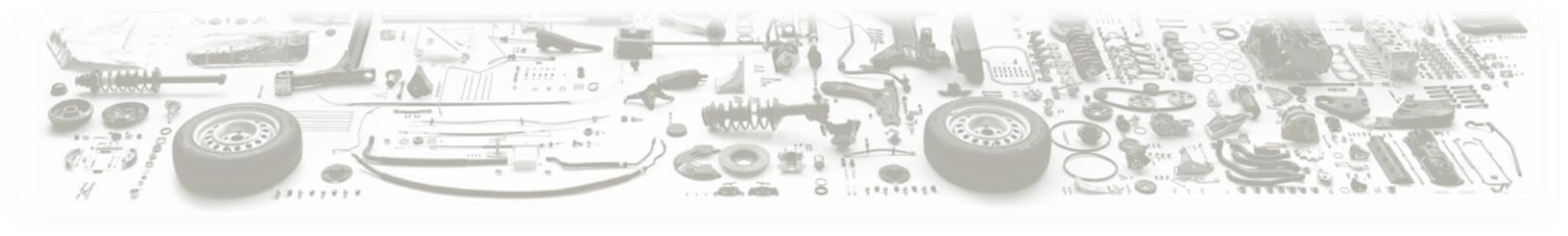
S5.4

Adult Mouse Cortical Cell Taxonomy by Single Cell Transcriptomics

Bosiljka Tasic, Vilas Menon, Thuc Nghi Nguyen, Tae Kyung Kim, Tim Jarsky, Zizhen Yao, Lucas Gray, Staci Sorensen, Tim Dolbeare, Boaz Levi, Michael Hawrylycz, Christof Koch, Hongkui Zeng

Allen Institute for Brain Science, Seattle, WA, United States

Nervous systems are composed of numerous cell types, but the extent of cell type diversity is poorly understood. Here, we construct a cellular taxonomy of one cortical region, primary visual cortex, in adult mice based on single cell RNA-sequencing. We identify 49 transcriptomic cell types including 23 GABAergic, 19 glutamatergic and 7 non-neuronal types that cover the transcriptional phenotypic landscape with a combination of discreteness and continuity. We also analyze cell-type specific mRNA processing and characterize many transgenic Cre lines for cell type composition. Finally, we show that some of our transcriptomic cell types display specific and differential electrophysiological and axon projection properties, thereby confirming that these single cell transcriptional signatures can be associated with specific cellular properties.



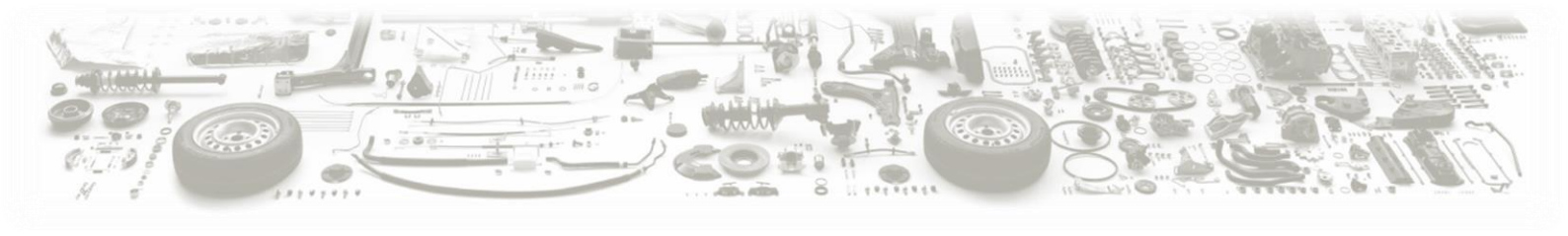
S5.7

Conserved principles of brain development revealed by comparative single-cell RNA-seq in mouse and human ventral midbrain

Gioele La Manno[1], Daniel Gyllborg[1], Enrique Toledo[1], Amit Zeisel[1], Simon Stott[2], Peter Lonnerberg[1], Roger Barker[2], Ernest Arenas[1], Sten Linnarsson[1]

[1]*Karolinska Institutet, Stockholm, Sweden*
[2]*University of Cambridge, Cambridge, United Kingdom*

With the advent of stem cell differentiation technology, molecular knowledge of the human developmental process has now the potential to directly improve disease modelling, drug discovery and cell replacement therapies. However, little is known molecularly about human development, even for tissues where the clinical relevance could be large. Here we present the first comprehensive molecular analysis of murine and human ventral midbrain development. We performed single-cell RNA-seq of more than two thousand five hundred single cells across different stages of VM development. To our knowledge this is the first time-resolved single-cell study of a complex developmental process performed in any mammalian tissue, and the first cross-species comparative single-cell RNA-seq dataset of any kind. This data allowed an unbiased classification of cell types that are involved in the process and the determination of a spatial and lineage-specific transcription factor code. We further described, for different lineages, the transcriptional dynamics corresponding to differentiation and maturation. We finally characterized developmental cues and environmental stimuli attributing them to the responsible cells types. This work not only offers insights on principles of development conserved across species but it has clinical implication establishing an important reference for differentiation protocols, cell replacement therapies and disease modelling of Parkinson's disease.



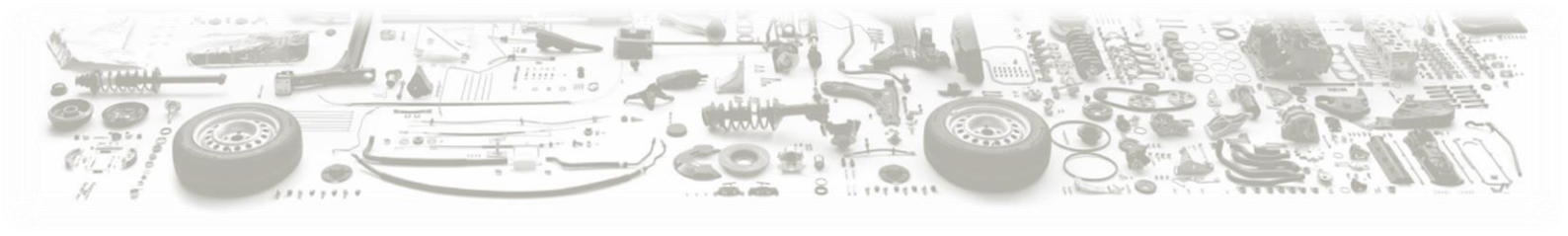
P001

A generic, cost-effective and scalable cell lineage analysis platform

Tamir Biezuner, Adam Spiro, Ofir Raz, Shiran Amir, Lilach Milo, Noa Chapal-Ilani, Veronika Berman, Rivka Adar, Yael Chagit-Fried, Elena Ainbinder, Ehud Shapiro

Weizmann Institute of Science, Rehovot, Israel

Advancements in single cell sequencing have led to tremendous improvements in understanding lineage relations among individual cells. Knowledge of cell lineage trees would allow for deep understanding of cellular dynamics in many biological domains, for example cancer progression and stem cell dynamics. Methods based on somatic mutations enable sub-population analysis and clonal inference among individual cells, however, current single-cell methods are costly and could be biased due to functional dependencies. Here we engineered a streamlined automated workflow for single-cell lineage analysis which begins with individual cells and ends with their corresponding cell lineage tree. We demonstrate the feasibility of this platform using a controlled ex vivo grown cell lineage tree. We extended the experimental results and mapped the cell lineage reconstruction feasibility landscape by modelling and simulating the underlying cellular dynamics under various biological and technical parameters. Its high resolution, together with its cost-effectiveness and high-throughput make this platform a powerful tool to study lineage and clonal relations among individual cells and serve as a proof-of-principle prototype which lays the biological, computational and architectural foundations for an envisioned large-scale human cell lineage discovery project.



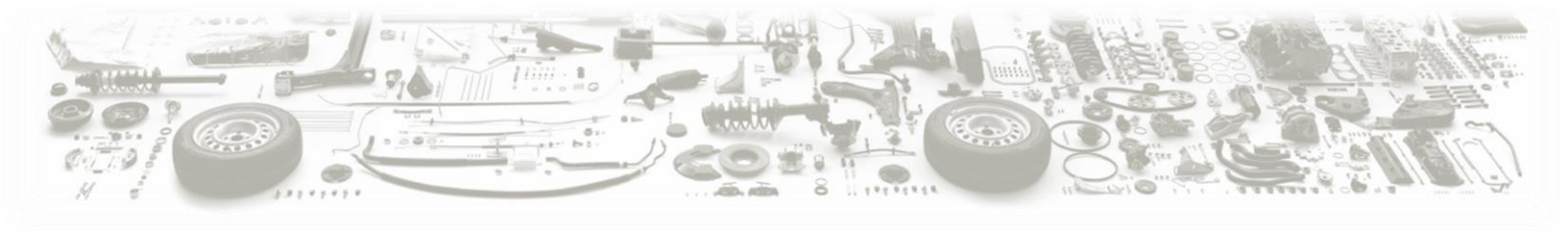
P003

Developing a rigorous quality control procedure for single cell transcriptomics data

Eshita Sharma, Quin Wills, Esther Mellado, Moustafa Attar, Karene Argoud, John Broxholme, Lorna Witty, Sarah Lamble, Amy Trebes, Paolo Piazza, David Buck, Rory Bowden, Helen Lockstone

Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

The Oxford Genomics Centre, based at the Wellcome Trust Centre for Human Genetics, has established a single cell service over the past year and is an integral part of the Oxford Consortium for Single Cell Biology. During this intensive development phase, data have been generated from a range of samples (primary cells, stem cells, cell lines), cell separation techniques (C1, FACS) and library preparation protocols (SMARTer1, SMART-seq2, SMARTer v3).
Here we present an overview of quality control (QC) for single cell transcriptomic data, in the context of comparing commonly used chemistries and platforms. We highlight relative strengths and describe informative QC metrics now applied in the standard Consortium workflow. Among our findings are technical effects causing preferential enrichment or loss of certain transcripts; these biases are sequence-specific but only partially explained by GC-content.
To assess overall quality, cDNA input concentration, number of detected genes, and proportion of reads mapping to the 500 most highly expressed genes for each cell are captured on a single plot. This separates failed and sub-optimal libraries from the majority that perform well. A strong cDNA effect is often observed (despite input normalisation) and is likely to arise from variation in library complexity between cells. This is an important factor influencing the counts obtained and subsequent interpretation. Low initial cDNA concentration tends to produce lower complexity libraries and failed samples invariably have low cDNA together with a reduced gene to spike-in ratio. QC results and recommendations are summarised in a bespoke report for each individual project.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

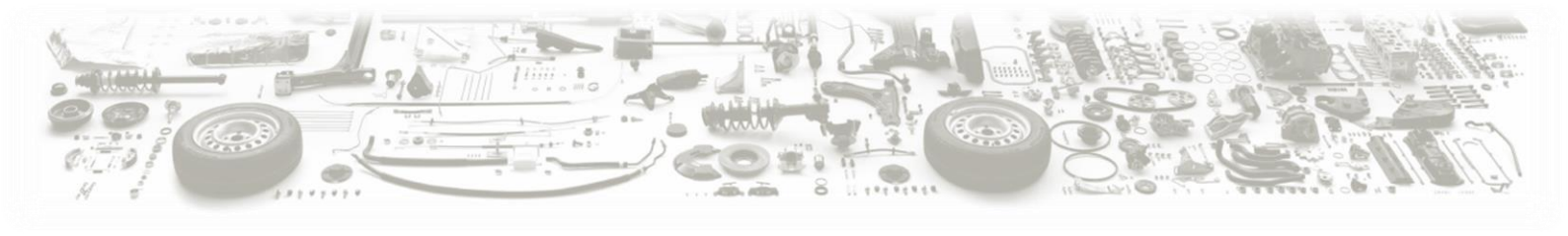
P004

Robust lineage reconstruction from high-dimensional single-cell data

Eugenio Marco, Gregory Giecold, Lorenzo Trippa, Guo-Cheng Yuan

Harvard University, Boston, MA, United States

Single-cell gene expression profiling technologies provide an excellent opportunity to uncover rare cell types and cell-fate transition events during development. However, it remains a major challenge to systematically identify rare cell types and the lineage relationships among different cell-types, mainly due to the technical difficulty of precise gene expression profiling in single cells. While several methods have been developed to reconstruct cell lineages from data, one common limitation is that their results are sensitive to measurement uncertainty and sample sizes. It is desirable to account for such uncertainty in subsequent functional validation and mechanistic investigations. Toward this end, we have developed a novel computational method using an ensemble-based technique. The results from different subsamples are assembled together to infer probabilistic information regarding the robustness of cell-type classification and lineage relationships. The method is fast and applicable for analysis of large datasets. We have applied this technique to several datasets, obtained from various developmental processes. Compared to existing methods, our method not only significantly improved the robustness of the lineage reconstruction, but correctly inferred the uncertainty of lineage relationships. Our method provides a powerful tool for systematic exploration of single-cell data.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P005

Single cell transcriptomics analysis of induced pluripotent stem cell-derived cortical neurons reveals frequent dual layer identity

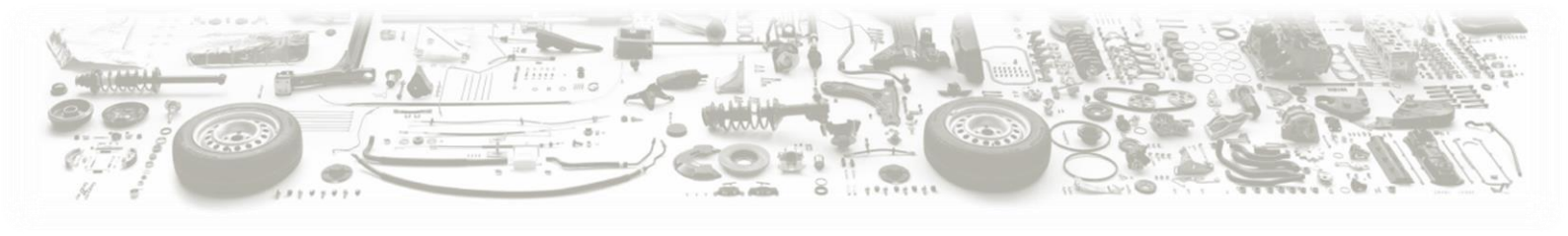
Adam Handel, Satyan Chintawar, Tatjana Lalic, Emma Whiteley, Jane Vowles, Alice Giustacchini, Karene Argoud, Paul Sopp, Rory Bowden, Sally Cowley, Sarah Newey, Colin Akerman, Chris Ponting, Zameel Cader

University of Oxford, Oxford, United Kingdom

Induced pluripotent stem cell (iPSC)-derived cortical neurons present a powerful new model of neurological disease. Previous work has established that differentiation protocols produce cortical neurons but little has been done to characterise these at cellular resolution.[1] In particular, it is unclear to what extent in vitro two-dimensional, relatively disordered culture conditions recapitulate the development of in vivo cortical layer identity. Single cell multiplex RT-qPCR was used to interrogate the expression of genes previously implicated in cortical layer or phenotypic identity in individual cells. Unexpectedly, 22.7% of neurons analysed frequently co-expressed canonical fetal deep and upper cortical layer markers, and this co-expression was also present at the level of translated protein. By comparing single cell RNA-seq data from our cells to available single cell RNA-seq data from human fetal and adult brain, we observed that this co-expression of layer markers was also seen in primary tissue.[2,3] These results suggest that establishing neuronal layer identity in iPSC-derived or primary cortical neurons using canonical marker genes transcripts is unlikely to be informative.

References:
[1] Shi Y et al. Nat. Protocol. 2012;7(10):1836–1846.
[2] Pollen AA et al. Nat. Biotechnol. 2014;32:1053–1058
[3] Darmanis S et al. Proc. Natl. Acad. Sci. 2015; ahead of print

Acknowledgements:
This project was supported by StemBANCC, Oxford NIHR BRC, and a Research Training Fellowship from the Wellcome Trust.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

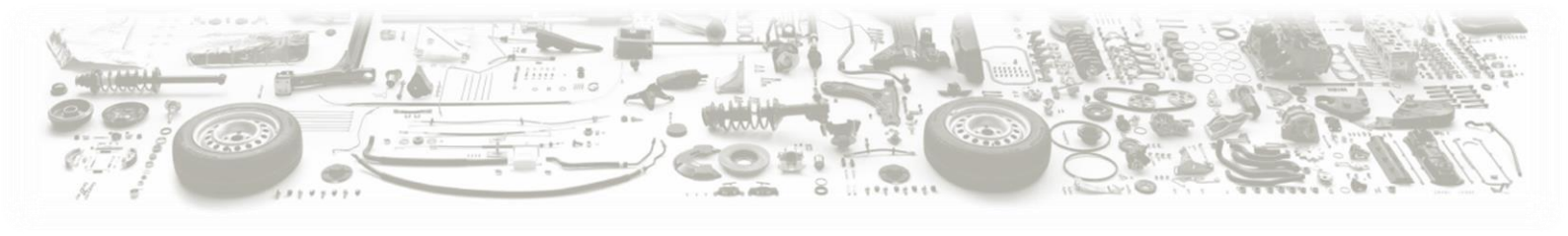
P006

Multi-parameter RNA and protein data from the same single cell

Kenneth J. Livak[1], Shuqiang Li[1], Caroline J. Gallant[2], Spyros Darmanis[3], Alexander Genshaft[4], Sanjay Prakadan[4], John J. Trombetta[5], Alex K. Shalek[4], Ulf Landegren[2]

[1]*Fluidigm Corporation, South San Francisco, CA, United States*
[2]*Uppsala University, Uppsala, Sweden*
[3]*Stanford University, Stanford, CA, United States*
[4]*Massachusetts Institute of Technology, Cambridge, MA, United States*
[5]*Broad Institute of Harvard and MIT, Cambridge, MA, United States*

Quantification of transcripts in single cells by qPCR or RNA-sequencing has demonstrated the power of single-cell correlations in identifying and cataloging distinct cell types. Determining how these transcriptional differences relate to function requires establishing similar correlations with the effectors of the genetic program, namely proteins. The specificity of the Proximity Extension Assay (PEA, Olink Bioscience) enables running 96 (or more) antibody assays simultaneously on a sample. By using the Script Builder[™] software with the C1[™] system (Fluidigm Corporation), the PEA protocol has been adapted for the analysis of single cells. In addition, the method has been expanded to include generation of a cDNA library from each single cell. This makes it possible to detect and quantify multiple proteins and multiple transcripts from the same single cell. The methodology has been applied to the cancer-derived cell line MCF7. Quantification of both RNA and protein was accomplished using qPCR. In order to explore dynamic response at the single-cell level, the cells were treated with the phorbol ester phorbol myristate acetate. Measuring the effect of this perturbation on individual MCF7 cells illustrates the interplay between transcript and protein levels in determining cell response.



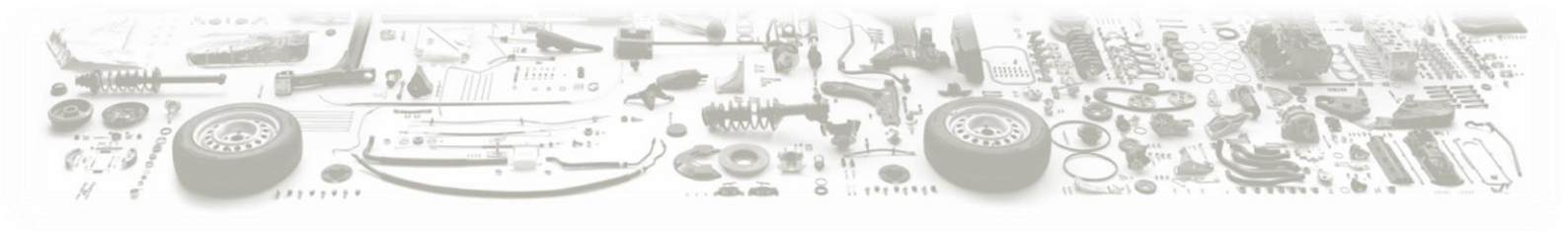
P007

Bursty gene expression in the intact mammalian liver

Keren Bahar Halpern, Shalev Itzkovitz

Weizmann Institute of Science, Rehovot, Israel

The question whether cells in intact mammalian tissues show bursty gene expression has been left unanswered for many years. Bursty Gene expression is referred to the process in which the promoter is open for short periods of time producing nascent mRNA followed by periods of promoter quiescence. Promoters tend to stochastically transition between a closed, transcription-prohibitive state and an open permissive state. This bursting phenomenon can create intrinsic variability in the mRNA content of isogenic cells. The expression level of genes in steady state is mainly affected by three parameters: transcription rate, degradation rate and the state of the chromatin. Same steady state mRNA levels can be achieved by different combinations of these parameters. Analysis of the gene expression parameters in complex and heterogeneous tissues such as the liver allow us to evaluate the expression signature of single cells inside its microenvironment. In our study we utilize the single molecule fluorescence in-situ hybridization (smFISH) technique in order to identify mature and premature mRNA molecules of endogenous genes and to extract the features of gene expression in complex tissues. We show that liver gene expression is highly bursty affecting the variability between neighboring cells and the localization of mRNA in the cell. Moreover, we show that temporal averaging and liver polyploidy serve to minimize the intrinsic variability associated with transcriptional bursts.



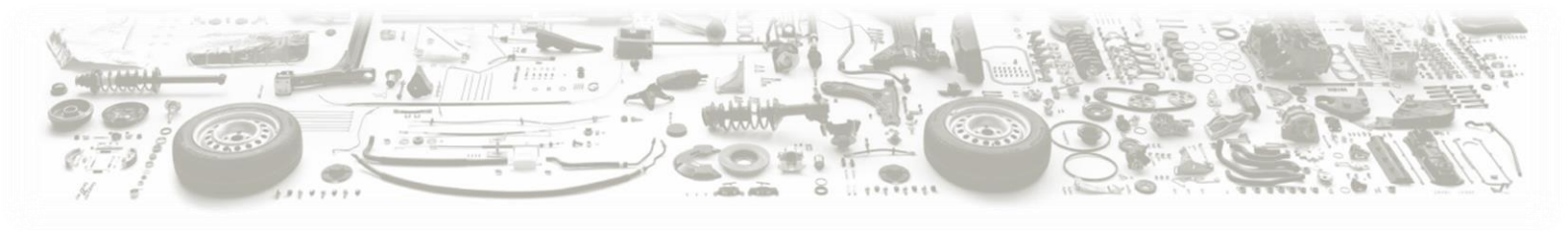
P008

Single-cell transcriptomes of fluorescent, ubiquitination-based cell cycle indicator cells

Michael Böttcher[1], Tsukasa Kouno[1], Elo Madissoon[2], Efthymios Motakis[1], Imad Abugessaisa[1], Sachi Kato[1], Harukazu Suzuki[1], Yoshihide Hayashizaki[1], Takeya Kasukawa[1], Piero Carninci[1], Timo Lassmann[3], Jay Woo Shin[1], Charles Plessy[1]

[1]*RIKEN Center for Life Science Technologies, Yokohama, Japan*
[2]*Karolinska Institutet, Huddinge, Sweden*
[3]*Telethon Kids Institute, Perth, Australia*

Using HeLa cells that report their cell cycle phase through fluorescent, ubiquitination-based cell cycle indicators (Fucci), we produced a reference dataset of more than 270 curated single cells for which each single-cell's transcriptome can be matched with cell cycle information via the fluorescence intensity of the transgenes in that cell. We developed a comprehensive open data management and quality control pipeline that enables users of our dataset to process all available sequence and image files in a highly reproducible way without prior knowledge of the underlying bioinformatical toolset. The final output of that pipeline is a customizable table with relevant metadata and quality information for each single cell. This metadata table can be easily used as input for sophisticated data analysis. Our workflow is also adjustable for usage with other single-cell datasets that consist of RNA-sequencing and fluorescence data. Currently, we use the Fucci dataset to create a model for cell cycle phase inferences, which can be applied to other single-cell transcriptomes without cell cycle phase reporting metadata.



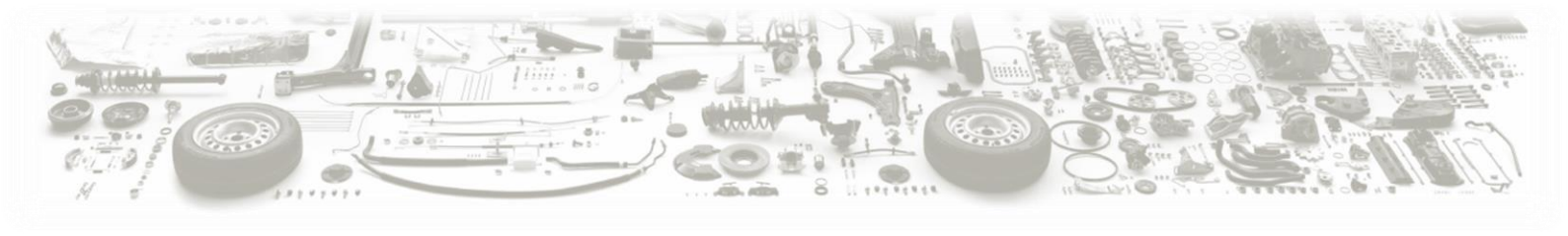
P009

Discrete Distributional Differential Expression (D3E) - A Tool for Gene Expression Analysis of Single-cell RNA-seq Data

Martin Hemberg, Mihails Delmans

Wellcome Trust Sanger Institute, Hinxton, United Kingdom
University of Cambridge, Cambridge, United Kingdom

The advent of high throughput RNA-seq at the single-cell level has opened up new opportunities to elucidate the heterogeneity of gene expression. One of the most widespread applications of RNA-seq is to identify genes which are differentially expressed (DE) between two experimental conditions. Here, we present a discrete, distributional method for differential gene expression (D3E), a novel algorithm specifically designed for single-cell RNA-seq data. We use synthetic data to evaluate D3E, demonstrating that it can detect changes in expression, even when the mean level remains unchanged. D3E is based on an analytically tractable stochastic model, and thus it provides additional biological insights by quantifying biologically meaningful properties, such as the average burst size and frequency. We use D3E to investigate experimental data, and with the help of the underlying model, we directly test hypotheses about the driving mechanism behind changes in gene expression.



P011

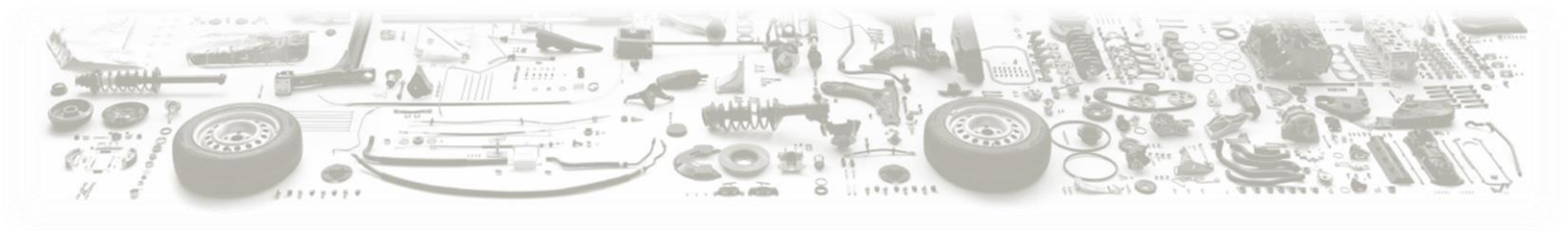
Predictive transcriptomics of cell death

Lars Velten, Chiara Baccin, Lars Steinmetz

EMBL, Heidelberg, Germany

When clonal populations of cells are exposed to a pro-apoptotic drug such as the cytokine TRAIL, a fraction of the cells die and a fraction survive. Available evidence (1) suggests that the death/live decision occurs before exposure to the drug, but studying the gene expression of more versus less apoptosis-prone cells is challenging: If the drug is applied, the transcriptome of apoptotic cells is profoundly altered. Our goal with predictive transcriptomics is to sequence cells before application of the drug, and still be able to assign a likely fate to each cell. Therefore, cells are imaged before and after drug application, and a machine learning classifier is trained to predict cell fate based on cellular morphology, population context, and cell history. In the next step, a fresh batch of cells is physically sorted based on classifier prediction by using adaptive feedback microscopy (2). Cells from the two subpopulations (apoptosis-prone and apoptosis-resistant) that have never been exposed to the drug can then be subjected to molecular profiling.

(1) Spencer et al., Nature 459: 428-32
(2) Conrad et al., Nature Methods 8: 246-9



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P012

Heterogeneity and lineage commitment in the common myeloid progenitor

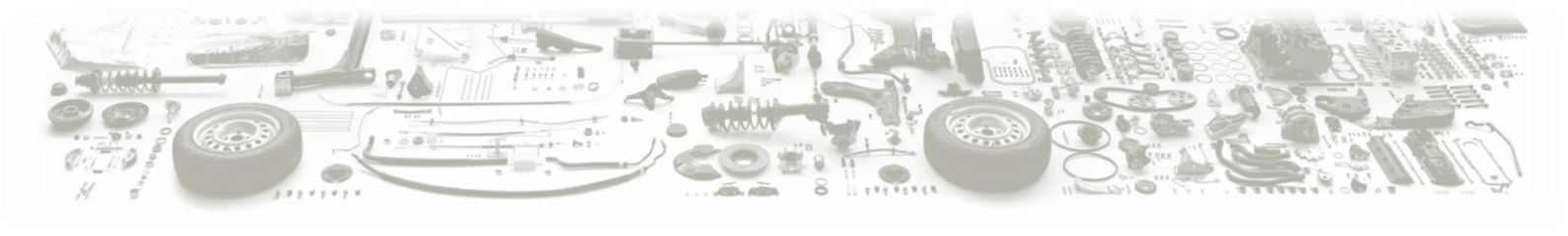
Franziska Paul[1], Yaara Arkin[1], Diego Adhemar Jaitin[1], Amir Giladi[1], Ephraim Kenigsberg[1], Hadas Keren-Shaul[1], Deborah Winter[1], David Lara-Astiaso[1], Assaf Weiner[1], Felicia Katherine Bratt Lauridsen[2], Steffen Jung[1], Andreas Trumpp[3], Bo Torben Porse[2], Amos Tanay[1], Ido Amit[1]

[1]*Weizmann Institute of Science, Rehovot, Israel*
[2]*Rigshospitalet; Biotech Research and Innovation Centre, Copenhagen, Denmark*
[3]*German Cancer Research Center, Heidelberg, Germany*

Within the bone marrow, hematopoietic stem cells differentiate and give rise to diverse blood cell types, including erythroid, myeloid and lymphoid lineages. The regulation and dynamics of intermediate progenitor populations are currently defined coarsely using a limited set of surface markers. Here we uncover unexpected transcriptional heterogeneity within myeloid progenitor compartments by studying in vivo transcriptional states at single cell resolution using MARS-Seq (1). We show that common myeloid progenitors (CMP) are highly heterogeneous, and identify at least five CMP subpopulations associated with erythrocytes, megakaryocytes, monocytes, granulocytes and CD8+ dendritic cells. Functional and chromatin assays (2) suggest that CMP transcriptional priming is coupled with in vivo developmental commitment. Computational analysis of transcription factor modules and follow-up knockout assays uncover some of the regulatory mechanisms underlying the commitment dynamics of myeloid progenitors. Single cell analysis thereby shows that lineage commitment in hematopoiesis occurs earlier and is less structured than previously appreciated.

We thank members of the Amit and Tanay labs for critical discussions. We thank G. Brodsky for artwork. Research was supported by the European Research Council and Israel Science Foundation (1782/11, 1050/12) grants to I.A. and A.T.; work in the Porse Laboratory was supported through a grant from the Novo Nordisk Foundation section for Stem Cell Biology in Human Disease.

1. D. A. Jaitin et al. Science 343, 776-779 (2014).
2. D. Lara-Astiaso et al. Science 345, 943-949 (2014).



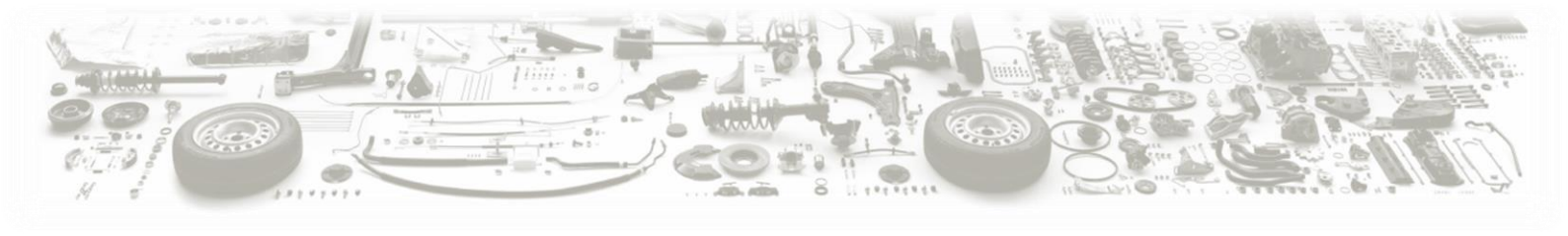
P013

Studying host-pathogen interactions using single-cell RNA-Seq.

Gal Avital[1], Roi Avraham[2], Amy Fan[2], Tamar Hashimshony[1], Deborah Hung[2], Itai Yanai[1]

[1]*Technion - Israel Institute of Technology, Haifa, Israel*
[2]*Broad Institute of Harvard and MIT, Cambridge, MA, United States*

While over 80 years have passed since antibiotics have been applied therapeutically, bacterial infections continue to be a major cause of death worldwide. Advances in transcriptomics enable us to systematically study the interaction between bacterial pathogens and their hosts. However, to date, no method has been able to capture the transcriptome of both the bacterial and the infected macrophage at the single-cell level, since single-cell RNA-Seq is typically limited to polyadenylated transcripts. Here, we introduce 'All-Seq', a method for single-cell RNA-Seq targeting all RNAs, thereby capturing the transcriptomes of both the host and pathogen. This method is highly multiplexed, uses unique molecular identifiers (UMI) thereby allowing single transcripts quantification, efficiently removes rRNA, and invokes in vitro transcription. We are in the process of querying for causal relationships among the expression of bacterial and host genes, by comparing with spiked-in molecules. The All-Seq method will also be instrumental for projects involving the study of non-polyadenyated transcripts or partially degraded samples, as well as for studying expression heterogeneity in bacterial populations of cells.



P014

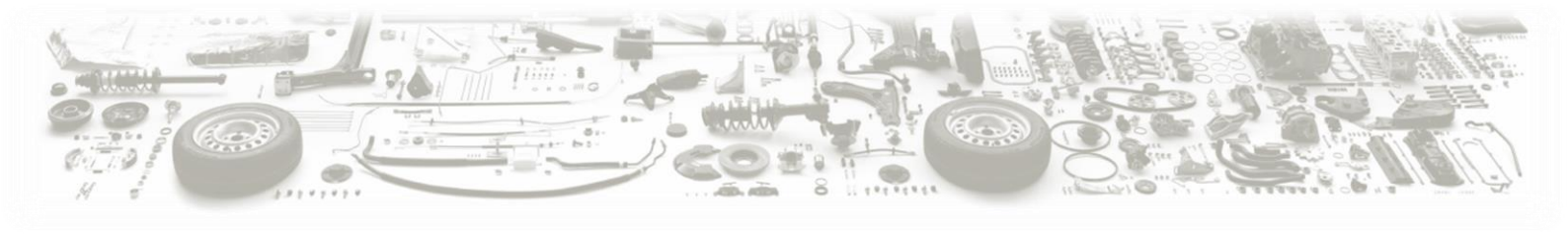
Single-cell resolution spatial transcriptomics and molecular characterization of cell type complement in marine annelid P. dumerilii.

Kaia Achim[1], Nils Eling[2], John Marioni[2], Detlev Arendt[1]

[1]*European Molecular Biology Laboratory, Heidelberg, Germany*
[2]*European Molecular Biology Laboratory, Hinxton, United Kingdom*

Understanding the cell type complement in any multicellular organism requires integration of the information of gene expression as well as the spatial organization of cells.
Featuring a mixture of numerous functional cell types and progenitor populations in a compact organism during a larval stage, and a key phylogenetic placement, marine annelid Platynereis dumerilii presents an excellent model for understanding the cell type complement in basal bilaterians. Gene expression profiling by mRNA in situ hybridization has been used previously to generate a gene expression atlas of P. dumerilii brain[1]. To complement this ISH-based gene expression database, we developed a high-throughput method for assaying cells by single-cell mRNA sequencing and subsequent identification of their spatial origin within the brain atlas[2], thus establishing a whole-transcriptome wide database of spatially referenced gene expression.
We are now extending the single-cell data analysis from spatial transcriptomics to whole-organism-wide characterization of functional cellular modules and cell type classification. By complementing the resulting P. dumerilii cell type catalogue with similar data from basal metazoans, we aim to unravel key characteristics of cell type identity, as well as events that lead to the emergence and diversification of cell types in animals.

References:
1. Tomer R, Denes AS, Tessmar-Raible K, & Arendt D (2010) Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. Cell 142(5):800-809.
2. Achim K, et al. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nature biotechnology 33(5):503-509.



P015

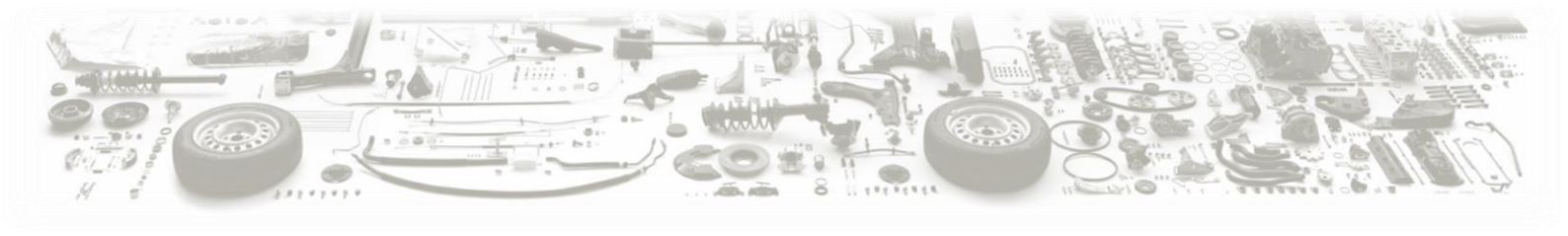
In situ DNA and RNA barcoding and amplification for same-sample sequencing and imaging

Bastiaan Spanjaard[1], Nicola Crosetto[2], Magda Bienko[2], Alexander van Oudenaarden[1], Laura Annaratone[3], Anna Sapino[3]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Karolinska Institutet, Stockholm, Sweden*
[3]*Azienda Ospedaliera Citta della Salute e della Scienza di Torino, Turin, Italy*

We describe T7 promoter-assisted In Situ DNA and RNA Barcoding and linear Amplification (TISBA), which enables sequencing of low-input, fixed cell and tissue samples while preserving them for downstream microscopy assays. In cells and tissue sections immobilized on a glass surface, we simultaneously measured copy number alterations and gene expression, followed by DNA fluorescence in situ hybridization in the same specimen. TISBA is a versatile and automatable approach for same-sample sequencing and imaging with potentially broad applications in research and diagnostics.

This work was supported by NWO (VICI Award) and ERC (ERC-AdG 294325-GeneNoiseControl) grants to A.v.O.



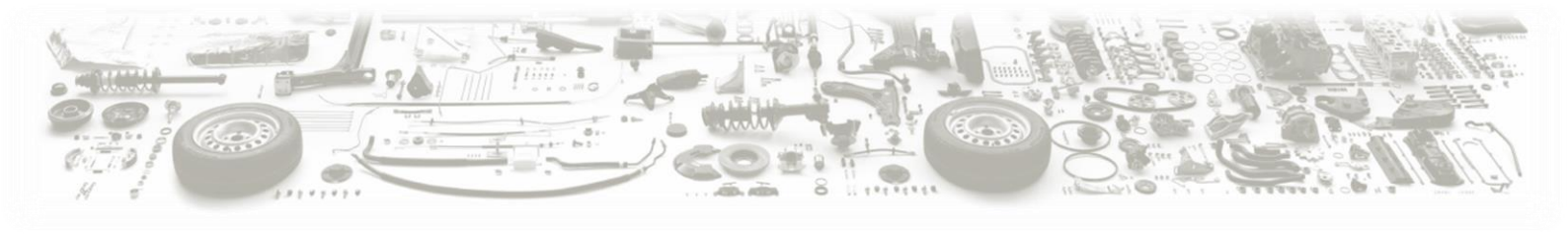
P016

Unravelling the molecular events leading to hematopoietic stem cell generation in the mouse embryo aorta

Chloe Baron, Lennart Anne Kester, Jean-Charles Boisset, Alexander van Oudenaarden, Catherine Robin

Hubrecht Institute, Utrecht, Netherlands

Hematopoietic Stem Cells (HSCs) are initially generated and expanded during embryonic life. They are first detected in the Dorsal Aorta (DA) at embryonic day (E)10.5 of mouse development. HSCs are part of clusters (Intra-Aortic Hematopoietic Clusters or IAHCs) that are tightly attached to the endothelial layer of the DA. IAHCs are not solely composed of HSCs since IAHCs start to appear in the DA before HSCs and in higher numbers. We have recently shown that IAHCs contain HSC precursors (pre-HSCs) able to progressively mature into HSCs in vivo (as shown by successful long-term multilineage reconstitution of primary neonates and secondary adult recipients). So far, the process of pre-HSC maturation remains unclear.
We are currently using a single cell RNA-sequencing approach to study the molecular events leading to HSC formation in IAHCs. We focused on two types of cell samples isolated based on their spatial and temporal appearance in the mouse embryo aorta. Single IAHC cells were isolated to purity based on c-kit expression by flow cytometry at different developmental time points (before [E10] and after [E11] HSC emergence). Single IAHCs were also isolated by a mechanical pick-up that we recently developed in our lab. This novel technique gives the unique advantage to isolate individual IAHCs from live embryo slices based on their precise location in the DA (ventral or dorsal) (HSCs are restricted to the ventral part of the DA).
The analysis and comparison of the RNA-Sequencing data should help to identify genes and signalling pathways involved in the process of pre-HSC maturation towards a HSC fate and also to assess the level of heterogeneity of IAHCs.



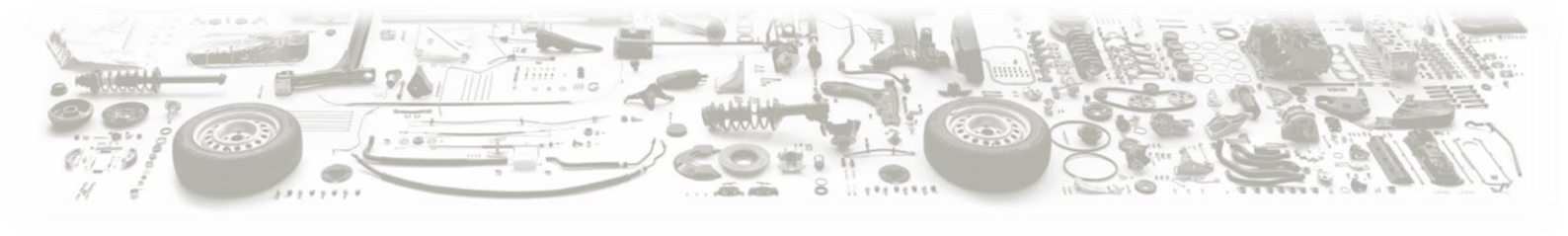
P017

Single-cell analysis of pancreatic development under normal and diseased state

Maayan Baron, Itai Yanai

Technion, Haifa, Israel

Loss or dysfunction of β-cells in the pancreas results in diabetes which affects over 200 million people worldwide. The pathways leading to this state are heterogeneous, ranging from internal factors and genetic susceptibility to external ones including environmental stressors such as viruses, toxins and food. These phenomena have led to the belief that the cure for diabetes will involve the introduction of new reservoirs of insulin-secreting cells. However, the near-exclusive focus of diabetes research on β-cells overlooks the complex cell population that makes up the pancreas and the interaction among its cell-types. Single-cell gene expression profiling now provides a basic tool for reading out the cellular function of the cell-types in a population. Our approach integrates our CEL-Seq protocol to characterize pancreatic cell transcriptomes at the single-cell level with novel transcriptomic methods will lead to a 3D gene expression atlas of the zebrafish endocrine pancreas. Mining this data, we will examine the identity and spatial-disposition of the cell-types comprising the endocrine system and unravel the influence of a cell's spatial context on its own gene expression.



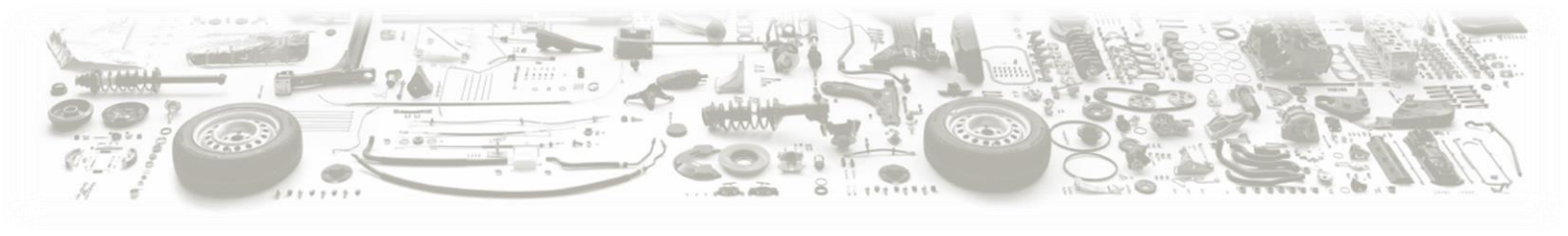
P018

Spatially Resolved Analysis of Differential Expression within Single Cells

Sanja Vickovic[1], Fredrik Salmen[1], Patrik Stahl[2], Sarantis Giatrellis[2], Jakub Westholm[3], Jose Fernandez Navarro[2], Jonas Frisen[2], Joakim Lundeberg[1]

[1]KTH, Solna, Sweden
[2]KI, Solna, Sweden
[3]SU, Solna, Sweden

Most of the current gene expression methods depend on an average view of a cell population and do not yield information about the profile of a single cell. We approach this question by analyzing spatially resolved gene expression information in both single cells and tissue sections by taking advantage of barcoded DNA microarrays.
The microarrays are subdivided into > 1000 features where the barcode sequence is varied in each feature. Each of those features then represents a potential site for a single cell to be profiled. Firstly, the actual cell location on the array is noted with immunohistochemistry. Secondly, mRNA is the captured using poly-A selection enabling short cDNA fragments to be generated. The isolated cDNA gives a spatially resolved footprint that is highly comparable to the cell it was derived from. This approach marks true biological variation and rises above technical noise of single cell libraries. Localized transcript reads belonging to single cells are then identified through a visualization tool (stVi) and downstream differential expression analysis of single cells in parallel is enabled.
Recently, we have incorporated indexed FACS sorting into our cell-processing pipeline giving rise to combining antibody affinity data with RNA-sequencing as well as prescreening cell populations to be profiled. Consequently, we could choose which cell we wanted to sort on top of each position on the barcoded array.
Spatial Transcriptomics is a method where localized transcripts originating from single cells are identified. FACS sorting, library preparation and normalization make this approach a state-of-the-art single cell-profiling platform.



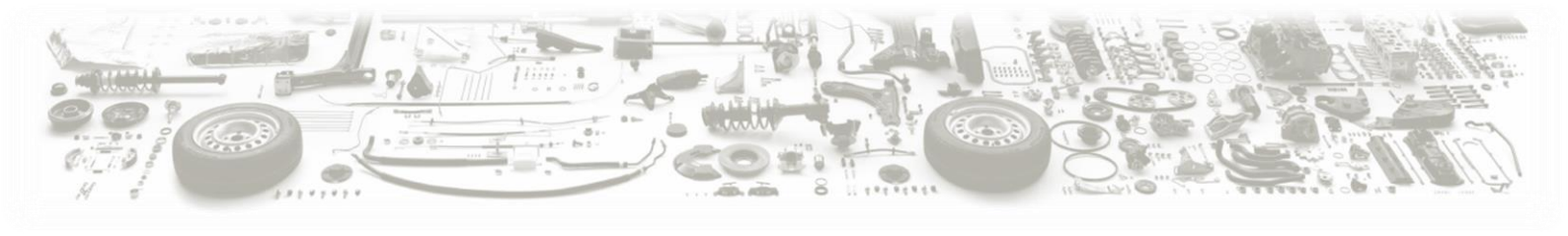
P019

Spatial Transcriptomics – Visualization and Quantification of Gene Expression in Whole Tissue Sections

Fredrik Salmén[1], Patrik Ståhl[2], Anna Lundmark[2], Sanja Vickovic[1], José Fernandez Navarro[1], Jens Magnusson[2], Jakub Orzechowski Westholm[3], Mikael Huss[3], Annelie Mollbrink[1], Paul Igor Costea[1], Pelin Akan[1], Jan Mulder[2], Olaf Ingemar Bergmann[2], Joakim Lundeberg[1], Jonas Frisén[2]

[1]*KTH Royal Institute of Technology, Stockholm, Sweden*
[2]*Karolinska Institutet, Stockholm, Sweden*
[3]*WABI, Stockholm, Sweden*

Here we present a novel method, Spatial Transcriptomics (ST), which combines gene expression analysis of multiple regions in intact tissue sections, with staining and imaging of the same sections. The protocol starts by placing a thin tissue section onto a glass slide, containing several thousands of small features. Each feature is built up by millions of nucleic acid capture probes, which all contain a specific sequence. The specific sequence is shared by all probes within a feature but varied between features. We refer to this sequence as a spatial barcode, something that is used to determine the location of each feature. The tissue section is stained and imaged in order to get the morphological information and to determine the location of each tissue region, in relation to the features. The imaging step is followed by a specific tissue treatment and complementary DNA (cDNA) synthesis. This enables the generation of cDNA libraries, where each transcript is attached to a capture probe. The finished cDNA libraries are sequenced using paired end sequencing, allowing us to get the spatial barcode sequence from one end and the transcript sequence from the other end. Since the location of all features and tissue regions are known, the spatial barcode sequence can be used to determine which region each transcript was derived from. Finally, the spatially located transcripts can be visualized in our software tool by combining the analyzed sequence data with the tissue image. The result is a high-resolution pattern of gene expression information across the analyzed tissue section, where the resolution is limited only by the size of the underlying features.



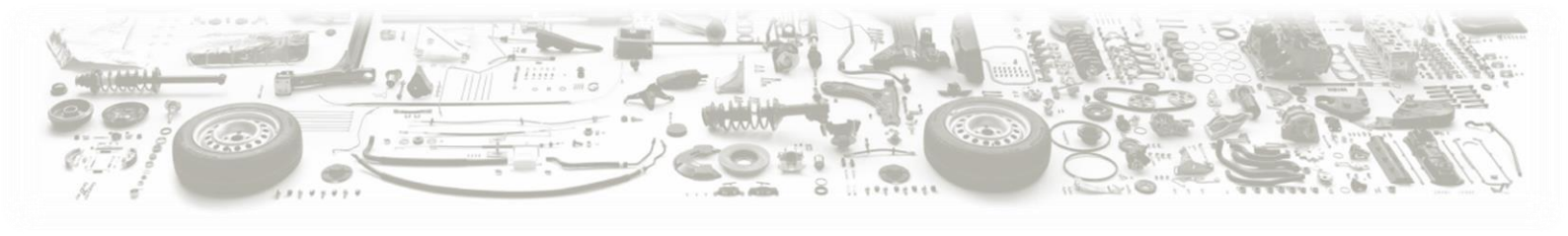
P020

Comparison of Different NGS Library Construction Methods for Single Cell Sequencing

Nan Fang, Rumeysa Akinci-Tolun, Isabell Haupt, Ioanna Andreou, Christian Korfhage

QIAGEN GmbH, Hilden, Germany

Recent advances in whole genome and whole transcriptome amplification technologies and next generation sequencing have enabled whole genome or transcriptome sequencing at the single cell level. Single cell sequencing studies have yielded new insights into the heterogeneity of the genome and transcriptome in individual cells. Such heterogeneity at the single cell level has been shown to be closely related to cellular function, differentiation, development, and diseases. A critical element of the single cell sequencing workflow is sequencing library construction step following whole genome or transcriptome amplification. An efficient library construction method is required to convert a high percentage of the DNA fragments to adaptor-ligated sequencing library and ensure high sequence complexity of the library. Furthermore, uniform representation of all genomic regions in a sequencing library is essential for retaining all important sequence information. Here we compared two library construction methods following REPLI-g MDA-mediated whole genome amplification (WGA) or whole transcriptome amplification (WTA): a ligation-based library construction method using GeneRead Library Prep Kit (QIAGEN); and a tagmentation-based method using Nextera DNA Sample Prep Kit (Illumina). Our results demonstrated that the Nextera library construction method can be directly used with the REPLI-g-amplified DNA following MDA reaction without the need for DNA purification. This could be beneficial if working with a high number of samples or if the complete workflow of single WGA/WTA and library construction should be automated. However, compared with the tagmentation method, the ligation-based library construction method is more flexible with the input DNA amount and delivers sequencing libraries with higher complexity and less bias, which is critical for sensitive applications such as identification of genomic variants or comprehensive profiling of transcriptome.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

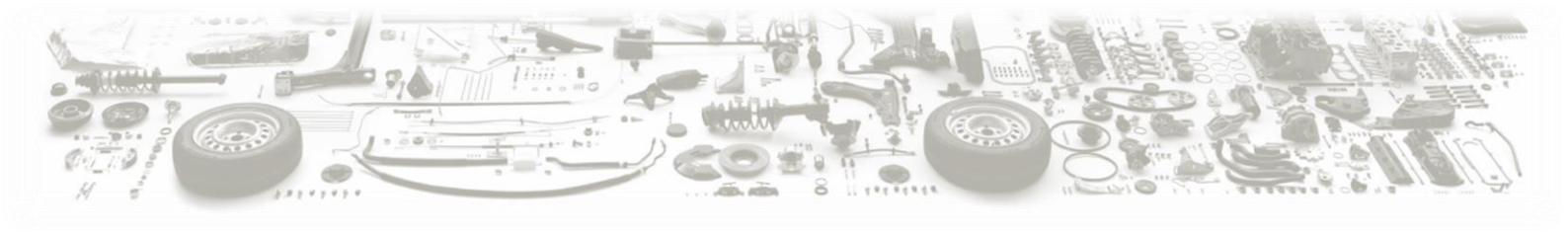
P021

Single-cell cytokine and transcriptome profiling of circulating Tfh cells in patients with systemic lupus erythematosus

Minsuk Kwak[1], Jin-Young Choi[2], Han Lin[1], Iva Xhangolli[1], Joseph Edgar Craft[2], Rong Fan[1]

[1]*Yale University, New Haven, CT, United States*
[2]*Yale University School of Medicine, New Haven, CT, United States*

Patients with systemic lupus erythematosus (SLE) exhibit activation and expansion of multiple pathogenic effector T cells that promote tissue injury via secretion of effector cytokines and dysregulated cytokine signaling. Polyfunctional cells, those capable of secreting multiple cytokines simultaneously, might contribute directly to SLE pathogenesis. Here, we applied a novel microfluidic single-cell multiplexed cytokine profiling technology to characterize aberrant cytokine programs and dissect pathogenic functions of circulating CD4+ T helper cells. We successfully quantified 10 potentially pathogenic cytokines and chemokines secreted by single T cells. We compared single-cell cytokine secretion by CXCR5+PD-1hiCCR7lo circulating follicular helper (cTfh) cells and CXCR5+PD-1loCCR7hi central memory (Tcm) cells from patients with SLE and healthy donors. Single-cell profiling revealed skewed cytokine programs with increased functional heterogeneity and dysregulated cytokine production among SLE T cell populations. Significantly, elevated production of pathogenic cytokines extended to typically more quiescent CXCR5+Tcm cells. Both cTfh and Tcm cells from SLE display a larger degree of polyfunctionality in terms of cytokine production. Using single-cell RNA-sequencing, we analyzed transcriptional heterogeneity and identified diverse cell types including potentially pathogenic cell subsets. Cytokine genes and immune-response-related transcriptional factors showed elevated and highly variable expression in SLE Tcm cells. The transcriptomes of both cTfh and Tcm were enriched in follicular helper T cell gene signatures including Bcl6, CXCR5, and IL-21. These results imply the major phenotypic shift of Tcm cells toward more effector-like T cells. In this study, our single-cell technology has served as the tool to investigate the cellular functions that may drive SLE pathogenesis.



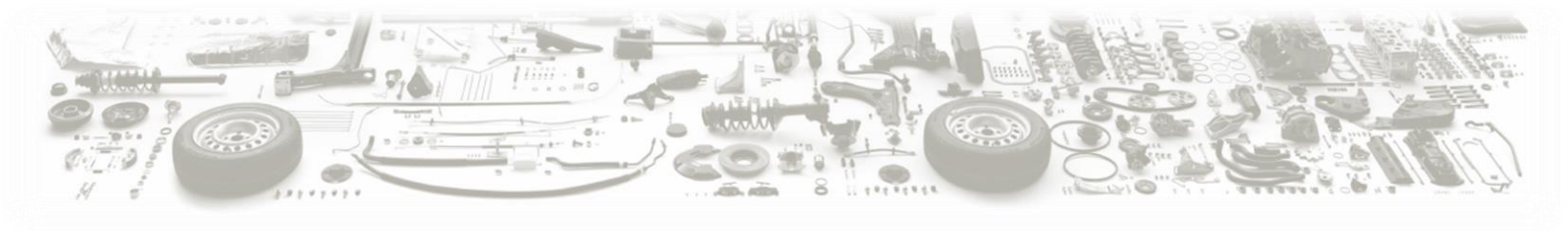
P022

Spatial mapping of single cell RNA-seq data to tissue sections, using multiplexed single-molecule fluorescence in situ hybridization.

Simone Codeluppi, Lars Borm, Amit Zeisel, Sten Linnarsson

Karolinska Institutet, Stokholm, Sweden

Recent technological advances in microfluidic and molecular biology have made characterization of the transcriptional profile of thousands of cells at high resolution feasible. However, single-cell protocols require mechanical or enzymatic disruption of the tissue source, removal of extracellular components and isolation of the cells. Therefore, the generated transcriptomes will lack of the anatomical information intrinsic of the tissue. Our goal is to add spatial information to transcriptome data by using multiplexed single-molecule fluorescence in situ hybridization (MsmFISH). We have used the mouse somatosensory cortex as model system, since its cell types were recently identified by single-cell RNAseq. Each identified cell type is characterized by a pool of differentially expressed genes that can be used as references to map the in silico clustered cell types on brain sections. To be able to determine the expression level and the position of the large amount of genes that characterize the identified cell types we developed an automated MsmFISH protocol that works on tissue sections. After molecular counting, normalization of the transcription profile of the MsmFISH identified cells was used to rank each cell in the RNAseq generated cell atlas. With the new MsmFISH protocol we have developed we are now able to add spatial information to sequencing data, bringing cell identity and function closer.



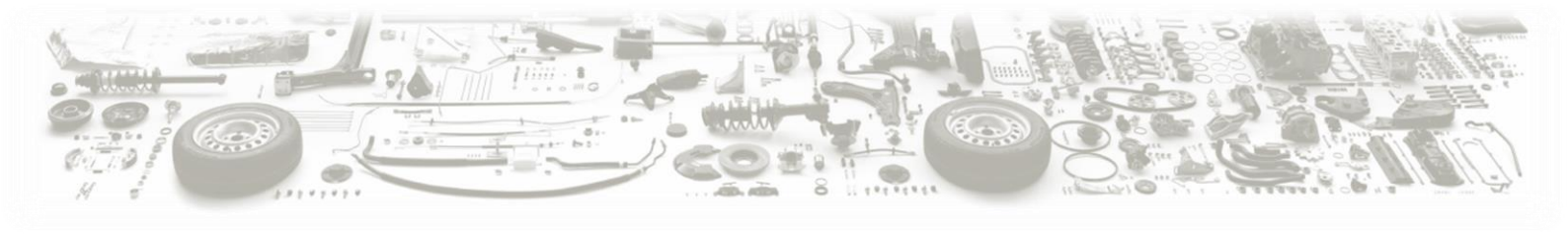
P023

De novo prediction of stem cell identity using single-cell transcriptome data

Dominic Grün[1], Mauro Muraro[2], Anna Lyubimova[2], Kay Wiebrands[2], Jean-Charles Boisset[2], Alexander van Oudenaarden[2]

[1]*Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany*
[2]*Hubrecht Institute, Utrecht, Netherlands*

Adult mitotic tissues like the intestinal epithelium or the bone marrow undergo constant turnover throughout the live of an organism. Differentiated cells have a limited lifetime requiring the existence of a somatic stem cell that can replenish all existing lineages within a tissue. To understand tissue homeostasis in a healthy organism and its aberrations upon diseases such as cancer knowledge of the stem cell identity is crucial. Single cell transcriptomics have recently proved to be a powerful tool for profiling the cell composition of tissues and organs. Here we develop a computational method for the derivation of differentiation trajectories from single cell transcriptome data. By scoring transcriptome characteristics that reflect the level of pluripotency we establish StemID, an algorithm that predicts the stem cell identity within a complex tissue. We demonstrate that StemID correctly identifies Lgr5 positive cells as the intestinal stem cells and recovers hematopoietic stem cells in the bone marrow. We applied StemID to predict candidate stem cell populations in the human pancreas, a tissue with largely uncharacterised turnover dynamics. We anticipate that StemID can leverage the full power of single cell transcriptomics for unraveling differentiation dynamics in a variety of systems.



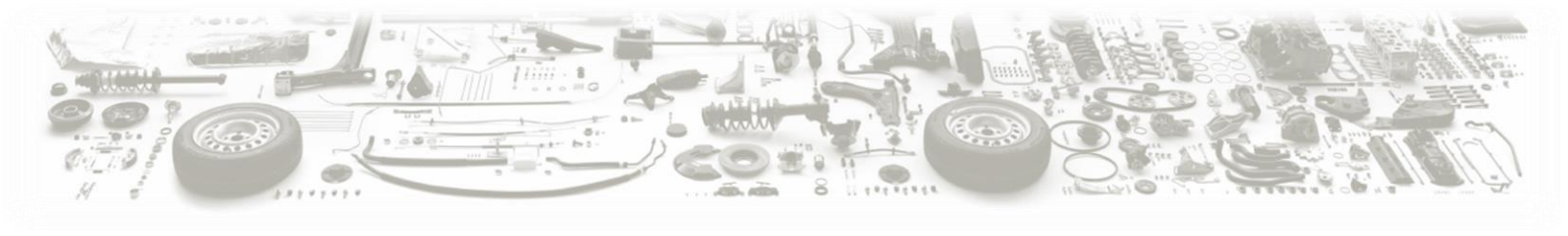
P024

Multiple Displacement Amplification-based Whole Transcriptome Amplification Enables Reliable Transcript Profiling From Limited Samples

Ioanna Andreou, Isabell Haupt, Annika Piotrowski, Nan Fang

QIAGEN GmbH, Hilden, Germany

RNA-Seq is developed to perform transcriptome profiling and provides a highly precise measurement of expression levels of transcripts and their isoforms. Normally RNA-Seq analysis requires at least 500ng -1µg of total RNA. While working with small biopsies, single cells such as CTCs, or other limited material, whole transcriptome amplification (WTA) is normally required. Various whole transcriptome amplification methods overcome limited RNA availability and enable transcriptome analysis from limited material or even single cells. In standard PCR-based WTA procedures, however, bias from uneven coverage of cDNA regions with high GC or AT content or base-copying errors can lead to the loss of transcripts.
We compare a standard RNA-Seq library prep method and RepliG RNA library protocol, which is a PCR-free protocol to efficiently generate RNA-Seq libraries from small amounts of RNA or single cell in 6.5–7 hours. The REPLI-g RNA library protocol uses Multiple Displacement Amplification (MDA), combined with an efficient library adaptor ligation procedure, to prepare RNA-Seq libraries from small RNA amounts with high fidelity, minimal bias and retention of samples transcriptional profile. Compared to the standard RNA-Seq library prep protocol, the REPLI-g RNA library protocol demonstrates similar reproducibility and sensitivity in transcript detection and quantification. Furthermore, we also demonstrated that the REPLI-g RNA library protocol can be reliably used for differential expression analysis.



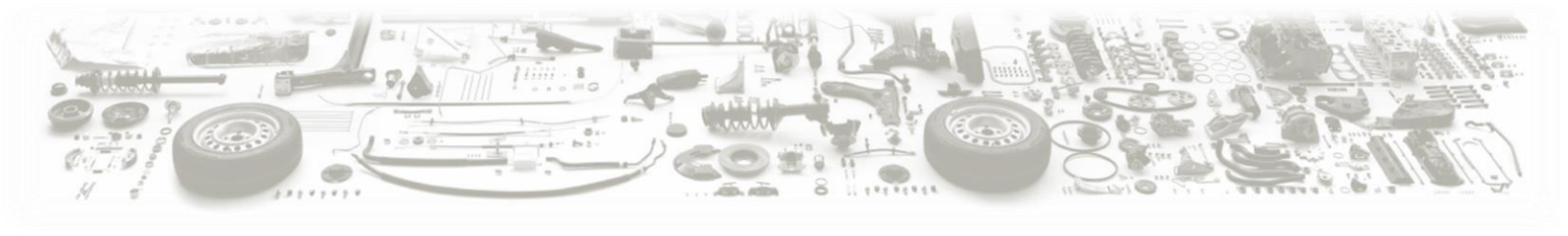
P025

Integrated DNA and RNA sequencing of single cells

Jiankui He

South University of Science and Technology of China, Shenzhen, China

Interest in single-cell whole genome analysis is growing rapidly in recent years, especially for profiling rare or heterogeneous populations of cells in cancer biology, neuroscience and stem cell studies. Various amplification protocols have been developed to sequence the genome or transcriptome of single cells. However, the current single-cell RNA-seq method or DNA sequencing method do not permit analysis of the transcriptome and genomic sequence of individual cells simultaneously. Here we described a new method to simultaneously quantify both the genome and transcriptome of the same cell by physically separating the nucleus from the cytoplasm using microinjection system. Using mouse oocytes as an example, we demonstrated that the integrated DNA and RNA sequencing of single oocytes can deliver both high coverage of genome and reproducible low-bias transcriptome. Further analysis also reveal significant RNA editing events in oocyte by using our single cell sequencing method. The technology developed in this study will be useful in clinical, when limited materials is available, such as in pre-implantation genetic diagnosis.



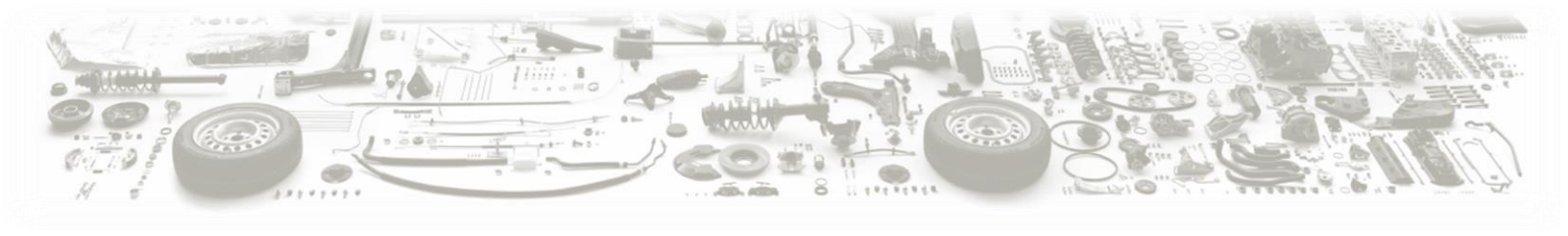
P026

Nephrectomy induced renal repair: an elegant tool to study the fate of injured cells

Benjamin Vervaet, Nathalie Le Clef, Patrick D'Haese

University Antwerp, Antwerp, Belgium

Acute renal injury typically affects the epithelial cells of the proximal tubules (PTCs). Despite its sensitivity, this cell type has a significant regenerative capacity and contributes to recovery of renal function by replacing deathly injured PTCs through proliferation. Yet, repair frequently fails leading to the development of chronic fibrosis. Recently, it has been shown that cell cycle arrested PTCs (at the G2/M stage) cause renal scar formation by excessively producing pro-fibrotic cytokines (Yang et al. Nature Medicine 2010;16(5):535-43). These authors also demonstrated that G2/M arrest in an injured kidney can be overcome by removing the uninjured contralateral kidney. If the latter kidney is left in place, repair is only marginal and the injured kidney turns fibrotic. It is yet unclear by which molecular mechanism a nephrectomy is able to alter the fate of injured kidney cells. As a first step, we here report the surgical optimization of a murine model in which the left kidney was injured by unilateral ischemia/reperfusion after which the contralateral kidney was either or not removed 3 days later. Six weeks after surgery, both gross histology and qPCR analysis of the pro-fibrotic genes Col1, Col4, TGFbeta and CCN2, confirmed that nephrectomy is able to strongly attenuate development of chronic fibrosis. Hence, we conclude this model to be relevant for further mechanistic investigations on renal repair, including fate control of proliferating PTCs. Central in such an approach is the future isolation and characterization of the PTCs by RNAseq.



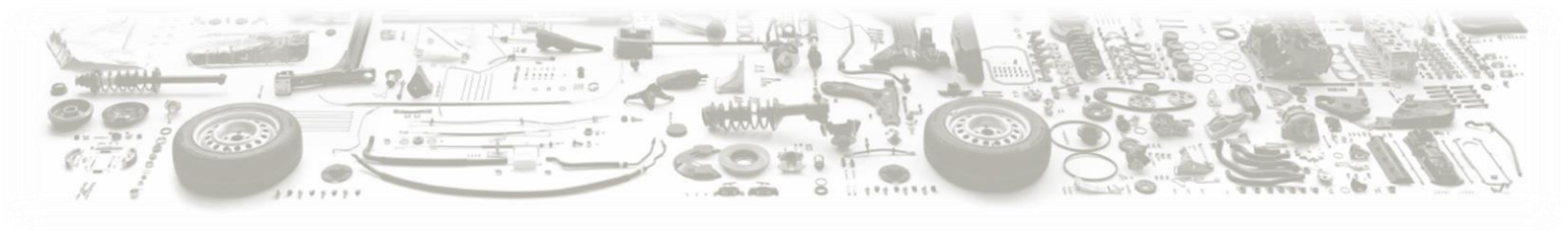
P027

The detection of complex structural variation in single cells with Strand-seq

Ewart Kuijk[1,2], Diana Spierings[3], David Porubský[3], Sjors Middelkamp[1,2], Edwin Cuppen[1,2], Peter Lansdorp[3], Wigard Kloosterman[4]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*University Medical Center Utrecht, Utrecht, Netherlands*
[3]*European Research Institute for the Biology of Ageing, University Medical Centre Groningen, Groningen, Netherlands*
[4]*Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, Netherlands*

Structural variations (SVs) are defined as rearrangements of large stretches of DNA, such as insertions, deletions, inversions, and translocations. Complex SVs have multiple breakpoints that cannot be explained by a single end-joining or DNA exchange event. A dramatic example of complex SV is chromothripsis, in which multiple chromosomes are broken and reassembled into mosaic chromosomes. Our knowledge on the origin of complex SVs such as chromothripsis is far from complete. The main difficulty in studying these events is that breakpoints of chromothripsis occur randomly, which necessitates genome-wide detection techniques. Furthermore, analysis of events occurring in single cells is necessary to understand the underlying mechanisms, rendering most genome-wide methods unsuitable. In Strand-seq, the parental DNA template strand from single cells is sequenced at low coverage. This method, initially developed to follow sister chromatids in daughter cells, can also be used to detect SVs by recurrent template switches. Here we investigated the suitability of Strand-seq to detect complex rearrangements in single cells. We performed Strand-seq on iPS cells from a patient with germline chromothripsis and both healthy parents. In the patient, the paternal chromosomes 1, 3, 7, and 12 have 17 breakpoints and the resulting fragments have been reassembled into 4 derivative chromosomes. Analysis of the Strand-seq data revealed that recurrent template switches could be readily detected at the breakpoint junctions in the chromothripsis iPS cells. These data demonstrate that Strand-seq is a powerful technique to identify complex SVs, which will help to elucidate the mechanisms that give rise to complex SVs.



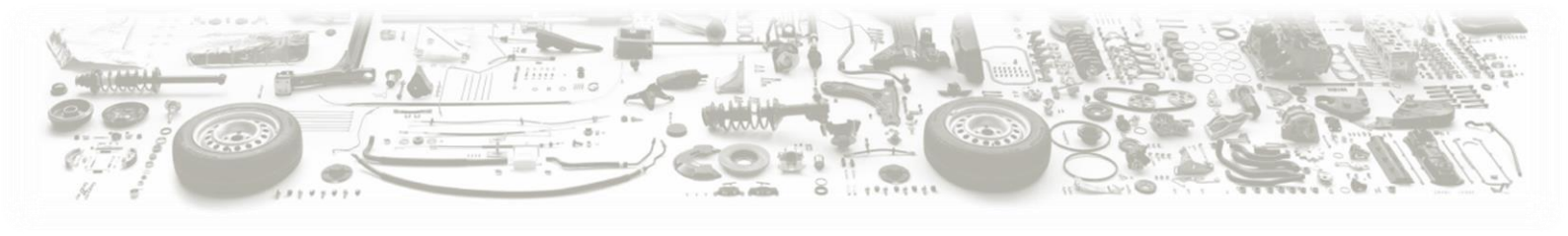
P028

Single cell microRNA sequencing

Michael VanInsberghe[1], David Knapp[2], Michelle Moksa[1], Connie Eaves[2], Martin Hirst[1], Carl Hansen[1]

[1]*University of British Columbia, Vancouver, BC, Canada*
[2]*British Columbia Cancer Agency, Vancouver, BC, Canada*

Measurements of populations at single cell resolution are critical to elucidating variable responses in immunity, measuring stochastic noise in gene expression, assessing the disease and developmental state of heterogeneous tissues, and can be used to identify potential mechanism. Such questions have driven the recent development of numerous genome-wide methods to interrogate single cell transcriptomes. These methods, however, are currently unable to measure microRNAs and other small RNAs - an important class of regulatory molecules.
Here we present a method and microfluidic platform capable of generating high-quality miRNA libraries from up to 96 single cells per device. The entire library construction process – from cell capture and lysis to indexing PCR – occurs on chip, only requiring a size-selection on the pooled libraries before sequencing. Library construction maintains both RNA integrity and relative miRNA abundance, and has a sensitivity of ~10-100 molecules. The technical noise has a coefficient of variation ranging from approximately 15% for highly-expressed miRNAs to 50% for those approaching the sensitivity limit, and is below the differences observed between single cells.
MicroRNA expression was first measured in 300 single cells sampled from a time-course retinoic acid differentiation of HL60 cells. In this model system, differential miRNA regulation dynamics were observed along the differentiation trajectory; for example miR-223 was found to be steadily upregulated, whereas miR-17 was sharply downregulated. In an effort towards understanding how miRNA change during hematopoietic stem cell (HSC) development, miRNA expression profiles were generated from more than 800 single cells from HSC, multipotent and committed progenitors, and mature cell fractions purified from Human cord blood. From these data, we were able to confirm previously-identified miRNAs associated with HSCs and early-progenitors, as well as identify new HSC-associated markers.
This technology fills a gap in single-cell genomic capabilities by enabling unbiased, genome-wide analysis of miRNA and other small RNAs, and has allowed us to investigate how miRNA change during Human hematopoietic development.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P029

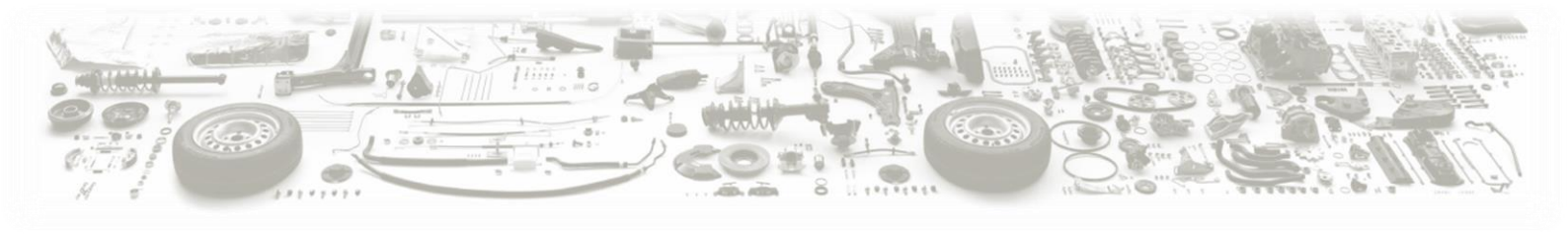
Identification of cellular biomarkers for HIV permissiveness through single-cell analyses

Simon Quenneville*[1], Antonio Rausell*[2], Sylvie Rato[1], Angela Ciuffi[1], Amalio Telenti[3]

[1]*Institute of Microbiology, University of Lausanne, Lausanne, Switzerland*
[2]*Vital-IT group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*
[3]*J. Craig Venter Institute, La Jolla, CA, United States*

Background: Variability to HIV infection can be observed at multiple levels: between individuals, between cell types and between individual cells of the same cell type. Here we used a single cell approach to study the transcriptome of individual cells in order to investigate cell heterogeneity and identify novel determinants responsible for susceptibility to HIV infection.
Methods: Whole transcriptome profiling of single cells was obtained using Fluidigm technology followed by single-cell RNA-Seq. Cells from two different donors that were differentially susceptible to HIV infection were investigated: high permissive versus low permissive. The criteria for candidate marker selection affecting HIV permissiveness included: expression at the cell surface, heterogeneous expression within donor and correlation between expression levels and permissiveness phenotype. The validation of candidate genes was performed by FACS sorting followed by HIV infection.
Results: We observed that transcriptional heterogeneity at single-cell level was mainly driven by heterogeneous cell activation state (TCR stimulation) and resulted in varying levels of cell permissiveness to HIV. Three markers, CD62L, TIM3 and CD40L, were able to enrich for susceptible cells and were thus confirmed as biomarkers of HIV susceptibility in this experimental setting. The combined use of the three markers was able to further improve the selection of susceptible cells.
Conclusions: Our data showed that cellular activation is a main factor of transcriptional heterogeneity across individual cells and determines susceptibility to HIV infection at the single-cell level. Single-cell analysis represents a general approach for biomarker identification of specific phenotypes, including latency establishment and reactivation.

* Equal contribution as co-first authors



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

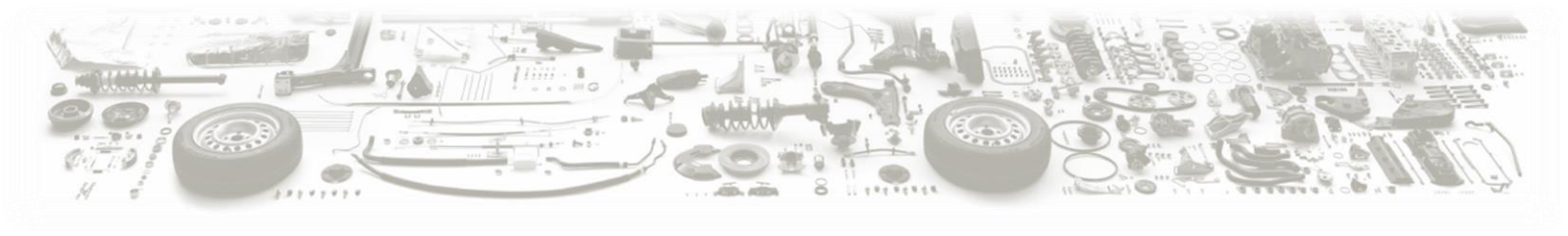
P030

Sincell: an R/Bioconductor package for the statistical assessment of cell-state hierarchies from single-cell RNA-seq data

Miguel Juliá[1], Amalio Telenti[2], Antonio Rausell[1]

[1]*Vital-IT group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland*
[2]*J. Craig Venter Institute, La Jolla, CA, United States*

Cell differentiation processes are achieved through a continuum of hierarchical intermediate cell-states that may be captured by single-cell RNA seq. Existing computational approaches for the assessment of cell-state hierarchies from single-cell data can be formalized under a general framework composed of i) a metric to assess cell-to-cell similarities (with or without a dimensionality reduction step), and ii) a graph-building algorithm (optionally making use of a cell clustering step). The Sincell R package implements a methodological toolbox allowing flexible workflows under such a framework. Furthermore, Sincell contributes new algorithms to provide cell-state hierarchies with statistical support while accounting for stochastic factors in single-cell RNA seq. Graphical representations and functional association tests are provided to interpret hierarchies. The functionalities of Sincell are illustrated in a real case study, which demonstrates its ability to discriminate noisy from stable cell-state hierarchies. Sincell is an open-source R/Bioconductor package available at http://bioconductor.org/packages/sincell. A detailed manual and a vignette are provided with the package.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

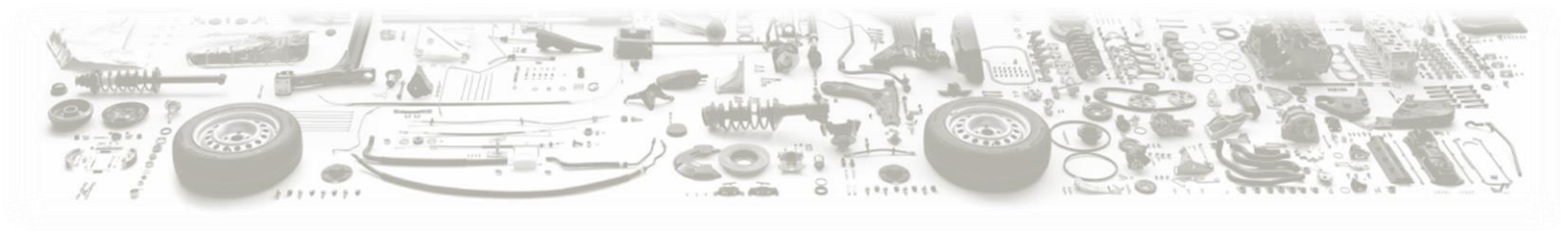
P031

Wishbone identifies bifurcating developmental trajectories in single cell data

Manu Setty[1], Michelle Tadmor[1], Shlomit Reich-Zeliger[2], Sean Bendall[3], Nir Friedman[2], Dana Pe'er[1]

[1]*Columbia University, New York, NY, United States*
[2]*Weizmann Institute of Science, Rehovot, Israel*
[3]*Stanford University, Stanford, CA, United States*

We present Wishbone, an algorithm for identification of bifurcating developmental trajectories. As input, Wishbone receives multi-dimensional single cell data and orders cells according to their developmental trajectory. Additionally, Wishbone labels each cell as belonging either to the trunk or one of two cell fates. Thus Wishbone can pinpoint where the bifurcation happens and characterize the differences following bifurcation. This allows us to trace the expression of lineage markers along the branching trajectory and characterize the decision making process for lineage commitment in single cells. Wishbone is robust to the free parameters used and significantly outperforms existing methods such as Monocle and Scuba in identification of both ordering and branching of cells.
We demonstrate Wishbone on T cell development in the mouse thymus, as this is an ideal system to study developmental trajectories with branching. In this system, CD4+ helper T cells and CD8+ cytotoxic T cells develop from lymphoid progenitors seeded from the bone marrow. We applied Wishbone to 42 channel mass cytometry data and accurately recovered the various known stages in T cell development including the bifurcation point. Further investigation revealed a preference for IL7 signaling activity along with Helios and Notch3 mediated signaling specifically in the CD8 lineage prior to negative selection.



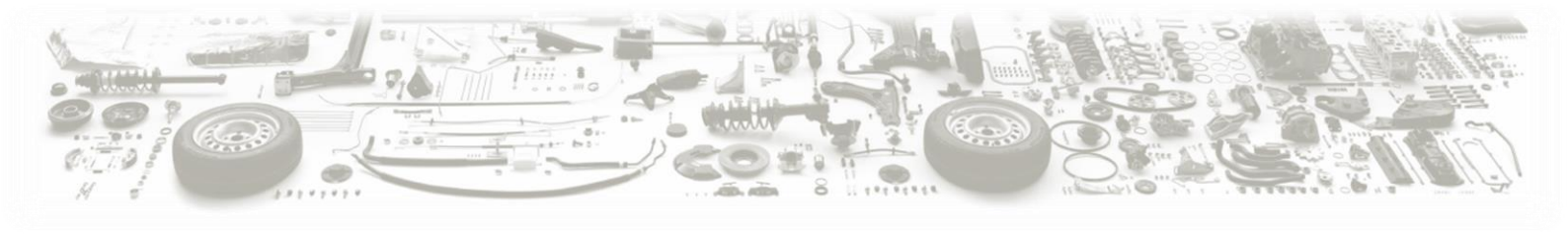
P032

Single cell DNA template strand sequencing reveals recombination hotspots in Bloom Syndrome cells

Niek van Wietmarschen, David Porubsky, Nancy Halsema, Inge Kazemier, Peter Lansdorp

University Medical Center Groningen, Groningen, Netherlands

Sister chromatid exchanges (SCEs) are the result of DNA double strand break repair via homologous recombination. As such, SCE rates are considered sensitive indicators for genomic instability. This is apparent in cells from patients with Bloom Syndrome (BS), a rare genetic disorder characterized by marked genomic instability, increased SCE rates, and a strong global cancer predisposition. However, the exact role of SCEs in genomic instability remains unclear.
We have recently described a new single cell sequencing technique called Strand-seq, which allows us to identify and map SCEs, and other genomic rearrangements, at high resolution. We used Strand-seq to identify SCEs and developed a bioinformatics pipeline to map them at a kilobase resolution. We analyzed hundreds of normal and Bloom Syndrome cells and mapped SCEs at a resolution a thousand times higher than previously possible.
Our analysis revealed a genome-wide increase in SCE frequency in BS cells, as well as several genomic regions where SCEs occur frequently across individual BS cells. Interestingly, each individual cell line appears to display a unique combination of hotspots, indicating a high level of variation between individual Bloom Syndrome patients.
We are currently investigating overall SCE locations and hotspot regions in order to elucidate the mechanism behind the high SCE rates in normal and BS cells. We expect that these results will lead to new insights into genomic instability in general, and specifically into the cancer predisposition associated with Bloom Syndrome.



P033

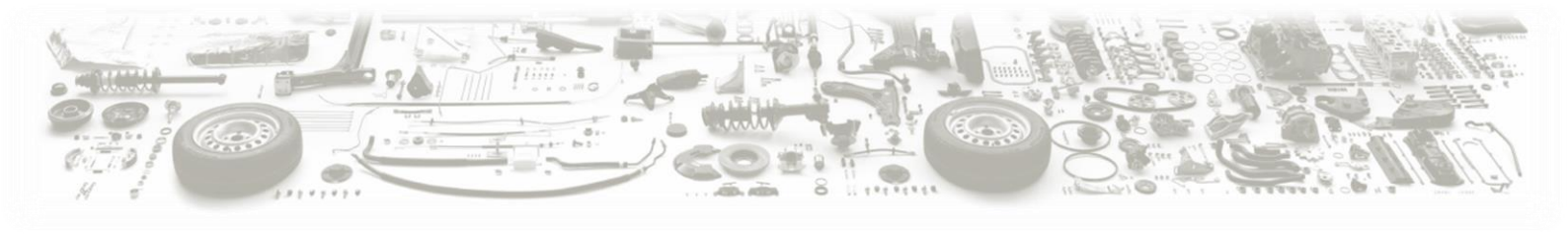
An integrated platform for: Real-time cell lineage tracing, Automated single-cell recovery, and Single-cell library construction

Marijn van Loenhout[1], Arezoo Alemzadeh Mehrizi[1], Véronique Lecault[1], David Knapp[2], James Piret[3], Connie Eaves[2], Carl Hansen[1]

[1]*Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, Canada*
[2]*Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada*
[3]*Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada*

Phenotypic cell-to-cell variability in clonal stem cell populations is able to drive cell fate decisions that enable self-renewal and the generation of differentiated progeny. Yet the mechanisms that are responsible for these fate decisions remain largely unknown. Elucidating these mechanisms requires knowledge of the 'non-genetic individuality' of cells, i.e. their developmental history as well as their molecular composition. Obtaining cell lineage information during clonal expansion and maintaining it throughout library construction and sequencing has, however, proven to be a major experimental hurdle. Here we present a platform that is able to trace lineages of expanding clones in real-time, selectively recover individual cells from those clones, and transfer them to a library construction platform that maintains their identity. An experiment starts with loading individual (Hematopoietic) cells into micro fabricated wells. Cell lineages are traced in real-time by automated image processing of time-lapse images as they are acquired. Individual cells can then be selected on the basis of lineage information and other phenotypic/morphological parameters. Selected cells are transferred to nanoliter sized droplets using an automated glass micropipette. Finally single cell RNAseq libraries are constructed using an STRT1 based protocol using a spotting robot to assemble the reactions.

Funding: This work was supported in part by a Banting postdoctoral fellowship and a KWF fellowship
1: Islam, S., et al. Nature Methods, 11(1), 163–6.



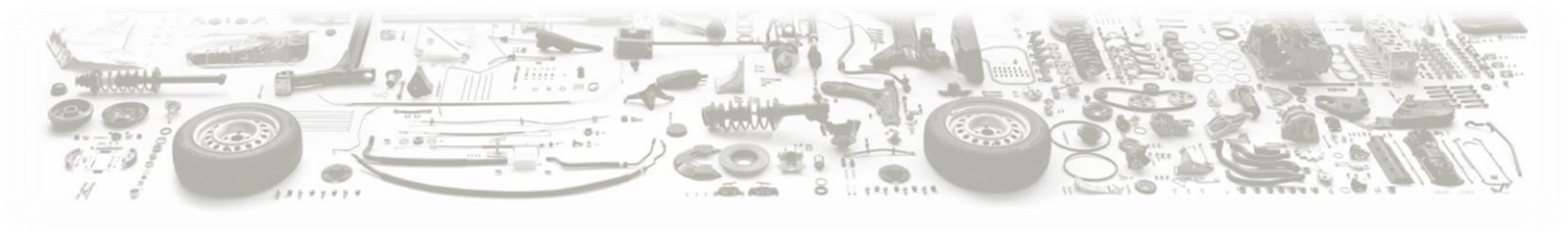
P034

A microfluidic device for single-cell genome-wide identification of copy number variation

Hans Zahn[1], Adi Steif[1,2], Sohrab Shah[1,2], Sam Aparicio[1,2], Carl Hansen[1]

[1]*University of British Columbia, Vancouver, BC, Canada*
[2]*BC Cancer Research Centre, Vancouver, BC, Canada*

A common feature of cancer is the genomic instability that results in increased rates of new lesions, including aneuploidy, large scale copy number alterations, translocations, and point mutations. As a tumour grows, these mutational events lead to the evolution of a hierarchy of lineages that have distinct genetic signatures. High-throughput sequencing technologies now make it possible to identify most of these mutations accurately, but bulk data does not reveal the co-occurrence of mutations within a given cell, allowing for only indirect inferences regarding the clonal structure. As a result it is difficult to identify distinct sub-populations of tumour cells that may drive cancer progression or be resistant to treatments.
Single-cell sequencing could address this shortfall by measuring the genotype of individual cells directly. Even though a tremendous effort is underway, only limited single cell datasets exist due to the fact that existing approaches have a limited sensitivity, are prohibitively expensive and labor intensive. To enhance the power of these single cell approaches, we are developing a flexible and scalable microfluidic platform for the construction of sequencing libraries from hundreds of single-cells. Our technology integrates the entire genomic library preparation workflow on a microfluidic device, including single cell sorting, DNA extraction, construction of indexed sequencing libraries, PCR amplification, and size selection using solid-phase DNA capture. Single cell sequencing results demonstrate that our platform yields genomes with low coverage bias suitable for genome-wide analysis of copy number variations. This technology will allow us to investigate the clonal evolution of different sub-populations of breast cancer cells and how they respond to treatment, knowledge that will be invaluable in the effort to improve clinical outcomes and develop more effective personalized medicine.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P035

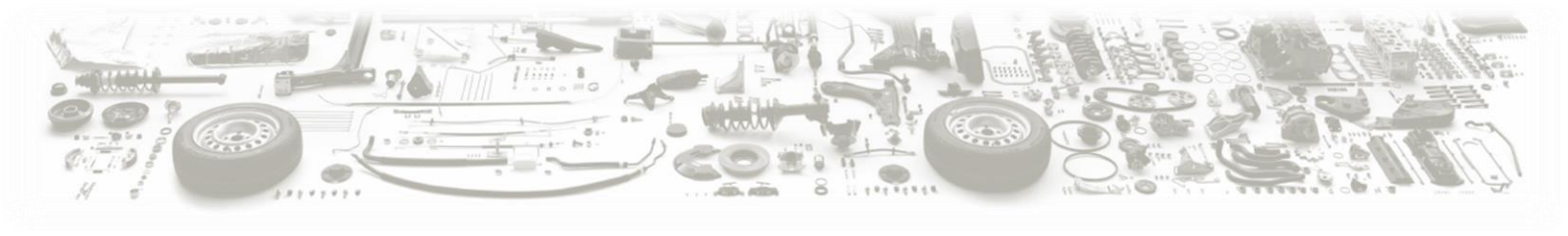
High-throughput single-cell experiments using droplet microfluidics

Anindita Basu[1], Evan Z Macosko[2], Christopher B Ford[1], Monika Kowalczyk[1], Oren Parnas[1], Karthik Shekhar[1], Rahul Satija[1], Itay Tirosh[1], James Nemesh[2], Alex K Shalek[1], Joshua R Sanes[3], Steven A McCarroll[2], David A Weitz[3], Aviv Regev[1]

[1]*Broad Institute, Cambridge, MA, United States*
[2]*Harvard Medical School, Boston, United States*
[3]*Harvard University, Cambridge, MA, United States*

We developed 'Drop-Seq' a high-throughput technique to profile transcriptomes of tens of thousands of single mammalian cells using a combination of DNA barcoding and emulsion microfluidics. This is accomplished by encapsulating and lysing one cell along with a uniquely barcoded microbead per emulsion droplet, that allows us to molecularly tag the RNA contents from each cell. This technique enables study of the transcription profiles of an unprecedented number of cells at single-cell resolution.
To demonstrate Drop-Seq's power to categorize cell types in complex tissues, we applied it to mouse retina. We analyzed ~45,000 single cells from the retina to computationally assemble an ab initio cell classification of the retina cells into 39 sub-types based on their transcription profiles. This classification faithfully reproduces existing molecular, physiological, and anatomical knowledge, while nominating novel classifications for further research. Since then we have applied Drop-Seq to a wide range of biological systems, including mouse and human bone marrow, murine bone marrow-derived macrophages and dendritic cells, and human glioblastomas.
We also use droplet microfluidics to study host-pathogen interaction at controlled multiplicity of infection (MOI) and high throughput, using macrophages and Candida Albicans (CA). We infect macrophage cells with CA at defined MOI in emulsion droplets, and use a combination of fluorescent reporters, live-cell imaging, and cell sorting to follow individual outcomes, identifying virulent sub-populations of CA, as well as instances where the infection is curbed successfully. We follow up on the sorted sub-populations using Drop-Seq.

Supported by the Klarman Cell Observatory at Broad Institute.



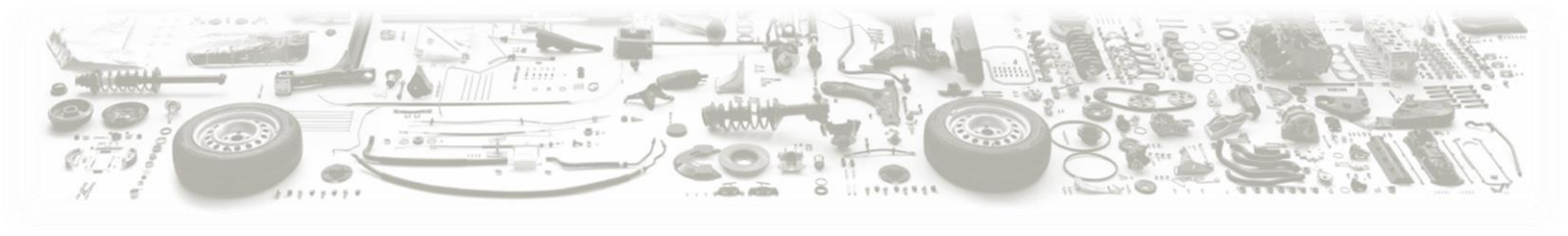
P036

A novel nanowell array system based on SmartChip™ technology for the isolation and processing of 1000s of single cells

Maithreyan Srinivasan

WaferGen Biosystems, Fremont, CA, United States

Advances in Next-Generation Sequencing have increased our ability to work with small/precious samples and grow our understanding of basic biological principles. Such biological insights are frequently derived from data "averaged" over bulk cell and/or tissue samples. However, single cell analyses suggest that cell-specific transcriptome differences may have profound functional consequences at the tissue level. While technologies to isolate and process single cells in a high throughput fashion have made progress, advancements in single cell biology require new technologies that can miniaturize and automate the isolation and processing of thousands of individual cells in a selective and flexible manner.
To satisfy this need, we have developed a high throughput single cell processing system to accommodate a number of downstream applications including RNA-seq. The system includes a 5184 nanowell array based on SmartChip technology coupled with a fluorescent microscope and a dispenser that can accurately dispense nanoliter volumes of cells and reagents into nanowells. An image-analysis program called CellSelectTM enables the selective processing of individual cell containing wells and helps minimize data analysis noise from wells containing multiple cells. Using Poisson dispense method, CellSelectTM can routinely identify about 1500 viable single cells as judged by staining for DNA and dead cells using Hoechst 33342 dye and propidium iodide, respectively. The output from the image analysis software is uploaded to the dispenser, which selectively adds reagents to wells of interest.
Five characteristics differentiate the nanowell array system from current market offerings: 1. High throughput – 1500 viable cells/array, 2. Flexibility – process up to 8 samples per array, 3. Scalability – isolate tens of thousands of cells, 4. Selectivity – ability to selectively process ONLY single-cell containing wells and 5. Cell type compatibility – ability to dispense a wide range of cell types including human tumor cells and neurons. We will present data that demonstrate the reproducible dispensing of individual cells in a Poisson dispense method using 5184 preprinted oligonucleotide-barcodes and process the cells for RNA-seq.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

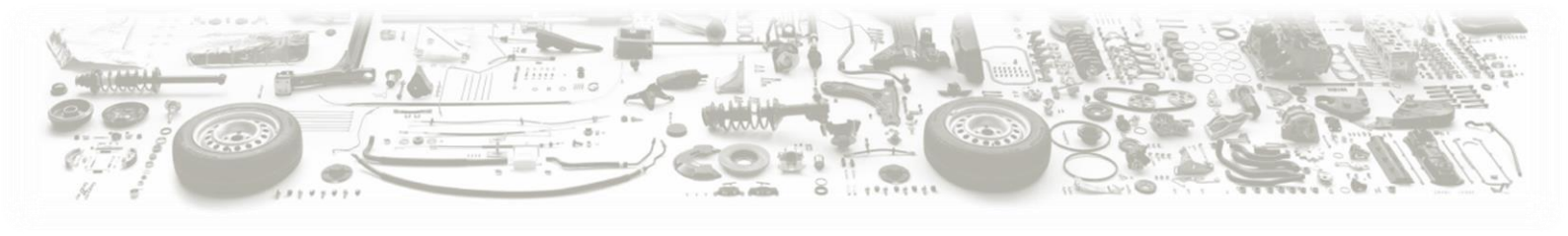
P037

Noise in gene expression is coupled to growth rate

Leeat Keren[1], David van Dijk[2], Dan Davidi[1], Shira Weingarten-Gabbay[1], Ghil Jonah[1], Adina Weinberger[1], Ron Milo[1], Eran Segal[1]

[1]*Weizmann Institute of Science, Rehovot, Israel*
[2]*Columbia University, New York, NY, United States*

Genetically identical cells exposed to the same environment display variability in gene expression (noise), with important consequences for the fidelity of cellular regulation and biological function. Although population average gene expression is tightly coupled to growth-rate, the effects of changes in environmental conditions on expression variability are not known. Here, we measure the single-cell expression distributions of ~900 S. cerevisiae promoters across four environmental conditions using flow cytometry, and find that gene expression noise is tightly coupled to the environment and is generally higher at lower growth rates. Nutrient-poor conditions, which support lower growth rates, display elevated levels of noise for most promoters, regardless of their specific expression values. We present a simple model of noise in expression that results from having an asynchronous population, with cells at different cell cycle stages, and with different partitioning of the cells between the stages at different growth rates. This model predicts non-monotonic global changes in noise at different growth rates as well as overall higher noise for cell cycle regulated genes in all conditions. The consistency between this model and our data, as well as with noise measurements of cells growing in a chemostat at well-defined growth rates, suggests that cell-cycle heterogeneity is a major contributor to gene expression noise. Finally, we identify gene and promoter features which play a role in gene expression noise across conditions. Our results show the existence of growth-related global changes in gene expression noise and suggest their potential phenotypic implications in noise-driven processes, such as persistence, survival after treatment and responsiveness to fluctuating environments.



P038

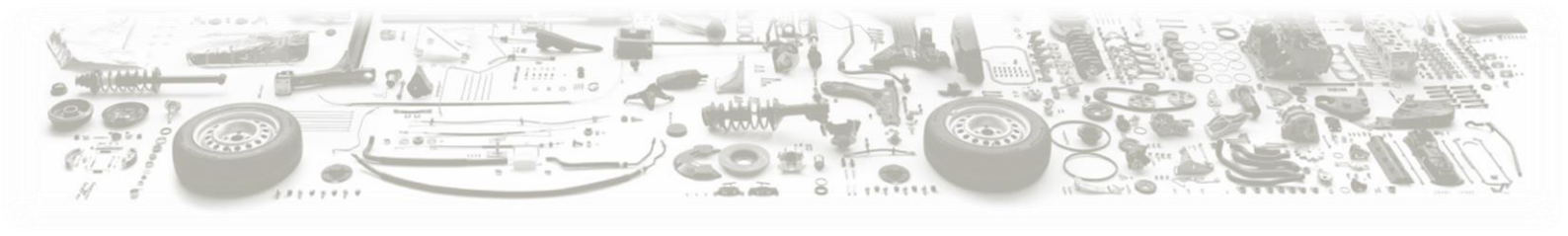
Single-cell 5-hydroxymethylcytosine sequencing reveals extensive chromosome-wide epigenetic heterogeneity among individual cells

Dylan Mooijman[1], Siddharth Dey[1], Jean-Charles Boisset[1], Nicola Crosetto[2], Alexander van Oudenaarden[1]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Karolinska Institutet, Stockholm, Sweden*

Recent advances in single-cell transcriptomics have enabled genome-wide quantification of cell-to-cell variability in gene expression. While these studies have revealed dramatic heterogeneity between the transcriptome of single cells, the upstream epigenetic mechanisms regulating this variability have been difficult to study and their role in tuning cellular heterogeneity is not well understood. Recently, the DNA modification 5-hydroxymethylcytosine (5hmC) has been shown to be present in various cell types and to have critical roles in development and gene regulation. Quantifying the heterogeneity of this epigenetic mark among individual cells is important towards understanding the dynamics and functional role of 5hmC.
We developed a genome-wide strand-specific 5hmC sequencing technology in single cells, which revealed pronounced cell-to-cell variability in the relative amounts of 5hmC on the two DNA strands of a given chromosome in mouse embryonic stem (mES) cells. Similar results in mouse embryos were used to determine that strand age is a major determinant of 5hmC variability between the two strands of DNA in a chromosome. Further, we developed a stochastic model that reproduces these findings and enables quantifying the rates of 5hmC turnover. The model predicted that this heterogeneity should decrease with increased turnover, which we experimentally confirmed by treating mES cells with vitamin C, a molecule previously shown to increase 5hmC turnover. Thus, our results demonstrate that even isogenic cell populations exposed to identical environments can display pronounced chromosome-wide epigenetic heterogeneity.

Acknowledgements: This work was supported by NWO (VICI award) and ERC (ERC-AdG 294325-GeneNoiseControl) grants.



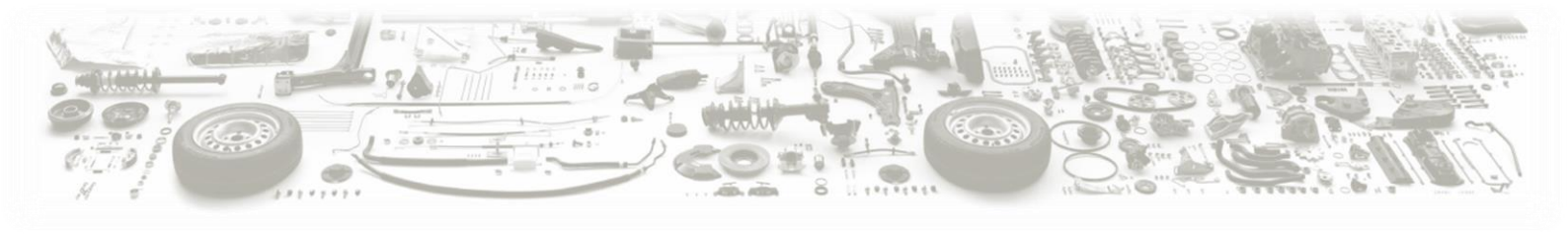
P039

Droplet transcriptomics applied to Xenopus laevis embryonic development

James Briggs[1], Allon Klein[2], Marc Kirschner[2]

[1]*Harvard University, Cambridge, MA, United States*
[2]*Harvard Medical School, Boston, MA, United States*

Specialized cell types underpin the complex physiology of multicellular organisms. Decades of histological, embryological, and genetic studies have revealed that this diversity of cell types emerges in the developing embryo through a series of hierarchical lineage decisions, which progressively restrict the developmental potential and specialize the individual behaviors of proliferating embryonic tissues. In the Kirschner and Klein laboratories, we are revisiting this account of cell type differentiation through the newly available lens of high-dimensional single cell analysis. We recently developed a novel droplet-based single cell sequencing methodology that we are now applying to profile the transcriptomes of tens of thousands of embryonic blastomeres sampled across a time-course of Xenopus laevis development, which connects the fertilized egg to a mature gastrula composed of ~12,000 cells. As others have begun to show, we are finding that single-cell sequencing is able to efficiently reveal the population structure of cell-types in an embryo, even in the absence of prior knowledge about its composition. As we increase the depth of our experiments, we expect to systematically reveal worlds of previously unappreciated cell-type diversity, for which we can evaluate and discover accurate marker gene sets. We are also developing high-dimensional data analysis methods that allow branching lineage trees to be inferred from time-course single-cell data. Ultimately we hope that, beyond generating detailed catalogues and maps, this analysis should in principle allow exploration of deep ontological questions about cell-types and their lineages. For example, are cell-types discrete, with the clear borders depicted in textbooks, or are they rather better thought of as continuums? Do cell-lineages explore individual, well-defined and deterministic paths between states, or do they rather stochastically sample from many available options? Answers to such questions have the potential to redefine the way we think about developmental biology. My poster presents progress towards addressing these questions from my recent thesis research.



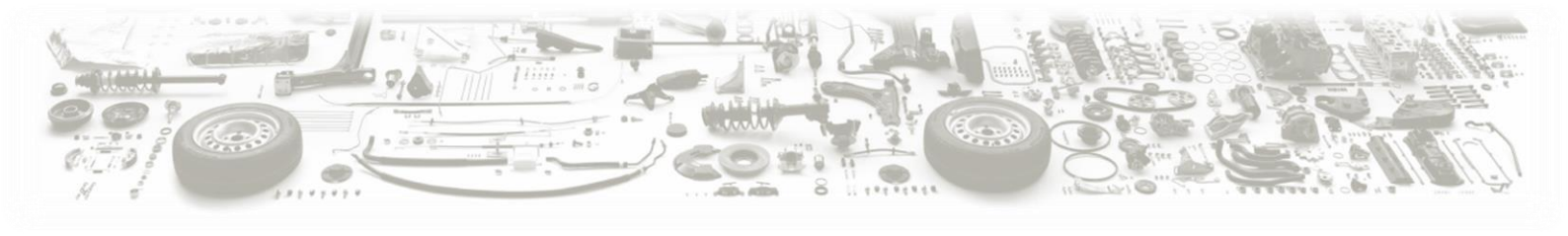
P040

Massively parallel single cell transcriptomics of the mouse retina

Karthik Shekhar

Broad Institute, Jamaica Plain, MA, United States

Dissecting cell types and cell states is essential to understanding how complex tissues are organized, and the emergence of cooperative
biological functions (e.g. immune response,visual perception). We recently developed a droplet microfluidics based technology that
enables massively parallel construction of single cell transcriptomic libraries (Drop-Seq). Applying Drop-Seq to the mouse retina and subsequent computational analysis revealed 39 cell types, recapitulating much of known retinal biology and nominating novel markers for uncharacterized types, validated using antibody staining. We are now applying Drop-Seq for directed profiling of retinal interneurons (bipolar cells, retinal ganglion cells, and amacrine cells) towards creating a comprehensive cell atlas for the murine retina. We are also interested in dissecting the cellular circuits underlying fate determination of developing retinal bipolar cells.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P041

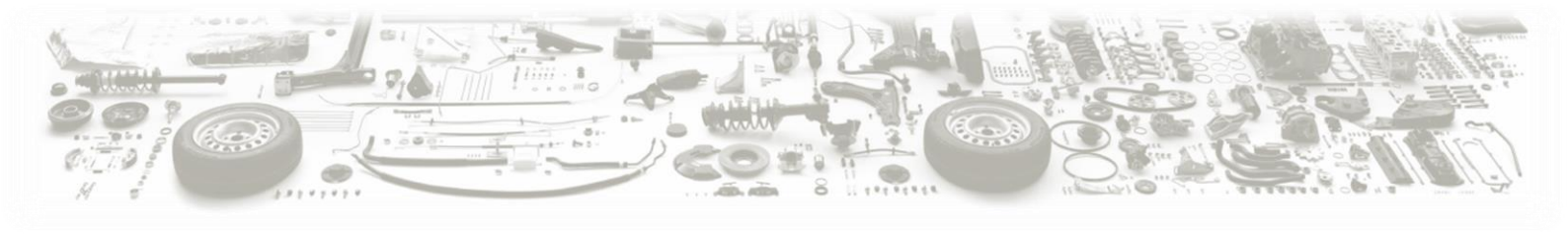
Analysis of Single Cell Fate Using Cellular Barcoding and tSNE

Jerry Gao, Shalin Naik

The Walter & Eliza Hall Institute of Medical Research, Melbourne, Australia

Cellular barcoding relies on differential tagging of precursor cells with unique heritable DNA barcodes so that subsequent quantification and barcode comparison between progeny cell types allows inference of lineage relationships. This approach has revolutionized the assessment of cell fate heterogeneity, but data derived can be highly complex and interpretation remains a major challenge. Recently, the t-Distributed Stochastic Neighbour Embedding (tSNE) dimensionality reduction algorithm has been successfully utilized for visualizing high-dimensional cytometry data. We therefore reasoned that tSNE may unmask complex traits and heterogeneity when assessing cell fate using cellular barcoding. We applied the algorithm and custom additions as a novel method for categorization of precursor cells with refined classification and intuitive biological meaning. Specifically, we further classify lymphoid-primed multipotent progenitors, previously thought to be a homogeneous population, into relatively discrete categories, and better classify haematopoietic progenitors in primates into two main types: several categories of lineage-restricted, sequential contributors, and a proportion of stable, multi-lineage contributing clones. Finally, by combining tSNE visualization across a time series in a video, we present a tool that captures a holistic overview of the dynamic and complex nature of single cell fate in longitudinal studies for more intuitive exploration and interpretation. This concept of augmenting tSNE visualizations of high-dimensional data can be widely applied across various fields of biology to explore data, perform systems-level analyses and generate novel hypotheses for subsequent validation.

This work was funded by GNT1062820 from the National Health & Medical Research Council, Australian Government.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P042

Pre-processing, quality control, and visualisation of single-cell RNA-seq data with `scater`

Davis McCarthy, Quin Wills

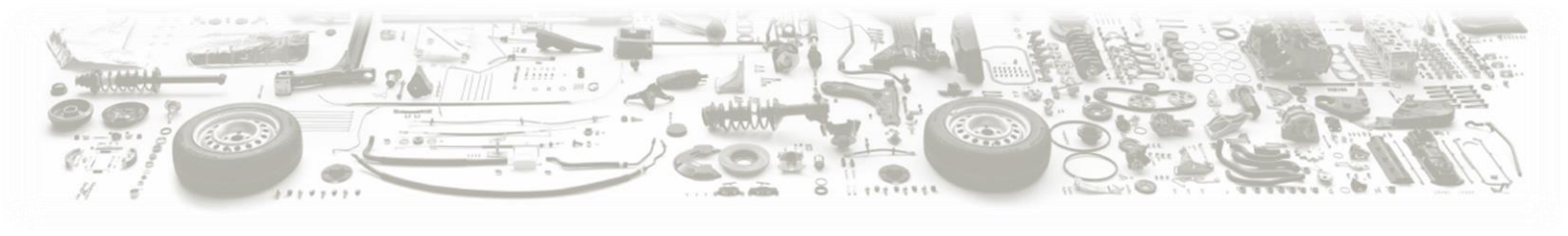
University of Oxford, Oxford, United Kingdom

Single-cell RNA-seq (scRNA-seq) is an exciting new technology that is rapidly gaining traction as a tool for understanding gene expression. We are still learning about its biases, artifacts, and other sources of unwanted variation, so producing nice, clean scRNA-seq datasets ready for statistical analysis currently takes a lot of work. The `scater` package ("single-cell analysis toolkit for expression in R") aims to ease this burden, by making rigorous pre-processing, quality control (QC), and visualisation easier so that we can proceed more quickly to downstream analyses.

At its core, `scater` provides a flexible data structure for scRNA-seq. Interoperability with other tools is a focus, and this data structure provides a foundation for other developers working on downstream analysis methods, as well as interacting smoothly with other R/Bioconductor tools. The package provides wrappers to `kallisto`, for fast quantification of transcript abundance.

The package provides convenient functions to compute a range of QC metrics. A substantial suite of plotting functions allows visualisation of QC metrics, gene expression levels, overviews of the dataset, reduced-dimension representations of cells, including principal components and t-distributed stochastic neighbour embedding, and more. These methods are useful for filtering genes and cells, and identifying variables that should be accounted for in normalisation or modelling.

We hope that `scater` fills a useful niche, enabling a workflow from raw RNA-seq reads to tidy data for downstream modelling entirely within R. The open-source code is available at http://davismcc.github.io/scater/, along with installation instructions and a vignette introducing users to the capabilities of `scater`.



P043

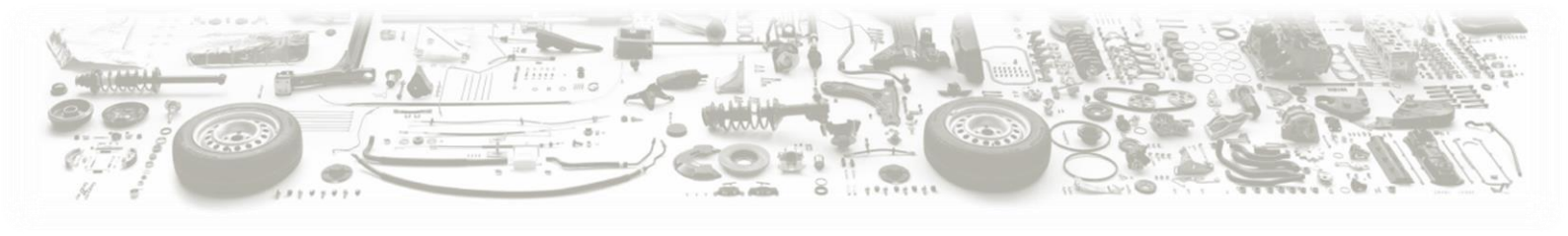
Spatially resolved RNA-Sequencing in the regenerating flatworm Macrostomum lignano

Philipp Weissert[1], Jan Philipp Junker[2], Alexander van Oudenaarden[2], Eugene Berezikov[1]

[1]*ERIBA, UMCG, Groningen, Netherlands*
[2]*Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, Netherlands*

Over the last years, RNA-Sequencing and single cell analysis techniques have been improved substantially, enabling more and more in-depth analyses of tissue samples. Recently, spatially resolved RNA-Sequencing has been refined to analyze gene expression of increasingly smaller subsets of cells down to even single cell level (1,2). We are employing the Genome-wide RNA Tomography technique (3) to spatially map gene expression over the full body-length of Macrostomum lignano, a powerful flatworm model organism for research on regeneration. One of the notable aspects of Macrostomum lignano, which distinguishes it from other flatworm models, is its very specific regenerative ability. Anterior regeneration is only possible anterior of the brain, whereas posterior regeneration is possible over a much wider range of the body (posterior to the pharynx). However, both regeneration-competent and in some cases also regeneration-incompetent wounds form a blastema (4). Although blastemal cells in various organisms have been examined by comparative analysis of RNA transcripts, spatial resolution has so far been limited and distinguishing between different subsets of blastemal populations difficult. Here, we show a genome-wide analysis of RNA transcript levels spatially resolved along the anterior-posterior axis of regenerating Macrostomum lignano, including the regeneration-competent tail blastema.

References:
(1) Combs PA, Eisen MB (2013), PLoS ONE 8(8)
(2) Junker JP and van Oudenaarden A (2014), Cell 157(1)
(3) Junker JP et al. (2014), Cell 159(3)
(4) Egger B et al. (2006), Dev Genes Evol. 216(10)
Funding sources and acknowledgements: This work is supported by ERC Starting grant to E.B.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

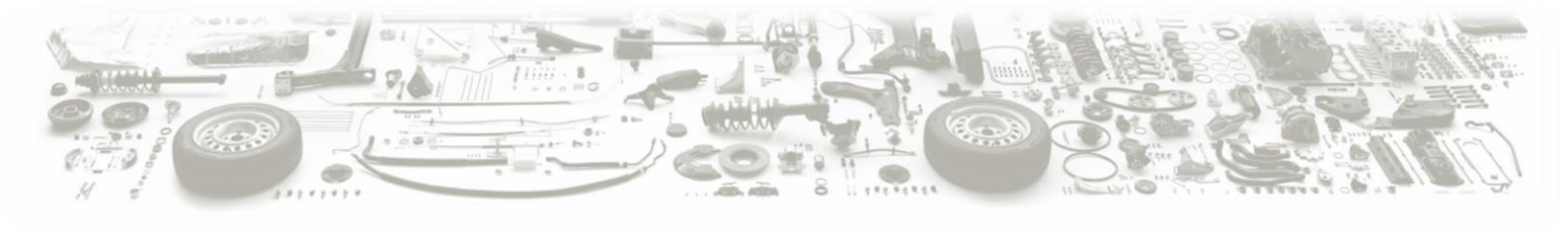
P044

Accounting for technical noise in unique molecular identifiers

Jong Kyoung Kim, John C. Marioni

EMBL-EBI, Hinxton, Cambridgeshire, United Kingdom

Single-cell RNA-seq (scRNA-seq) protocols require linear or exponential amplification of the minute amount of mRNA present in an individual cell, which is the main source of high levels of technical noise. To eliminate amplification-induced technical noise in scRNA-seq protocols, a recent approach called unique molecular identifiers (UMIs) has been proposed to count the absolute number of transcripts per gene within a single cell. Several studies have shown that UMIs reduce the overall levels of technical noise by eliminating technical noise introduced by amplification. Here we describe two UMI-specific sources of technical noise, which might lead to the overestimation of technical noise in UMI-based scRNA-seq data if not correctly accounted for: 1) the stochastic labelling process of UMIs and 2) amplification and sequencing errors in the molecular barcodes. By statistically modelling these UMI-specific technical noise, we provide practical guidelines and solutions to reduce the overall levels of technical noise in UMI-based scRNA-seq data.



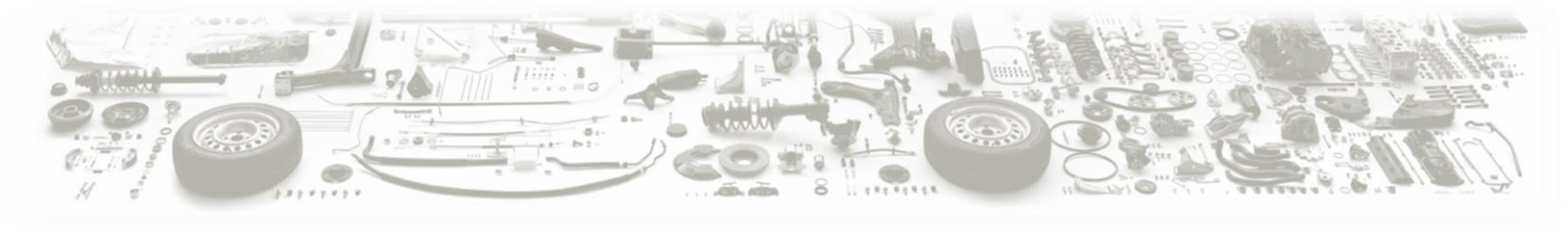
P045

A PCR-free protocol for single-cell DNA library construction

Ioanna Andreou[1], Nan Fang[1], Annika Piotrowski[1], Mary von Langsdorff[1], Devika Mathur[1], Mary von Langsdorf[1]f, Xuanyan Xu[2]

[1]*QIAGEN, Hilden, Germany*
[2]*QIAGEN, Venlo, Netherlands*

Single-cell genomic analysis enables researchers to gain novel insights across a diverse set of applications in developmental biology, tumor heterogeneity, and disease pathogenesis and progression. Typically, conducting single-cell genomic analysis using next-generation sequencing (NGS) methods is challenging because the amount of genomic DNA present in a single cell is very limited. PCR-based whole genome amplification methods tend to have high error rates, low coverage uniformity, extensive allelic drop-outs and limited amplification yields.
We describe a streamlined workflow for single-cell NGS library construction that uses Multiple Displacement Amplification (MDA) to amplify the whole genome with high uniformity and fidelity, combined with high adaptor ligation efficiency library construction. The entire procedure generates high-quality sequencing libraries without PCR amplification, thereby eliminating PCR-related bias and errors and reducing handling steps. Our data demonstrate that this PCR-free method for single-cell sequencing library preparation affords highly uniform sequence coverage and high library complexity.



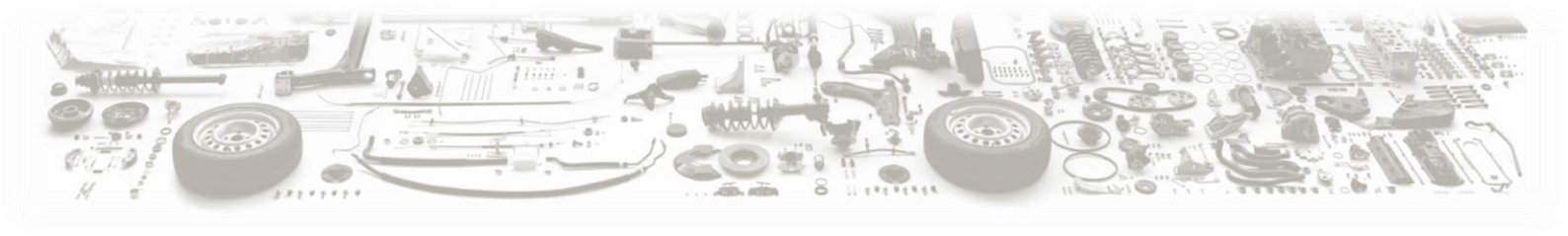
P046

Sequential Massively Parallel Single-cell RNA-Seq

Hadas Keren-Shaul, Ephraim Kenigsberg, Diego Adhemar Jaitin, Franziska Paul, Amos Tanay, Ido Amit

Weizmann Institute of Science, Rehovot, Israel

Single cell analysis can transform biology and medicine by uncovering the molecular and cellular dynamics of whole organisms; nevertheless, routine single cell analysis is challenging. Methods for high-throughput measurement and analysis of single cell RNA-seq are advancing rapidly, however a streamlined pipeline for experimental and accurate analysis of large-scale single cell transcriptomic measurements is currently lacking. Here we present Mars-seq 2.0, an experimental and analytical pipeline for routine biological experimentation in single cell resolution. We will discuss multiple improvements to MARS-seq including optimization of enzymes, robotic processes, lysis, primers and reaction volumes which significantly lower cost, improve stability and reduce the noise by 10-fold. We will also discuss a streamlined analytical pipeline with multiple layers of error detection and correction, graphically presenting key statistics of library complexity, noise distribution and sequencing saturation. Applications of our single cell RNA-seq flow can be used in any laboratory or industrial research thus providing biologists and physicians with an advanced user-friendly application for single cell research.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

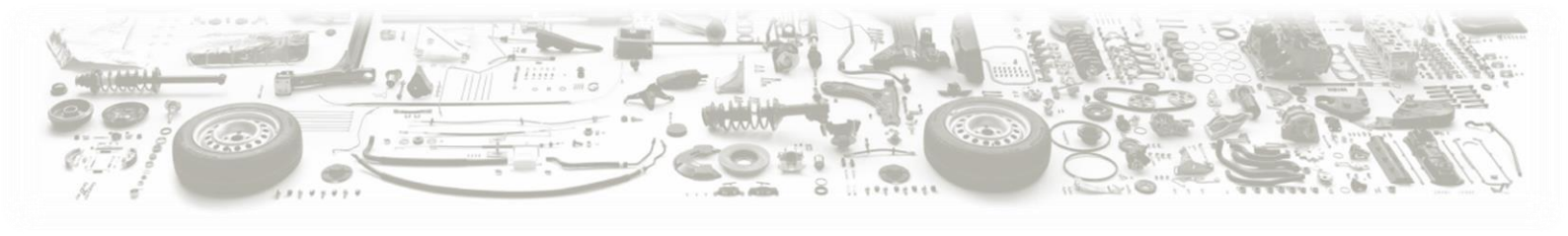
P047

A PCR-free protocol for single-cell RNA library construction

Ioanna Andreou[1], Nan Fang[1], Isabell Haupt[1], Mary von Langsdorff[1], Devika Mathur[1], Xuanyan Xu[2]

[1]*QIAGEN, Hilden, Germany*
[2]*QIAGEN, Venlo, Netherlands*

Recent advances in single-cell RNA sequencing technologies have revealed high heterogeneity in gene expression profiles of individual cells. Various whole transcriptome amplification (WTA) methods overcome limited RNA availability and enable transcriptome analysis of single cells. In standard PCR-based WTA procedures, however, uneven coverage of cDNA regions with high GC or AT content and base-copying errors can lead to the loss of data from a subset of actively expressed transcripts.
We describe a PCR-free protocol to efficiently generate RNA-seq libraries from a single cell or as little as picogram amounts of RNA in 6.5–7 hours. This protocol uses innovative Multiple Displacement Amplification (MDA), combined with an efficient library adaptor ligation procedure, to prepare RNA-seq libraries that retain the unique transcriptional profile of a single cell. All enzymatic steps are optimized for the efficient processing of RNA and amplification of cDNA with high fidelity and highly uniform coverage. This streamlined, PCR-free library construction procedure delivers high-quality libraries that enable whole transcriptome analysis from single-cell input.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P048

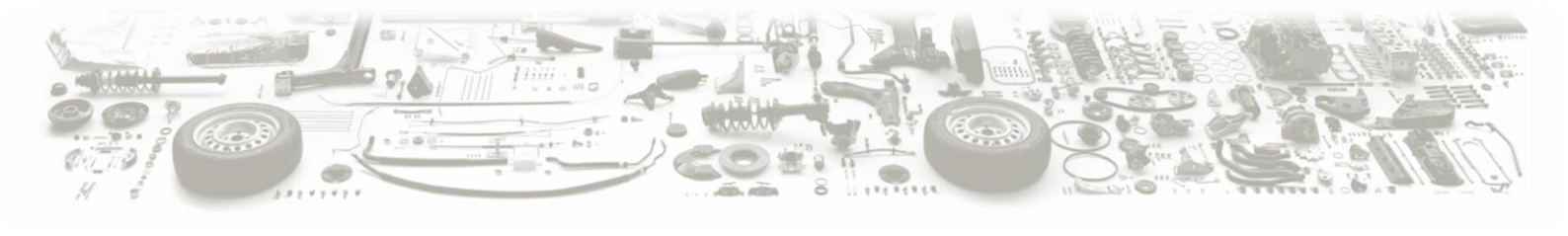
Single-cell transcriptomics meets epigenomics: heterogeneous Active/Polycomb promoter states during neuronal differentiation

Elena Torlai Triglia[1], Carmelo Ferrai[1], Andrea Piccolo[2], Tiago Rito[1], Aleksandra Kolodziejczyk[3], Gozde Kar[3], Sarah Teichmann[3], Mario Nicodemi[2], Ana Pombo[1]

[1]*Max Delbrück Center for Molecular Medicine, Berlin, Germany*
[2]*Universitá di Napoli, Naples, Italy*
[3]*EMBL-EBI, Cambridge, United Kingdom*

In mouse Embryonic Stem Cells (ESCs), developmental regulator genes are marked by Polycomb and associated with poised RNA Polymerase II (RNAPII), which keeps these genes primed for activation. Polycomb repression also acts on genes important for signalling and energy metabolism, which fluctuate between Active and Polycomb repressed states in ESCs (Brookes et al., 2012). We call these genes Active/Polycomb.
To investigate whether the Active/Polycomb state is ESC specific, we mapped genome-wide the occupancy of RNAPII and Polycomb along differentiation, capturing early exit from pluripotency and extending to immature and mature neuronal states. Our results show that the Active/Polycomb state is a common and dynamic promoter state, present in all the analysed differentiation steps. Active/Polycomb genes are characterised by intermediate expression levels, suggesting a role for Polycomb in fine-tuning their expression.
To explore whether the fluctuations between Active and Polycomb repressed states occur in different cells or alleles, we used two approaches that dissect transcriptional profiles at single-cell resolution. We applied stochastic modelling (Armond et al., 2014) to infer the expression in single cells from population-based transcriptomic data. Moreover, we produced single--cell mRNA--seq datasets in ESCs and differentiating cells using Fluidigm C1. By integrating epigenomics with single-cell transcriptomics, we identify a striking relationship that connects Active/Polycomb states with lower frequency of expression across the cell population.

Armond et al., 2014. Scientific Reports, 4, 3692
Brookes et al., 2012. Cell Stem Cell, 10(2), 157–70
We thank the Helmholtz Foundation (DE), BBSRC (UK), MRC (UK) for support.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P049

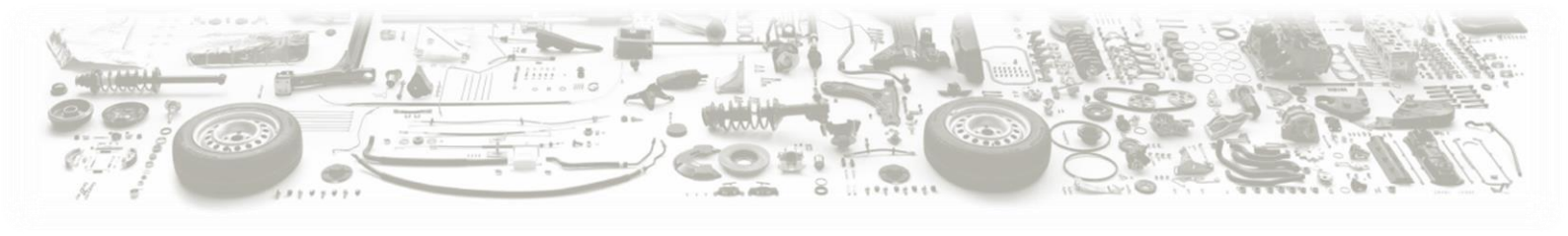
Probing vascular smooth muscle cell plasticity with single-cell RNA-seq

Erin Oerton[1], Jennifer Harman[2], Phoebe Oldach[1], Helle Jørgensen*[2], Mikhail Spivakov*[1]

[1]*Babraham Institute, Cambridge, United Kingdom*
[2]*University of Cambridge, Cambridge, United Kingdom*

Vascular smooth muscle cells (VSMCs) possess a remarkable capacity to change phenotype in response to injury, termed phenotypic switching. In healthy vessels, these cells exist in a 'contractile' state where they regulate blood flow and blood pressure, but reversibly switch into a so-called 'synthetic' state in order to repair the vessel wall. Misregulation of this process is a hallmark of many vascular diseases including atherosclerosis. During phenotypic switching, VSMCs downregulate the contractile differentiation markers displayed in healthy tissue. At the same time, these cells show increased migration, proliferation and secretion of proinflammatory cytokines, thereby contributing directly to disease-associated vascular inflammation and vessel narrowing.
It is unclear whether all VSMCs have equal potential for phenotype switching or if this plasticity is restricted to a specific sub-set of cells. This question has become particularly pertinent after recent claims that stem cell populations residing within the vessel wall may also contribute to disease. To address these issues, we are using single-cell RNA-seq to profile gene expression in individual ex vivo VSMCs. We are particularly interested in learning whether phenotype switching is a ‘binary’ process and whether VSMC heterogeneity underlies their differential capacity to undergo phenotypic switching.

* Senior authors on this study.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

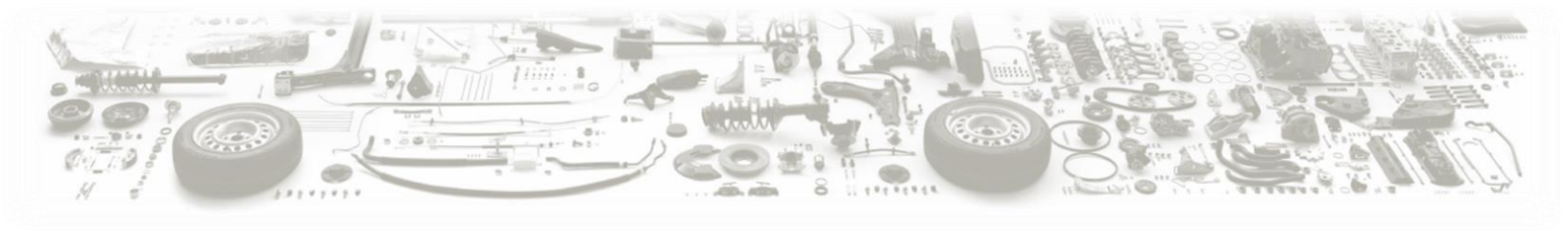
P050

Building complete chromosomal haplotypes using single-cell sequencing

David Porubsky[1], Ashley Sanders[2], Niek van Wietmarschen[3], Nancy Halsema[3], Inge Kazemier[3], Marianna Bevova[3], Victor Guryev[3], Peter Lansdorp[3]

[1]*European Research Institute for the Biology of Ageing, Groningen, Netherlands*
[2]*BC Cancer Agency, Vancouver, BC, Canada*
[3]*University Medical Center Groningen, Groningen, Netherlands*

Genomes of diploid organisms are organized in pairs of chromosomes, one inherited from the father and the other from the mother. Each homologous chromosome harbors specific parental alleles, and the relative position of alleles to each other has proven to be clinically important. Assignment of these alleles to corresponding homologous chromosomes is referred to as phasing and defines the unique sequence of alleles residing on a single homologue, called a haplotype. Knowing the haplotype of each homologue of a patient is invaluable for studying gene-disease associations, disease susceptibility and disease inheritance patterns. While current sequencing technologies can easily detect positions of variant sites in a genome, the relative position of these sites is hidden in the mix of reads representing both paternal and maternal homologues. Here we use the single-cell sequencing technique, called Strand-seq, to read single-stranded template DNA of each homologous chromosome. The beauty of this technique stems from retaining directionality of each sequencing read, which allows us to map each read to a single parental chromosome, and phase a complete homologue in a single cell. Combining the sequencing data from several single cell allows us to reconstruct an individual's genome while maintaining phase and haplotype information. We have used this technique to build chromosome-length haplotypes for all members of a well-studied HapMap family trio. Comparison of our results with HapMap reference showed high concordance, demonstrating the accuracy and sensitivity of our approach. Moreover, since we do not require multiple generations to build accurate haplotypes, we were able to more precisely map the parental homologues in this trio, to study meiotic recombination events in the child of this family. Taken together, implementing Strand-seq and our analysis pipeline offers a powerful, high-throughput approach to rapidly assemble haplotypes, which we predict will open up new possibilities for genetic association studies and personalized medicine.



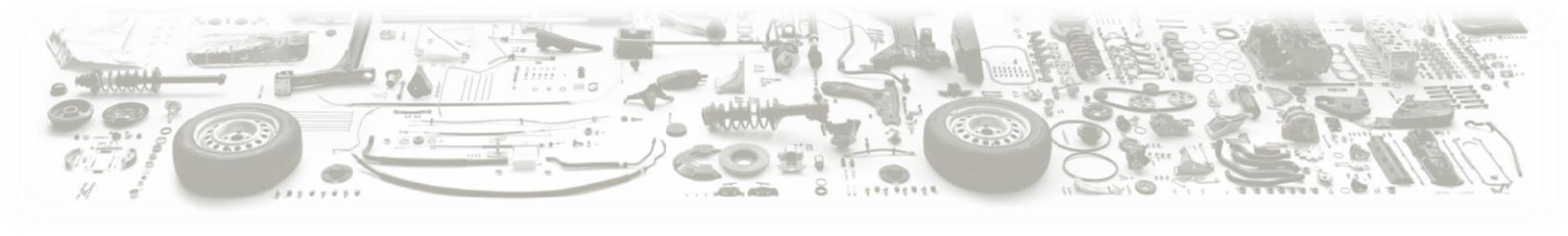
P052

Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow

Andreas Schlitzer[1], Sivakamasundari V[2], Jinmiao Chen[3], Michael Poidinger[3], Shalin Naik[4], Paul Robson[5], Florent Ginhoux[3]

[1]*University of Bonn, Bonn, Germany*
[2]*Genomics Institute Singapore, Singapore, Singapore*
[3]*Singapore Immunology Network, Singapore, Singapore*
[4]*WEHI, Melbourne, Australia*
[5]*Jackson Laboratory for Genomic Medicine, Farmington, CT, United States*

Mouse conventional dendritic cells (cDCs) can be classified into two functionally distinct lineages: the CD8a+ (CD103+) cDC1 lineage, and the CD11b+ cDC2 lineage. cDCs arise from a cascade of bone marrow (BM) DC-committed progenitor
cells that include the common DC progenitors (CDPs) and pre-DCs, which exit the BM and seed peripheral tissues before differentiating locally into mature cDCs. Where and when commitment to the cDC1 or cDC2 lineage occurs remains poorly understood. Here we found that transcriptional signatures of the cDC1 and cDC2 lineages became evident at the single-cell level from the CDP stage. We also identified Siglec-H and Ly6C as lineage markers that distinguished pre-DC subpopulations committed to the cDC1 lineage (Siglec-H−Ly6C− pre-DCs) or cDC2 lineage (Siglec-H−Ly6C+ pre-DCs). Our results indicate that commitment to the cDC1 or cDC2 lineage occurs in the BM and not in the periphery.



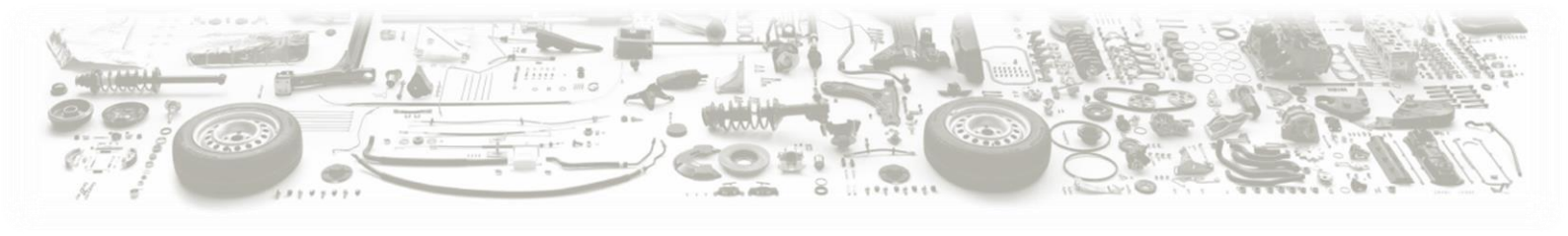
P053

Regulation of mRNA translation during mitosis

Marvin Tanenbaum[1], Noam Stern-Ginossar[2], Jonathan Weissman[3], Ron Vale[3]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Weizmann Institute, Rehovot, Israel*
[3]*UCSF, San Francisco, CA, United States*

Entry into and exit from mitosis is driven by precisely-timed changes in protein abundance, and involves transcriptional regulation and protein degradation. However, the role of translational regulation in modulating cellular protein content during mitosis remains poorly understood. Here, using ribosome profiling, we show that translational, rather than transcriptional regulation is the dominant mechanism for modulating protein synthesis at mitotic entry. We find both a weak global translational repression, as well as a highly potent, gene-specific repression of a small subset of mRNAs (~200). Almost all translationally repressed mRNAs are re-activated in G1 phase, indicating that the regulation is rapidly reversed as cells exit mitosis. The rapid kinetics and reversibility of regulation make translational control the ideal mode of regulation for this cell cycle transition. One of the most pronounced translationally repressed genes in mitosis is Emi1, an inhibitor of the anaphase promoting complex (APC), which is degraded during mitosis to allow subsequent APC activation. We find, however, that Emi1 protein degradation is insufficient for full Emi1 inhibition. Rather, combined inhibition of protein synthesis and protein degradation is required for complete Emi1 inhibition and APC activation. These results show that translational control is the dominant mode of regulation during mitotic entry and exit, and suggest that translational repression may be used to enhance post-translational protein inactivation.



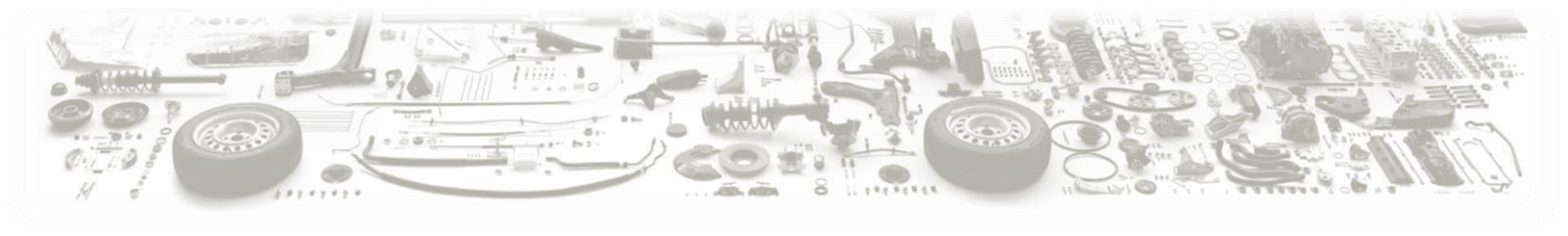
P054

Unsupervised Ordering of Single-Cell along Developmental Path using Sorting Points into Neighborhoods (uSPIN)

Mai Chan Lau, Chen Hao, Michael Poidinger, Jinmiao Chen

Singapore Immunology Network, Singapore, Singapore

Recent studies carried out at single-cell resolution have revealed that a seemingly homogenous cell population often comprises of cells at various proliferating and differentiating stages. Being able to accurately order these cells into their developmental path allows for better understanding of cellular dynamics, whereby key regulatory genes can be identified effectively. Improvements on existing sorting methods for single-cell RNA-seq are desirable as they either require prior knowledge in selecting cell sorting genes, or generate sub-optimal local cell ordering. In this work, we propose a Sorting Points into Neighborhoods (SPIN)-based cell ordering method, called unsupervised SPIN (uSPIN). SPIN provides a means to reveal linear and nonlinear relationship through ordering pairwise distance matrices. SPIN was originally applied to bulk transcriptomic data for mapping progression of colon cancer [1]. Here, we adopted and modified SPIN to derive single-cell developmental trajectory from single-cell RNA-seq data. uSPIN is an automatic cell sorting method with unsupervised gene selection and improved local ordering. uSPIN achieves unsupervised gene selection through comparing cells located at the earliest and latest developmental stages as identified by SPIN. Furthermore, uSPIN improves accuracy of cell ordering via local sorting. We applied uSPIN to single-cell RNA-seq data of macrophage–DC progenitor (MDP), common DC precursor (CDP) and pre-dendritic cells (pre-DC), wherein uSPIN was able to recapitulate the developmental cascade from MDP through CDP to pre-DC and revealed different developmental stages of pre-DCs.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P055

Mpath maps multi-branching single-cell trajectories revealing dendritic cell subset commitment of progenitors in mouse bone marrow

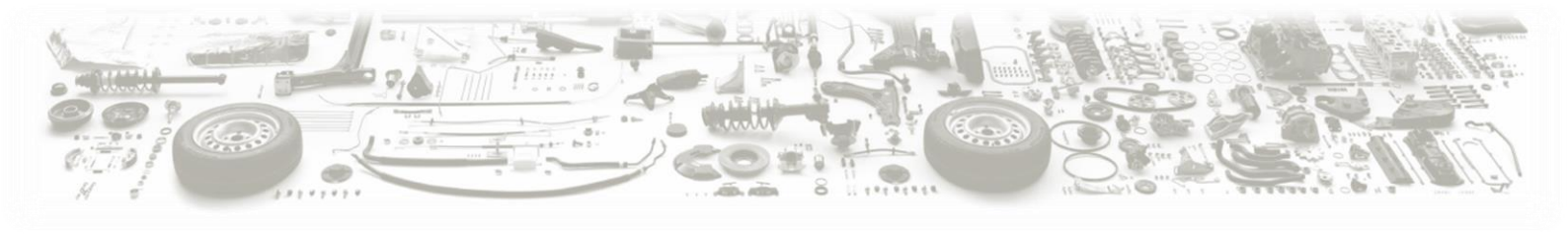
Jinmiao Chen, Andreas Schlitzer, Florent Ginhoux, Michael Poidinger

Singapore Immunology Network, Singapore, Singapore

Single-cell RNA sequencing offers unprecedented resolution of the continuum of transitional cell states during cellular differentiation and development. However, to date solutions for determining multi-branching developmental trajectories have been limited. Here we present Mpath, an algorithm that derives multi-branching developmental trajectories by neighborhood-based cell state transition mapping from single-cell RNAseq data. Applied to mouse dendritic cell (DC) progenitors, Mpath constructed multi-branching trajectories spanning from macrophage DC progenitors (MDP) through common DC progenitor (CDP) to pre-dendritic cells (pre-DC). This trajectory detected a branching event at pre-DC stage revealing pre-DC subsets that were exclusively committed to cDC1 or cDC2 lineages, wherein regulatory markers of early DC subset commitment were identified. Re-ordering cells along developmental trajectories reveals phase-dependent waves of gene regulation and a switch from cell cycle, metabolism and adhesion to cell differentiation during DC maturation and subset commitment. This study produces results that both support and expand our previous study on the use of single cell RNAseq to examine DC development[1]. Applied to a different dataset, Mpath recapitulated time course of primary human myoblast differentiation without explicitly given time point information, and isolated a branch of non-differentiating cells. Our study suggests Mpath will be a useful tool to examine a wide range of other similar processes.

References

1. Schlitzer, A. et al. Nature immunology 16, 718-728, doi:10.1038/ni.3200 (2015).



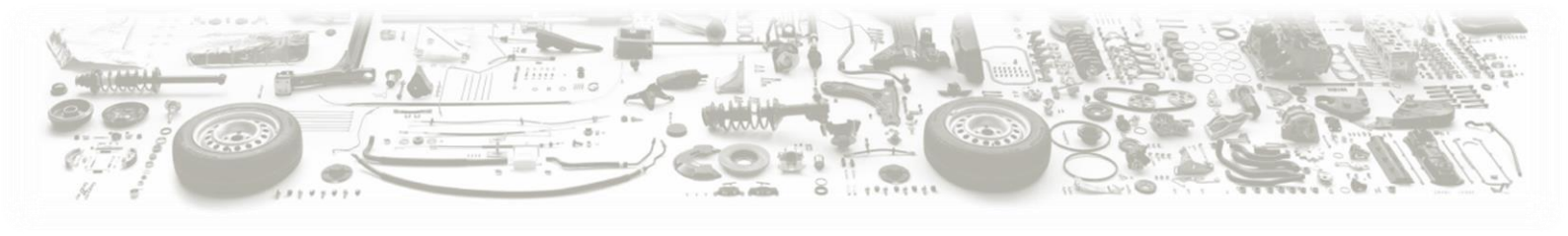
P056

Ultra-accurate genome sequencing of single cells by SIngle-Stranded Sequencing in micrOfluidic Reactors (SISSOR)

Wai Keung Eric Chu, Ho Suk Lee, Xiaohua Huang, Kun Zhang

University of California San Diego, La Jolla, CA, United States

Confident identification of somatic mutations in single mammalian cells is challenging. Current single-cell genome sequencing approaches generate tens of thousands of false positive calls per genome, which is manageable for calling millions of germline variants but greatly outnumbers the somatic mutations per genome and results in an unacceptable level of false discovery rate. We developed a strategy for improving the accuracy by at least 1000 folds by separating megabase-size Watson and Crick DNA strands in a single cell for whole genome amplification and sequencing, and removing errors based on the consensus calls in conjunction with haplotype-based correction.
To this end we designed a microfluidic processor that can separate long single-stranded DNA strands in a single mammalian cell, randomly partition megabase-size single-stranded molecules into multiple nanoliter-size micro-reactors for unbiased amplification, and sequencing library construction. Harvested amplicons are barcoded and the resulting sequencing libraries were sequenced with Illumina SBS chemistry. Several designs of microfluidic chips have been evaluated in terms of cell capture and lysis, completeness of DNA strand dissociation, degree of DNA fragmentation/degradation, partitioning of single-stranded DNA molecules, whole-genome amplification of sub-haploid fractions, and biochemistry of library construction. Low coverage single-cell sequencing has provided clear evidence of successful separation and amplification of single-stranded DNA with N50>500k within the device. With single human fibroblasts, a haplotype contig coverage of ~1.8x has been achieved , which allowed us to phase the Human Leukocyte Antigen (HLA) locus. We envision this sample preparation device, with further integration of sequencing libarary construction functionality, would enable routine clinical sequencing of single cells for diagnostic purposes.



P057

FastProject: A software package for the exploration of single-cell RNA-Seq data

David DeTomaso, Nir Yosef

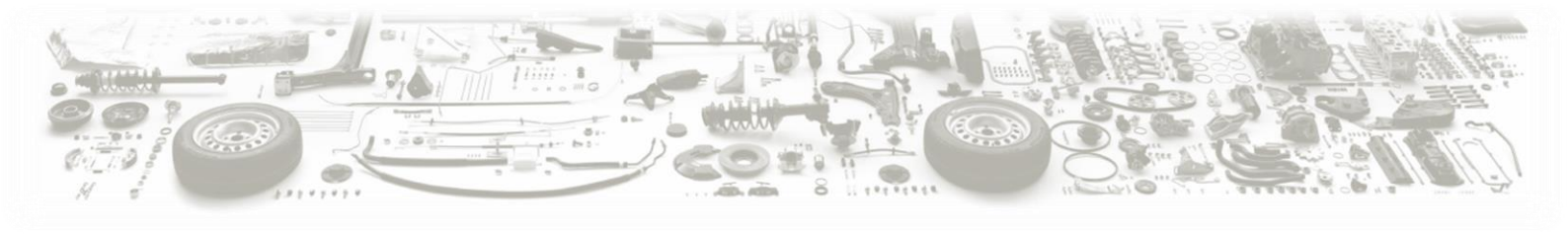
UC Berkeley, Berkeley, CA, United States

A key challenge in the emerging field of single-cell RNA-Seq is to visualize how cells relate to one another and characterize phenotypic diversity in an informative manner. To this end, many dimensionality reduction techniques can be used to generate visualizations of single-cell data. However, applying dimensionality reduction methods naively may be misleading as the variability between cells can arise from technical factors (e.g. differences in cell integrity and detection efficiency), and the projected axes may not necessarily capture the most important biological variation.

To address these issues, we developed the software package FastProject. To visualize single-cell data, FastProject incorporates linear dimensionality reduction techniques (e.g. PCA) as well as several non-linear techniques such as ISOMAP and tSNE.

To aid in interpreting the results of dimensionality reduction, FastProject utilizes a novel functional annotation approach to relate single-cell profiles to legacy genomic signatures or other genomic data. Signatures and profiles from either bulk population data or other single-cell data are used to annotate the cells under examination, select between projections, and highlight potential biological variation.

To lessen the effects of technical variation between samples, FastProject models the probability that an unexpressed transcript was missed due to technical error and reduces the contribution of unreliably expressed transcripts when annotating cells with legacy signatures. Furthermore, strategies to filter out uninformative genes, are included both as a preprocessing convenience and as a method to explore the effect of gene filtering on projections.



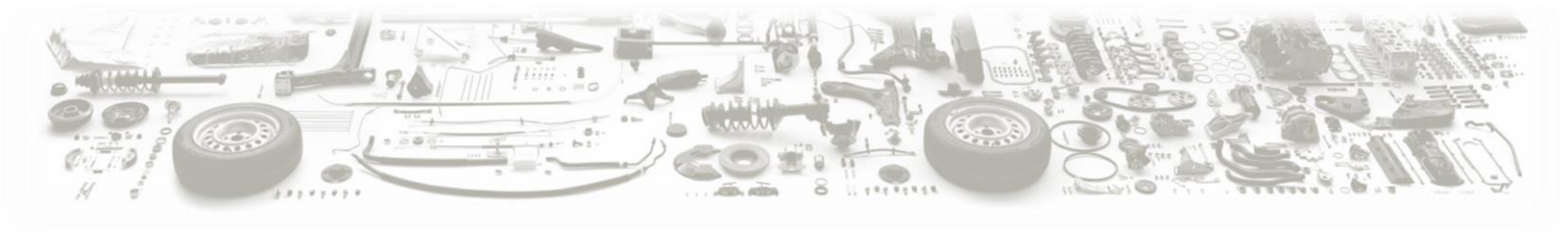
P058

Cell-cycle classification at the single-cell level with Random Forest

Mickael Mendez[1], Efthymios Motakis[2], Michael Böttcher[2], Tsukasa Kouno[2], Elo Madissoon[3], Piero Carninci[2], Timo Lassmann[4], Jay Shin[2], Charles Plessy[2]

[1]*RIKEN Yokohama, Yokohama, Japan*
[2]*RIKEN Center for Life Sciences Technologies, Yokohama, Japan*
[3]*Karolinska Intitutet, Stockholm, Japan*
[4]*Telethon Kids Institute, Perth, Australia*

Over the last two decades there have been considerable developments towards the elucidation of the molecular mechanisms that control cell cycle progression. On the one hand, recent sequencing technologies have enabled the study of the transcriptome at the single-cell level. On the other hand Fluorescent ubiquitination-based cell cycle indicator (Fucci) technologies have provided valuable information on the spatiotemporal dynamics governing multicellular cell cycle progression. Both technologies have opened up new perspectives for the analysis of the cell cycle mechanism.
In this work we follow a multi-platform approach that combines the merits of both technologies. Fucci unravels the cell cycle transition dynamics by the expression of phase specific protein markers. Its results are subsequently combined with the single cell transcriptome data to obtain a machine learning (Random Forests) predictive model of the cell cycle. Our goal is to provide a predictive model able to identify in which specific cell cycle phase a single cell is. The raw data for this study have been collected through a series of experimental and analytical steps with detailed quality control, image processing and in-house developed methodologies.



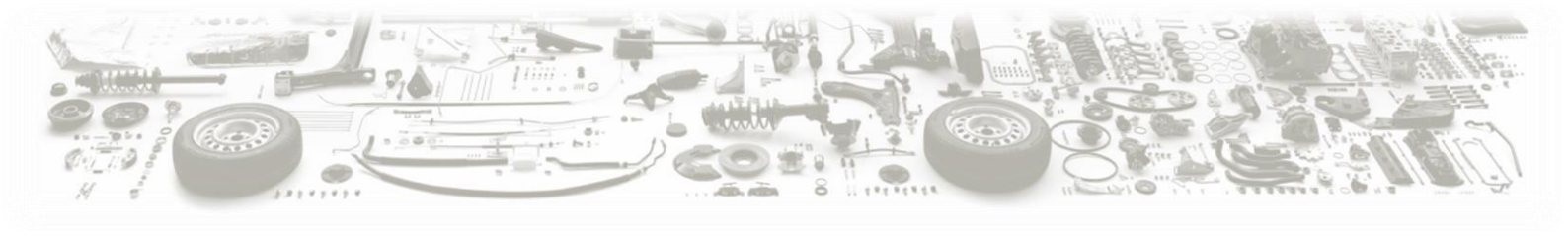
P059

Resolving the Cellular Hierarchy of Thrombocyte Lineage Development Using Single Cell RNA-Seq

Iain Macauley[1], Charlotte Labalette[1], Valentine Svensson[2], Lauren Ferreira[1], Fiona Hamey[3], Sarah Teichmann[4], Ana Cvejic[1]

[1]*Wellcome Trust Sanger Institute, Cambridge, United Kingdom*
[2]*EMBL, Cambridge, United Kingdom*
[3]*Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, Cambridge, United Kingdom*
[4]*EMBL-EBI, Cambridge, United Kingdom*

The prevalent model for haematopoiesis assumes a stepwise set of obligatory steps through which progeny of haematopoietic stem cells pass during lineage development. Our understanding of the transcriptional programs that govern lineage differentiation has mainly been advanced by population-level analysis of these discrete steps. Therefore, the current approaches had limited success in depicting the continual nature of fate decisions that the cell makes during differentiation. Here we used single cell RNA-Seq of cd41:EGFP cells to reconstruct the cellular hierarchy of thrombocyte lineage development in zebrafish. We computationally ordered cells according to their most likely developmental chronology, with each cell representing a distinct time point along a continuum. Our model predicts a process associated with marked decrease in total mRNA content and the number of expressed genes.
We identified genes that are highly correlated with transcriptional states along the process as well as genes that are dynamically regulated during differentiation of thrombocytes. Computational analysis revealed novel patterns of usage of duplicated genes during thrombopoiesis in zebrafish.



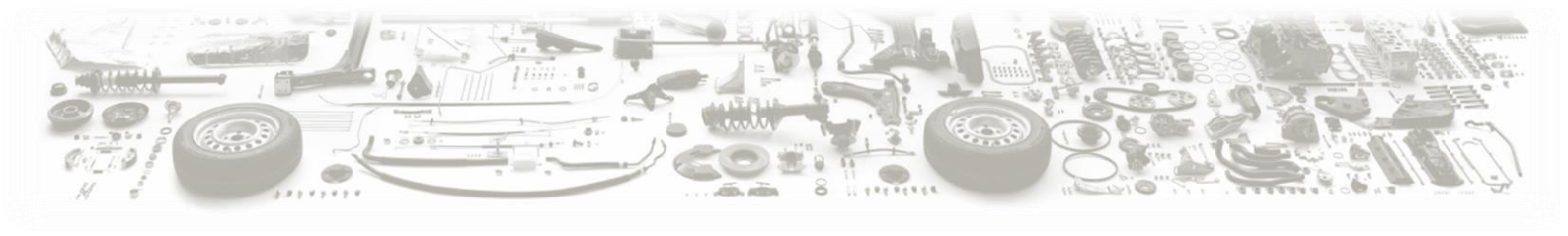
P060

Using Computational Topology to Study Cell Cycle in single-cell RNA-seq Data

Valentine Svensson[1], Sarah Teichmann[2]

[1]*EMBL, Cambridge, United Kingdom*
[2]*EMBL-EBI, Cambridge, United Kingdom*

Some processes in cells are inherently cyclical, the most obvious example being the mitotic cell cycle. Cells sampled evenly from a cyclic process will give rise to certain patterns of gene expression regulating the different phases of a cycle. Neither linear projections such as PCA nor most non-linear methods can faithfully represent cyclic latent variables due to their underlying assumptions. We therefore apply methods from computational topology to study cyclic patterns using persistent cohomology. In the end we arrive at statistical tests for determining the degree of cycling, and a way to sort cells over a cyclic process. We apply the method to investigate cell cycling rates in ESC to NPC differentiation series, and CD4 T-cells from mice during malaria infection, and found that the method can detect changes in cell proliferation rates which we would expect in these experiments.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

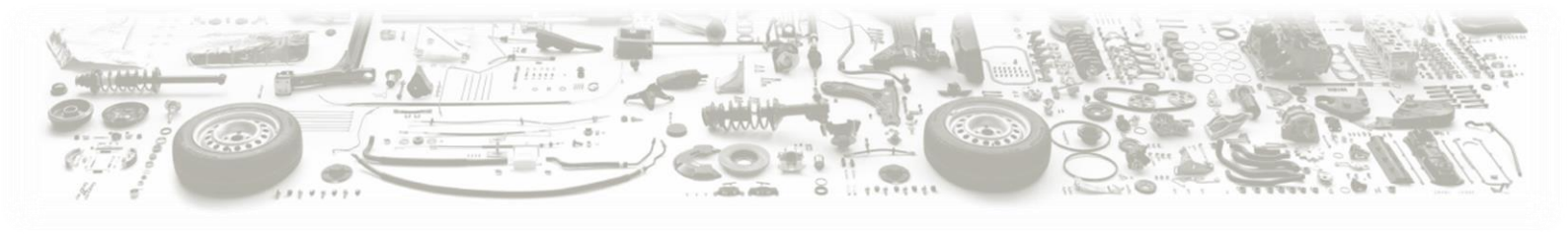
P061

Dissecting the direct reprogramming path from fibroblasts to neurons using single-cell RNA-seq

Barbara Treutlein[1], Qian Yi Lee[2], J. Gray Camp[1], Marius Wernig[2], Stephen R. Quake[2]

[1]*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[2]*Stanford University, Stanford, CA, United States*

Direct lineage reprogramming represents a remarkable conversion of cellular and transcriptome states. However, the intermediates through which individual cells progress are largely undefined. Here we used single cell RNA-seq at multiple time points to reconstruct the direct reprogramming path from mouse embryonic fibroblasts (MEFs) to induced neuronal (iN) cells mediated by the transcription factor Ascl1. Compared to reprogramming towards pluripotency, the process is fairly synchronous suggesting a deterministic mechanism, yet we still observed marked heterogeneity at each time point. When ordered by transcriptome similarity rather than time we find that cells transition through graded states at each step along the reprogramming path. Ascl1 overexpression results in a well-defined initialization causing cells to exit the cell cycle and re-focus gene expression through distinct neural transcription factors. We find that alternative non-neuronal fates can emerge during reprogramming through mechanisms attributed to Ascl1 silencing. These data provide a high-resolution approach for understanding transcriptome states during lineage differentiation.



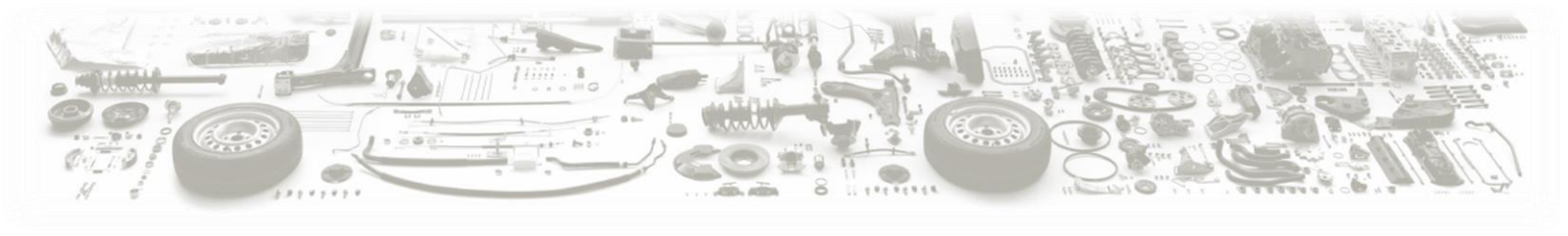
P062

Applying single cell genomics to reconstruct uniquely human cerebral functions

J. Gray Camp

Max Planck Institute of Evolutionary Anthropology, Leipzig, Germany

An understanding of what constitutes the biological basis for modern humans, especially in the realm of cognition, is of fundamental importance for understanding how modern humans came to dominate the biosphere. We have cataloged the genomic changes that define the modern human condition compared to our closest living (the great apes) and extinct (Neanderthals and Denisovans) relatives. We are introducing some of these human-specific genetic changes into mice and use single cell transcriptomics to understand regulatory programs underlying human-specific cerebral functions. FOXP2 is a transcription factor thought to be involved in the acquisition of language in the human lineage that carries two amino acid substitutions that are unique to humans. Mice carrying a humanized version of FOXP2 show increased procedural learning and have altered ultrasonic vocalizations, which have been linked to changes in cortico-basal ganglia circuits in these mice. The gene regulatory mechanisms controlling these behavioral changes are unknown. Using single cell RNAseq, we find that Foxp2 is expressed in a subset of medium spiny neurons in the striatum and pyramidal neurons in layer 6 of the frontal cortex. We compare these Foxp2-expressing cells in wild type and humanized mice to understand how the change of two amino acids alters cell type-specific gene regulatory networks. In parallel experiments, we are establishing an in vitro system to model the effect of ancestralizing the Foxp2 gene in human neurons. Together, these approaches present a strategy to use single cell transcriptomics to understand uniquely human biology.



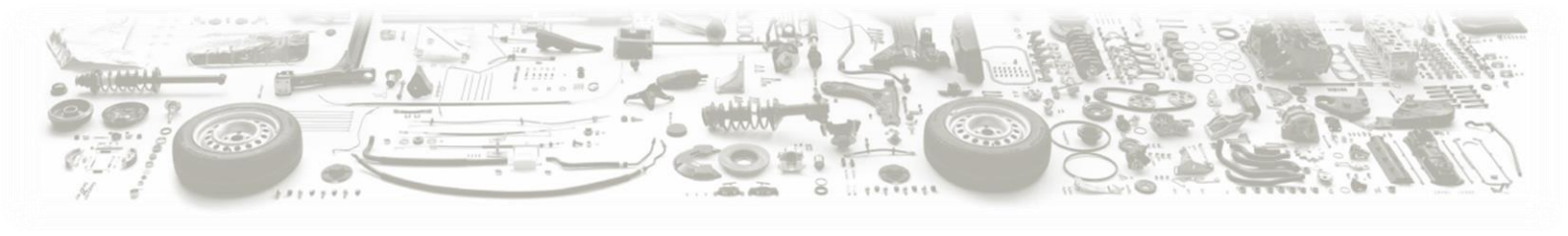
P063

Single cell sequencing to define subpopulations of cancer associated fibroblasts in a mouse model of breast cancer

Michael Bartoschek, Sebastian Braun, Kristian Pietras

Lund University, Lund, Sweden

Fibroblasts play various roles in tumorigenesis, tumor growth and metastasis. Resident tissue fibroblasts have been shown to suppress tumor growth. During tumor progression resident fibroblasts transition to cancer associated fibroblasts (CAF) and change their behavior leading to a tumor supportive signature. Further CAFs were shown to originate from recruited bone marrow mesenchymal stem cells. Previous experiments indicate that different populations of CAFs can be distinguished by the expression of markers such as platelet-derived growth-factor-receptor alpha (PDGF-Rα). The populations are connected to different function and origin. We will use the mouse mammary tumor virus polyoma virus middle T antigen (MMTV-PyMT) mouse model of breast cancer for our studies. The aim is to isolate CAFs from tumors and analyze them by single cell sequencing. With that we aim to identify and further characterize the known CAF populations and eventually define even more populations. Understanding of CAF populations and their functional differences might enable to specifically target tumor supportive CAFs or enable to induce a shift from tumor supportive to tumor suppressive CAF populations.



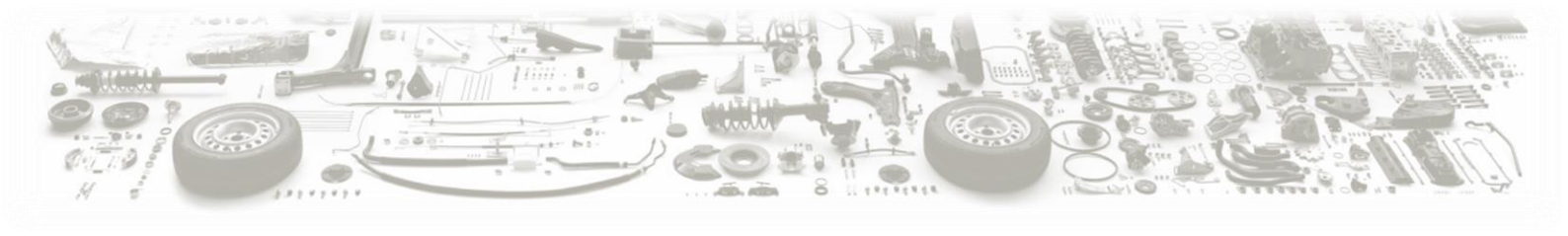
P064

Nfix regulates temporal progression of skeletal muscle acute and chronic regeneration

Giuliana Rossi[1], Stefania Antonini[1], Chiara Bonfanti[1], Mattia Bastoni[1], Stefania Monteverde[1], Shahragim Tajbakhsh[2], Giulio Cossu[3], Graziella Messina[1]

[1]*University of Milan, Milan, Italy*
[2]*Institut Pasteur, Paris, France*
[3]*Institute of Inflammation and Repair, University of Manchester, Manchester, United Kingdom*

Nfix is part of a family of four highly conserved proteins that act as transcriptional activators and/or repressors of cellular and viral genes. We previously demonstrated that Nfix regulates the transcriptional switch from embryonic to fetal myogenesis. Here we show that Nfix is necessary for proper timing of satellite cell differentiation and muscle regeneration, exerting some of his functions during regeneration by directly repressing the Myostatin promoter. nfix null mice display a delayed regeneration after injury, and this deficit is reversed upon Myostatin down-regulation in vivo. Moreover, mice lacking Nfix show a reduced myofiber cross sectional area and a predominant slow twitching phenotype.
Muscular Dystrophies (MDs) are clinically and molecularly heterogeneous diseases, characterized by primary muscle wasting that compromises patient mobility and, in the most severe cases, respiratory and cardiac functions, leading to wheelchair dependency, respiratory failure and premature death. Since fast muscle fibers are predominantly affected in MDs, we crossed our slower-twitching and regenerating nfix null mice with the α-sarcoglycan null mouse model for Limb-Girdle MD 2D. Double nfix:α-sarcoglycan null mice are characterized by an impressive amelioration of the dystrophic pathology, from the histological and functional point of view, leading us to propose Nfix inhibition as a therapeutic tool for the treatment of MDs.



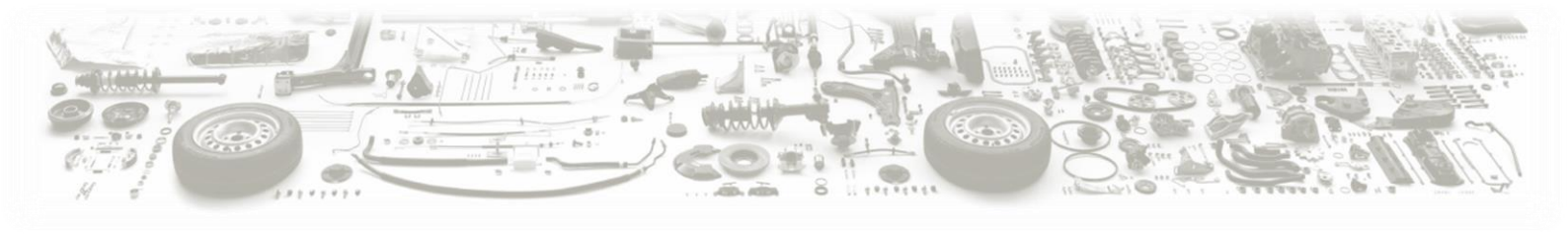
P065

Modelling the impact of growth on cellular heterogeneity and stochastic noise

Lorenzo Ficorella[1], Malika Saint[1], Samuel Marguerat[1], Vahid Shahrezaei[2]

[1]*MRC Clinical Sciences Centre, Imperial College London, London, United Kingdom*
[2]*Imperial College London, London, United Kingdom*

Stochastic cell-to-cell variability in gene expression within isogenic cell populations is commonly called noise; it affects cellular behaviour and contributes to the development of phenotypic heterogeneity in cell populations. Phenotypic heterogeneity and noise can have both positive and negative impact on fitness; therefore, cells have evolved to buffer or exploit them (e.g. “bet hedging” strategies). Recent evidence suggests that the interplay among cell growth, gene expression, and cellular metabolism can generate phenotypic heterogeneity in isogenic cell populations. Here we use mathematical modelling of quantitative phenotypic data in order to study the interplay between growth and phenotypic heterogeneity in the fission yeast Schizosaccharomyces pombe. We have acquired high-throughput quantitative phenotypic measurements of fission yeast cells growing in different conditions by flow microscopy. We use these static images of asynchronous cells to build models of growth rate variability as a function of environmental conditions. Moreover, by studying how noisiness of phenotypic features changes in different experimental conditions and as a function of the cell cycle, we aim at defining common rules of the regulation of phenotypic heterogeneity. Using a Bayesian modelling framework we investigate the connection between quantitative features, and ask how the strength and direction of interactions change with experimental growth conditions. Taken together our analysis will define how phenotypic variability depends on environmental conditions in fission yeast, and uncover potential growth–related “variability check-points”.



P066

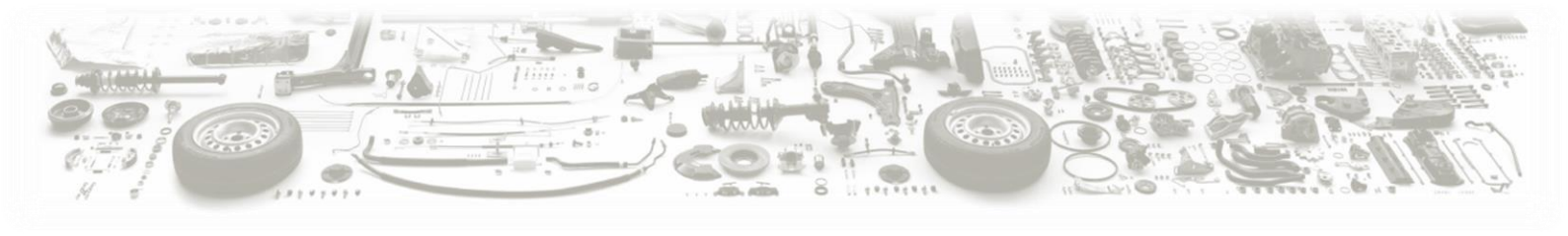
Strategies for comprehensive and sensitive single-cell RNA sequencing

Omid Faridani[1], Rickard Sandberg[1,2]

[1]*Ludwig Institute for Cancer Research, Solna, Stockholm, Sweden*
[2]*Karolinska Institutet, Stockholm, Sweden*

Single-cell transcriptomics open doors to study heterogeneity of gene expression in individual cells within populations, e.g. the cell types present within complex tissues. Furthermore, it enables investigations of expression dynamics of various transcripts across single cells. For these questions, it will be important to improve upon existing methods so that they capture a larger fraction of the RNA molecules present within cells, with lower biases and losses during library construction. There are a number of published methods that provide next-generation sequencing compatible cDNA libraries including Smart-seq2 [1]. Here, we introduced an improved method to comprehensively profile gene expression in individual cells. Preliminary analyses of single HEK293T and embryonic stem cells show improved sensitivity and gene detection, and further validations have been performed on purified and diluted RNA. Altogether, we identify and demonstrate how a modified Smart-seq2 protocol can achieve a more comprehensive characterization of the transcriptional landscape within cells.

Reference:
1. Picelli, S. et al. Nature methods 10, 1096–8



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P067

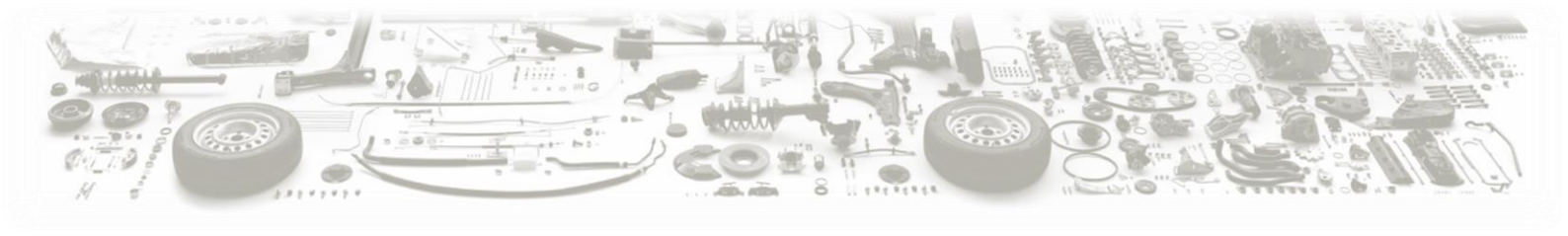
Re-inventing statistical methodology for the challenges of single cell genomics

Justina Zurauskiene[1], Emma Pierson[2], Suchen Jin[2], Richard Savage[3], Christopher Yau[4]

[1]*University of Oxford, Oxford, United Kingdom*
[2]*Department of Statistics, Oxford, United Kingdom*
[3]*Systems Biology Centre, Coventry, United Kingdom*
[4]*Wellcome Trust Centre for Human Genetics and the Department of Statistics, Oxford, United Kingdom*

Single cell genomics presents novel statistical challenges due to data characteristics that are unique to single cell measurements. Although classical statistical analysis tools are routinely applied to single cell data, they are sub-optimal with respect to this type of data, and cannot account for specific features such as dropout events, complex cell state hierarchies, and non-trivial high-dimensional correlation structure. We have developed a suite of statistical tools (https://github.com/cwcyau) that attempts to overcome these challenges with a focus on single cell gene expression analysis. The first is pcaReduce a hierarchical clustering approach that exploits the mathematical connection between k-means clustering and principal components analysis (PCA) to associate cluster identities with principal components of variation. The second, Zero-Inflated Factor Analysis (ZIFA) extends the probabilistic frameworks for PCA and Factor Analysis to incorporate dropout events; and the third, scge – a simulation toolkit for generating realistic single cell gene expression data that can be used to aid statistical methods development.

This work is funded by the UK Engineering and Physical Sciences and Medical Research Councils, the Wellcome Trust and the Oxford-Stanford Big Data for Human Health programme.



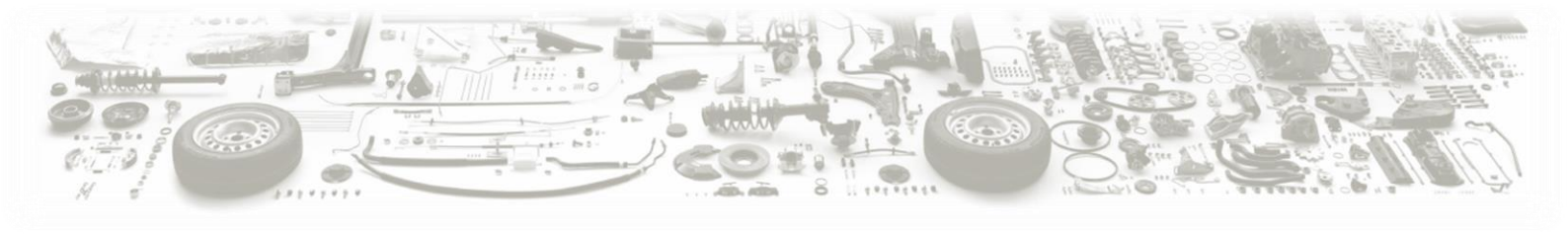
P068

Early commitment and robust differentiation in intestinal crypts

Beata Toth-Cohen, Shani Ben Moshe, Avishai Gavish, Naama Barkai, Shalev Itzkovitz

Weizmann Institute of Science, Rehovot, Israel

Intestinal crypts are a classic model for studying differentiation of stem cells. A small number of stem cells in the bottom of the crypts constantly produce progenies that clonally expand and differentiate into either secretory or absorptive cells. Exact mechanisms that maintain robust proportions of differentiated cell types are unknown. Using clonal fate mapping and single-molecule transcript counting, we demonstrate that individual Lgr5+ stem cells commit to their fate before clonal expansion. This early fate commitment is volatile, as small variations may be exponentially amplified by the proliferation process, leading to large variations in differentiated cell proportions among crypts. Two distinct mechanisms are used by colonic crypts for minimizing this variability – Delta-Notch lateral inhibition, operating in a confined crypt zone, and dispersive goblet cell migration. Our analysis reveals extensive overlap of differentiation and proliferation and defines a new crypt layer sharply above the stem cells, the 'commitment zone', where clonal fate is determined.



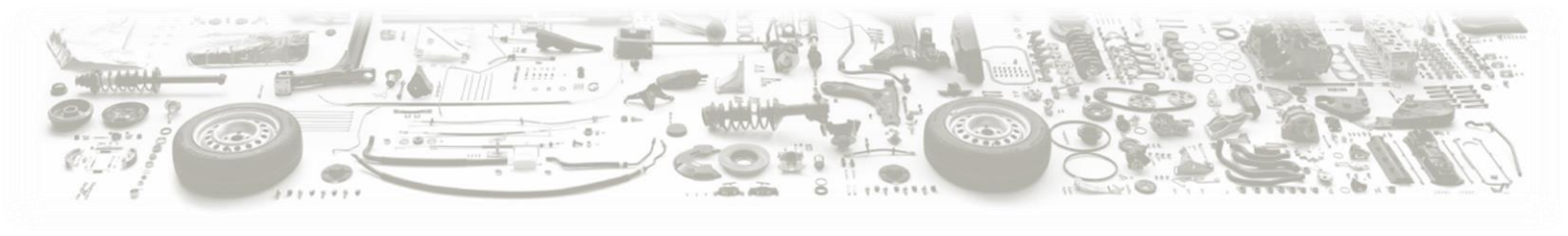
P069

A physical cellular interaction network reveals new niches in the mouse bone marrow

Jean-Charles Boisset, Dominic Grün, Alexander van Oudenaarden

Hubrecht Institute, Utrecht, Netherlands

All blood cells originate from the bone marrow (BM). The organization of the different blood cell types in the BM influences their respective function. This phenomenon has been mainly described in the case of the hematopoietic stem cell interacting with its niche, also composed of diverse cell types. However, the relative organization of the different BM cells is poorly known. Here we used a systematic approach to construct a physical interaction network of all BM cells. This was achieved by manually picking doublets (or other structures composed of a few cells that can be easily distinguished) from mildly dissociated BM. These physically interacting cells were further micro-dissected into single cells, and their individual transcriptome were sequenced, retaining the information of which cells interacted with which. Based on the single cell transcriptomes and clustering algorithms, we were able to identify the different cell types present in the BM, and notably to describe the differentiation path of the neutrophil lineage at the single cell level. We then reconstructed a cellular network by analyzing the frequency of interaction in between the different cell types. We found that megakaryocytes and terminally differentiated neutrophils, B-lymphocytes and neutrophil precursors as well as macrophages and erythroblasts (already described as the erythroblastic island) interacted more frequently than would be expected by chance. These results provide the first description of a cellular interaction network, and lay down the basis for further niches identification.



P070

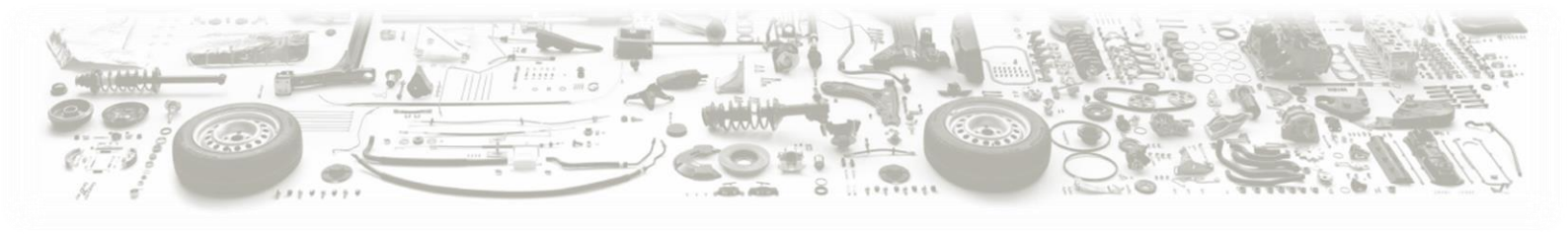
Linking cell lineage and the transcriptome on a single cell level

Tino Frank, Stavroula Skylaki, Martin Etzrod, Timm Schröder, Savas Tay

ETH Zurich, Basel, Switzerland

Cell division leads to the redistribution of a cell's content into two sister cells. This reorganization can create heterogeneity and differentiation of the mother cell's state. It is currently unclear, how symmetrical this division process is and how asymmetry influences cell fate. Also, the lineage development of single cells is hardly understood.
Studying cell division and differentiation processes experimentally on a single cell level requires sophisticated integrated multistep technology that includes lineage tracing and genomic profiling. Cells need to be cultured, tracked and profiled singly. While there are routines for each single task, the combined and integrated process is not reliably realized in one shot due to technical challenges. Maintaining single cell identity is missed by statistical approaches, such as well-mixed cell suspensions for flow-cytometry or barcoding.
We present a microfluidic platform that overcomes this particular caveat and is able to culture, track, retrieve and profile single cells after several divisions. Our system accomplishes this without loosing identity, by combining integrated microfluidic circuits and video microscopy. We culture cells and track them on single cell level. After a few generations we singly sort the cells on chip and use single cell qPCR without loosing the linking information. This powerful tool allows us to link cell dynamic behavior, lineage information and genomic profiling on true single cell resolution.

This work is funded by SNF NCCR Molecular Systems Engineering.



P071

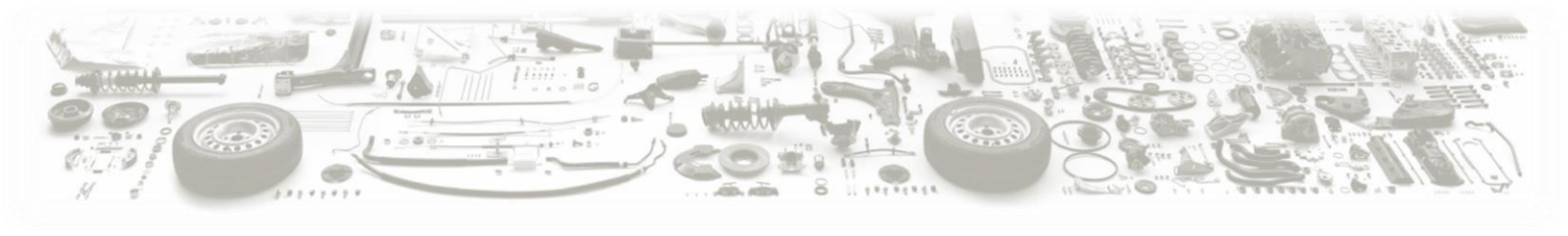
Linking high resolution secretion profiles to dynamic inputs in single cells

Alicia Kaestli, Michael Junkin, Savas Tay

ETH Zurich, Basel, Switzerland

The dynamics of cytokines received and released from cells is important in understanding how cells communicate with each other, and more specifically, how the immune response is synchronized [1]. Pulsing TNF over a population of cells transforms NF-κB oscillations from non-entrained to entrained oscillations with more efficient gene transcription as a result [1]. Nevertheless, the dynamic profile of cytokines released from both single and small populations of cells - given a dynamic input - is not well understood. To investigate these questions, we developed an automated microfluidic chip that is able to reconstruct secretion profiles from single and small populations of cells. The device can dynamically stimulate cells with multiple inputs, and measure the secretion of each cell chamber over ten time points with 60 minute resolution. We can culture 64 single cells or cell populations in subnanoliter-sized chambers for over three days. The chip has individual control over cell seeding and stimulation of each cell chamber. As stimulation frequency can be used to synchronize cell populations [1], the chip can apply different input frequencies to groups of 16 cell chambers. The chip has the ability to work with rare cell types – like primary T-cells in cerebral spinal fluid - by seeding single cells at low concentrations. With the rise of personalized medicine, the analysis of rare cell types is becoming increasingly important for discovering, diagnosing, and treating novel disease subsets.

This work is supported by an ERC Starting Grant (SingleCellDynamics).
[1] Kellogg. Cell 160, 3 (2015)



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

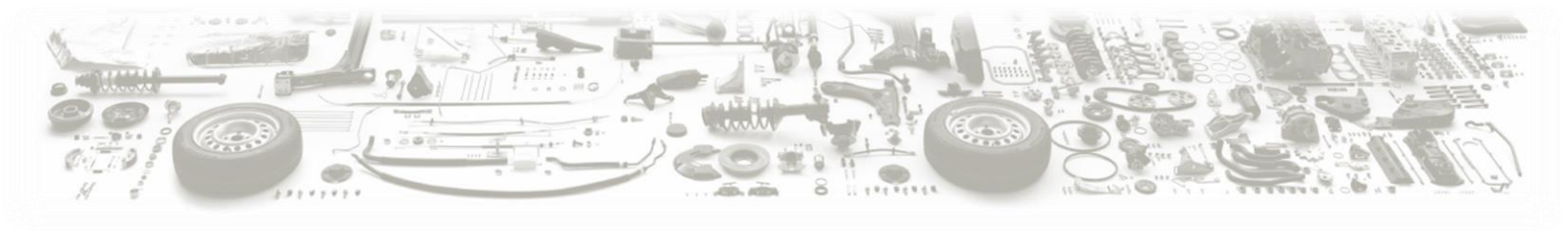
P072

Characterisation of isoform-expression patterns in single-cell RNA-sequencing data

Trung Nghia Vu[1], Quin Wills[2], Krishna Kalari[3], Liewei Wang[3], Yudi Pawitan[1], Mattias Rantalainen[1]

[1]*Karolinska Institutet, Stockholm, Sweden*
[2]*University of Oxford, Oxford, United Kingdom*
[3]*Mayo Clinic, Rochester, NY, United States*

Transcriptional heterogeneity is common in seemingly homogenous populations of cells. Single-cell gene-level expression variability has been characterized by RNA-sequencing in multitudes of biological context to date, but few studies have focused on heterogeneity at isoform-level expression.
Here we investigated the commitment of individual cells to expressing specific isoforms by studying the co-variability of expression levels of pairs of isoforms from the same genes in single-cell RNA-sequencing data (Fluidigm C1) from a triple-negative breast cancer cell line (N=384 cells). We propose a novel method “ISOform-Patterns (ISOP)”, based on mixture modeling, to characterize dependency between isoforms from the same genes. Based on this method we defined a set of principal patterns of isoform expression relationships and demonstrated that these patterns were present in multiple single-cell data sets.
We also assessed to what extent expression patterns of isoform pairs were associated with intrinsic factors, such as different transcriptional start sites, the number of annotated isoforms of the gene and average gene expression level. We found that the frequencies of isoform patterns were different between isoforms with the same and different transcriptional starting site, while isoform patterns were not directly associated with average gene expression level. We also investigated the effect on the distribution of isoform expression patterns under a small-molecule perturbation. Applying ISOP for analysis of isoform expression patterns allowed us to discover isoform patterns associated with the perturbation, including novel findings not discovered through conventional differential expression analysis.



P073

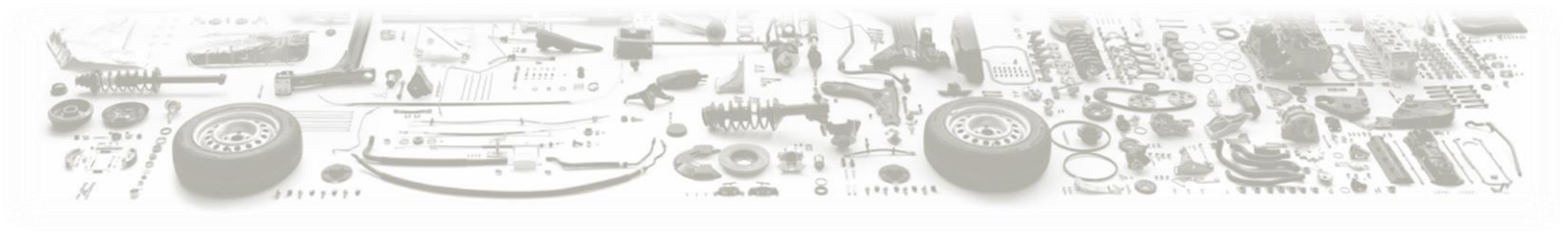
Single-cell RNA-seq reveals maturation stages of the Paneth cell lineage

Kay Wiebrands, Dominic Grün, Anna Lyubimova, Nobuo Sasaki, Hans Clevers, Alexander van Oudenaarden

Hubrecht Institute, Utrecht, Netherlands

Paneth cells (PCs) are long-lived secretory cells that reside at the bottoms of small intestinal crypts. Besides serving as niche cells for the neighboring Lgr5-positive stem cells, PCs secrete granules containing a broad spectrum of antimicrobial proteins, including lysozymes and defensins. Here, we have used single-cell RNA sequencing (1) to explore PC differentiation. We found a maturation gradient from early secretory progenitors to mature PCs, capturing the full maturation path of PCs. Moreover, differential expression of a subset of defensin genes, e.g. Defa20, in lysozyme-high PCs, reveals at least two distinct stages of maturation. Lineage tracing of Lgr5-dsecendant cells showed that shorter-traced cells are enriched in the Defa20 low state, while longer traced cells primarily fall into the subclass with high Defa20 levels. PCs derived from small intestinal organoids show a similar pattern, though less pronounced. These data show that single cell sequencing can be used to trace the differentiation of a specific cell lineage with high resolution and reveal the succession of transcriptome changes during maturation of PCs.

1. Hashimshony, T. Cell Rep. 2, 666–73 (2012).



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

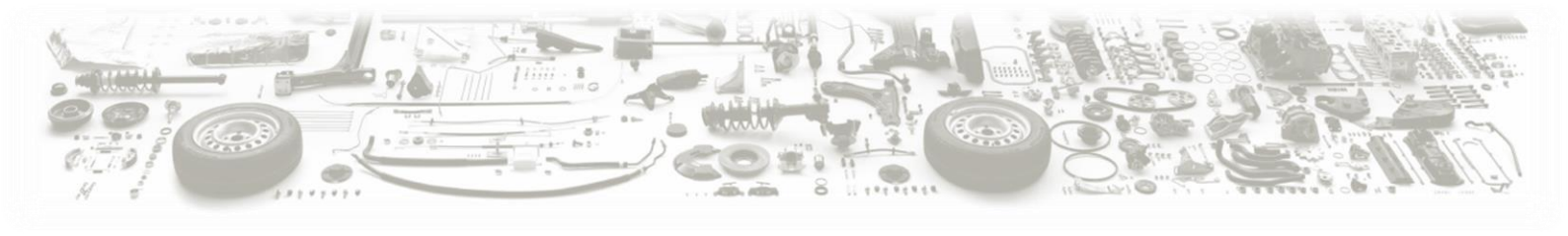
P074

Single-cell printing for the genetic analysis of cancer cells

Julian Riba[1], Jean Christoph Niemöller[1], Stefan Zimmermann[1], Jonas Schoendube[2], Sabine Bleul[1], Peter Koltay[1], Roland Zengerle[1], Rainer Claus[1], Heiko Becker[1], Andre Gross[2]

[1]*University of Freiburg, Freiburg, Germany*
[2]*cytena GmbH, Freiburg, Germany*

Cancer initiation and evolution is based on the sequential acquisition of mutations in single cells leading to genetically heterogeneous cell populations. We used the Single-cell Printer (SCP) for the isolation of single cancer cells followed by genetic analyses. Similar to an inkjet printer, the SCP uses a piezo-driven dispenser chip to generate free-flying microdroplets. A camera system coupled with computer-assisted image processing enables to detect cells in the chip nozzle and to print droplets containing exactly one cell. The SCP allows for single cell deposition efficiencies > 80% and viability rates > 90% for mammalian cells (Gross et al. 2013 JALA 18:504–518). Here, we used the SCP for printing single cells of the human osteosarcoma cell line U2OS in a 384-well microtiter plate. 20 U2OS cells were printed in 20 wells, each preloaded with only 1 µl PBS in addition to 20 cells printed in 20 dry wells resulting in a total single-cell deposition efficiency of 98% (39/40) according to the SCP images. Whole genome amplification (WGA) was performed on the 20 cells printed in PBS, and on five cells printed in dry wells using the REPLI-g™ Single Cell Kit (Qiagen) with reduced reagent volumes. The DNA yield as measured by Qubit™ quantitation was 3.9±0.2 µg for the cells in PBS, and 3.8±0.1 µg for the cells in dry wells, respectively. A subsequent PCR on LINE1 retrotransposons revealed positive results in all WGA samples, and the c.1538G>T-mutation in the SLC34A2 gene, pre-described in U2OS cells, was detected in all of eight cells analyzed. In conclusion, we established a highly-efficient workflow allowing for the high-throughput analysis of genetic aberrations in single cancer cells.



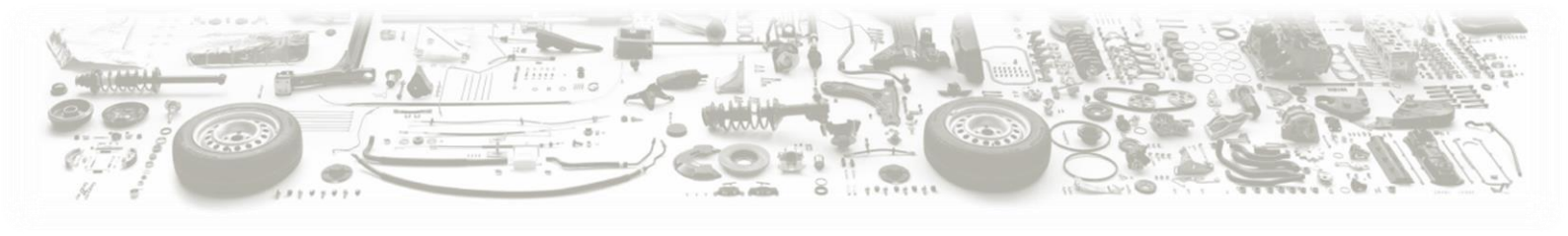
P075

Single-cell medicine: Transcriptional analysis of normal and diabetic human pancreatic islets

Åsa Segerstolpe[1,2], Athanasia Palasantza[1,2], Pernilla Eliasson[3], Magnus Bjursell[3], Carina Ämmälä[3], Dave Smith[3], Rickard Sandberg[2]

[1]*Ludwig Institute for Cancer researc, Uppsala, Sweden*
[2]*Karolinska Institute, Stockholm, Sweden*
[3]*AstraZeneca, Mölndal, Sweden*

The pancreas is vital for controlling metabolic processes through the secretion of exocrine digestive enzymes and endocrine hormones, like insulin and glucagon. The majority of the pancreas is composed of exocrine ductal and acinar regions and intermingled are the endocrine islets of Langerhans. Five endocrine cell-types have been described as islet components; the alpha (secreting glucagon), beta (insulin), delta (somatostatin), PP (pancreatic polypeptide) and epsilon (ghrelin) cells. Loss, or reduced function, of the insulin producing beta cells leads to metabolic disorders, most prominently diabetes.
We have analysed the cellular composition of human pancreatic islets from five donors, including healthy and diabetic. Transcriptional profiles for 1400 cells were obtained with the Smart-seq2 method and the cellular heterogeneity decomposed. Pure cell-type transcriptional profiles were acquired that recapitulated known features of endocrine cell identities. We also analysed the cell-type resolved transcriptional aberrations in pancreatic islets of type 2 diabetic individuals. The single-cell technology is very promising for analyses of pathological tissues. It allows the distinction of cell-type compositional effects from changes in cellular identities, yielding a high-resolution understanding of pathological transcriptional changes.



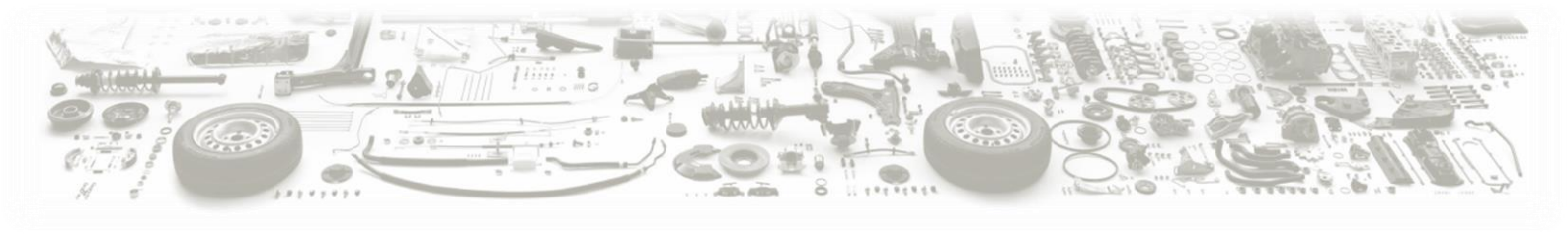
P076

ESpresso: an single cell gene expression database of mouse embryonic stem cells

Aleksandra Kolodziejczyk[1], Jong Kyoung Kim[1], Tomislav Ilicic[1], Jason Tsang[1], Johan Henriksson[2], Kedar Natarajan[1], Alex Tuck[1], Marc Bühler[3], Pentao Liu[1], John Marioni[1], Sarah Teichmann[1]

[1]*EMBL-EBI, Hinxton, United Kingdom*
[2]*Karolinska Institutet/EBI, Huddinge, Sweden*
[3]*FMI, Basel, Switzerland*

Cell culture conditions for embryonic stem cells are important in order to maintain their self-renewal potential. Our lab has dissected the transcriptomes of individual mESCs cultured under three conditions (serum, 2i and alternative 2i) using full-transcript single cell RNA-sequencing. To allow others to further investigate this data, we have made it publicly available through our database ESpresso together with other previously published single-cell data. This dynamic interface allows one to look up things such as the distribution of gene expression over cells and conditions. Further, expression levels can be compared between different (sub)populations and clustering can be performed online. The noisiness (stochasticity) of gene expressions and groups of genes can also be assessed. ESpresso is open source and can thus be applied to any other single-cell RNA-seq/rt-qPCR dataset. We believe this resource will make single-cell data more accessible to the community as any gene can be looked up in a matter of seconds without expert knowledge of RNA-seq.



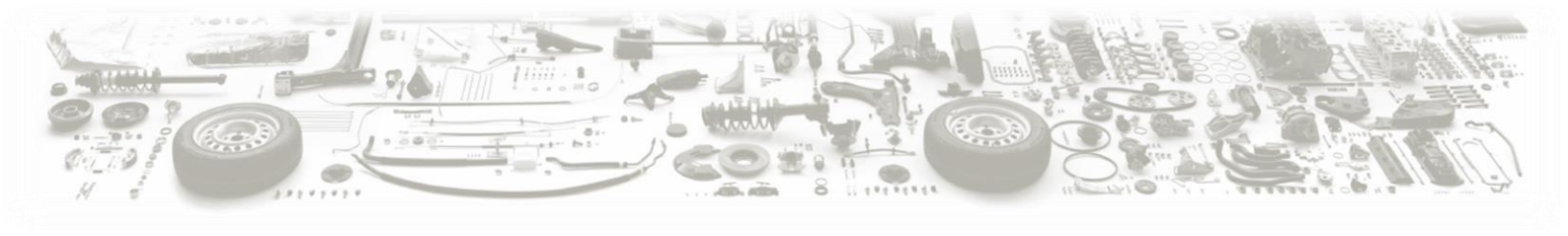
P077

Single cell expression profiling of rare subpopulations in acute leukemia

Christoph Ziegenhain[1], Binje Vick[2], Swati Parekh[1], Beate Vieth[1], Sarah Ebinger[2], Erbey Özdemir[2], Irmela Jeremias[2], Wolfgang Enard[1]

[1]*Ludwig-Maximilians University Munich, Martinsried, Germany*
[2]*Helmholtz Zentrum München, Munich, Germany*

Leukemia consists of heterogeneous cells with distinct mutational or functional properties. In acute myeloid leukemia (AML), cancer cells undergo a clonal evolutionary process and accumulate mutations. In order to characterize heterogeneity on the level of gene expression within cases of AML, we performed single-cell RNA sequencing combined with a genotyping strategy. We obtained cells from patient-derived xenograft mouse models, where subclonal mutations were previously found by exome-sequencing. Differential expression analysis for the first time shows clearly distinct cellular states between subclonal populations.
We next searched for a subpopulation of treatment resistant, relapse inducing cells, which determine patients' outcome in acute lymphoblastic leukemia (ALL). Patient samples were again obtained from a xenograft mouse model initiated with primary leukemia cells, but cells were now separated into rapidly growing and a rare, dormant subpopulation by labeling slowly dividing cells. Single cell RNA sequencing revealed a distinct expression profile in dormant cells that could be isolated at minimal numbers. In accordance with the dormant phenotype, we found a downregulation of cell cycle genes in this subpopulation, whereas cell adhesion was increased. By comparing our data with published expression profiles, we could show that dormant AML cells resemble high-risk subpopulations in patients' leukemia specimens and are more homogeneous than their rapidly proliferating counterparts [1].
Our data suggest that the dormant ALL cells isolated from the mouse model could represent a treatment-resistant cancer stem cell subgroup which might be important in the clinic for the prognosis of patients with acute leukemia.



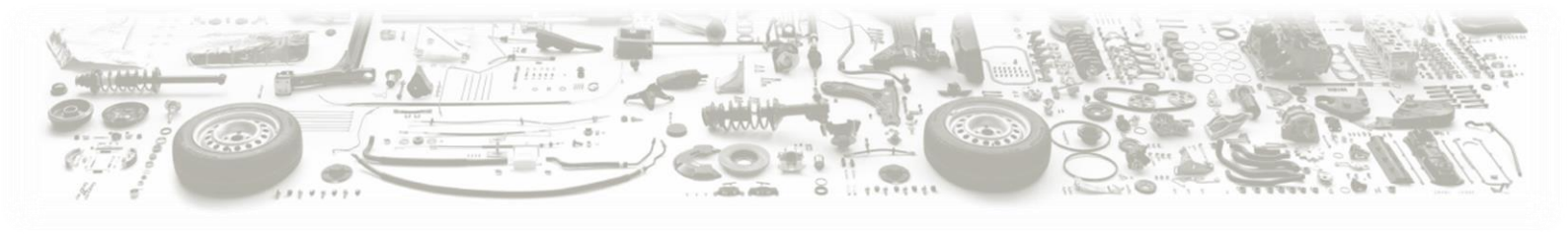
P078

Aneuploidy and CNV detection using single-cell-sequencing

Aaron Taudt[1], Bjorn Bakker[2], David Porubsky[2], Hilda van den Bos[2], Diana Spierings[2], Peter Lansdorp[2], Floris Foijer[2], Maria Colomé-Tatché[2]

[1]*Eriba, Groningen, Netherlands*
[2]*UMC Groningen, Groningen, Netherlands*

Whole chromosome duplications and deletions (aneuploidies) and copy number variations (CNVs) are a hallmark of cancer cells. In general, multiple aneuploidies and CNVs coexist in the same tumor, and this level of heterogeneity may influence the metastatic potential of cancers and determine the success of specific treatments. Therefore, quantifying the level of karyotype heterogeneity in cancers is crucial for understanding the relationship between chromosomal instability and disease, and will help in the design of personalized treatments.
Single-cell-Next Generation Sequencing (NGS) techniques offer a powerful way to study CNVs in a high-throughput fashion at a single cell level with high resolution. Here we present a method for the accurate and reliable mapping of CNVs from single-cell-NGS data. Our tool explicitly models the data structure of single-cell-NGS experiments by using a Hidden Markov Model where every distinct copy-number is assigned to a hidden state, and where read count distributions are modeled as negative binomial distributions. The resolution of this approach is only limited by sequencing coverage.
We validate our approach by analyzing samples with known karyotype, and use it to study karyotype heterogeneity in a mouse model for chromosomal instability in the epidermis.
This tool is written in C++ and compiled as an R package, offering a combination between a user friendly interface and fast runtimes, allowing for downstream bioinformatics analysis.



P079

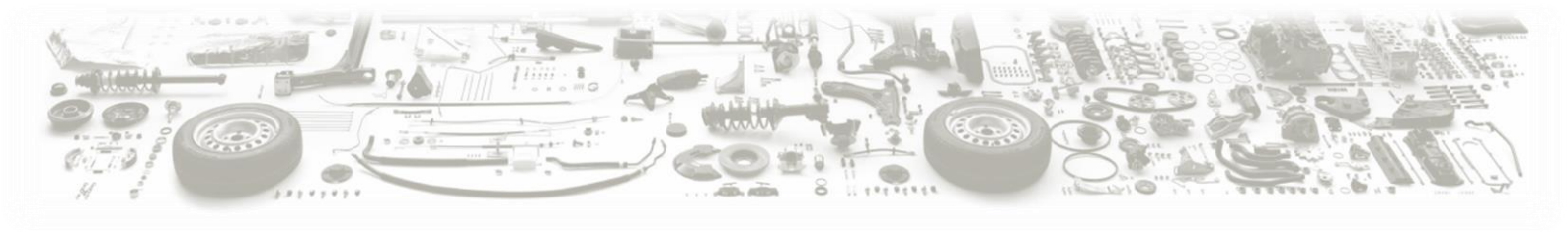
Single-cell mRNA sequencing of type 2 diabetes human pancreas reveals cell type-specific and disease related genes

Mauro Javier Muraro, Gitanjali Dharmadhikari, Dominic Grün, Eelco de Koning, Alexander van Oudenaarden

Hubrecht Institute, Utrecht, Netherlands

The human pancreas is composed of a number of different cell types that secrete digestive enzymes, regulate glucose homeostasis and are linked to type 2 diabetes (T2D). In order to fully characterize these cells and their contribution to T2D, it is important to have cell type-specific gene expression profiles. However, no comprehensive transcriptome data exists for human pancreatic cells, mostly due to the lack of reliable cell-surface markers. To overcome this challenge, we sequenced mRNA from single pancreatic cells of healthy and T2D human donors using CEL-seq, a method for amplifying and barcoding mRNA from single cells (1). We obtained well-defined clusters of cells that express marker genes for all major pancreatic cell types. This allowed us to explore differences in gene expression between closely related but distinct cell types such as alpha and beta cells. Furthermore, we found heterogeneity within known cell types in the form of sub-populations of acinar and beta cells. Lastly, we identified differentially expressed genes between healthy and T2D pancreatic cell types. Therefore, this dataset can be used as a cell type-specific resource to study the human pancreas at single cell resolution.

1. Hashimshony, T. Cell Rep. 2, 666–73 (2012).



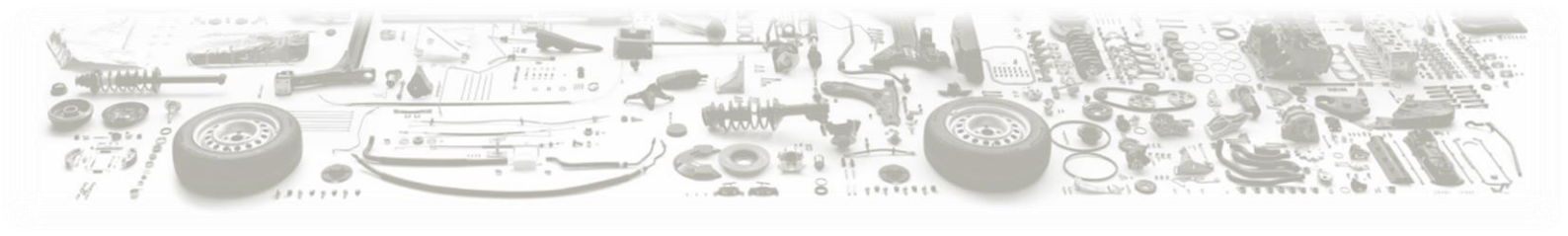
P080

Long non-coding RNA signatures in tumor infiltrating lymphocytes

Marco De Simone[1], Grazisa Rossetti[1], Alberto Arrigoni[1], Paola Gruarin[1], Elena Provasi[1], Claudia Politano[1], Valeria Ranzani[1], Raoul Bonnal[1], Valentina Vaira[2], Alessandro Palleschi[2], Silvano Bosari[2], Sergio Abrignani[3], Massimiliano Pagani[3]

[1]*INGM - Istituto Nazionale Genetica Molecolare-Romeo ed Enrica Invernizzi, Milan, Italy*
[2]*Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
[3]*INGM - Istituto Nazionale Genetica Molecolare, Università degli Studi di Milano, Milan, Italy*

An in-depth molecular characterization of tumor infiltrating lymphocytes (TILs) may lead to the isolation of specific diagnostic markers and to the comprehension of TILs role in tumor immune escape. Moreover new therapeutic targets for the effective modulation of these cells in cancer might be identified through this kind of analysis.
Long non-coding RNAs (lncRNAs) are a novel class of regulatory RNAs with a high cellular specificity of expression involved in cell fate determination and maintenance of cell identity in diverse biological contexts including the human immune system.
Since very little is known on both the coding genes and lncRNAs expression profile in TILs, we isolated CD4+ Th1, Th17 and Tregs cells infiltrating both tumor and healthy tissue as well as lymphocytes from lymphoid tissues and peripheral blood of Non-Small-Cell-Lung cancer patients. We analysed these cells by RNA-seq, performed de-novo transcriptome reconstruction and defined a set of coding genes and lncRNAs that are specifically expressed in TIL subsets. In order to check for the expression variability of these signatures among the TILs we are now validating them at single cell level using the FluidigmC1 platform



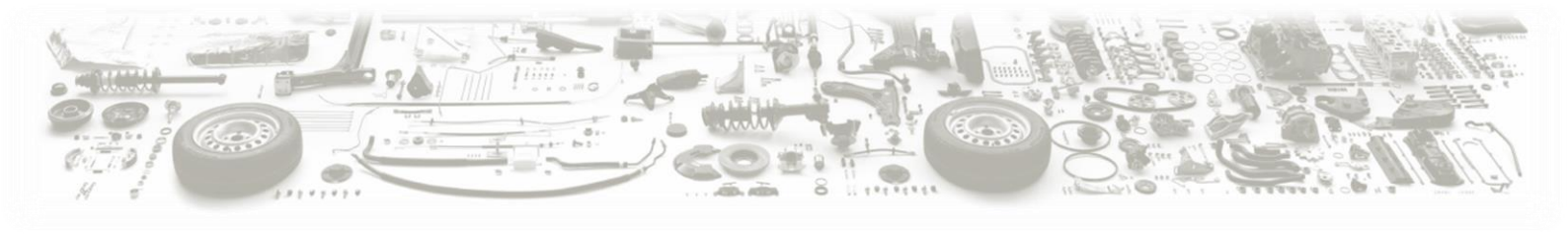
P082

Single cell mRNA-sequencing of mESCs reveals cell-to-cell variation in pluripotency and cell cycle genes

Aleksandra A Kolodziejczyk[1,2], Jong Kyoung Kim[2], Tomislav Ilicic[1], Cheuk-Ho Tsang[1], Johan Henriksson[1], Kedar N Natarajan[1], Alex Tuck[3], Marc Bühler[3], Pentao Liu[1], John C Marioni[2], Sarah A Teichmann[2]

[1]*Wellcome Trust Sanger Institute, Cambridge, United Kingdom*
[2]*EMBL European Bioinformatics Institute, Cambridge, United Kingdom*
[3]*Friedrich Miescher Institute for Biomedical Research, Basel, Swaziland*

Cell culture conditions for embryonic stem cells are important for maintaining their self-renewal potential. We dissected the transcriptomes of individual mESCs cultured in three conditions (serum, 2i and alternative 2i (Shimizu et al., 2012)) using full-transcript single cell RNA-sequencing. Firstly, we show that the transcriptomic signature of cells grown in these media is distinct, with cells grown in 2i media being most similar to the blastocyst cells of the early embryo. Secondly, we found that global levels of intercellular heterogeneity in gene expression are indistinguishable between conditions. At the same time, specific groups of genes (pluripotency genes in serum, cell cycle genes in 2i) do differ in their noise levels across culture conditions. Differences in cell cycle genes' noise profiles correlate with proliferation rate, where slowly-cycling cells have broader more noisy expression profiles and clearer separation between cells in G1/S and G2/M phases. Moreover, we identified 2C-like population (Macfarlan et al., 2012) in 2i cultured cells, we characterized them and compared their transcriptomes to in vivo data from early stages of mouse embryo development (Deng et al., 2014). We observed that these cells globally have more similar transcriptomes to blastocyst cells than cells from 2-cell stage of the embryo.



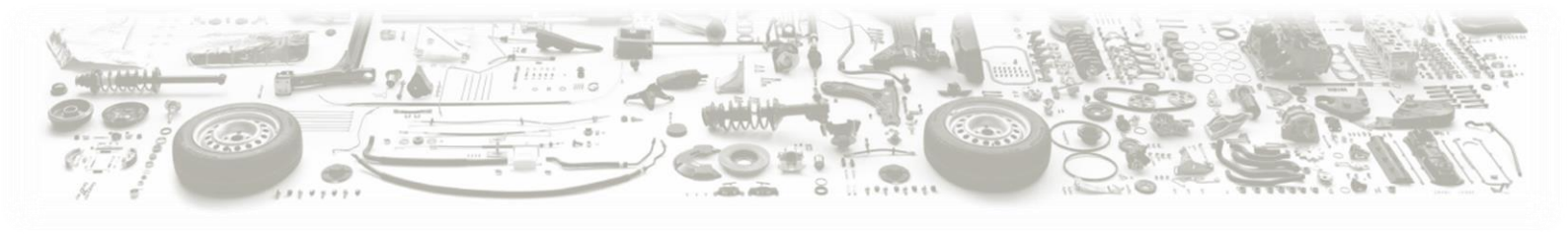
P083

Single-cell 5-hydroxymethylcytosine sequencing reveals extensive strand bias among individual cells

Dylan Mooijman[1], Siddharth Dey[1], Jean-Charles Boisset[1], Nicola Crosetto[2], Alexander van Oudenaarden[1]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Karolinska Institutet, Stockholm, Sweden*

Recently a new DNA modification 5-hydroxymethylcytosine (5hmC) was shown to be present in various cell types and to have critical roles in development and gene regulation. Understanding the dynamics of this epigenetic mark is crucial to understanding the specific role of 5hmC. Here, we have developed a single-cell, strand-specific 5hmC sequencing technology that allows us to quantify hydroxymethylation variability in single mouse embryonic stem cells and the developing mouse embryo.
Using strand-specific information, we quantify the amount of 5hmC on the sense versus the anti-sense strand and find extensive strand bias between cells. Using mouse embryos, we show that the bias is induced by cell division. Inspired by these results, we developed a quantitative stochastic model to find that low rates of hydroxymethylation can cause the bias we detect.
These results indicate that single-cell sequencing can reveal enzymatic rates and provide novel insights on the possible role of 5hmC in embryonic stem cells.



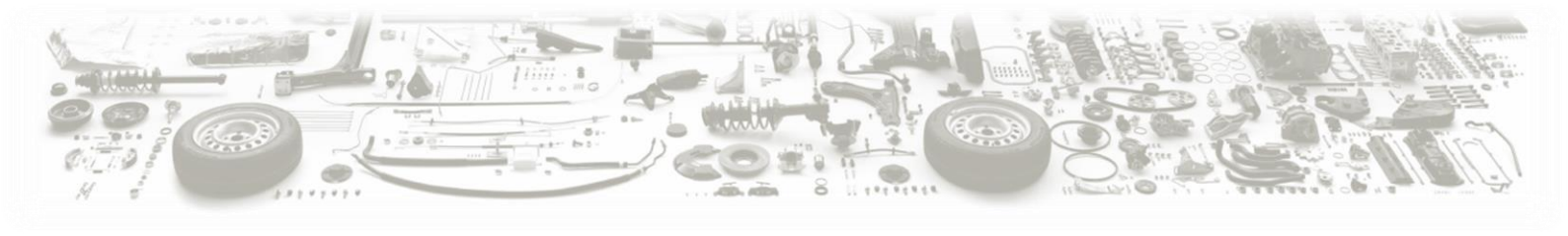
P084

Understanding human specific cognitive traits by dissecting the neuronal diversity of humans and great apes

Sabina Kanton, Malgorzata Gac, Tobias Gerber, Gray Camp, Svante Pääbo, Barbara Treutlein

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Humans exhibit remarkable cognitive skills compared to their closest living relatives, the great apes which has its cellular basis in the neuronal diversity in the human brain. However, the heterogeneity of neuronal cell types in the human brain is still far from being understood, especially in the light of potential cell subtypes that might be specific to humans compared to great apes. Here, we seek to understand the neuronal diversity in humans and great apes by using single nuclei mRNA-Sequencing on cortical tissue of humans, chimpanzees, gorillas, orang utans and macaques. mRNA-Sequencing of single nuclei isolated from frozen brain tissue by means of fluorescent activated cell sorting offers a detailed transcriptional readout which has been shown to correlate well with transcriptomes of whole cells. First data reveal the applicability of isolating single nuclei from human and great ape frozen cortical brain tissue, thereby offering great potential to characterize the diverse transcriptional neuronal landscape in these different species.
Using these data, we aim at dissecting the neuronal diversity in humans and great apes. Comparing the neuronal diversity offers the potential to identify orthogonal as well as novel human specific transcriptional signatures which eventually represent the cellular basis of human specific cognitive abilities.



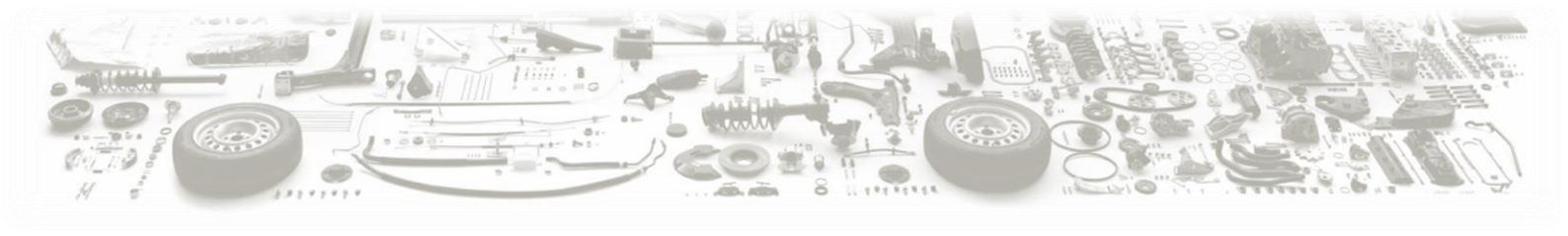
P085

TruePrime™, a superior technology for single cell whole genome amplification based on TthPrimPol

Angel Picher[1], Oliver Wafzig[2], Luis Blanco[3], Armin Schneider[2]

[1]*SYGNIS Biotech SLU, Madrid, Spain*
[2]*SYGNIS Bioscience GmbH & Co KG, Heidelberg, Germany*
[3]*CBM, Madrid, Spain*

TruePrime™ is the name of a novel technology dedicated to the amplification of genomic DNA. While the current gold standard MDA (multiple displacement amplification) relies on short oligonucleotides to start off the amplification, TruePrime™ is based on a combination of Phi29 DNA polymerase with the recently discovered primase/polymerase TthPrimPol. In this setup, TthPrimPol synthesizes the DNA primers needed for Phi29 DNA pol, which allows for the exponential amplification of genomic DNA. TthPrimPol is a monomeric enzyme (34 kDa) that displays a potent primase activity, preferring dNTPs as substrates unlike conventional primases. This DNA primase activity can be activated by magnesium or manganese ions, having wide sequence specificity for template recognition. Key advantages of the TruePrime™ technology for amplification of single cell genomes include complete absence of primer artefacts, superior sensitivity down to the femtogram range, high reproducibility, and relative insensitivity to external contaminations. Moreover, the TruePrime™ workflow is easy and reaction products work well with Illumina or IonTorrent platforms. Analyses on genomic DNA amplified from single Hek293 cells in comparison to non-amplified DNA reveal superior genome coverage with little bias and minimal introduction of error such as allelic dropout or chimera formation.



P086

Multiscale pattern detection in spatially resolved gene expression data

Joseph Herman, Peter Kharchenko

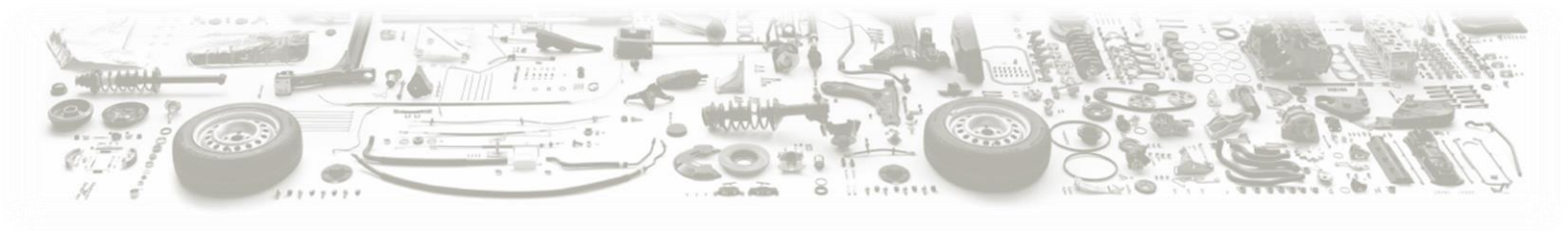
Harvard Medical School, Boston, MA, United States

The advent of multiplexed in situ transcriptomics opens up the exciting possibility of a highly detailed characterisation of tissue architecture based on gene expression. However, the sparse, high-dimensional nature of this data, combined with various technical sources of noise, necessitates the development of advanced statistical methods in order to reliably identify patterns of spatial variability.

Previous approaches have focused on comparing expression levels between individual cells or predefined regions of tissue. However, variations in expression may occur simultaneously at multiple different scales, from the subcellular level, to gradients between groups of cells; hence, it may not be clear at which scale the analysis should be carried out. Analysis of smaller regions provides higher resolution to detect localised features, but lower power and higher susceptibility to noise; conversely, comparison between larger regions results in higher power, but loss of resolution, and the boundaries of the relevant regions will usually not be known in advance.

In order to address these challenges, we developed a multiscale approach that simultaneously carries out differential expression analysis at all scales and locations. Our method enables identification of multiple (potentially overlapping) gene sets, and patterns of spatial heterogeneity in gene expression, combining information from spatially coordinated genes in order to detect more subtle variations in expression at lower coverage.

The method scales efficiently to datasets with measurements for thousands of genes, and application to FISSEQ data demonstrates the ability to highlight key patterns of spatial variability, as well as sets of spatially coordinated genes.



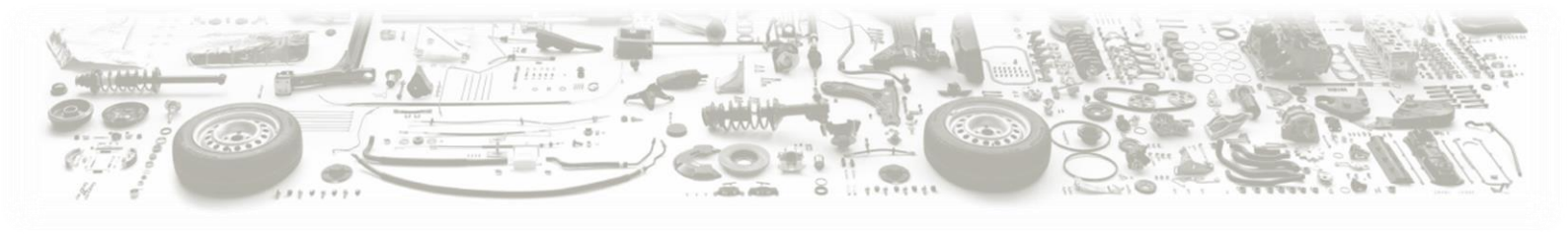
P087

Bingo-DNA: how to 3-fold increase qPCR logarithmic accuracy with a self-assembled DNA nanodevice

Lucia Coral[1], Alex Stopar[1], Matteo Castronovo[2]

[1]*University of Trieste, Trieste, Italy*
[2]*University of Udine, Udine, Italy*

Nucleic acid quantification strategies involving quantitative (q)PCR analysis can be biased by amplification reactions and, in the case of RNA, by retro-transcription reaction. Moreover, the detection threshold of 2-fold variation in DNA concentration with qPCR assays is a major hurdle in genomics and is typically circumvented with more complex approaches such as digital-PCR or next-generation sequencing. Our work aims at coupling qPCR with a self-assembled nanosensor, which can help overcome the aforementioned difficulties without requiring updates to traditional qPCR instrumentation.
Components of such sensor are three consecutive “foot-loop” DNA probes, each carrying a target-complementary sequence in the loop. Target can be any short DNA or RNA sequence such as miRNA. Probes are assembled over a common scaffold that joins their “feet”. Each “foot” carries a restriction site, and upon hybridization of three copies of the same target molecule on the respective loops, the site of each foot is destabilized (“bingo” configuration). Only in this case, the whole scaffold is protected from enzymatic cleavage and the full sequence of the “bingo-scaffold” can be amplified with PCR.
Target and full sequence bingo-scaffold concentrations are correlated by power function that allows transforming a two-fold variation of target concentration into an approx. 8-fold variation of bingo-scaffold concentration. In other words, the currently nearly indiscernible ΔCt =1 of the target turns into ΔCt =3 of the bingo-scaffold. This expected result increases standard qPCR sensitivity 3-folds. In this presentation we will report on recent results and expected results of the above-described project.



P088

Single-cell RNA-sequencing reveals two subpopulations of satellite cells

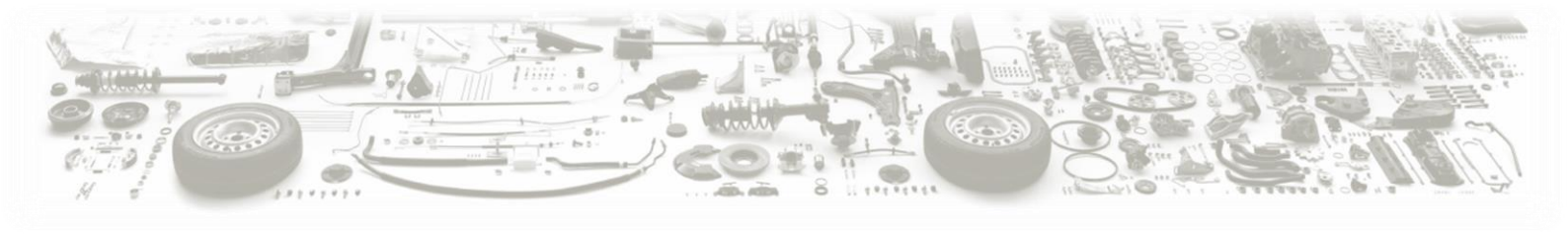
Susanne Carina van den Brink[1], Fanny Sage[1,2], Bastiaan Spanjaard[1,2], Catherine Robin[1,2], Alexander van Oudenaarden[1,2]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Royal Netherlands Academy of Arts and Sciences, Utrecht, Netherlands*

Regeneration of skeletal muscle in adults depends on the activation of otherwise quiescent muscle stem cells, the satellite cells. We applied single-cell RNA-sequencing (CEL-Seq) to satellite cells isolated from uninjured tibialis anterior muscles of Pax7nGFP mice(1) to characterize their heterogeneity and found two subpopulations of satellite cells. The smaller subpopulation is characterized by the expression of Fos, Jun, Socs3, Cxcl1, heat shock proteins and other genes that are known to be upregulated in muscles upon injury(2). We could not validate the in vivo expression of Fos, Socs3 and Cxcl1 in satellite cells of uninjured muscles with single molecule fluorescent in situ hybridization (smFISH) experiments, suggesting that the expression of these genes is induced artificially by the collagen-based muscle dissociation that precedes our CEL-Seq experiments. Therefore, we believe that a subset of the quiescent satellite cells interprets the dissociation as muscle injury, leading to their activation. To test this hypothesis we are currently performing both dissociation time-courses and smFISH experiments on injured muscles.

References:
(1)Sambasivan, R., Developmental Cell 16, 810-821, 2009
(2)Warren, G. L., J Physiol 582.2, 825-841, 2007

Acknowledgements: This work was supported by the European Research Council Advanced grant (ERC-AdG 294325-GeneNoiseControl) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) Vici award. Pax7nGFP mice were kindly provided by Sharagim Tajbakhsh.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

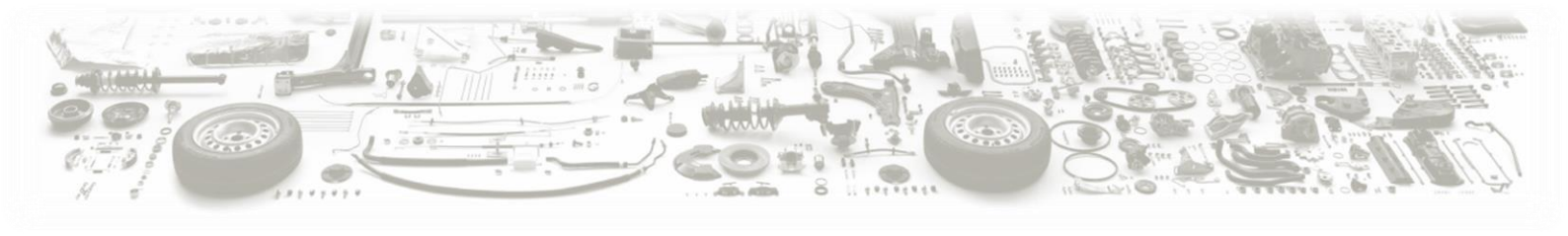
P089

Droplet Microfluidics for Barcoding Single-Cell Transcriptomics

Linas Mazutis[1], Allon Klein[2], Ilke Akartuna[3], Dave Weitz[3], Marc Kirschner[2]

[1]*Vilnius University, Vilnius, Lithuania*
[2]*Harvard Medical School, Boston, MA, United States*
[3]*Harvard University, Cambridge, MA, United States*

We present a high-throughput droplet microfluidics approach for barcoding the RNA from thousands of single cells for subsequent analysis by next-generation sequencing. The individual cells are loaded into aqueous droplets along with a set of uniquely barcoded DNA primers enabling single-cell transcriptomics of a large number of cells in a heterogeneous population. The method shows a surprisingly low noise profile and is readily adaptable to other sequencing-based assays. We use the platform to analyze mouse embryonic stem cells, revealing in detail the population structure and the heterogeneous onset of differentiation after leukemia inhibitory factor (LIF) withdrawal. The reproducibility of these high-throughput single-cell data allowed us to deconstruct cell populations and infer gene expression relationships.



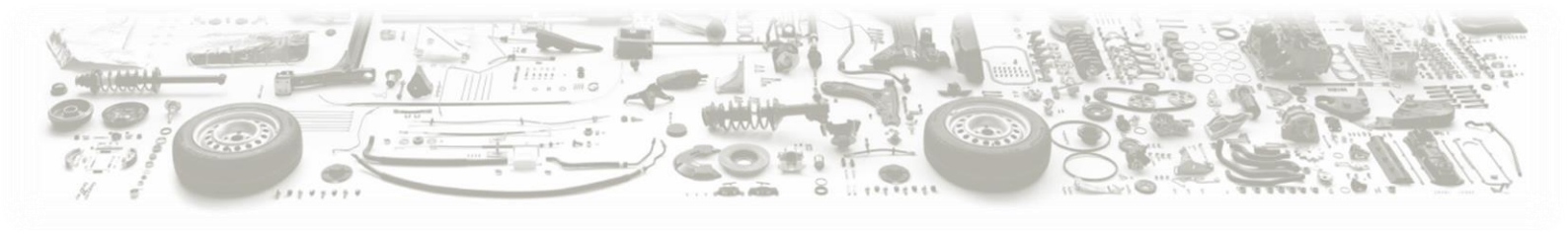
P090

High-resolution pseudotemporal ordering of single-cell gene expression profiles

Kieran Renfrew Campbell, Caleb Webber, Chris Ponting

University of Oxford, Oxford, United Kingdom

The transcriptomes of single cells undergoing biological processes - such as differentiation or apoptosis - display remarkable heterogeneity that is not captured through bulk sequencing. However, single-cell sequencing itself offers only a snapshot of dynamic and evolving cellular processes. Consequently, one outstanding challenge in single-cell transcriptomics is to identify an ordering of individual transcriptomes that best captures the progression of cells through a given process, while remaining robust to the high levels of technical and biological noise that is characteristic of single-cell data. Here we present an approach to assigning pseudotime to individual cells that first uses laplacian eigenmaps - a nonlinear manifold learning algorithm - to reduce the dimensionality of the single-cell data, before fitting non-parametric principal curves in the reduced space. We further present a method able to detect switch-like differential expression of genes along pseudotime that takes into account the excess of zeros typically present in single-cell RNA-seq data.



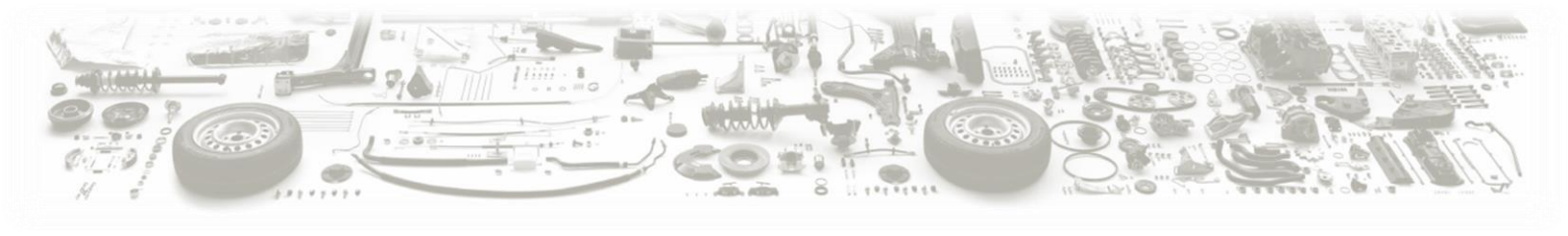
P091

Single Cell Phylogenetic Fate Mapping

Christopher Wei, Kun Zhang

University of California, San Diego, CA, United States

Questions regarding cellular growth, migration, and death within both healthy and diseased human tissue remain due to the difficulties in tracing the developmental lineages of individual cells. High resolution cell lineage trees can help answering fundamental questions such as how healthy human tissue develop, undergo repair and turnover, or how diseases such as cancer progress and spread. Fate mapping methods based on low-frequency somatic point mutations in general are costly and have limited resolution. Microsatellites are highly mutable and more informative. However, currently only tens of loci can be typed per single cell.
We present a method for constructing single cell phylogenetic fate maps based on simultaneously typing thousands of microsatellite loci. Single cell genomes are isolated and amplified in high-throughput, followed by padlock capture and deep sequencing of microsatellite loci. A computational algorithm was developed for accurate calling of microsatellite alleles from Illumina single-end sequencing reads. We are currently applying this method to construct phylogenetic fate maps of single cortical neurons in human post-mortem brains. With 2-5 million sequencing reads per neuron, we obtained microsatellite calls on over 300 loci shared among different neurons, which is sufficient to resolve single cells at a resolution 10-50x higher than that achieved by previous methods. Ultimately, our low cost, high throughput single cell microsatellite genotyping method will allow for high resolution fate mapping of any human normal and disease tissue.



P092

Single cell profiling of an in vivo, Ras V12 based tumour in the Drosophila eye primordium

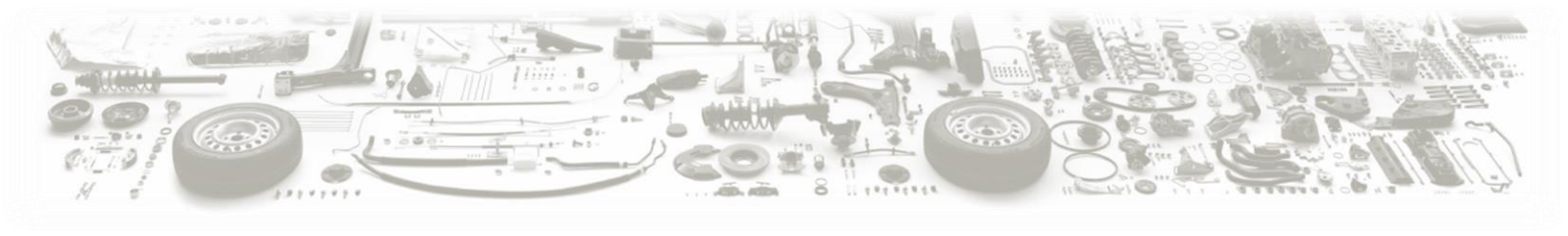
Kristofer Davie, Jelle Jacobs, Valerie Christiaens, Stein Aerts

KU Leuven, Leuven, Belgium

The Ras V12/scrib-/- tumour model has been widely used in Drosophila to study Ras-dependent oncogenesis and epithelial-mesenchymal transition (EMT). Recently, using a combination of ATAC-seq, FAIRE-seq and RNA-seq, we identified regulatory regions and two key transcription factors (AP-1 and Stat92E) involved in the development of these tumours (1). Using data from the whole tumour, it is however impossible to determine whether these factors are part of the same gene regulatory networks (GRN) in a cell, or if they are controlling separate GRNs in different cells. In an attempt to address this question of tumour heterogeneity at the gene regulatory level, we used the C1 single cell auto-prep system (Fluidigm) and performed single-cell RNA-seq (SMART-seq2 (2)) and single-cell ATAC-seq (3) on dissociated in vivo tumours. We envision that this approach will help in unraveling cellular state transitions downstream of oncogenic Ras and enable understanding GRNs at the single cell level.

References:
1. Davie and Jacobs et al, 2015, doi: 10.1371/journal.pgen.1004994,
2. Picelli et al, 2013, doi: 10.1038/nmeth.2639,
3. Buenrostro et al, 2014, doi: 10.1038/nature14590

Funding provided by: IWT (141509 to KD, 131287 to JJ), OT (OT/13/103 to SA), FWO (G.0704.11N, G.0640.13, and G.0791.14 to SA)



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P093

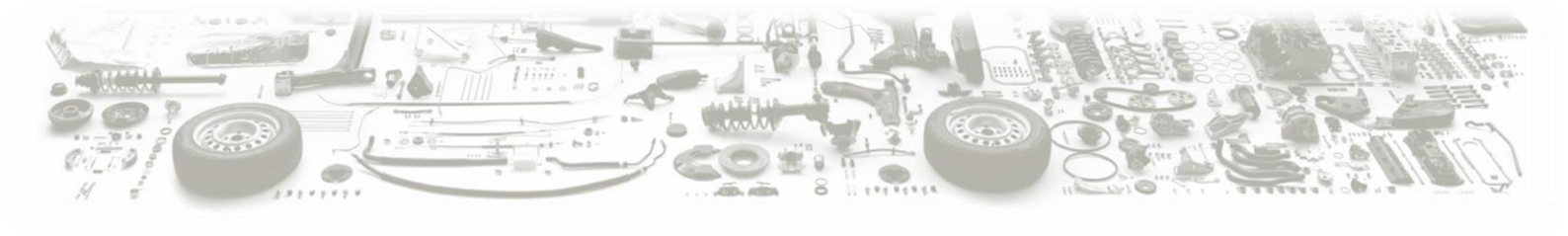
Single-cell ATAC-seq provides a new dimension to profiling cellular regulation and variation

Jason Buenrostro[1], Beijing Wu[1], Ulrike Litzenburger[1], Dave Ruff[2], Michael Gonzales[2], Michael Synder[1], Howard Chang[1], William Greenleaf[1]

[1]*Stanford University, Stanford, CA, United States*
[2]*Fluidigm Corporation, South San Francisco, CA, United States*

We report the development of a method to profile cell-to-cell variation in individual cells by illuminating the landscape of functional genomic elements using the single-cell Assay for Transposase-Accessible Chromatin (scATAC-seq). Employing the C1™ system and programmability provided by the C1 Script Builder™, we unveil a robust workflow and reagent system for generating genome-wide DNA accessibility profiles from individual mammalian cells. Our initial survey across 254 GM12878 single-cells was compiled and combined in an ensemble dataset, pooled scATAC-seq profiles closely fit ATAC-seq and DNase genome-wide hypersensitivity maps obtained from bulk cellular samples comprised of tens of thousands and millions of cells, respectively. The scATAC-seq datasets produce an average of 73,000 mapped fragments to genomic regions and in a typical cell, our scATAC-seq libraries cover 9.4% of promoters. We extended this technique to interrogate the epigenomes of eight cell types comprising 1,632 C1 capture chambers: tier 1 ENCODE cell lines H1, K652, GM12878, as well as V6.5 mouse ESCs, EML, TF-1, HL-60 and BJ fibroblasts. Using a data analysis tool kit focused on genomic features, we quantify regulatory variation, identify transcription-factor motifs and structured cis features associated with epigenomic and cellular variability. The application of scATAC-seq to cellular populations champions a new experimental paradigm to unravel mechanisms of heterogeneity within single-cells, which promises new insights into cellular plasticity.

The authors acknowledge support from NIH, Lifespan Extension Foundation, Rita Allen Foundation, Baxter Foundation, HHMI and the NSF.
References: Buenrostro et.al. Nature 2015 Jun 17.doi:10.1038/nature14590, Buenrostro et. al. Nat Meth 10, 1213-1218 (2013).



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

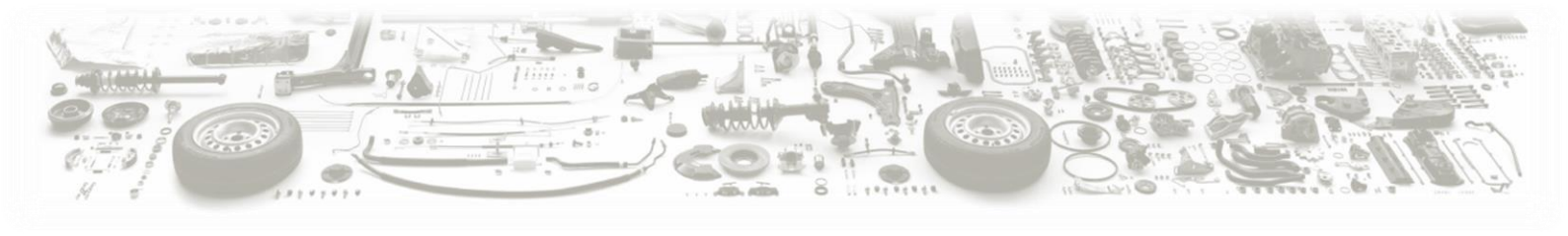
P094

Molecular programs sustaining outer radial glia during human corticogenesis

Alex Pollen[1], Tomasz Nowakowski[1], Jiadong Chen[1], Hanna Retallack[1], Carmen Sandoval-Espinosa[1], Joe Shuga[2], Siyuan Liu[1], Michael Oldham[1], Aaron Diaz[1], Daniel Lim[1], Anne Leyrat[2], Jay West[2], Arnold Kriegstein[1]

[1]*UCSF, San Francisco, CA, United States*
[2]*Fluidigm, South San Francisco, CA, United States*

Radial glia, the neural stem cells of the neocortex, are located in two niches: the ventricular zone and outer subventricular zone. Although outer subventricular zone radial glia may generate the majority of human cortical neurons, their molecular features remain elusive. By analyzing gene expression across single cells, we find that outer radial glia preferentially express genes related to extracellular matrix formation, migration, and stemness, including TNC, PTPRZ1, FAM107A, HOPX, and LIFR. Using dynamic imaging, immunostaining, and single cell clonal lineage analysis, we relate these molecular features to distinctive behaviors of outer radial glia, demonstrate the necessity of STAT3 signaling for their cell cycle progression, and establish their extensive proliferative potential. These results suggest that outer radial glia maintain the subventricular niche through local production of growth factors, potentiation of growth factor signals by extracellular matrix proteins, and activation of self-renewal pathways, thereby enabling the developmental and evolutionary expansion of the human neocortex.



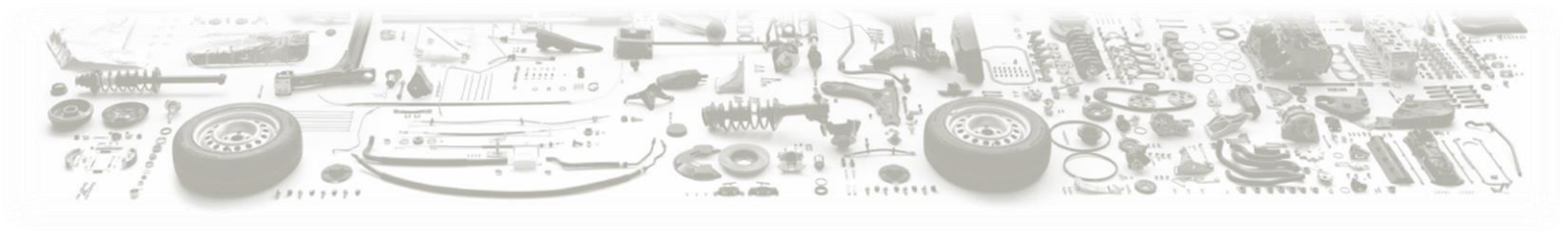
P095

Decoding Amplified taRgeted Transcripts with Fluorescence In Situ Hybridization (DARTFISH) for RNA mapping in human brains

Matthew Cai[1], Ho Suk Lee[1], Jian-Bing Fan[2], Kun Zhang[1]

[1]*University of California San Diego, San Diego, CA, United States*
[2]*Illumina, Inc, San Diego, CA, United States*

Technology to analyze gene expression of a cell population with single-cell resolution and localization is critical for understanding the heterogeneity of structured tissues such as human brain and tumors. Advancements in single-cell sequencing allow the full transcriptome of isolated cells to be profiled but it neglects the native spatial context of the cell. While several methods have been reported, in situ RNA mapping in complex tissues, including post-mortem human specimens, remains to be demonstrated. We have developed a highly multiplexed method, called DARTFISH, for mapping of an arbitrary subset of RNA transcript in situ. To achieve robust detection in complex tissue samples that has significant background autofluorescence, DARTFISH adopted an in situ cDNA transcription and rolling circle amplification strategy similar to FISSEQ. We leverage the multiplex capability and high specificity of padlock probes to capture thousands of targets in situ and include a hybridization-based combinatorial barcode scheme that allows amplicons to be decoded with quick reaction kinetics and under isothermal conditions.
As a proof of concept we performed DARTFISH on human culture fibroblast cells and cortical sections from human post-mortem brains. With a probe set targeting 240 genes, we detected 800 amplicons per cell in fibroblast monolayer and 140 amplicons per cell in 10µm human cortical sections. In a 0.6mm^2 cortical section, we decoded 27,812 amplicons mapping to 235 of the 240 genes. We also demonstrated that amplicon copy number can be used to quantify transcript abundance.



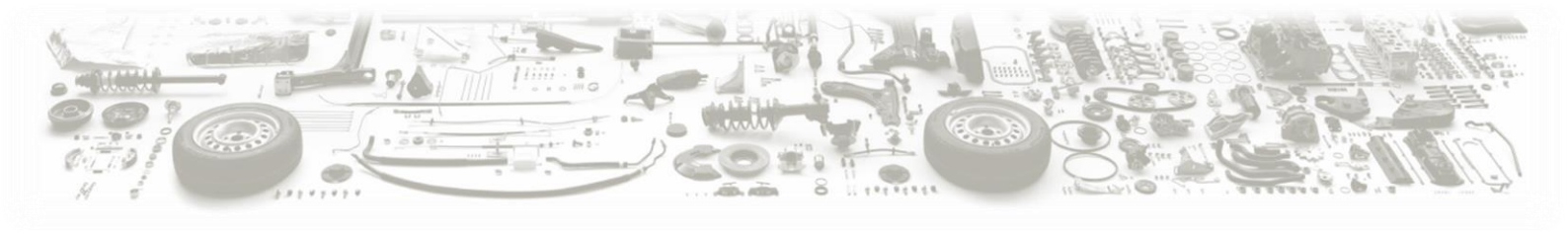
P096

DNA-microscopy: Sequence-readouts of tissue microstructure

Joshua Weinstein

Broad Institute, Cambridge, MA, United States

Many complex cell populations, from the brain to the adaptive immune system, rely on diverse gene variants, somatic mutations, and expression patterns for some of their most essential functions. However, rather than only endowing intrinsic properties to individual cells, genetic heterogeneity often operates at the level of cellular interactions. Where a cell is relative to other cells is therefore a dimension of its phenotypic state that is no less important than any single gene it expresses. The growing toolset for spatial transcriptomics aims to close the gap between these two types of measurements. Nevertheless, sequence accuracy, resolution, and ease-of-use remain obstacles to wide adoption. We present a new and experimentally simple method, called DNA-microscopy, for encoding spatial distributions of gene sequences and expression directly into DNA-sequencing libraries. We demonstrate that these distributions can then be decoded accurately in proof-of-principle experiments.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

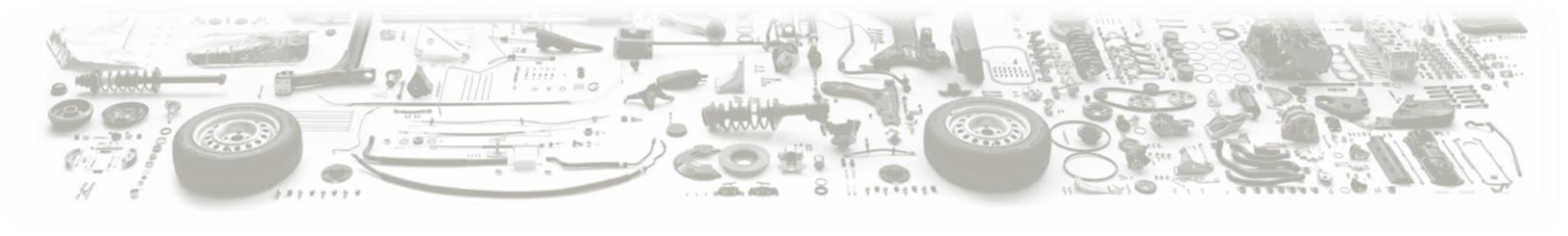
P097

Integrated transcriptome and methylome based single-cell approach to interrogate early mouse embryonic development

Hisham Mohammed[1], Aurora Savino[1], Iain Macaulay[2], Antonio Scialdone[3], Stephen Clark[1], Tamir Chandra[1], Carla Mulas[4], Wendy Dean[1], John Marioni[3], Jenny Nichols[4], Wolf Reik[1]

[1]*The Babraham Institute, Cambridge, United Kingdom*
[2]*Wellcome Trust Genome Campus, Cambridge, UK, Cambridge, United Kingdom*
[3]*European Molecular Biology Laboratory, EBI, Cambridge, UK, Cambridge, United Kingdom*
[4]*Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom*

The early mouse embryo rapidly establishes itself from a single fertilized zygote, into distinct cell types with specialized roles. The embryo progressively moves from a state of pluripotency to more differentiated cell states in a seemingly well-orchestrated manner.
Subsequent to delineation of the trophectoderm at the blastocyst stage, primitive and visceral endoderm lineages emerge as well as the epiblast which eventually gives rise to the three primary germ layers – endoderm, ectoderm and mesoderm. To understand the regulatory processes involved, we employ an integrated method that allows RNA-seq and Bisulphite-seq from the same cell. Results obtained begin to reveal the detailed relationships between the transciptome and methylome in early embryonic development.
Analysis of the transcriptome reveals key transcriptional networks potentially controlling epiblast differentiation. Precursors of the primitive endoderm and visceral endoderm are identified in the stages preceding their proper establishment. The early post-implantation embryo shows prevalence of repressive transcription factors. Further, increased heterogeneity in the epiblast is observed as the embryo progresses towards gastrulation, which then leads cells to identify lineage and commit. This transition also corresponds with an increase in DNA methylation, which thus potentially plays a role in this process.
This dataset thus begins to map the transcriptional network across early development, allowing the detailed study of decision making processes during differentiation.



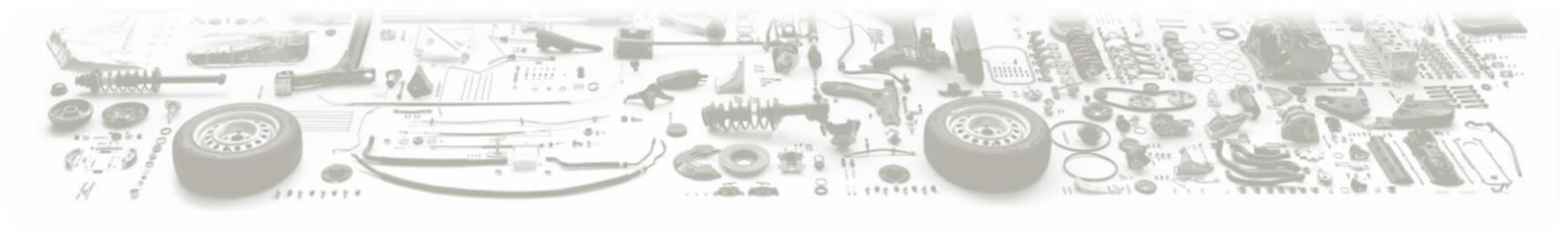
P098

Linking heterogeneity in gene expression patterns to cell history and fate in the early mouse embryo

Mubeen Goolam[1], Antonio Scialdone[2], Iain Macaulay[2], Sarah Graham[1], Agnieszka Jedrusik[1], Anna Hupalowska[1], Thierry Voet[2], John Marioni[3], Magdalena Zernicka-Goetz[1]

[1]*University of Cambridge, Cambridge, United Kingdom*
[2]*Wellcome Trust Sanger Institute, Hinxton, United Kingdom*
[3]*EMBL-EBI, Hinxton, United Kingdom*

The first cell fate decision in mammalian development leads to specification of pluripotent embryonic lineage and differentiating extra-embryonic lineages. How and when this decision is initiated remains unknown.
Two alternative models have been proposed: the first model suggests that all cells are equally likely to contribute to any cell lineage and have the same developmental potential until they adopt different cell positions during blastocyst formation. On the other hand, according to the second model, heterogeneity between cells is established before differential cell position and plays a role as it leads to developmental bias.
We addressed this controversy by determining the extent of transcriptional heterogeneities between all individual cells in the mouse embryo at different stages. By relating gene expression to the spatial context of each cell within the mouse embryo during progression towards lineage specification, we identified Sox21 as an early regulator of pluripotency. Cells establishing the pluripotent lineage initiate Sox21 expression at the 4-cell stage, while cells leaving totipotency to enter a differentiation programme do not. In agreement, depleting Sox21 leads to premature expression of Cdx2 to initiate differentiation.
Our results indicate that cell division history is linked to cellular heterogeneity in gene expression patterns at the 4-cell stage and that this has a functional role in guiding cell fate in the mouse embryo.



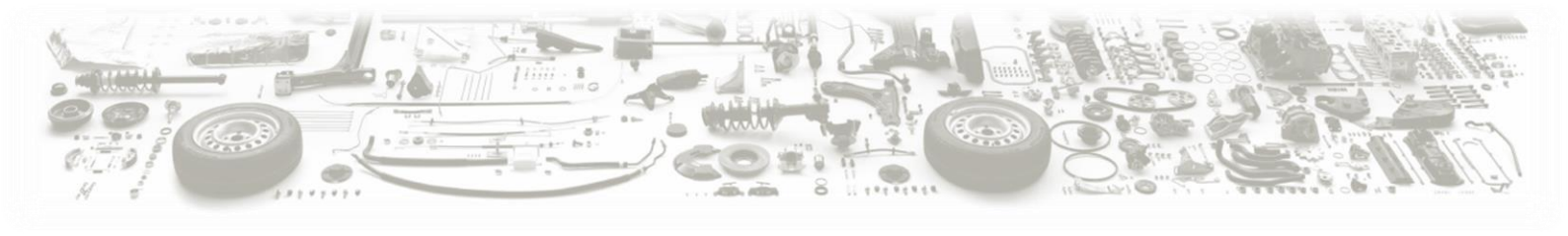
P099

Single cell view of host virus interactions in the marine environment

Shilo Rosenwasser, David Pilzer, Ron Rotkopf, Assaf Vardi

Weizmann Institute of Science, Rehovot, Israel

Marine photosynthetic single-cell microorganisms (phytoplankton) are the basis of marine food webs. Phytoplankton blooms are ephemeral events of exceptionally high primary productivity that regulate the flux of carbon across marine food webs. The cosmopolitan coccolithophore Emiliania huxleyi is a unicellular eukaryotic alga, responsible for the largest oceanic algal blooms covering thousands of square kilometres. Annual E. huxleyi spring blooms are frequently terminated by infection of a specific large dsDNA virus (E. huxleyi virus, EhV). We have recently demonstrated that viral infection is achieved by rapid viral-induced remodelling of several fundamental host metabolic pathways such as lipid metabolism, autophagy and meiosis. These results raise the hypothesis that heterogeneity in cell metabolic state will shape susceptibility to viral infection and allow the co-existence of host and virus within a genetically identical population. To investigate the level of heterogeneity between infected cells and its interplay with host metabolic state, we isolated single cells during different phases of infection and profiled simultaneously viral and host gene expression on a single-cell level. Based on viral genes expression and their association to different phases of infection, we differentiated and clustered cells in different viral infection-states within infected population. Expression of host genes encoded for metabolic pathways were induced specifically at distinct phases of viral infection which points at possible metabolic strategies employed during hijacking of the host metabolism. We propose that examining host-virus interactions at a single cell resolution will provide novel insights into the cellular mechanisms that govern the “arms race” in the marine environment.



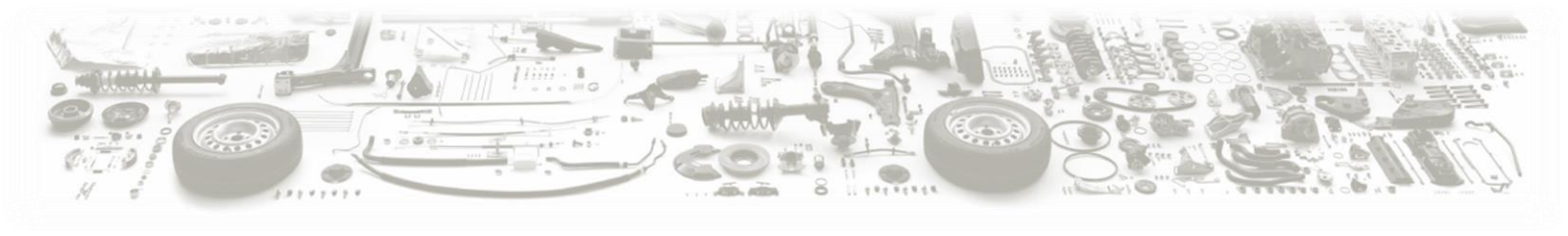
P100

Genome-wide allelic expression reveals incomplete reprogramming in Human Female Germ cells

Abel Vertesy[1], Wibowo Arindrarto[2], Matthias S. Roost[3], Björn Reinius[4], Rickard Sandberg[4], Susana M Chuva de Sousa Lopes[3], Alexander van Oudenaarden[1]

[1]*Hubrecht Institute, Utrecht, Netherlands*
[2]*Sequencing Analysis Support Core, Leiden, Netherlands*
[3]*Leiden University Medical Center, Leiden, Netherlands*
[4]*Karolinska Institute, Stockholm, Sweden*

In the germline, epigenetic reprogramming equalizes the paternal and maternal (epi)genome before meiotic entry. Key aspects of epigenetic reprogramming in germ cells are the erasure of genetic imprints and the reactivation of the previously inactivated X chromosome. As a result, allelic expression of imprinted genes and X-linked genes changes from being restricted to one parental allele to being permissive for both alleles. The maturation of female germ cells in human embryos is suggested to be an unsynchronized process, unlike in mice, therefore, requiring a single cell approach.
Here, we investigated this first by performing exome sequencing of five human embryos and their mothers. That allowed us to identify the unique set of individual informative SNPs and more importantly to determine the parental allelic specificity on a genome-wide scale. Next, using single-cell RNA sequencing we investigated allele-specific expression to determine the transcriptional maturation signature as well as the dynamics of (epigenetic) reprogramming of individual germ cells from female human embryos of first and second trimester.
By using both male and female somatic cells and autosomes as negative and positive control, respectively, we find that all (female) germ cells analysed showed some degree of reprogramming (X reactivation and imprint erasure) before entering meiosis. However, most (non-meiotic) germ cells still show a strong expression bias towards the original active parental allele, even in the second trimester. Surprisingly, we observed that the erasure of genomic imprints and chromosome X reactivation at the single cell level appear as two, at least partially, independent processes.



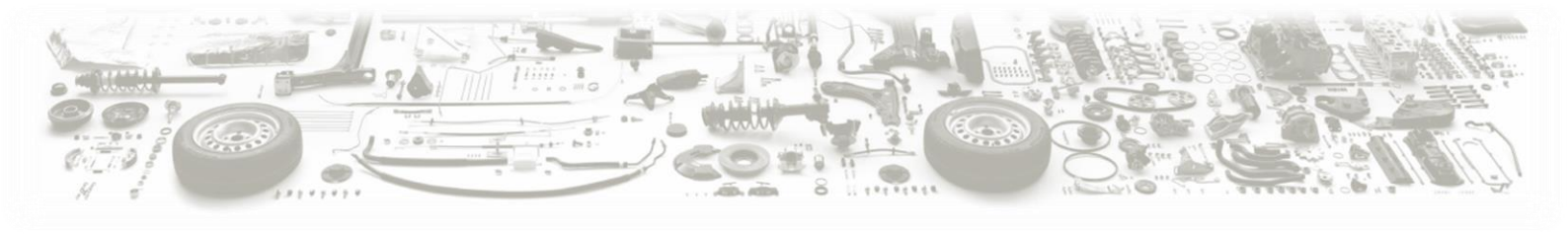
P101

Single cell transcriptional profiling of mouse embryonic stem cells across cell-cycle stages

Kedar Natarajan, Sarah Teichmann

Wellcome Trust Sanger Institute, European Bioinformatics Institute, Cambridge, United Kingdom

Cell cycle is fundamental process occurring in most cells and progression through cellcycle stages (G1, S and G2M) impacts transcription, gene regulation, chromatin architecture, cellular response and cell fate decisions. We profile 333 Hoechst-stained, single mES cell transcriptomes (2i-LIF) across different cellcycle stages using Fludigm C1 system, after quality control and normalization. Cellular size (Hoechst-staining) is directly linked to the ratio of sequencing reads to ERCC reads across single cells. Principal Component Analysis(PCA) captures the cell cycle contribution predominantly across PC1(14.2% variance) and PC2(5.3%) and projecting PCA-loadings to the published Quartz-Seq data revealed strong correlation between both datasets (albeit the different media conditions). We observe heterogeneous expression of several G1 (Hspa4, Slbp, Cdkn3), S (Ccd14, Insig2) and G2M (Plk2, Pknox1) markers that are masked in bulk studies as well as consistent differential expression (SCDE and DESeq) and Gene Ontology enrichment for ‘Mitosis’ (6.2x10-8), ‘Cytokinesis’ (1.7x10-5) for G2M cells and ‘transcriptional regulation’ (3.4x10-4) in G1 cells. Psuedotime ordering of single cells (using Monocle) achieves high correlation between identified states and pre-determined cellcycle classification. Remarkably cells are arranged in cyclic path starting and finishing at G1 cells, spanning from G1 to S to G2M stages. We identify known and novel G1 and S marker genes with peak expression with roles in metabolism (Dlat1), chromatin remodeling (Suv420h1) and embryonic development (Klhl1, Gmcl1). For target validation, we are developing new mES cell-line expressing Cas9 protein and FUCCI reporters for targeted gene knockout and impact on cell cycle length, pluripotency and differentiation potential. The identification of new functional markers and their cell cycle roles will allow researchers to better manipulate stem cells, stem cell differentiation as well as for cellular reprogramming.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

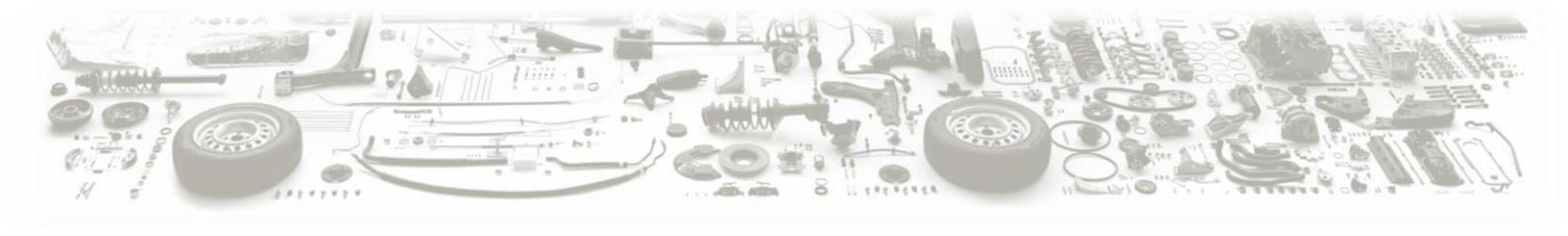
P102

Single nucleus RNA-seq reveals cell types and dynamics of adult neurogenesis in the Hippocampus

Naomi Habib[1], Yinqing Li[1], Lukasz Swiech[1], John Trombetta[1], Matthias Heidenreich[1], Aviv Regev[1,2], Feng Zhang[1,2]

[1]*Broad Institute, Cambridge, MA, United States*
[2]*Massachusetts Institute of Technology, Cambridge, MA, United States*

The cellular heterogeneity in the brain limits our ability to assess effects of in vivo state on gene expression from bulk tissue samples. Recent advances in RNA-seq technology enable measurements in single cell resolution. However, application of existing techniques to the brain is limited because it requires harsh cell dissociation of live cells. To overcome this obstacle we have developed a high-throughput single nucleus RNA-seq method, and applied it to study the adult mouse Hippocampus region. We show that nuclear RNA can be used to study cellular diversity, discover cell types, assign neurons to brain regions and identify molecular markers. Moreover, we combine single nucleus RNA-seq with a new unbiased labeling method of newborn cells to reveal the transcriptional dynamics during adult neurogenesis. This unique dynamic view of the process enabled us to identify cellular pathways and regulators involved in key steps of the neurogenesis process. Taken together, we show that single nucleus RNA-seq can be used to study cellular heterogeneity as well as dynamic processes in any heterogenic tissue.



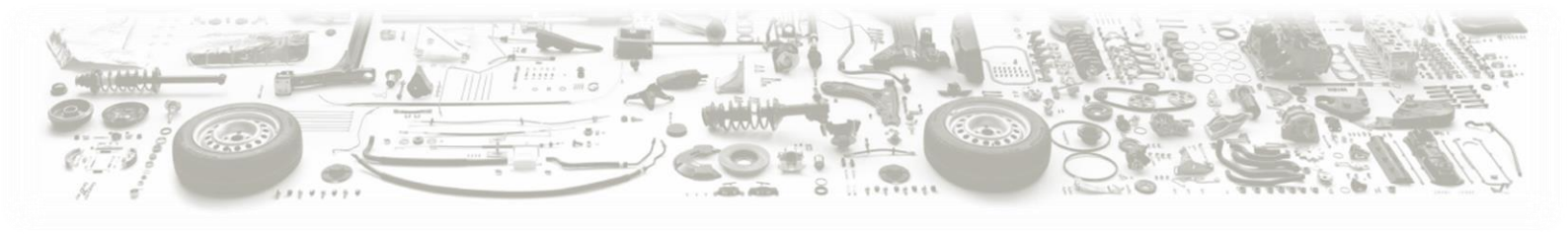
P103

A study of mRNA sorting by single cell sequencing of early stage Xenopus laevis embryos

Helena Storvall[1], Björn Reinius[1,2], Chika Yokota[1], Rickard Sandberg[1]

[1]*Karolinska Institutet, Stockholm, Sweden*
[2]*Ludwig Institute for Cancer Research, Uppsala, Sweden*

The early embryonic development of the African clawed frog, Xenopus laevis, is characterized by a strong visible polarity in the oocyte and a late zygotic genome activation. The polarity is thought to serve an important function for the later patterning of the embryo, and there are several mRNAs known to be localized to either pole before fertilization. The sorted mRNAs have been shown to have repetitive motifs in the 3'UTRs that serves as sorting signals. In this study we use a single cell sequencing technique to get a global view of maternal mRNA sorting in the early cleavage stages, what sorting signals are involved, and resolve how the initially sorted mRNAs move. To compensate for the incomplete genomic sequence and annotation for Xenopus laevis we have also assembled a transcriptome from early embryonic cells. Using a combination of our transcriptome, known Xenopus genes and the genomic sequence we have identified and annotated both previously known and novel sorted genes, and used motif finding algorithms to find consensus sorting signals in their UTRs.



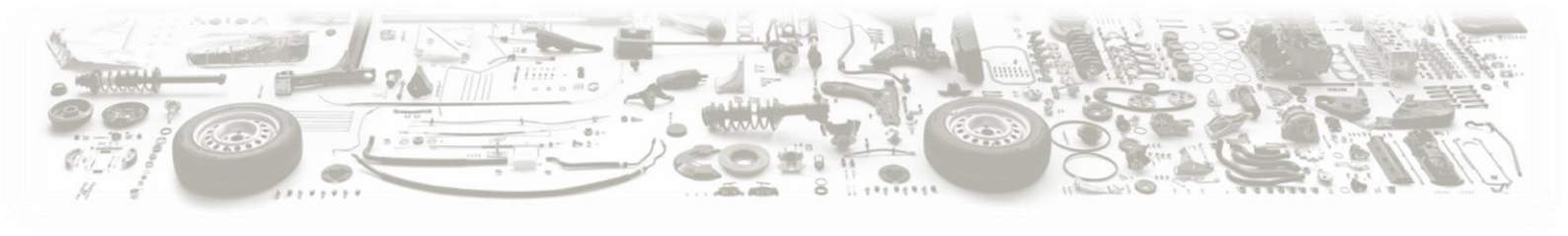
P104

CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq

Tamar Hashimshony[1], Naftalie Senderovich[1], Gal Avital[1], Agnes Klochendler[2], Alison Cole[1], Yaron de Leeuw[1], Leon Anavy[1], Dave Gennert[3], Shuqiang Li[4], Kenneth J Livak[4], Orit Rozenblatt-Rosen[3], Yuval Dor[2], Aviv Regev[3], Itai Yanai[1]

[1]*Technion, Haifa, Israel*
[2]*Hebrew University, Jerusalem, Israel*
[3]*Broad institute, Canbridge, MA, United States*
[4]*Fluidigm, San Francisco, CA, United States*

Single-cell transcriptomics requires a method that is sensitive, accurate, and reproducible. We previously described CEL-Seq, a method using in-vitro transcription for amplification of RNA. CEL-Seq also features early barcoding, which greatly reduces hands-on time and library preparation costs. Here, we present CEL-Seq2, a modified version of our CEL-Seq method, with higher sensitivity, lower costs, and even less hands-on time. We use CEL-Seq2 to analyze every cell of the developing C. elegans embryo, a feat that could not have been achieved without the protocol's increased sensitivity due to the small size of the cells. We also implemented CEL-Seq2 on Fluidigm's C1 system, thereby providing its first single-cell on-chip barcoding method. This allows for a single library to be prepared for each C1 chip. Using this technology, we analyzed cycling mouse fibroblasts with a GFP marker for S/G2/M, and detected the gene expression changes that accompany the progression of the cell-cycle.



P105

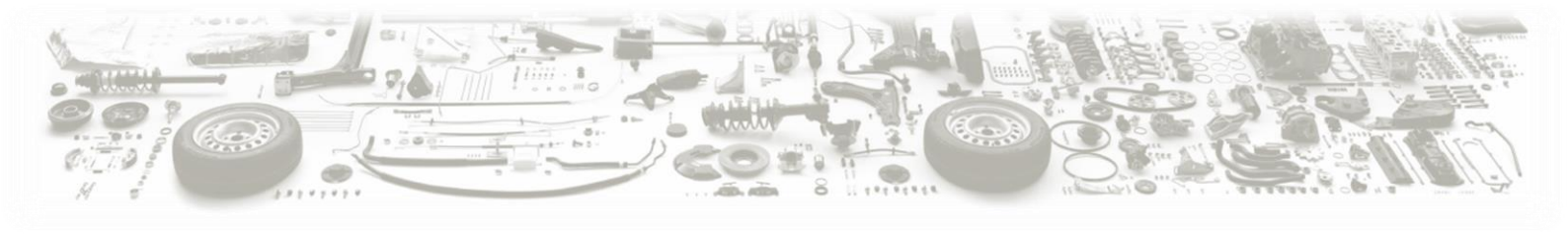
Tracking Edge Rewiring During the Epithelial-To-Mesenchymal Transition

Roshan Sharma, Smita Krishnaswamy

Columbia University, New York, NY, United States

The epithelial-to-mesenchymal transition (EMT) is a process by which epithelial cells transform into mesenchymal cells by losing their rigid adherence neighbors, acquiring motility, and gaining the ability to invade the basal lamina. This is thought to mimic the process of cancer metastasis. Because epithelial cells differ vastly from mesenchymal cells, we hypothesize that the signaling network is dynamically rewired during EMT. We use the Py2T mouse breast cancer cell line as a model system for EMT and study signaling through single-cell mass cytometry data measured at 3 and 5 days after TGFB treatment and also under a panel of drug inhibitions. Our goal is to derive key changes in the signaling network that drives the EMT transition and characterize perturbations that halt the process.

We use a single snapshot of individual cells to recover a trajectory of EMT-time and map how key signaling edges change during this process. First, we characterize subpopulations of cells in various stages of EMT through multidimensional clustering and trajectory-mapping to derive a continuous EMT-timestamp for each cell. Then, we use recently developed computational tools, conditional-Density Rescaled Visualization (DREVI) and conditional Density Resampled Estimate of Mutual Information (DREMI) to mathematically characterize signaling relationships and score strengths of interactions in the signaling network. We extend these techniques to higher dimensions to track how the DREVI and DREMI of an edge change with time, by using EMT-time as a third dimension. As we track edge-dynamics along EMT-time, we observe strengthening and waning of influences in the signaling network and identify key features of edges that drive EMT. We use these features to predict inhibitions that partially halt EMT. Computationally, we provide a framework for the use of cell-to-cell heterogeneity in the EMT transition process to recover system dynamics of a cellular transformational process.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P107

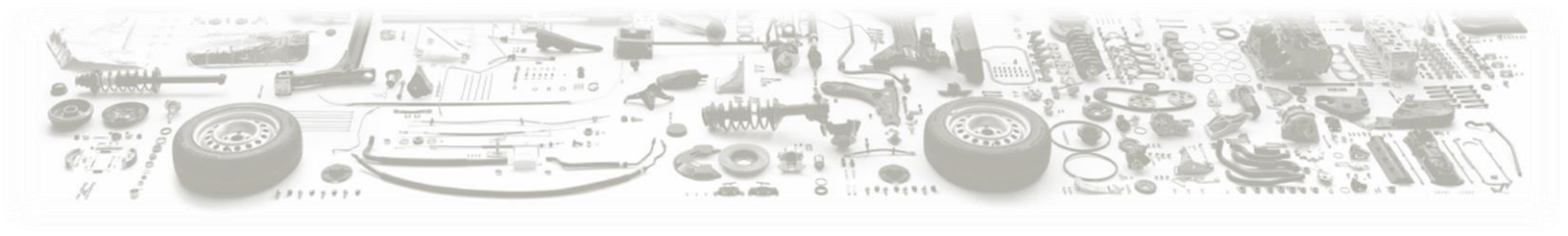
Towards transforming cell-to-cell variations into biological insights: From single cells to networks to human populations

John S. Tsang[1,2], Manikandan Narayanan[1,2], Andrew J. Martins[1,2], Yong Lu[1,2], Bethany Fixsen[1,2], Thorsten Prustel[1,2], Angelique Biancotto[2], Naisha Shah[1,2], Pamela L. Schwartzberg[2,3], Rachel Gottschalk[1,2], William Lau[2], J. Philip McCoy[2,4], Cindi Pfannkoch[1,2], Katherine Wendelsdorf[1,2], Foo Cheung[2]

[1]NIAID, Bethesda, MD, United States
[2]NIH, Bethesda, MD, United States
[3]NHGRI, Bethesda, MD, United States
[4]NHLBI, Bethesda, MD, United States

Increasing amounts of single-cell expression and phenotypic data are being generated for studying cellular heterogeneity. There are several pressing challenges in transforming such data into biological insights, including: 1) how to disentangle biological versus technical variations for assessing meaningful changes in cell-to-cell expression variation (CEV) across conditions; 2) how to utilize CEV to help infer biological mechanisms and functions; 3) what are the functions (or functional correlates) of CEV in different cell types and conditions. Here we highlight integrated computational and experimental efforts in our lab to help address these challenges. To robustly assess biological variation vs. that induced by assay-sensitivity limitations often prevalent in single-cell profiling, we develop a strategy for inferring CEV parameters and their changes across conditions via Bayesian integration of simultaneously obtained expression profiles from single and random pools of k (e.g., 10) cells. Using human macrophage activation as a model, we show how single- and k-cell assessment of CEV enables robust inference of condition-specific rewiring of regulatory networks. Our analysis uncovered an unexpected, condition-specific signaling and transcriptional circuit that underlies the macrophage response to Interleukin (IL)-10 and revealed that differences in signaling activity within cells exposed to distinct environments can alter information propagation behavior among genes. Finally, we discuss time-resolved quantification of CEV in distinct immune cell subsets in a human cohort and how such a population-based approach can reveal functional correlates of CEV, including examples linking CEV to human aging and genetic variants associated with disease.

Funded by Intramural Program of NIAID/NIH.



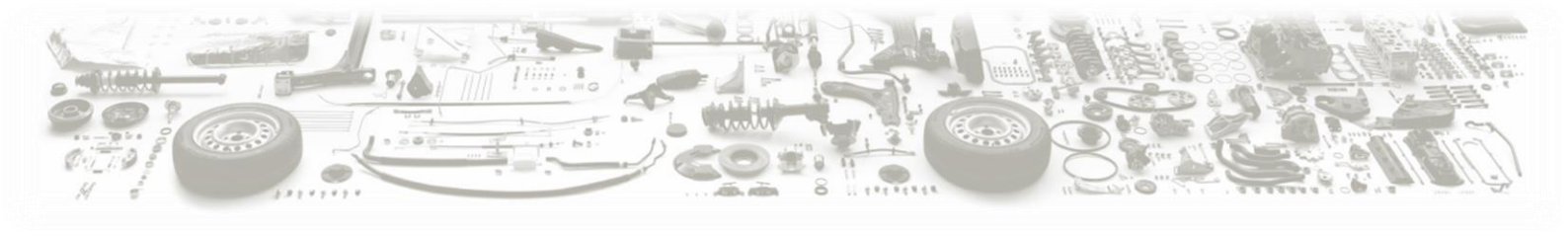
P108

Transcriptional profiling of the adolescent mouse dentate gyrus

Hannah Sophie Hochgerner, Amit Zeisel, Sten Linnarsson

Karolinska Institutet, Stockholm, Sweden

The dentate gyrus is a sub-structure of the hippocampus and amongst others involved in the formation of new episodic memories. It mainly receives input from the entorhinal cortex (perforent path) and projects to the hippocampus CA3. Anatomically it consists of three layers (i) a dense granule cell layer, (ii) a molecular layer mainly featuring granule dendrites and fibres, and (iii) the hilus or polymorphic layer in the center of the structure. Today, the dentate gyrus is arguably most studied for its maintained neurogenesis in the adult rodent and human brain. Here, we aim to characterize the dentate gyrus' cellular diversity beyond major classified groups. We performed single-cell RNA-seq on 2300 cells, isolated from CD-1 mice of postnatal days 21-28. We identified a small number of inhibitory, interneuron-type cells and a portion of excitatory cells located to the hilus and neighbouring CA3. The majority of the neuronal population was made up of excitatory granule neurons, sharing a general neuronal expression profile without distinct granule markers; likely representing postnatally-born, immature granule neurons. We further found several classes of the major glial populations (astrocytes, oligodendrocytes and microglia), as well as endothelial cells and several cycling cells.



P109

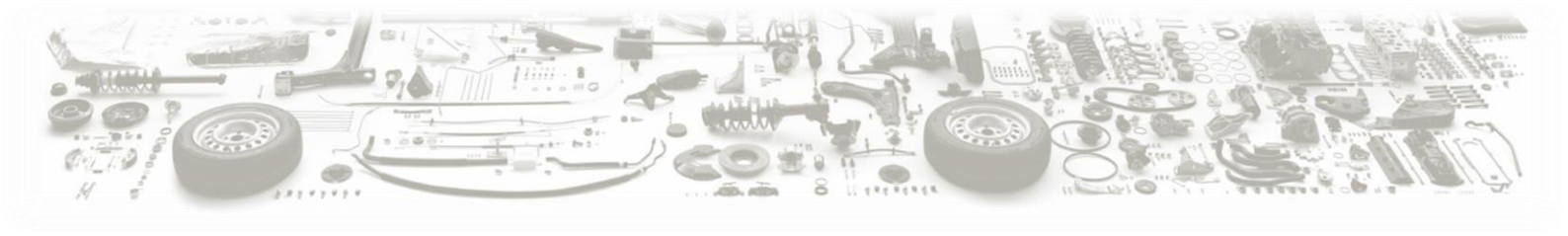
Single cell isoform and alternative promoter usage determined by PacBio long read sequencing and STRT single cell RNA seq

Kasper Karlsson, Sten Linnarsson

Karolinska Institutet, Stockholm, Sweden

Single cell transcriptomics has been focused on finding more and more precisely defined cell types in various tissues. Gene quantity has been studied extensively but less attention has been given to isoform variation within single cells (Velten et al), and no attempt has been done to study splice variation with a long read technology nor has a genome wide promoter mapping study on single cells been done to our knowledge.
We used PacBio sequencing technology to profile 6 single cells from three different cell types. To make sense of the highly amplified material Unique Molecular Identifiers (UMI) were used for error correction.
The STRT RNA library preparation method targets the 5' prime end of the RNA transcript and each transcript read can therefore be assigned to a specific promoter. We used this property of the STRT method to study alternative promoter usage using previously published data from 3000 single cells from the somatosensory cortex and hippocampal CA1 region (Zeitsel et al).
We found that alternative isoform usage is common for genes within single cells, both for exon inclusion/exclusion events and heterogeneity in exon start and end positions.

References
Velten et al, Mol Syst Biol. (2015) 11: 812
Zeitsel et al, Science (2015) Vol. 347 no. 6226 pp. 1138-1142



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P110

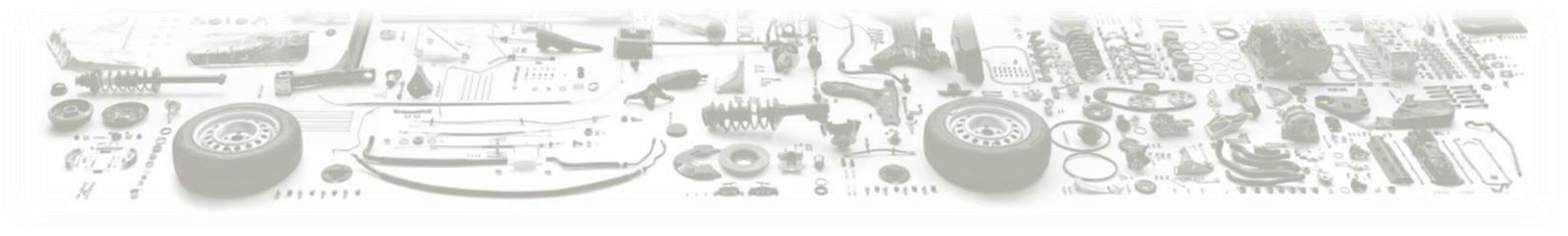
Identification of adipogenic precursor populations through single-cell RNA-seq

Petra Schwalie[1], Nassila Akchiche[2], Julie Russeil[1], Christian Wolfrum[2], Bart Deplancke[1]

[1]*EPFL, Lausanne, Vaud, Switzerland*
[2]*ETHZ, Zurich, Switzerland*

Adipose depots consist of a heterogenous and dynamic mix of cell types, which makes their molecular and developmental characterisation highly challenging. However, a comprehensive identification of the types of cells forming distinct fat tissues is essential for understanding the etiology of obesity and its associated metabolic pathologies. Single-cell transcriptomics already provided unprecedented insight into the molecular composition of complex systems, for instance by the in-depth characterization of immune* and brain cell types*. Here, we employed single-cell RNA-seq to molecularly dissect adipogenic precursor populations resident in the mouse subcutaneous stromal vascular fraction. Using the SMART-seq protocol, we successfully enquired 264 adipose stem cells, detecting over 4,000 expressed genes per cell. Overall, the expression estimated by aggregating single cells was highly similar to matched population RNA-seq replicates. We used dimensionality reduction and clustering techniques to group cells according to their gene expression similarity, delineating three distinct subpopulations and their transcriptional markers, including transcription factors, cell surface receptors and secreted factors. Specifically, population A shows high expression of adipogenic genes including Pparg, Fabp4 and Cd36, population B enriches for complement factors and vasculature-related genes, and population C for stemness and differentiation-blocking genes. To understand functional differences between these distinct cell groups, we complemented our computational analyses with extensive phenotypic characterizations in terms of cell-to-cell signaling and differentiation capacity. Overall, our study demonstrates that the adipogenic stem cell fraction can be subdivided into distinct cell populations with defined molecular properties, which interplay to mediate the complex responses of fat depots to metabolic stimuli.

* Jaitin et al. Science. 2014 Feb 14;343(6172):776-9
* Zeisel et al. Science. 2015 Mar 6;347(6226):1138-42



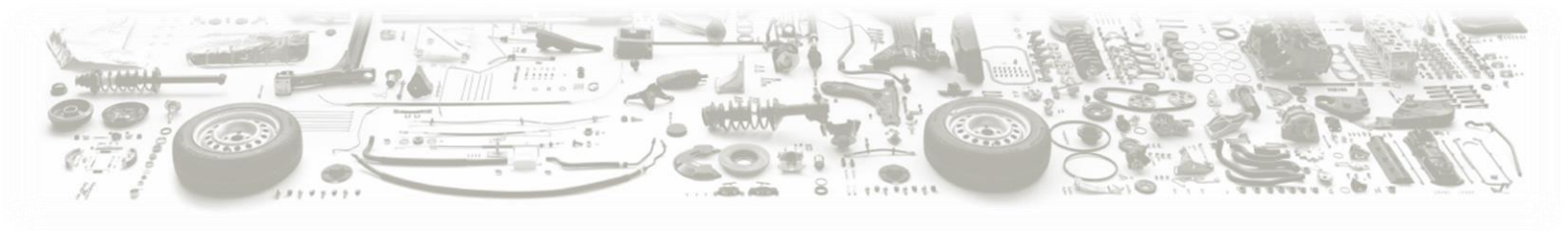
P111

Allele Specific Expression Analysis of Human iPS Cells at Single Cell Level

Yasutaka Mizoro, Masahiro Nakamura, Akira Watanabe

Kyoto University, Kyoto, Japan

Increasing of the single cell RNA sequencing technology has enabled to describe the heterogeneity of gene expression in each cell. These heterogeneous expressions are considered as the essential to maintain the pluripotency of the pluripotent stem cells (ES/iPS cells), but the mechanisms of gene regulation remains unknown. Recent studies using fluorescent in situ hybridization (FiSH) technology reported imbalanced allelic expressions of several genes that may related in the heterogeneity of pluripotent gene expressions, but the throughput was very limited. In this study, we aim to illustrate a overviews of gene regulation dynamics in each human iPS cell and have performed single cell RNA sequencing. Then we established the method for the allele specific expression (ASE) analysis and calculated the ASE of each iPS cell. We observed many of genes exhibited imbalanced and fluctuated allelic usages and this fluctuation was significantly correlated with the variations of gene expression levels. These findings imply significance of allelic gene regulations in the heterogeneous gene expressions and provide a new insight into how cells maintain their property.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P112

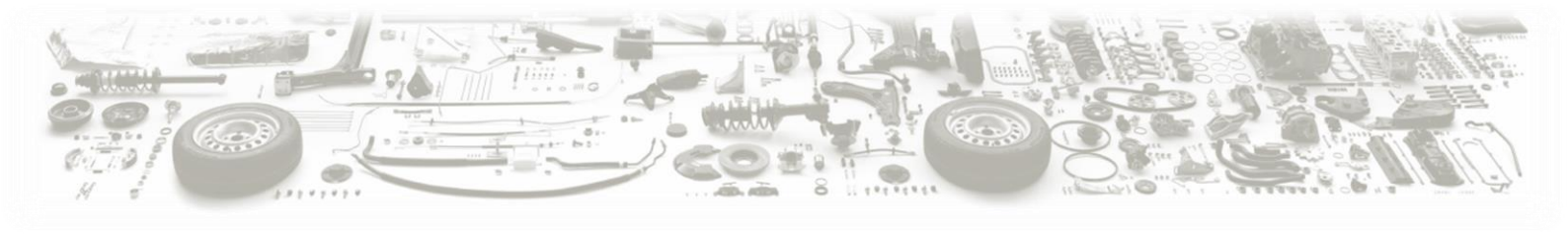
Functional and molecular characterization of human haemopoietic stem and progenitor cells

Bilyana Stoilova[1], Dimitris Karamitros[1], Zahra Aboukhalil[1], Georg Otto[1], Marina Samitsch[1], Andreas Reinisch[2], Lynn Quek[1], Nicolas Goardon[1], Ravindra Majeti[2], Paresh Vyas[1,3]

[1]*Weatherall Institute of Molecular Medicine, Oxford, United Kingdom*
[2]*Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford, CA, United States*
[3]*University of Oxford, UK, Oxford, United Kingdom*

Haemopoiesis is an excellent model to study mechanisms regulating key processes in tissue hierarchies (e.g. self-renewal, lineage commitment and terminal differentiation). Most blood cells in steady state murine haemopoiesis are generated from progenitors downstream of haemopoietic stem cells (HSC)1,2. Human early progenitor populations are poorly defined. One key progenitor we previously purified is the human lymphoid-primed multipotent progenitor (LMPP) that generates both the lymphoid and granulocyte-macrophage lineages3. Importantly, in human Acute Myeloid Leukemia (AML), leukemia-propagating stem cells are often arrested at an LMPP-like stage3.
Here we describe the functional (by in vivo transplantation and in vitro liquid culture and colony assays to assess lineage potential both qualitatively and quantitatively) and transcriptional programs (by bulk and single cell (~800 cells) RNA sequencing) of seven purified human hematopoietic stem and progenitor populations (HSPCs).
Our results show: (i) human HSPC populations are distinct at a global transcriptional level in a manner that mirrors their distinct functional potentials in vivo and in vitro (ii) single cell RNA-Seq analysis identifies transcriptional heterogeneity within a cell population, supplying tools for further purification of human HSPCs.
The data provide a cellular framework to understand molecular mechanisms regulating normal progenitor cell fate choice during haemopoiesis and identify the common mechanisms that lead to arrested differentiation at specific points in haemopoiesis that result in AML.

1 Sun, J. et al. Nature 514, 322–327 (2014)
2 Busch, K. et al. Nature 518, 542-6 (2015)
3 Goardon N. et al. Cancer Cell 19, 138–152 (2011)



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P113

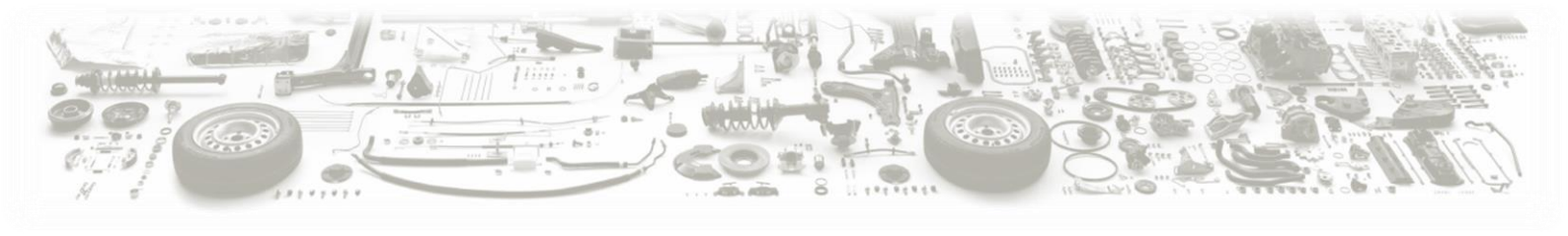
Identifying cellular origins of Barrett's oesophagus and oesophageal cancer using single-cell RNA-seq

Richard Owen[1], David Severson[1], Michael White[1], Barbara Braden[2], Adam Bailey[2], Elizabeth Bird-Lieberman[2], Mark Middleton[1], Ben Schuster-Böckler[1], Xin Lu[1]

[1]*University of Oxford, Oxford, United Kingdom*
[2]*Oxford University Hospitals, Oxford, United Kingdom*

Oesophageal adenocarcinoma (OAC) is highly fatal and increasingly common. Barrett's oesophagus (BO), characterised by a switch from squamous to intestinal-type epithelium at the gastro-oesophageal junction, is the main risk factor for OAC. The cellular origin of this switch, and how it relates to cancer-initiating cells, is not known. Whole-tissue RNA-seq of BO demonstrates an unexpectedly mixed epithelial transcriptomic signature, conflicting with current mechanistic theories of stem cell transcommitment initiating BO. Single-cell RNA-seq is thus necessary to characterise transcriptomic signatures in BO to identify cell-of-origin candidates and enable prediction of malignant progression.
Samples were obtained from BO, with controls from the duodenum, stomach and normal oesophagus. Viable single epithelial cells were isolated using fluorescence-activated cell sorting and cDNA prepared using smart-seq2. Single cells and matched whole tissue from the same biopsy underwent RNA-seq. Data were validated by correlating pooled cell and whole-tissue gene expression. Cell types were identified by transcriptomic profiling using hierarchical clustering methods and differential expression analysis.
Correlation between pooled cell and whole-tissue gene expression was strong. Both hierarchical clustering and principal component analysis grouped normal cells into expected origin tissue types. High cellular heterogeneity was observed in pathological tissues, consistent with known and unknown cellular sub-types. A unique gene expression signature was identified in BO compared to control epithelial types.
Further analysis and validation of our data will provide insight to the phenotype defining transcriptional features of the upper gastrointestinal tract and how transcriptionally dysregulated cell-subtypes link to malignancy.

Funded by Oxford BRC, CRUK and Ludwig Cancer Research.



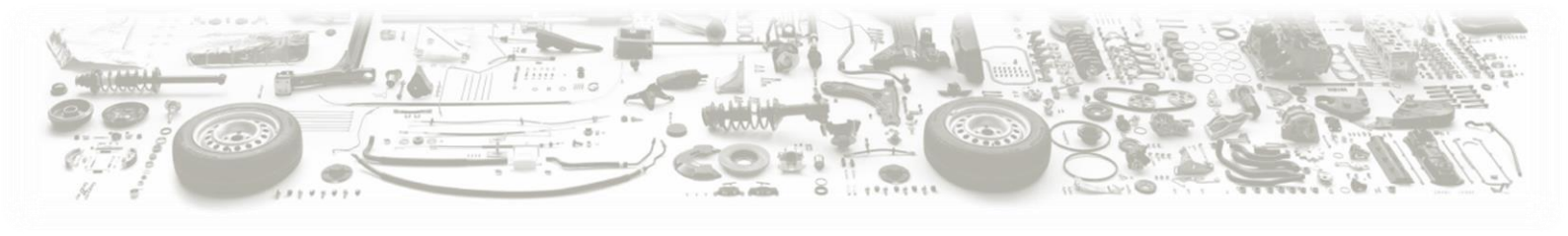
P114

Heterogeneity comparison with scRNA-Seq of human embryonic stem cells in two different states

Aurora Savino, Tamir Chandra, Ferdinand von Meyenn, Wolf Reik

Babraham Institute, Cambridge, United Kingdom

Embryonic Stem Cells (ESCs) have the ability to give rise to all cell lineages present in an organism. Despite being a fundamental concept of developmental biology the way this plasticity is achieved is not fully understood. Studies on mouse ESCs revealed that they are transcriptionally heterogeneous and that transcriptional differences are reflected in diverse differentiation potentials.
Here, we seek to assess and dissect heterogeneity in human ESCs. For this, we collected single-cell RNA-seq data on hundreds of hESCs in two conditions: the classical “primed” state, and the recently described KLF2 and NANOG induced "naive" state. These data allowed us to compare the variability of the two conditions, revealing the higher transcriptional homogeneity of the naive state, and to retrieve subpopulations based on gene expression levels, especially evident in the primed state. Moreover, we observed particularly high expression variability of endogenous retroviruses, especially ERV1 elements that have been linked with pluripotency and have been found to be upregulated during embryonic genome activation of human embryo development.
We expect that the comparison with publicly available data on hESCs will allow us to uncover the relationship between expression variability and epigenetic features, such as DNA methylation, chromatin state and histone modifications. Importantly, the parallel study of “primed” and “naïve” condition will allow the characterization of differences in gene network properties between the two states and therefore find important regulators of the ground state of pluripotency.



P115

Single cell transcriptomes from mouse hippocampus and cerebellum as strategic data in the Human Brain Project

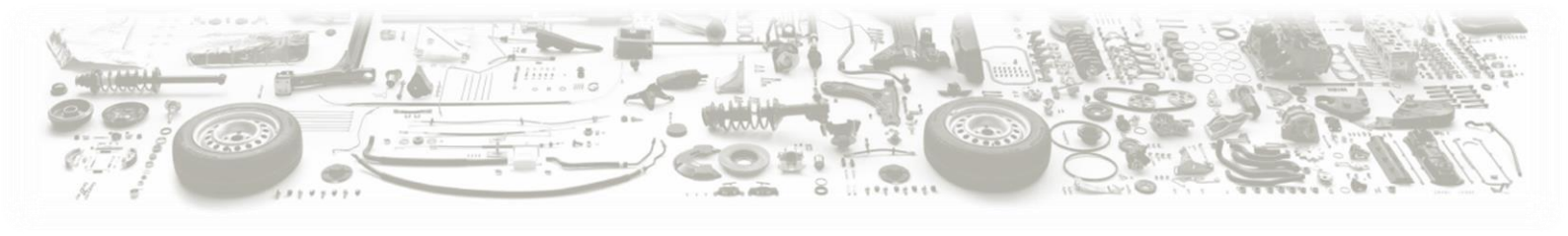
T. Grant Belgard, Juan F. Montiel, Peter L. Oliver, Chris P. Ponting

University of Oxford, Oxford, United Kingdom

The Human Brain Project aims to generate single cell transcriptomes in order to classify neurons on the basis of their transcriptomes, map these classifications back to known types on the basis of their gene expression patterns, and then provide an other-than-uniform prior on the relative proportions of key physiological molecules across neuronal types for modeling.
Enzymatically dissociated cells were processed using an adapted version of Smart-seq2 (Picelli et al. 2013 & 2014) on the Fluidigm C1 platform. We generated full-length transcriptomes from single cells of the young adult mouse hippocampus and cerebellum. We used ERCC RNA spike-ins to assess technical noise and implemented a QC pipeline to evaluate sample quality and cell identity.
Different size- and dissection-based cellular isolation methods gave dramatic differences in the proportion of cells that were neurons (ranging from less than 10% to 50%) allowing us to refine cellular enrichment as required. These full-length transcriptome data in hippocampus and cerebellum complement a recent digital gene expression dataset using STRT in mouse cortex (Zeisel et al.). We have additionally piloted these methods in human with freshly resected human cerebral cortex.
To the best of our knowledge this is the first time that full length transcripts have been derived from multiple brain regions of mouse and human with a standardized technology.

References:
Picelli et al. Nature Methods 10, 1096-1098 (2013)
Picelli et al. Nature Protocols 9, 171-181 (2014)
Zeisel et al. Science 10.1126/science.aaa1934 (2015)

Acknowledgements:
The Human Brain Project is funded by the European Commission.



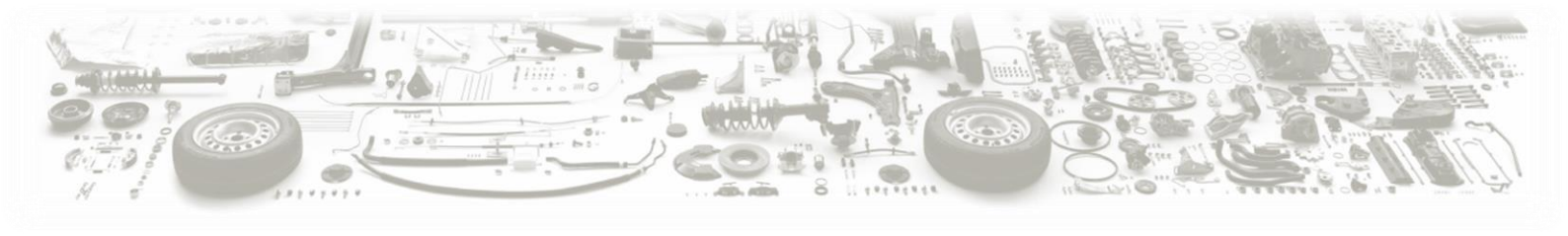
P116

Assessment of miRNA functions in steady-state conditions

Marcel Sauerbier[1], João Frade[2], Silvina Catuara[2], Luis Zapata Ortiz[2], Diego Barcena Menéndez[2], Maria Méndez-Lago[2], Marc Riemer Friedländer[1]

[1]*Stockholm University, Stockholm, Sweden*
[2]*Centre for Genomic Regulation, Barcelona, Spain*

microRNAs (miRNAs) are ~22nt long regulatory non-coding RNAs that can each target hundreds of mRNAs. Some strongly down-regulate their targets during dynamic processes such as development. However, the functions of miRNAs in steady-state conditions are not well understood and many deeply conserved miRNAs only have subtle effects on their target transcripts. It has been suggested that the function of these miRNAs is to buffer expression oscillations of their target transcripts. However, if these oscillations are not synchronized, they risk being averaged out when pools of cells are profiled. Therefore, to assess these miRNA functions we will apply transcriptomics to measure expression in hundreds of individual wild-type cells and mutant cells that are void of miRNAs.
In our study it is crucial to reduce sources of technical noise as they could mask transcriptional variance. We have applied tamoxifen treatment and clonal expansion to conditional knock-out cells to ablate Drosha, an essential factor in miRNA biogenesis. Hence we know that our wild-type and mutant cells are genetically identical, except for the deleted gene. Further, we will purify cells in G2/M to reduce cell cycle related transcript variation. Last, both Smart-Seq2 and MARS-Seq technology will be applied to address platform specific errors and biases, and spike-ins will facilitate estimation of technical and biological noise.
As the profiled transcripts will be miRNA targets and non-targets it will be possible to directly compare these groups. Argonaute CLIP-seq data and sequence analysis will be integrated to obtain high-confidence classification of miRNA targets and non-targets.



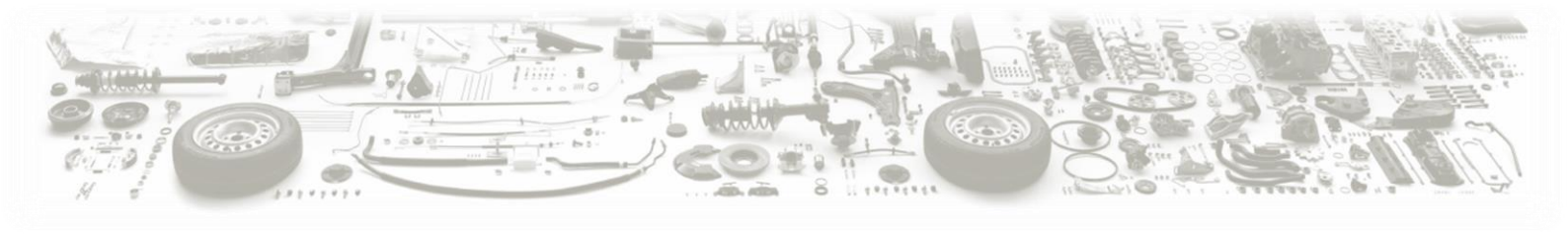
P118

Single-cell DNA methylome sequencing and bioinformatic inference of epigenetic cell state dynamics

Matthias Farlik, Nathan C. Sheffield, Angelo Nuzzo, Paul Datlinger, Andreas Schönegger, Johanna Klughammer, Christoph Bock

Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

Single-cell genome and transcriptome sequencing are broadly contributing to our understanding of cellular heterogeneity, whereas methods for single-cell epigenomics are still in their infancy. Here we describe a whole genome bisulfite sequencing assay that enables DNA methylation mapping in very small cell populations (µWGBS) and single cells (scWGBS). Our assay is optimized for profiling many samples at low coverage, and we developed a bioinformatic method that analyzes collections of single-cell methylomes to infer cell state dynamics. Using these technological advances, we studied epigenetic cell state dynamics in three in vitro models of cellular differentiation and pluripotency, observing characteristic patterns of epigenome remodeling and cell-to-cell heterogeneity. The described method enables single-cell analysis of DNA methylation in a broad range of biological systems (for example in embryonic development and stem cell populations), and it can be used to establish composite methylomes that account for cell-to-cell heterogeneity in complex tissues and tumors.



P119

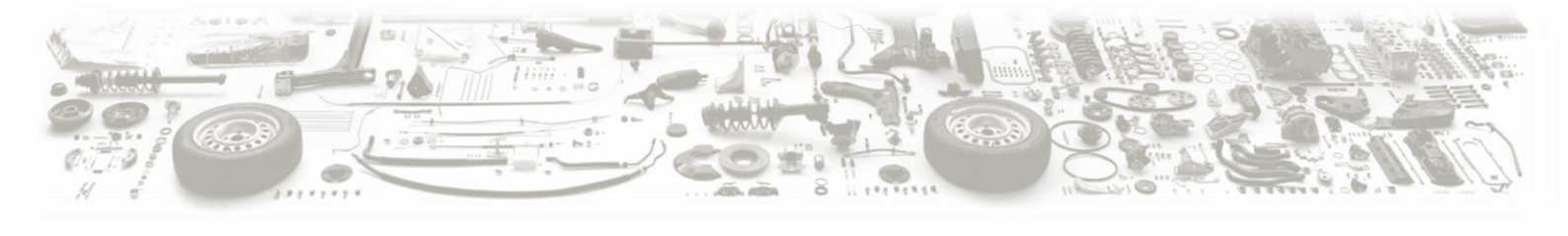
Simultaneous analysis of targeted genes and transcripts of thousands of single tumor cells in parallel

David Redin, Sanja Vickovic, Erik Borgström, Afshin Ahmadian

Science for Life Laboratory, Solna, Stockholm, Sweden

Exploring the relationship between genomic and transcriptomic states of an individual cell, at the same time point, provides the most definitive way of understanding how genetic variation influences gene expression. We have developed a targeted assay(1) where genomic sequences and their corresponding transcripts can be analyzed simultaneously for thousands of single cells in parallel, enabling relative quantification in heterogeneous samples. We use this method to identify and quantify circulating tumor cells amongst a high abundance of normal cells, and to probe the relationship between genotype and phenotype in these cells.

1. Borgström, E. et al. Nat. Commun. 6:7173 doi: 10.1038/ncomms8173 (2015).



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

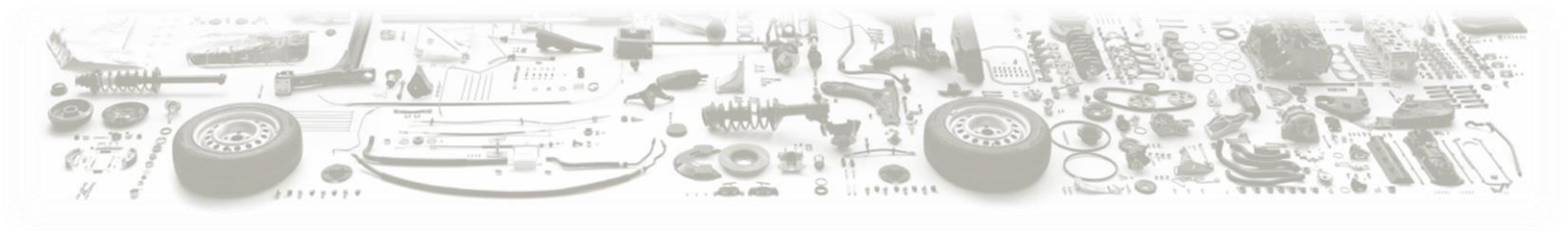
P120

Identification of novel regulators of Th17 cell pathogenicity by single-cell genomics

Nir Yosef[1], Jellert Gaublomme[2], Youjin Lee[3], Rona Gertner[4], Li Yang[5], Pier Paolo Pandolfi[3], Tak Mak[6], Rahul Satija[7], Alex K. Shalek[8], Vijay K. Kuchroo[3], Aviv Regev[2], Hongkun Park[4]

[1]*UC Berkeley, Berkeley, CA, United States*
[2]*Broad Institute, Boston, MA, United States*
[3]*Harvard Medical School, Boston, MA, United States*
[4]*Harvard, Boston, MA, United States*
[5]*East Carolina University, Greenville, NC, United States*
[6]*Princess Margaret Hospital, Toronto, ON, Canada*
[7]*NYU, New York City, NY, United States*
[8]*MIT, Boston, MA, United States*

IL17-producing T helper cells (Th17) fulfill critical functions in the immune system, such as clearing fungal and extracellular infections, but can also induce autoimmune diseases. Mirroring, this duality, researchers have predicted that Th17 cells can exist in several states, some more pathogenic than others, and hence more prone to induce autoimmune disease. However, it has been difficult to determine the specific aspects of the Th17 circuit that control pathogenicity, because pathogenic cells are mixed with – and hence largely obscured by – non-pathogenic ones.
Here, we use single-cell RNA-seq to investigate the molecular mechanisms governing heterogeneity and pathogenicity of murine Th17 cells isolated from the central nervous system (CNS) and lymph nodes (LN) at the peak of autoimmune encephalomyelitis or polarized in vitro. We have developed a number of computational tools to analyze the ensuing data, including a normalization scheme that accounts for “nuisance” factors that affect cellular variation (e.g., differences in detection efficiency), and an annotation scheme to interpret the observed variation using legacy signatures derived from bulk-population genomic data.
Our analysis reveals that Th17 cells span a spectrum of states in vivo, including a self-renewal state in the LN, and Th1-like effector/memory states in the CNS. Relating these states to in vitro differentiated Th17 cells, we discovered novel genes governing pathogenicity and disease susceptibility, and validated the crucial role in Th17 cell pathogenicity of four of these genes using knockout mice. Cellular heterogeneity can thus be leveraged for selective suppression of pathogenic Th17 cells while sparing non-pathogenic tissue-protective ones.



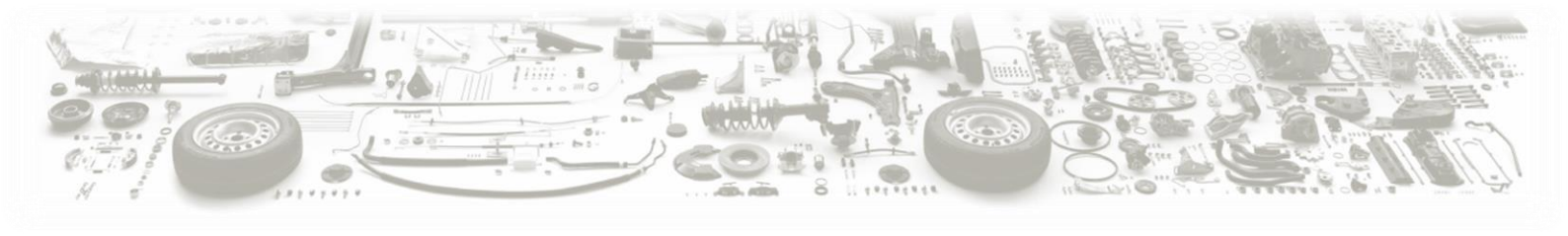
P121

Single-cell RNAseq of the three-layered cerebral cortex of turtles and lizards

Maria Antonietta Tosches, Tracy M Yamawaki, Georgi Tushev, Gilles Laurent

Max Planck Institute for Brain Research, Frankfurt am Main, Germany

The cerebral cortex, implicated in higher cognitive functions in mammals, remains a poorly understood brain area despite seemingly invariant structural, and possibly computational, features. To facilitate the study of cortical function we have turned to a simpler model, the reptilian cortex. Reptiles and mammals evolved from a common amniote ancestor in which cortex comprised only three layers. This organization has been retained throughout the cortex of extant reptiles. In contrast, in mammals the three-layered structure is only present in some areas, the “paleo-“ and “archi-cortices”, while the “neocortex” contains an elaborated six-layer structure. By studying reptilian cortical circuits, we hope to shed light on the original, and possibly fundamental, design principles of cerebral cortex.
Reptilian cortex is also thought to contain fewer cell types than found in mammalian cortex, based on morphological and histochemical studies. However, the true extent of cortical cell-type diversity in reptiles is unknown. We have chosen two reptilian species representing key phylogenetic branches in the reptilian tree: a turtle (Trachemys scripta) and a lizard (Pogona vitticeps).
In order to identify the building blocks of these cortical circuits, we are using single-cell RNAseq. In contrast to prior approaches, single-cell sequencing allows the unbiased de novo discovery of cell types. It is therefore ideal for characterizing cell types in non-conventional animal models. These data will provide a molecular entry point for further functional characterization of these simple cortical models.



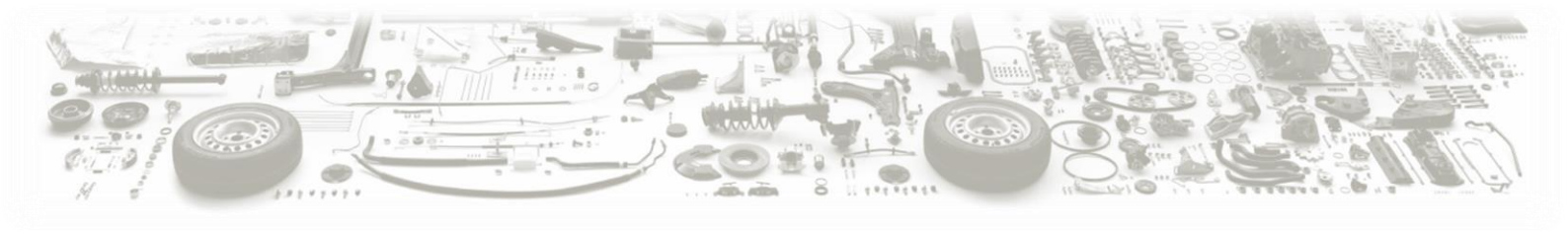
P122

The spatial mapping of single cells reveals new gene expression patterns in the developing brain of Platynereis dumerilii

Nils Eling[1], Kaia Achim[2], Detlev Arendt[2], John Marioni[1]

[1]*European Bioinformatics Institute, Cambridge, United Kingdom*
[2]*European Molecular Biology Laboratory, Heidelberg, Germany*
[2]*European Molecular Biology Laboratory, Heidelberg, Germany*
[1]*European Bioinformatics Institute, Cambridge, United Kingdom*

Single-cell RNA-sequencing (scRNAseq) led to the discovery of novel cell types and cell functions within heterogeneous cell populations. One major drawback of scRNAseq experiments is the loss of spatial information due to tissue disruption during single cell collection therefore restricting the understanding of cellular behaviour. Emerging methods for spatial mapping of single cells based on marker gene expression have opened up the field of transcriptome-wide expression profiling across tissues and organs. Previously, cells were mapped into the developing brain of the annelid Platynereis dumerilii by comparing the cells' mRNA profiles with spatial gene expression profiles. Integrating the spatial position of the cells with their transcriptome-wide expression profiles enables the identification of novel cell types, subregions and expression patterns in the annelid's brain at 48 hours post fertilization (hpf). After quality control and centre correction of the mapping results, expression profiles of 68 correctly mapped cells reveal locally restricted gene expression patterns potentially connected to the anterior-posterior patterning of the developing brain. In parallel, hierarchical clustering of 385 cells taken from the whole embryo at 48hpf show groups of cells that map locally to the annelid's brain. Pax6+/Wnt8+ cells mapped to the annelid's developing mushroom body appear to be proliferative without the expression of additional marker genes while cholinergic photoreceptor neurons map to the anterior part of the brain. Interestingly, a group of head neurons located in the anterior part of the brain cluster closely with muscle cells indicating a developmental connection between the embryo's head and trunk.



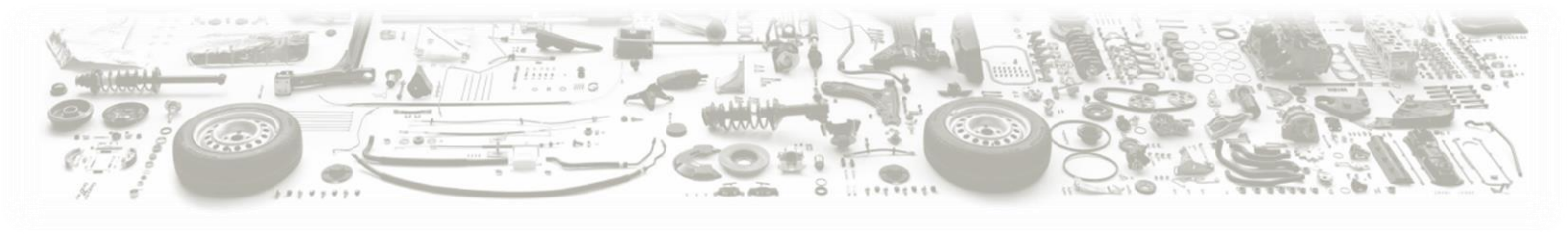
P123

Impact of Growth Dependent Variability in Gene Expression on Cellular Phenotypes

Malika Saint[1], Lorenzo Ficorella[1], Vahid Shahrezaei[1], Samuel Marguerat[1]

[1]*MRC CSC, London, United Kingdom*
[2]*Imperial College London, London, United Kingdom*

Genetically identical cells grown in uniform conditions exhibit heterogeneous phenotypes. This is because gene expression is not only a consequence of genetic and environmental factors but also of complex features such as the cell metabolic status or cellular growth rate. In addition, the stochastic nature of the biochemical reactions that underlie gene expression is per se a source of cell-to-cell variability and phenotypic heterogeneity. The full extent of phenotypic and metabolic heterogeneity in microorganism populations and its dependence on growth conditions is far from being understood. To explore this question we use high-dimensional phenotyping by RNA-seq of single fission yeast cells (Schizosaccharomyces pombe) grown asynchronously in uniform conditions. mRNA libraries from single cells were constructed incorporating unique molecular barcodes to individually tag RNA molecules and correct for amplification bias. In addition, cell specific barcodes introduced during reverse transcription allowed to sequence large numbers of cells in pools reducing processing costs. Our analysis revealed a series of distinct expression profiles in cells growing at steady state in uniform conditions. Further co-variance analysis was used to inform on regulatory networks underlying phenotypic heterogeneity. Finally we integrate simple imaging data from each cell with its respective transcriptomic measurements in an attempt to connect cellular morphological features with gene expression. Eventually, this analysis will help unravelling the interplay between growth rates and phenotypic heterogeneity in a simple eukaryote.



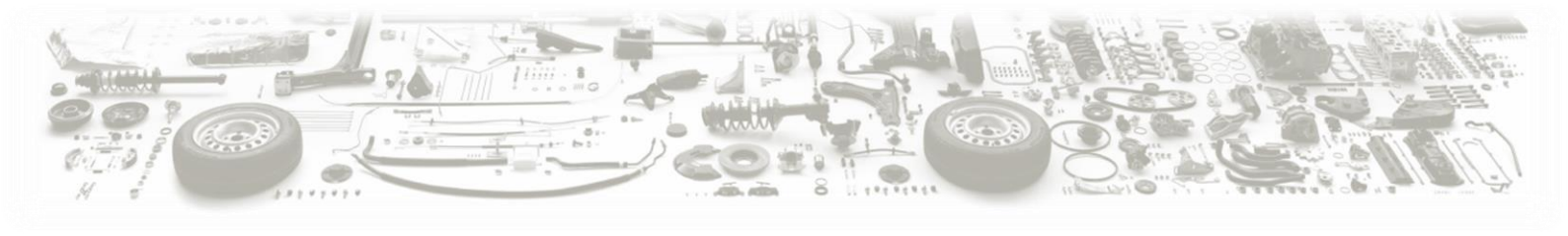
P124

No evidence for increased aneuploidy in the brain of Alzheimer's patients

Hilda van den Bos, Diana Spierings, Nancy Halsema, Inge Kazemier, Karina Hoekstra-Wakker, Aaron Taudt, David Porubský, Maria Colomé-Tatché, Erik Boddeke, Peter Lansdorp

University Medical Center Groningen, Groningen, Netherlands

Recent studies have suggested that the human brain might contain aneuploid cells, which have lost or gained one or more chromosomes, and that such chromosomal copy number variations could contribute to neurological disorders. High rates of aneuploidy were found in the developing human brain (around 33% of cells), with lower rates in the adult brain using fluorescence in situ hybridization (FISH). Interestingly, the brains of patients with Alzheimer's disease (AD) were reported to contain a higher number of aneuploid cells using FISH. Because the reported rate of aneuploidy in adult brains ranged widely (from 0% to 40%) it remains to be determined whether aneuploidy is a normal feature of the human brain and whether (increased) aneuploidy is involved in neurological disorders. In the current study we used single cell whole genome sequencing (WGS) to assess aneuploidy in isolated single neurons from the frontal cortex of individuals with and without AD (Braak stage IV and I respectively). In contrast to the previous FISH methods our single cell sequencing approach allows determination of the karyotype of individual, non-dividing cells. Nuclei were isolated from frozen tissue sections and single NeuN-positive nuclei were sorted by fluorescence activated cell sorting. Importantly, WGS libraries for sequencing on an Illumina platform were prepared without upfront whole genome amplification. Applying our method to single NeuN-positive nuclei of a Down syndrome sample confirmed the presence of 3 copies of chromosome 21 in all nuclei analyzed, demonstrating the reliability and accuracy of our method. Preliminary analysis revealed no difference in aneuploidy rate between neurons from AD patients and normal controls .



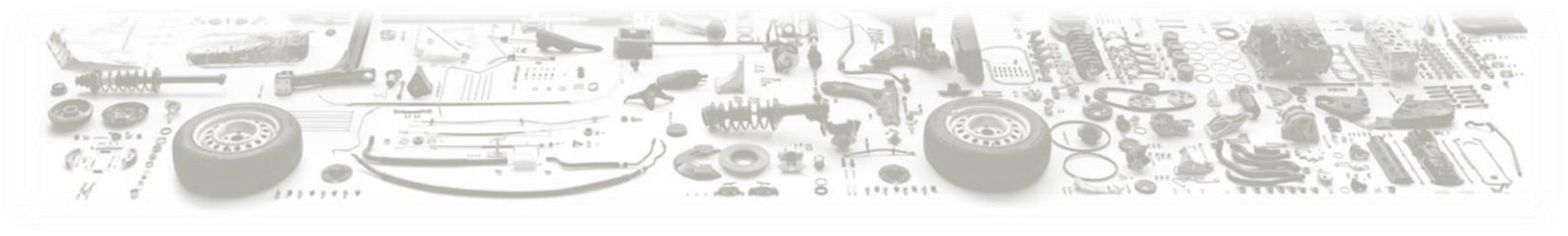
P126

G&T-seq: Separation and parallel sequencing of the genomes and transcriptomes of single cells

Iain Macaulay[1], Wilfried Haerty[2], Parveen Kurmar[3], Yang Li[2], Tim Xiaoming Hu[2], Mabel Joey Teng[1], Mubeen Goolam[4], Magdalena Zernicka-Goetz[4], Chris Ponting[2], Thierry Voet[3]

[1]*Wellcome Trust Sanger Institute, Cambridge, United Kingdom*
[2]*University of Oxford, Oxford, United Kingdom*
[3]*KU Leuven, Leuven, Belgium*
[4]*University of Cambridge, Cambridge, United Kingdom*

Advances in genome or transcriptome sequencing from many single cells are offering a unique perspective from which to investigate cellular heterogeneity in development and disease. Here we present G&T-seq which permits simultaneous sequencing of the genome and the transcriptome from the same single cell.
G&T-seq provides whole genome amplified genomic DNA and full-length transcript sequence and with automation, 96 samples can be processed in parallel.
Using cancer cell lines and other models, we confirmed a relationship between DNA copy number and gene expression dosage at the single cell level. From single cells from the breast cancer cell line HCC38 and matched normal control cells, several thousand transcripts were detected per cell, while low coverage genome sequencing demonstrated that copy number variants observed in bulk were preserved in the genomic analysis of single cells. Furthermore, integrated analysis of the genome and transcriptome of single cells also allows validation of genomic single nucleotide variants in transcripts, as well as the detection of fusion transcripts and their associated genomic rearrangements.
In addition, we present further applications of the method in primary cell models of development and disease.



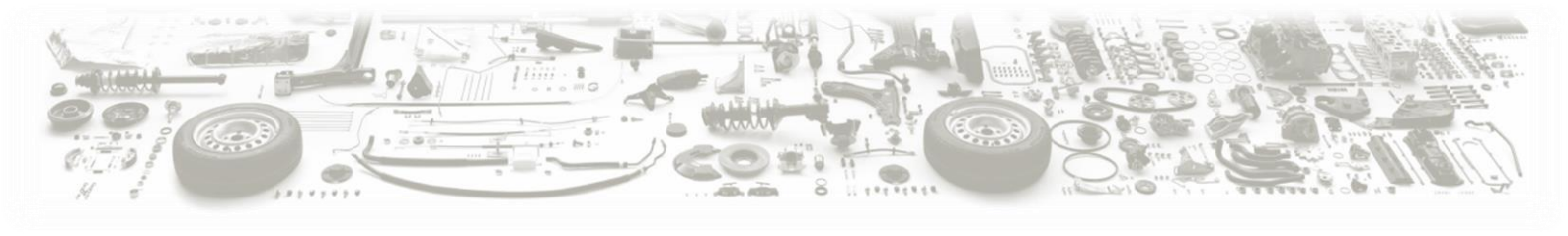
P127

Profiling of DNA-replication by single-cell sequencing

Parveen Kumar, Niels Van der Aa, Masoud Zamani Esteki, Ligia Mateiu, Ryo Sakai, Koen Theunis, Jan Aerts, Thierry Voet

University of Leuven, Leuven, Belgium

Single-cell genome sequencing is paramount for understanding genetic heterogeneity in normal biological and disease processes, and for enabling genome-wide screens of rare cells in the clinic. Often, the individual cells are isolated randomly regardless of their cell cycle state. Here we establish metrics for identifying cells in S-phase within a population of whole-genome sequenced single lymphoblastoid cells using analyses of sequence read depths. We show that DNA-replication perturbs conventional copy number analyses, leading to the detection of DNA losses and gains that may be falsely interpreted as genuine genetic variation in the cell. In contrast, by developing new sequence analyses, we were able to reveal the DNA-replication program ongoing within a single cell at high resolution, enabling novel routes to investigate cell-to-cell variation in replication dynamics genome wide. The established workflow is of value to single-cell analyses in general and DNA replication research in particular.



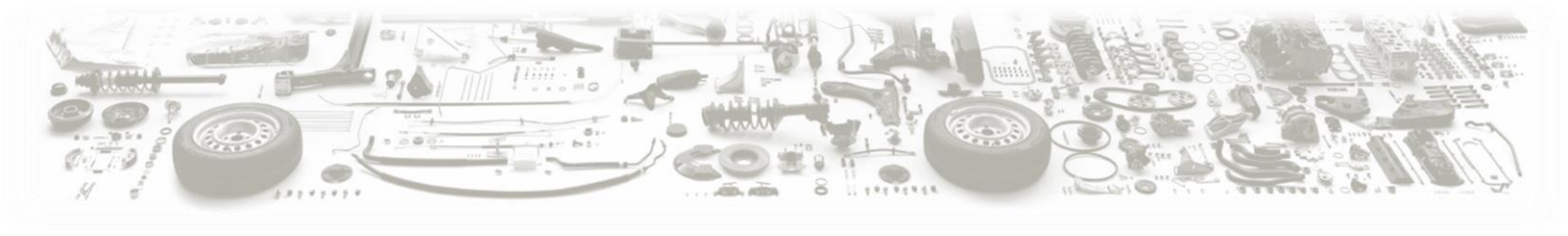
P128

Leapfrogging genome assemblies by sequencing single cells

Mark Hills[1], Ester Falconer[1], Kieran O'Neill[1], Ryan Brinkman[1], Peter Lansdorp[2]

[1]*BC Cancer Agency, Vancouver, BC, Canada*
[2]*European Research Institute for the Biology of Ageing, Groningen, Netherlands*

The mouse and human genome assemblies have revolutionized the way we conduct molecular research, rapidly expanding our understanding of the biology of these organisms and the diseases associated with them. These references are of high quality but still contain fragments which have not been localized to specific genomic regions and large regions that are incorrectly orientated with respect to the rest of the chromosome. Using Strand-seq, a single cell next-generation sequencing technique, we identified errors in the latest assemblies of mouse and humans with the goal of improving and refining these genomes. Strand-seq involves only sequencing the parental template strands in single cells, preserving the directionality of DNA. Across multiple single cells, orientation errors were readily detected as changes in the strand directionality across a chromosome, while unmapped fragments were mapped to specific regions as they share the same strand directionality as the region where they located. Applying this further, we have expanded this method to rapidly and inexpensively assemble genomes from a variety of model organisms at varying stages of completion, from early-draft builds consisting of thousands of fragments, to more complete assemblies. We have created improved assemblies for zebrafish, rat and pig, correcting orientation errors which encompassed up to 30 % of the genome sequence in some references. In addition, for reference genomes at the scaffold-stage (which consist of tens of thousands of unmapped fragments), we have clustered these fragments into individual chromosomes, and into a relative order, vastly improving the genome assemblies for the Tasmanian devil, guinea pig, ferret and Xenopus. We demonstrate how this strategy allows us to leapfrog intermediate builds to rapidly and efficiently create better reference genome assemblies.



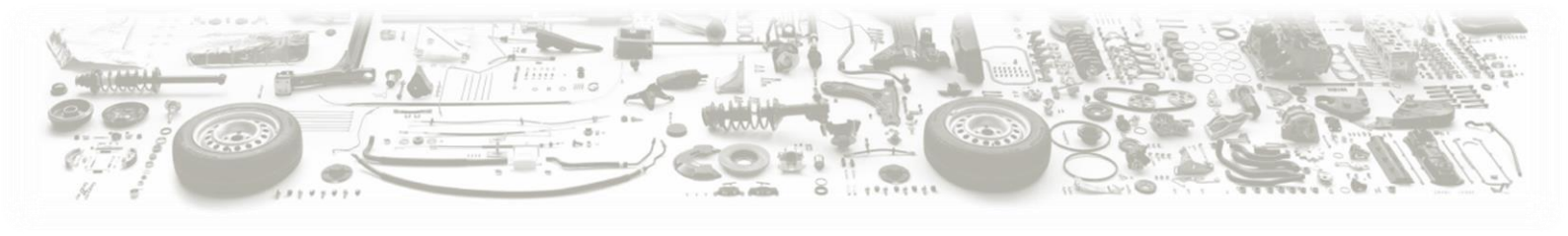
P129

Dissecting the transcriptional heterogeneity of murine epidermis at single cell resolution

Simon Joost, Amit Zeisel, Tina Jacob, Xiaoyan Sun, Gioele La Manno, Peter Lönnerberg, Sten Linnarsson, Maria Kasper

Karolinska Institutet, Stockholm, Sweden

The murine epidermis with its hair follicles (HFs) represents an invaluable in vivo model system for tissue regeneration and stem cell research. Intriguingly, many of the epithelial skin functions are carried out by what is commonly considered one cell type: the keratinocyte. Here we used single-cell RNA sequencing to reveal how heterogeneity and plasticity of keratinocytes is tuned at the transcriptional level. We analysed the transcriptomes of 1423 cells randomly picked from the murine telogen epidermis. Unsupervised clustering revealed 13 main populations of interfollicular and follicular cells and provides novel markers for previously known and newly defined populations. In-depth analysis further uncovered several distinct HF-subpopulations, including six bulge and six upper HF populations, and one population lining the HF opening. Reconstruction of the differentiation process at single cell resolution showed that epidermal stratification is governed by two transcriptional waves and suggests novel regulatory factors involved in this process. Finally, our data propose that most transcriptional heterogeneity in the murine telogen epidermis can be explained along two dimensions: one linked to a universal differentiation program, and one specific to local niches. This study provides the first unbiased and systematic view of transcriptional organization of an adult epithelial tissue and highlights how cellular heterogeneity of one cell type can be orchestrated in vivo to assure tissue homeostasis.



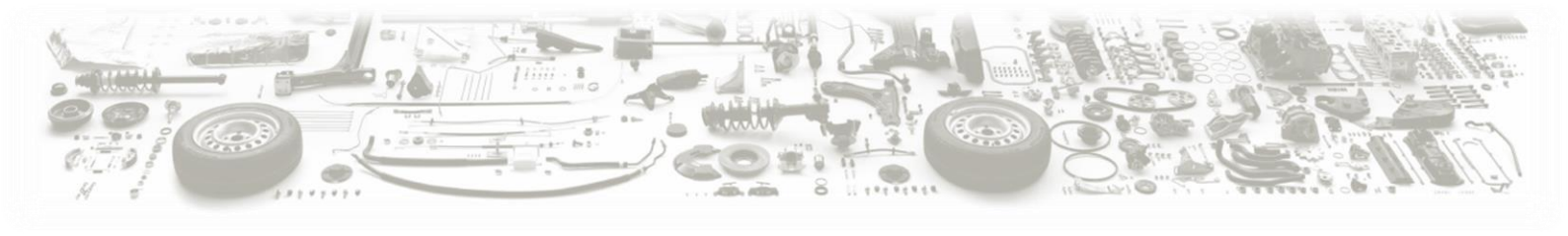
P130

Dynamic Cellular Heterogeneity Resulting from Random Monoallelic Gene Expression in Somatic Cells

Björn Reinius[1], Jeff E Mold[2], Daniel Ramsköld[2], Qiaolin Deng[1], Jakob Michaëlsson[2], Jonas Frisén[2], Rickard Sandberg[1]

[1]*Ludwig Institute for Cancer Research, Stockholm, Sweden*
[2]*Karolinska Institutet, Stockholm, Sweden*

Cellular heterogeneity can emerge from the expression of only one parental allele. The nature of random monoallelic expression (RME) of autosomal genes in somatic cells is however largely undetermined. Here, we applied allele-sensitive single-cell RNA-seq on individual clonal mouse fibroblasts and in vivo expanded human T lymphocytes to concurrently distinguish clonally inherited (fixed) and cell-to-cell variable (dynamic) RME in somatic cells. Fixed autosomal RME occurred only at low levels (~0.5% of genes) in fibroblasts and in T cells, whereas dynamic RME was highly abundant (dozens of %). Furthermore, we determined the correlation of dynamic RME with cell size and cellular RNA content. In conclusion, autosomal RME is generally dynamic rather than fixed, and phenotypic consequences thereof will occur transiently in scattered cells rather than in stable and spatially restricted patterns of clonally related cells.



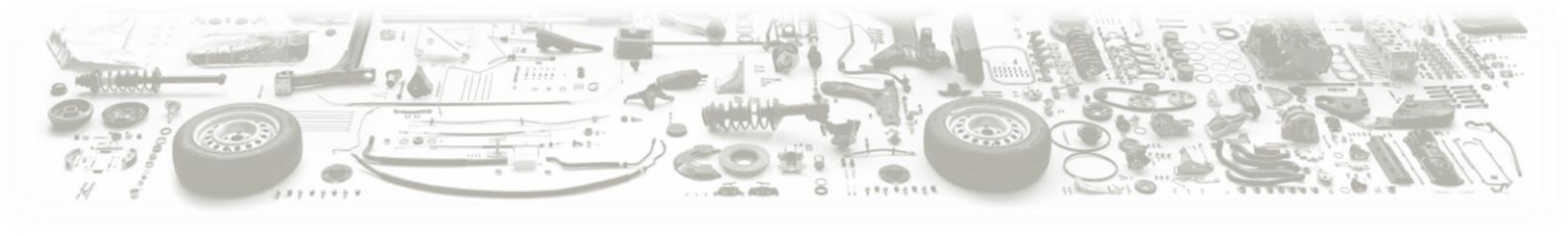
P131

Phasing of single DNA molecules by massively parallel barcoding

Erik Borgström

Science for life laboratory, Solna, Sweden

High-throughput sequencing platforms mainly produce short-read data, resulting in a loss of phasing information for many of the genetic variants analysed. For certain applications, it is vital to know which variant alleles are connected to each individual DNA molecule. We developed a method for massively parallel barcoding and phasing of single DNA molecules. First, a primer library with millions of uniquely barcoded beads is generated. When compartmentalized with single DNA molecules, the beads can be used to amplify and tag any target sequences of interest, enabling coupling of the biological information from multiple loci. We recently applied the assay to bacterial 16S sequencing and up to 94% of the hypothesized phasing events were shown to originate from single molecules. The method enables use of widely available short-read-sequencing platforms to study long single molecules within a complex sample, without losing phase information.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

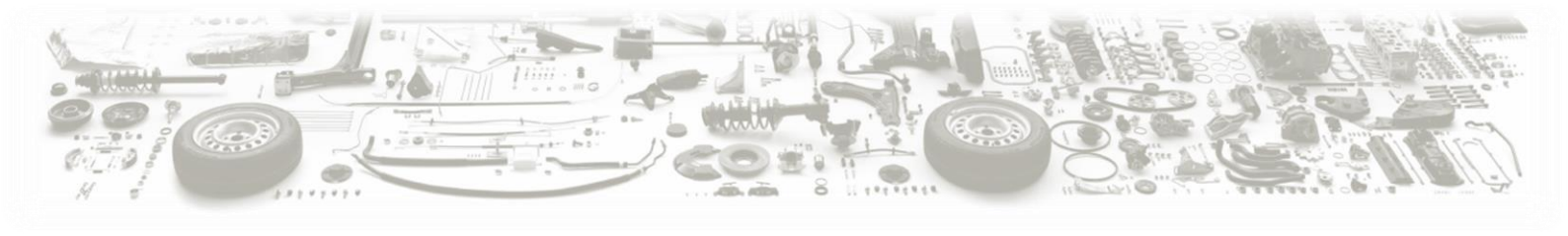
P133

Linking T-cell receptor sequence to single cell transcriptome analysis using short paired-end RNA-sequencing

Shaked Afik[1], Kathleen Yates[2], Jernej Godec[2], Michael Cole[1], Arlene Sharpe[2], Nicholas Haining[2], Nir Yosef[1]

[1]*University of California, Berkeley, CA, United States*
[2]*Harvard Medical School, Boston, MA, United States*

The ability to link antigen specificity of T-cell receptors (TCR) to single cell transcriptome data is a major technical challenge, as the hypervariability of the sequences that encode the antigen recognition region makes TCR identification using standard transcriptome analysis a challenging task. Current methods for TCR identification in single cells allow expression analysis of only a limited set of genes. Single cell RNA-sequencing is a promising approach that can be easily applied using long read length, which is exceedingly costly.
We present “TCR Reconstruction Algorithm for Paired-End Single cells” (TRAPES), a fast and efficient algorithm to reconstruct TCR sequences in single cells using short (~25bp) paired-end reads. By combining principles of de novo assembly with the genomic structure of TCRs, TRAPES reconstructs the TCRs using an iterative dynamic programing algorithm. Moreover, TRAPES annotates the antigen recognition region and produces a clear visualization and statistical analysis of the reconstructed sequence. We applied our algorithm to different T-cell populations in human and mouse, using both tetramer sorted and naïve T-cell populations. TRAPES successfully reconstructs the TCRs, observing only few clones in tetramer sorted populations and many distinct TCRs in the naïve population. We validated our results by sequencing the same cells using 150bp reads, achieving similar success rates. Applying TRAPES to sequenced cells enables us to combine the study of T-cell repertoires along with functional analysis of the cell’s transcriptome. Using TRAPES would result in gaining a greater understanding of cell-to-cell variability within and between clones and naïve T-cells



P134

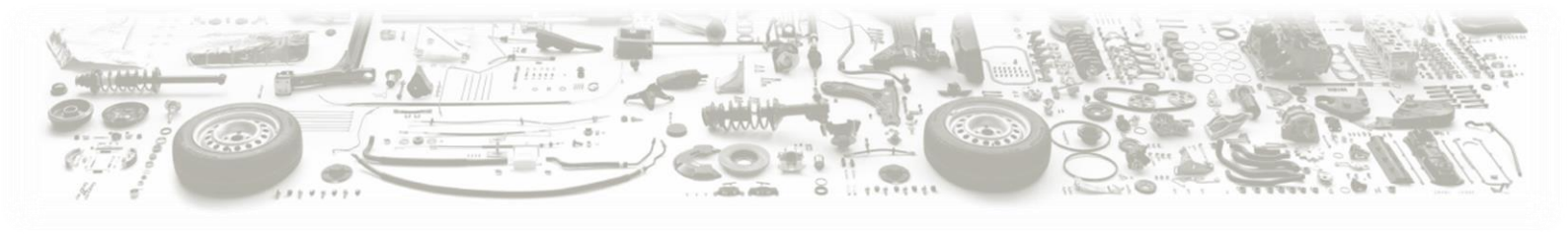
Achieving unparalleled sensitivity and reproducibility in single-cell transcriptomics

Magnolia Bostick, Cynthia Chang, Andrew Farmer

Clontech Laboratories, Mountain View, CA, United States

Gaining the ability to identify and quantify the mRNA from a single cell has been a substantial benefit to many scientific fields, especially cancer research, development, neurobiology, and immunology (where homogeneous populations are elusive). Unfortunately, the small amount of mRNA present in each cell and the large biological variation across cells provide a significant challenge: to make this technology progressively more sensitive, reproducible, and high throughput. With each advance in technology, lower-expressed genes can be identified with more confidence and differential expression can be more accurately ascertained.
Ideally, researchers would like to analyze hundreds or thousands of individual cells for any given condition. Unfortunately, many of the protocols for transcriptome library production from single cells are time consuming and not readily amenable to scaling experiment size. By modifying adapters, adding cellular and molecular indexes, and pooling, we have developed a simplified protocol that allows for parallel library production for mRNA quantification in a high throughput manner. Additionally, the absolute number of molecules captured from each cell can be measured. Finally, this method also allows for more libraries to be sequenced together, decreasing the cost of discovery.
While developing this protocol and utilizing some of the modifications developed in Smart-seq2 (Picelli et al.), we improved the sensitivity of the SMART® (Switching Mechanism at the 5' end of the RNA Template) single cell technology. We have improved the number of genes identified and decreased the background present in the Smart-seq2 protocol. Additionally, technical reproducibility and accuracy have been improved in this method, giving higher confidence in differential expression analysis.

Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, et al. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10: 1096–1098.



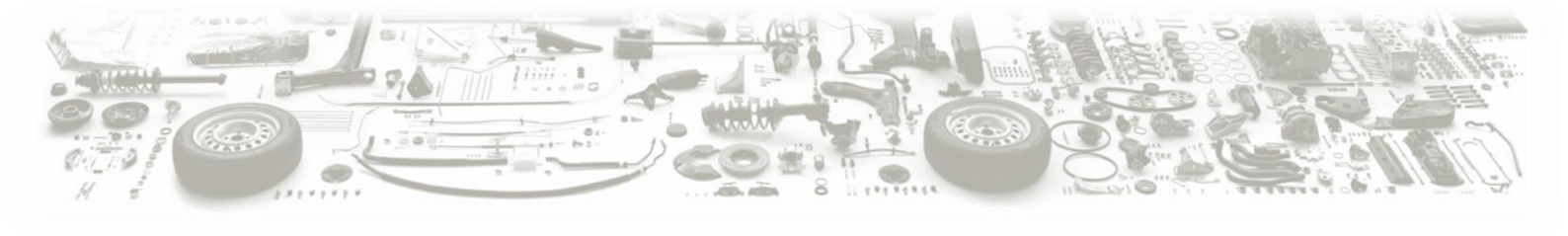
P135

Single-cell profiling of the mouse vascular system

Michael Vanlandewijck[1], Bàrbara Lavina Siemsen[2], Konstantin Gaengel[2], Marco Castro[2], Christer Betsholtz[2], Rickard Sandberg[3]

[1]*ICMC, Karolinska Institute, Huddinge, Stockholm, Sweden*
[2]*Uppsala University, Uppsala, Sweden*
[3]*Karolinska Institute, Stockholm, Sweden*

The vascular network of the human body is of paramount importance for the proper function of the entire organism, as it is responsible for organ homeostasis. It is therefore not surprising that the circulatory system plays a role in almost all diseases, ranging from minor infections to life-threatening illnesses like cancer and Alzheimer's disease. Although a vast research effort is underway to fully document and understand the vascular system, much knowledge is still lacking. Endothelial cells, the main structural component of the vessels of the vascular system, are currently studied mostly in vitro since the process of blood vessel growth is highly dynamic and very challenging to study in an in vivo setting. Recently, the Betsholtz lab has developed a new transgenic mouse line that strongly expresses GFP under control of the Claudin5 promoter, a highly specific endothelial cell marker. Not only does this mouse model allow for advanced imaging of angiogenesis in an in vivo setting, it also allows allows for fast and highly specific isolation of endothelial cells using fluorescent assisted cell sorting (FACS). In addition to this mouse model, the lab also developed another transgenic mouse model, engineered to study pericytes, the enigmatic cells lining the capillaries of the circulatory network. As there are no reliable markers known for pericytes, the mouse expresses two different fluorophores; GFP under control of the PDGFRbeta promoter, and DsRed under control of the NG2 promoter. Cells that express these two fluorophores simultaneously are almost exclusively pericytes. The aim of the project is, for the first time ever, to profile both the endothelial cells and the pericytes of every major organ in the mouse on the transcriptional level. We have chosen to use single cells RNA sequencing as the method of choice, as this technique will allow us to differentiate sub-populations of endothelial cells and pericytes between and within different organs. When an integral database of the mouse vasculature is obtained, the transcription profile of the vasculature of mouse disease models can be investigated and compared to this database, opening up new insights and possible avenues of treatment for a plethora of diseases.



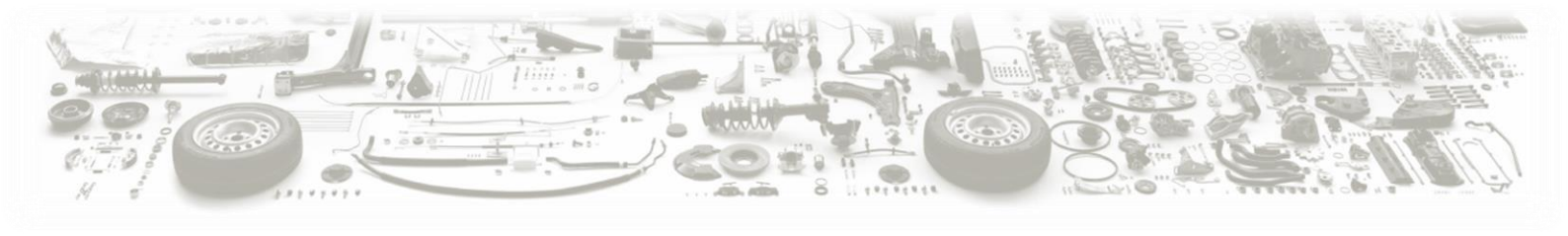
P136

Reconstructing cell lineage trees from whole-genome single-cell sequencing data

Jakub Truszkowski[1,2], Nick Goldman[1], Simon Tavare[2]

[1]*EMBL- European Bioinformatics Institute, Hinxton, United Kingdom*
[2]*Cancer Research UK, Cambridge, United Kingdom*

Single-cell sequencing provides an opportunity to survey the genomic heterogeneity of cells within an organism. Most cell divisions introduce mutations as a result of DNA replication errors. Consequently, the history of cell divisions in the organism is encoded inthe genomes of individual cells, and could be reconstructed by phylogenetic methods.
We present a method for reconstructing evolutionary histories of single cells from somatic single-nucleotide variants (SNVs) while accounting for high levels of allelic dropout often found in single-cell sequencing data. Our method models mutations and sequencing errors as two separate processes, which enables us to correctly identify the majority of allelic dropout events. Under this model, the problem of inferring the most likely evolutionary history can be reduced to finding a series of graph cuts in a certain graph. Through simulations, we show that our method outperforms standard phylogenetic methods for this task, particularly when mutation rates are low. In most cases, our method can find accurate phylogenies for data sets comprising hundreds to thousands of single cells within hours on a standard desktop computer.



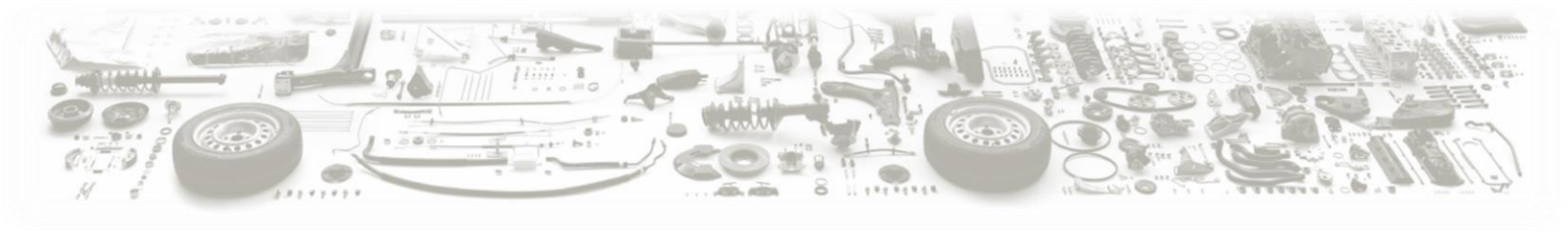
P138

Transcriptional analyses of plasmodium falciparum erythrocytic development at cellular resolution

Mtakai Ngara, Sven Sagasser, Mia Palmkvist

Karolinska Institute, Stockholm, Sweden

The erythrocytic phase of malaria infection is responsible for the disease pathology. Investigating the temporal regulation of gene expression during this phase is important in understanding the disease and for identifying potential drug and vaccines targets. Studies to date have analyzed the temporal programs through transcriptional profiling of a large number of synchronized infected cells. Although single-cell RNA-seq on human cells can reveal heterogeneity and provide an understanding of phenotypes at cellular resolution, its utility for infectious diseases remains largely unexplored. For transcriptional studies of individual malaria parasites, the minute RNA content pose challenges to RNA-sequencing.
To this end, we have adopted the Smart-seq2 protocol to sequence polyA+ transcriptomes of individual and populations of 10,000 Plasmodium-infected erythrocytes, sampled at critical time points throughout the erythrocytic phase. We based our analyses on
165 P. falciparum-infected scRNAseq samples, in which we detected at least 200 genes per individual parasite. Importantly, developmental markers peaked in expression at the expected stages and we observed patterns of stage-specific gene expression in both bulk and single-parasite.
Transcriptional activity varies across the development cycle, with lower activity during the early phases of each stage (ring, trophozoite and schizont) and reached higher levels at the respective late stages. The distinct subgroups emerging during the given life stages provide evidence for the existence of a bet-hedging strategy by the parasite.
Our study provides a proof-of-principle and demonstrates the utility of single-cell RNA-seq for transcriptional analyses of infectious diseases at unprecedented resolution. The single-cell resolution will also be extremely important for understanding parasite adaptation, where the individual parasite variability can be monitored to reveal parasite strategies for escaping immune response, chemo protection and evolution of drug resistance.



P139

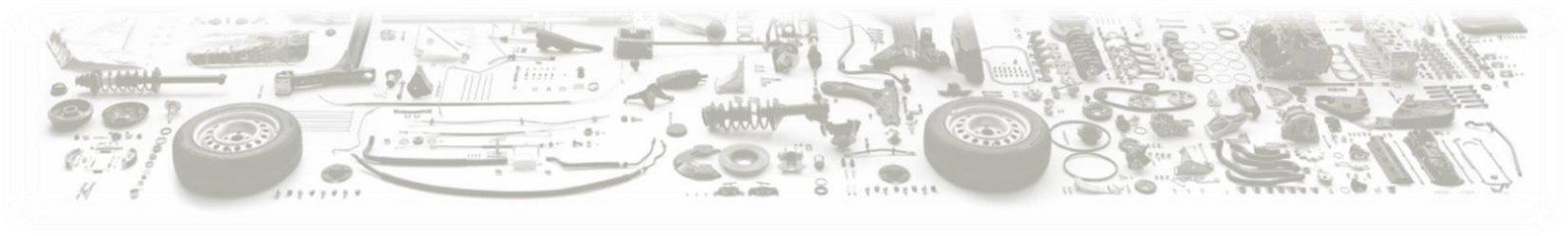
High throughput microfluidic imaging analysis and visualisation integrated with single-cell sequencing

Quin Wills, Rory Nolan, Esther Mellado, Xènia Massana-Muñoz, Sergi Padilla-Parra, Stephen Taylor, Rory Bowden

University of Oxford, Oxford, United Kingdom

While microfluidics have helped spark the current growth in single-cell sequencing, they've not yet been used to their full imaging potential. Being able to image cells is not only vital for quality control (QC) but also useful to stratify cells based on imaging phenotypes. For this reason we have developed SCEnIC, open source R server software to support C1 chip studies within the Oxford Consortium for Single Cell Biology. Using a simple web interface, SCEnIC supports brightfield and fluorescence image analysis, including live cell (time-series) imaging, confocal imaging and Forster-based Energy Transfer (FRET) ratiometric analysis. SCEnIC uses an imaging algorithm developed to zoom in on cells and automate the QC by failing unusual fluorescence features. After QC and analysis, SCEnIC outputs fluorescence results as a simple spreadsheet for integration with sequencing results. SCEnIC also integrates with Zegami for high throughput image exploration [1]. Here we demonstrate SCEnIC's and Zegami's features to support imaging, including more challenging work such as perturbation (time-series) studies, and viral-fusion FRET imaging. [1] www.zegami.com

Funding: The Oxford Consortium for Single Cell Biology is funded by a UK Medical Research Council Clinical Research Capabilities and Technologies Initiative grant. SP is funded by an NDM leadership fellowship.



P140

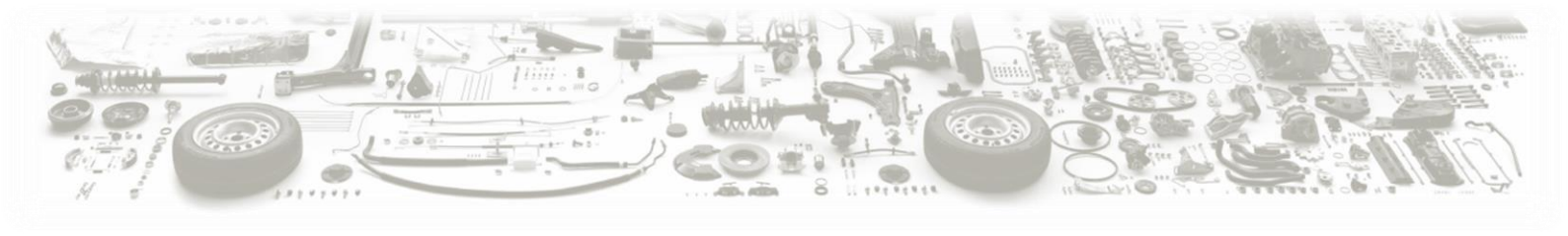
Powerful latent variable models for decomposing cell-to-cell heterogeneity in single-cell RNA-seq studies

Florian Buettner, Oliver Stegle

European Bioinformatics Institute, Cambridge, United Kingdom

A major challenge in the analysis of single-cell RNA-seq studies is to decompose the sources of variation that underlie transcriptional heterogeneity into technical noise, the core biological processes of interest and confounding factors. We have recently shown that one or few such factors, such as the cell cycle, can be effectively accounted for using a single-cell latent variable model (scLVM)[1], thereby improving interpretation and power of single-cell analyses. However, general solutions to identify and account for the full breadth of the factors that cause variation are lacking. To address this, we have generalized scLVM to model an arbitrary number of processes, leveraging prior knowledge in form of the gene sets that annotate biological processes of interests. Technically, we use sparse group factor analysis methods to fit this model and enable scaling such analyses to tens of thousands of cells and hundreds of factors. This generalized model (scLVM2) efficiently copes with overlapping annotations and facilitates both the identification as well as visualization of specific, interpretable processes that drive the cell-to-cell variation - including core processes of interest and confounders.
After validating the model, we applied scLVM2 to 96 hematopoietic stem cells (HSCs), containing long-term functional HSCs as well as short-term HSCs [2]. Based on 50 annotated hallmark gene sets (GESA/ MSigDB), scLVM2 identified Kras signaling and cell cycle as driving factors, consistent with the importance of both processes for HSC function. We show the power of our approach by generating interpretable visualizations and identifying new marker genes for these processes.

References:
[1] Buettner et al. Nature biotechnology 33.2 (2015): 155-160
[2] Wilson et al. Cell stem cell, 2015.



P141

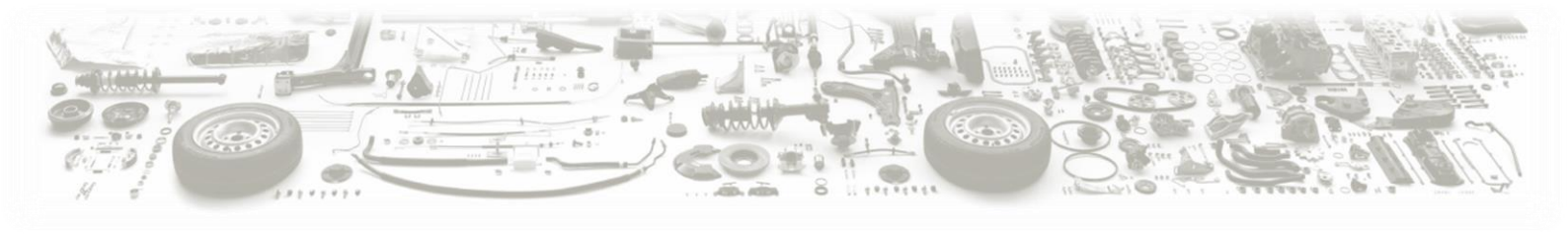
Integrated single-cell transcriptomics and imaging to study human-virus interactions: HIV fusion and the EBV life cycle

Quin Wills[1], Daniel Jones[1], Xènia Massana-Muñoz[1], Esther Mellado[1], Vito Ricigliano[2], Rory Nolan[1], Adam Handel[1], Zameel Cader[1], Eshita Sharma[1], Helen Lockstone[1], John Broxholme[1], Rory Bowden[1], Chris Holmes[1], Marco Salvetti[2], Sergi Padilla-Parra[1]

[1]*University of Oxford, Oxford, United Kingdom*
[2]*Sapienza, University of Rome, Rome, Italy*

Studying the functional genomics of viral infections requires the ability to tease apart different viral states, different cell states, and the interaction of these host-pathogen states. We present two studies: (i) lymphoblasts from identical twins discordant for multiple sclerosis that have been lab-infected with EBV, (ii) human primary T cells and HeLa cells that have been lab-infected with HIV-1. The first study considers the putative role of EBV in multiple sclerosis B cell dysfunction, while the second considers T cell factors involved with HIV fusion. In both studies, fluorescence/FRET (Forster Resonance Energy Transfer) imaging of the cells on microfluidic chips was used to differentiate known cell states prior to human and viral gene RNA-seq. The use of imaging to first stratify the cells validated the studies by revealing known viral biology in the cell transcriptomes (CD40-mediated inflammatory response with EBV, and kinesin dependent uncoating of HIV). The transcriptomic data was then used to further dissect cell sub-states based on the expression of viral genes, the human cell cycle and consensus clustering of gene network topologies. Using these two studies we present a general framework for modelling host-virus interactions, which uses software that we have developed for high-throughput imaging analysis, combined with our ROC (RObust Coexpression) approach for identifying novel biology in this challenging sparse data. The goal in these ongoing projects is to better understand modifiable interactions that have an impact on disease risk.

Funding: The Oxford Consortium for Single Cell Biology is funded by a UK Medical Research Council Clinical Research Capabilities and Technologies Initiative grant. SP is funded by an NDM leadership fellowship. MS is funded by Fondazione Italiana Sclerosi Multipla (Italian MS Society).



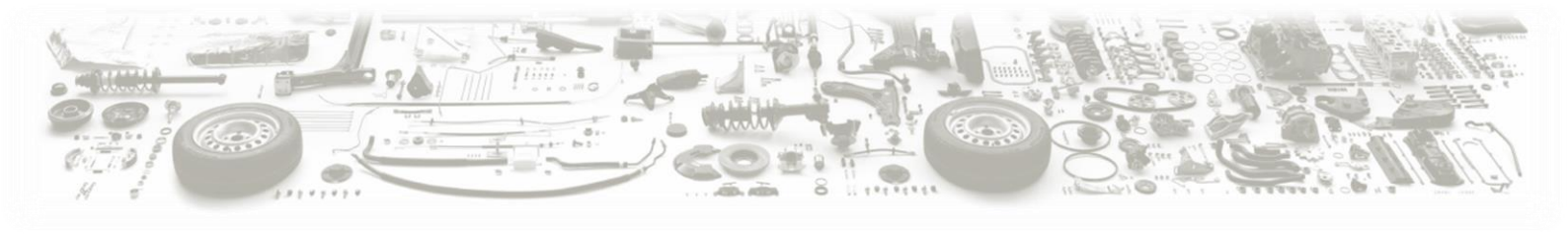
P142

Host-pathogen interactions of toxoplasma gondii at the single-cell level

Philip Brennecke, Yuhsin Huang, Kerry Buchholz, Wu Wei, Sarah Ewald, John Boothroyd, Lars Steinmetz

Stanford University, Stanford, CA, United States

Toxoplasma gondii is a single-celled obligate intracellular parasitic protozoan that causes the disease toxoplasmosis. It is estimated that 30-50% of humans have been exposed to the parasite and might be chronically infected. In immunocompromised individuals infection can cause fatal illness. Upon invasion of host cells, toxoplasma gondii rewires host transcription substantially. However, molecular players involved in the process of transcriptional rewiring remain largely elusive. Here we use single-cell RNA-sequencing to obtain both host and parasite transcriptomes from single infected host cells. This approach will allow us to gain mechanistic insights into transcriptional rewiring by looking at gene-gene correlations between host and parasite in individual infected host cells. Moreover, the single-cell resolution of the assay allows us to assess heterogeneity of both infection and host response.



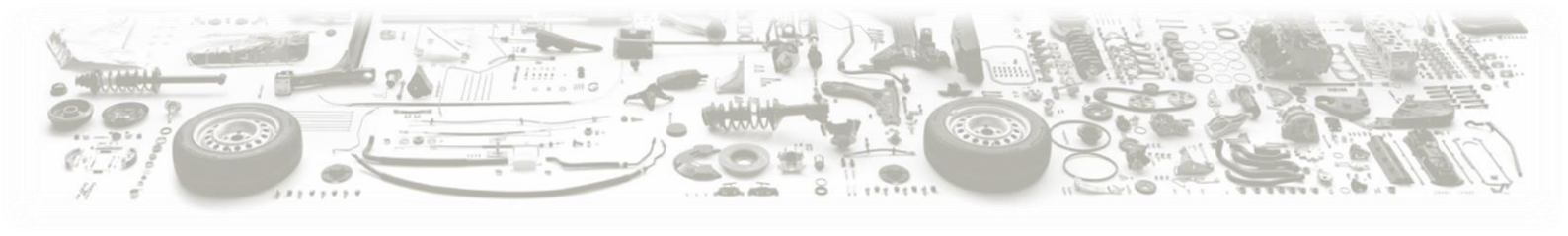
P143

High-throughput single-cell genomics at a large genome centre

Iraad Bronner, Joanna Cartwright, Andrew Sparkes, Stephan Lorenz

Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

Improvement of NGS chemistries increased sequencing output and provided a significant reduction of costs per base. These improvements in sequencing technology, with significant breakthroughs in methods to isolate and amplify the transcriptome and genome of individual cells enabled the expansion of single cell genomics. To understand the organisation and function of cell populations and the involvement and contribution of individual cells to tissue and organ function, during development and/or disease, scientists aim to examine thousands of cells.
Delivering these experiments on such large scale, however, is a challenging task.
The Wellcome Trust Sanger Institute has now established a single cell library preparation and sequencing core facility in order to provide standardised high-throughput single cell genome, transcriptome and epigenome amplification and sequencing services.
Our facility demonstrates, how flow cytometry, the most matured single cell technique, state-of-the-art liquid handlers and low-volume dispensers can be utilized to deliver thousands of high-quality single-cell genomes and full-length transcriptomes per day with a turnaround time of 2 days without compromising data content and quality. We use a high-speed nanoliter dispenser to set up many low-volume amplification reactions, while a more conventional Hamilton STARplus platform is used for more demanding single cell protocols like G&T seq or bisulfite sequencing. These systems feed into a sub-microliter pipeline that is built around an Access Laboratory Workstation with integrated ECHO525 acoustic dispenser for sample quality control and library preparation. These devices are linked together by further liquid handlers to allow parallelised sample cleanup and pooling. This is complemented by a tailor-made laboratory information management system to keep track of this highly-automated workflow and collect meaningful metadata for single cell experiments.
Other than droplet systems that pool 10,000s of cells into one sequencing reaction, our approach allows us to track every single cell from isolation to sequencing and revisit individual cells for deeper sequencing or validation experiments, adding additional value to our scientist's work to understand cell heterogeneity at a large scale.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P144

Integrative analysis of single cell genome and transcriptome sequencing reveals the effect of genomic copy number at single cell level

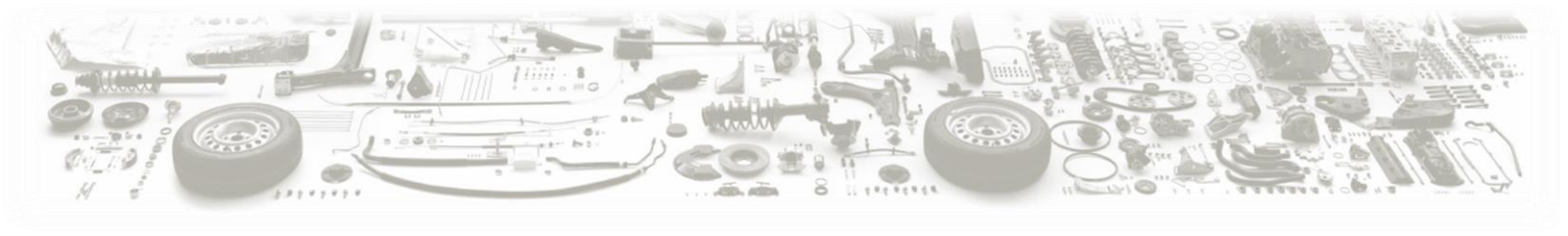
Tim Xiaoming Hu[1], Wilfried Haerty[1], Mabel Teng[2], Parveen Dabas[2], Chris P. Ponting[1], Thierry Voet[2], Iain C. Macaulay[2], John C. Marioni[3]

[1]*Oxford University, Oxford, United Kingdom*
[2]*Sanger Institute, Cambridge, United Kingdom*
[3]*EMBL-EBI, Cambridge, United Kingdom*

The effect of altering a gene's copy number on its transcriptional output is still unclear. Studies that investigate this phenomenon using samples from different individuals are confounded by genetic, environmental and cell cycle effects: any observed perturbation of expression could be the consequence not just of gene copy number, but also of a cell's genotype and its extra- and intra-cellular environment. To eliminate such confounding factors, isogenic cells grown in the same culture, with the same cell cycle state are required.

With G& T-Seq [1] the genome and transcriptome of the same single cell can be sequenced in parallel. Performing G&T-Seq on hundreds of cells separately enabled us to capture a naturally occurring, yet rare, aneuploidy event. The aneuploidy event generated a viable and stable aneuploidy sub-clone of lymphoblastoid cell line that is otherwise isogenic to its progenitor. By contrasting the rare aneuploid sub-clone to its euploid progenitor, we recapitulated the general positive correlation between DNA copy number and mRNA abundances. More interestingly, we identified genes that are strikingly discordant in DNA and mRNA dosage: a 50% increase in gene copy number results in near-complete ablation of these genes' expression. To the best of our knowledge, this is the first observation of DNA-mRNA-copy number discordancy in isogenic and iso-environmental single cells.

1. Macaulay, I.C., et al., G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods, 2015. 12(6): p. 519-22.



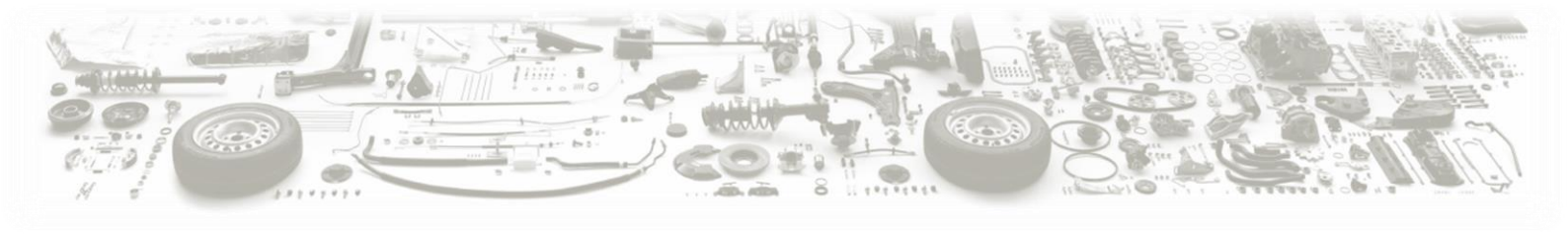
P147

Dynamics of the cell cycle revealed by parallel genome and transcriptome sequencing of single cells

Mabel Joey Teng[1], Iain Macaulay[1], Tim Xiaoming Hu[1], Niels Van der Aa[2], Parveen Kumar[2], Thierry Voet[1]

[1]*Sanger Institute, Cambridge, United Kingdom*
[2]*KU Leuven, Leuven, Belgium*

Single cell sequencing technologies allow the analysis of the genomes and transcriptomes of thousands of individual cells, and thus the exploration of the degree of heterogeneity within populations of cells. We applied single cell genome and transcriptome sequencing (G&T-seq) to explore, in parallel, the dynamics of DNA replication in the genome and changes in gene expression in cells isolated from different phases of the cell cycle. We FACS sorted cells from a human lymphoblastoid cell line (HCC38-BL) using DNA content staining to distinguish different phases of the cell cycle, and the genome and transcriptome were examined using the G&T-seq.
By isolating cells from G1, early, mid and late S-phase and G2/M phase, we identified clear genomic and transcriptional signatures of each phase, and for the first time, were able to annotate transcriptomic events during cell cycle and associate directly with genomic replication. These data provide insight into the relationship between gene expression and DNA replication during the cell cycle, and may provide the basis to build models to account for the effect of cell cycle on single cell genome and transcriptome sequencing data.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

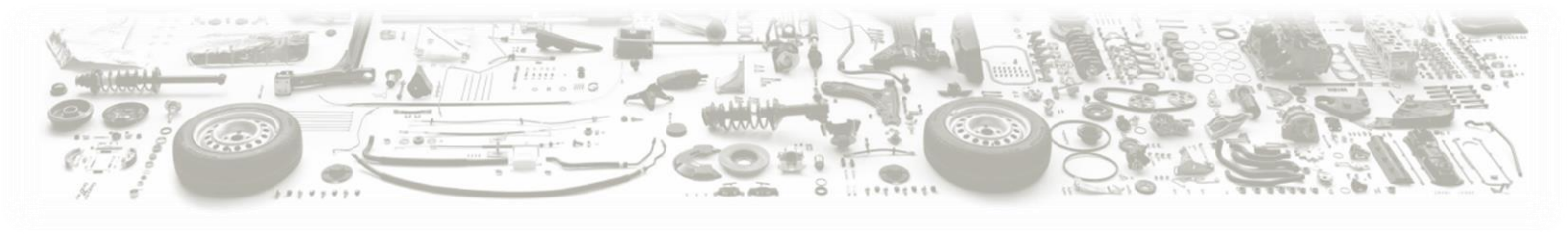
P148

Discovery of dendritic cell sub-populations in human blood by single cell RNA-sequencing

Alexandra-Chloe Villani[1], Rahul Satija[2], Christopher B. Ford[3], Morgane Griesbeck[4], Weibo Li[3], Philip De Jager[5], Aviv Regev[6], Nir Hacohen[1]

[1]*Broad Institute of MIT & Harvard; Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Cambridge, MA, United States*
[2]*NYU Center for Genomics and Systems Biology, New York Genome Center, New York City, NY, United States*
[3]*Broad Institute of MIT & Harvard, Cambridge,MA, United States*
[4]*The Ragon Institute of MGH, MIT and Harvard, Cambridge, MA,United States*
[5]*Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, United States*
[6]*Broad Institute of MIT & Harvard; Massachusetts Institute of Technology; and Howard Hughes Medical Institute, Cambridge, MA, United States*

Dendritic cells (DC) play a critical role in a host's response to pathogens and in the immune responses characterizing cancer, inflammatory and infectious diseases. To unbiasedly discover DC subtypes, which have historically been defined through surface marker analysis, we used single cell RNA-seq (Smart-seq2) to profile the transcriptome of 1056 single human blood DCs isolated from a healthy individual. While supervised analysis of DC markers effectively classified the 4 known populations (BDCA2+ (IL3RA), BDCA1+ (CD1C), BDCA3+ (CLEC9A), CD16+ (FCGR3A)), unsupervised analysis re-discovered all 4 subsets through a 512-discriminative gene signature in addition to highlighting novel heterogeneity within subsets, which clarifies unexplained DC heterogeneity observed in disease. Cross-referencing the 512-gene set with susceptibility loci identified specific subsets contributing to disease, with, for example, TCF7L2 (type II diabetes) uniquely expressed in CD16+ and CARD11 (ulcerative colitis) in BDCA2+. Multi-dimensional classification analysis of the first 384 single cells sequenced identified 26 outlier cells not clustering with any of the 4 known subsets. These outliers displayed a unique expression signature and a shared signature with BDCA2+ and BDCA3+. Using cell surface markers identified by scRNA-seq, we sorted and validated the existence of these cells in 10 additional healthy donors, showing they represent 0.06% of the PBMCs population in the blood and 2-3% of the DC populations across all 10 donors tested. Isolation, in vitro experimentations and profiling of an additional 600 single cell outliers enabled a deeper characterization of their phenotype. Together these analyses provide a comprehensive view of the DC landscape in blood.



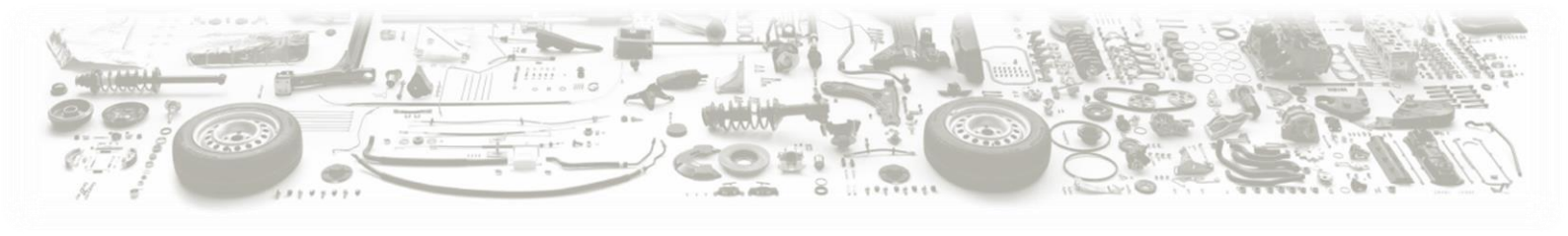
P149

Delineating cell lineage and pluripotency during human blastocyst formation by single cell transcriptional analysis

Daniel Edsgärd, Sophie Petropoulos, Björn Reinius, Qiaolin Deng, Rickard Sandberg, Fredrik Lanner

Karolinska Institutet, Stockholm, Sweden

Characterizing embryonic development during the first and arguably the most critical week of human development is of great importance. During the first 7 days the human embryo divides to form the first three lineages: trophectoderm (TE), primitive endoderm (PE), and pluripotent epiblast cells (EPI). To date, limited information exists pertaining to the transcriptional landscape and what drives lineage segregation and pluripotency in humans. Using Smart-seq2 single-cell RNA-sequencing, we have established a transcriptional road map consisting of ~1600 cells from 71 embryos, covering 8-cell stage to mature blastocyst at embryonic (E) day 7.
Our data describe the temporal progression of lineage segregation where TE vs ICM segregation occurs in the transition from E4 to E5. We identify genes linked to cell-cell junctions and epithelial polarization within the TE progenitors and genes linked to stem cell maintenance, protein kinase cascade and WNT signaling in the ICM cells.
Concurrently at E5 we do detect a priming of ICM cells towards EPI and PE segregation, which progressively mature during E6 and E7. During this progression, we identify genes linked to cell migration, epithelial morphogenesis and endoderm formation in emerging PE cells whereas the EPI cells express genes linked to stem cell maintenance and the Nodal, Notch and WNT signaling pathways.
The fundamental knowledge elucidated from this study is crucial in identifying mechanistic pathways underlying lineage segregation and the establishment of pluripotency, thus being of great importance for understanding human development and regenerative medicine.



SINGLE CELL GENOMICS

SOMETIMES THE SUM OF THE PARTS IS GREATER THAN THE WHOLE

P150

Exploring chromosomal conformations in cells among and between cell cycle phases using an improved version of the single-cell Hi-C protocol

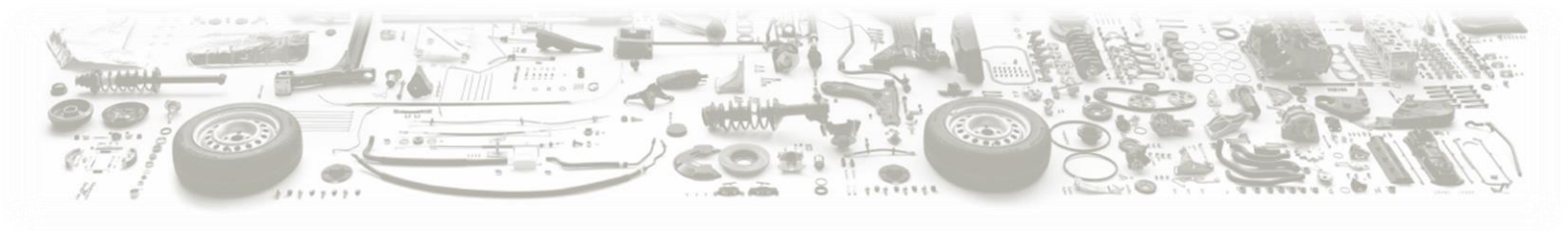
Takashi Nagano[1], Yaniv Lubling[2], Carmel Dudley[2], Csilla Varnai[1], Wing Leung[1], Steven W. Wingett[1], Amos Tanay[2], Peter Fraser[1]

[1]*The Babraham Institute, Cambridge, United Kingdom*
[2]*Weizmann Institute of Science, Rehovot, Israel*

Genomic architecture is a key player in the regulation of many cellular processes such as transcription, DNA replication and repair. The chromosomal conformation capture (3C) technique and its derivatives have contributed immensely to our understanding of chromatin organization and how it affects function. However, these techniques query a large population of cells, and therefore represent ensembles of possibly conflicting structures. We1 previously presented a single-cell variation of the Hi-C protocol. Single-cell Hi-C contributed to our view of chromosomal domains as stable structural units and showed that inter-domains interactions are non-random yet highly variable, but it allowed limited throughput and depth per cell. We recently modified the single-cell Hi-C protocol to allow routine profiling of hundreds of cells at depth of 40,000-200,000 contacts per hapolid nucleus. We report the analysis of around 1000 single nuclei, and demonstrate how single-cell Hi-C dissects major sources of variation in chromosome conformation. Specifically, we are showing how to decouple cell-to-cell variation that is associated with replication and mitosis, from general variation of the conformation of topological domains, chromosomal territories, gene promoters and more. Single-cell Hi-C is therefore shown to be a surprisingly robust and low-noise platform to study mechanisms of gene regulation at the single-cell level.

References

1. Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A & Fraser, P. (2013). Nature, 502(7469), 59–64. doi:10.1038/nature12593



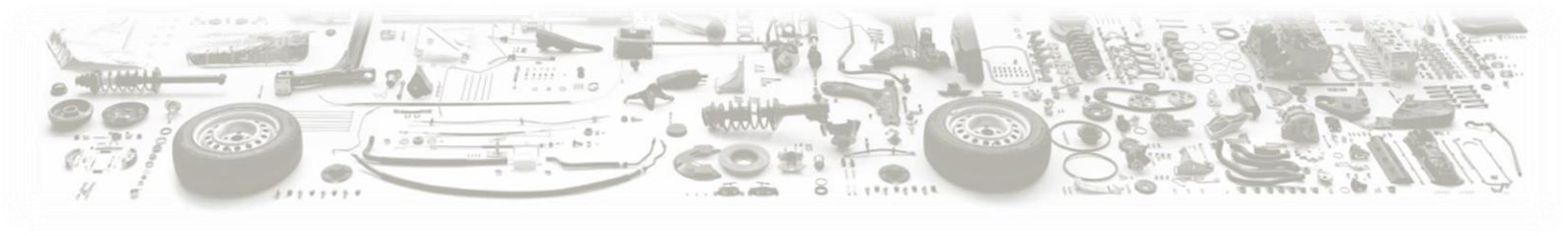
P152

Single cell transcriptional signatures of bone marrow failure in patients with Shwachman-Diamond Syndrome

Cailin Joyce[1], Lan Jiang[1], Sheng Li[2], Inga Hofmann[3], Chad Nusbaum[4], Colin Sieff[3], Christopher Mason[2], Guo-Cheng Yuan[1], Carl Novina[1]

[1]*Dana-Farber Cancer Institute, Boston, MA, United States*
[2]*Wiell Cornell Medical College, New York, NY, United States*
[3]*Children's Hospital Boston, Boston, MA, United States*
[4]*Broad Institute, Cambridge, MA, United States*

Shwachman-Diamond Syndrome (SDS) is a congenital syndrome caused by mutations in the SBDS gene. SDS primarily affects the pancreas, hematopoietic and skeletal systems, with hematopoietic manifestations including neutropenia, thrombocytopenia, pan-anemia and myelodysplasia with progression to acute myeloid leukemia. Because the affected bone marrow cells are rare and heterogeneous, the altered genetic networks in vivo remain unknown. Single cell genomic technologies enable unprecedented analyses of intractable cell types. We applied single cell RNA sequencing to CD34+ hematopoietic stem and progenitor cells (HSPC) from normal donors (n=95) and SDS patients (n=176) and established a hematopoietic ontogeny at single cell resolution based on the expression of 250 signature genes identified in population-level studies. Model based clustering revealed six clusters corresponding to hematopoietic stem cell, multipotent progenitor, common myeloid progenitor, megakaryocyte-erythroid progenitor, granulocyte-monocyte progenitor, and an unexpected CMP-GMP intermediate. Within-cluster differential gene expression analyses between normal and SDS cells revealed significant gene expression changes in every cluster, although many of the altered genes and pathways were different. Our data suggest that closely related HSPC subpopulations are variably affected by SBDS mutations, which may contribute to complex and unstable hematopoietic symptoms in patients.



P153

Using single-cell transcriptome analysis to unravel the patterns and mechanisms of promiscuous gene expression in medullary thymic epithelial cells

Alejandro Reyes[1], Philip Brennecke[2], Sheena Pinto[3], Kristin Rattay[3], Michelle Nguyen[2], Rita Küchler[3], Wolfgang Huber[1], Bruno Kyewski[3], Lars Michael Steinmetz[2]

[1]*European Molecular Biology Laboratory, Heidelberg, Germany*
[2]*Stanford University, Palo Alto, CA, United States*
[3]*German Cancer Research Center, Heidelberg, Germany*

Self-tolerance induction is a hallmark of the adaptive immune system and, in case this process fails, various autoimmune diseases have been shown to develop. Self-tolerance of T cells relies in the thymus, where maturing T-cells with high affinity to self-antigens are negatively selected. Among the various thymic antigen-presenting cells (APCs), medullary thymic epithelial cells (mTECs) stand out due to their ability to express genes whose regulation outside of the thymus is tightly controlled in time and space, such as genes coding for tissue-restricted antigens (TRAs). This unique expression program is known as promiscuous gene expression (pGE). pGE results in mosaic expression patterns, with each TRA being present in only 1-3% of mTECs. Although the transcriptional regulation of pGE is poorly understood, the transcriptional activator Aire represents an exception. Aire is known to promote the expression of a large fraction of TRA-encoding genes by binding to silent chromatin marks. Nevertheless, the underlying rules that govern pGE patterning at the single-cell level remain unclear. Do single mTECs express random sets of TRA-encoding genes or are there patterns of coordinated expression of TRA-encoding genes?
To answer this question, we performed scRNA-seq on 203 mature mTECs of mice as well as 3 mTEC subsets that were selected for the expression of particular TRAs. We found that mTECs represent a cell-type that is highly heterogeneous at the level of individual cells, and yet collectively express most of the genome. Furthermore, we demonstrated that pGE occurs in groups of co-expressed gene sets that are generally expressed at low cell frequencies. Co-expressed genes cluster in the genome and their promoters show enhanced chromatin accessibility. Our findings characterize pGE as a highly coordinated process, which ensures representation of the full diversity of self-antigens in the mTEC compartment by expressing recurring and complementing co-expression clusters in individual cells.