

Silesian University of Technology

INVESTIGATING SOURCES OF ZEROS IN 10X SINGLE-CELL RNASEQ DATA

Michał Marczyk

Department of Data Mining Silesian University of Technology Poland

Single-cel RNA sequencing

Single-cell sequencing is a resolution revolution in transcriptomics and genomics.



Lun et al. 2017

Lahnemann et al., 2020

"Zeros" in single-cell RNA sequencing data

- Some genes could be highly expressed in one cell but not expressed in another.
- For some genes, we see more cells with expression (green) than not (red), but overall, there are much more zeros in the data.
- Only 50 genes have non-zero values in all cells.

Cells

a de la constante de la constan La serie de la constante de la c

"Droplet scRNA-seq is not zero-inflated"

- The number of zeros observed is consistent with the theoretical models.
- Additional zero values in the data are likely due to biological variability.
- The most important factor that determines the number of zeros in the scRNA-seq data is the depth of sequencing (total UMI) per cell.

But other important factors could also exist!



Data



- 2 biological replicates (1a and 1b).
- DNA and RNA for all experiments was extracted on the same day.

Single-cell vs bulk RNA-seq



The technological differences are greater than the biological differences between samples.

"Zeros" from scRNA-seq in other modalities



Dropout rate vs signal from 3 different modalities

Higher expression and chromatin openness, and lower methylation level relate to lower dropout rate.

f(x) = c +



Functional analysis of 5PL model residuals

Paths with genes with fewer zeros than expected

PATI

WNT SI

DILATED

GNRH SIG

VIBRIO CHC

CHRONIC M

GLUTATH

MAPK SI

GLYCOLYSIS GL

ECM RECEP

AUTOIMMUNE

UBIQUITIN MEDIA

VEGF SIG

LINOLEIC

PYRIM

ADF

ARGININE AND PRC

FC GAMMA R MEDIATE

CYSTEINE AND METHIO

SNARE INTERACTIONS IN VESI

GLYOXYLATE AND DICARBOXY

LEUKOCYTE TRANSENDOTI

NEUROTROPHIN SIG

CARDIAC MUSC

OXIDATIVE P

HUNT

ALZ

PAR

REGULATION OF ACTI

MATURITY ONSET DIABET

PATHOGENIC ESCHERICH

VASOPRESSIN REGULATED WATE

PU

RENAL

HYPERTROPHIC CARD

ANTIGEN PROCESSING AI

PANTOTHENATE AND C

ATHWAYS IN CANCER -	***	***			*	**
SIGNALING PATHWAY -	*				*	*
MELANOMA	*	*				*
RDIOMYOPATHY HCM -	*	*			*	*
D CARDIOMYOPATHY	***	***			*	*
AL CELL CARCINOMA	**	**			*	
BNA DEGRADATION -					*	*
AND PRESENTATION -					*	*
SIGNALING PATHWAY	**	*		*		
COA BIOSYNTHESIS -					*	*
HOLEBAE INFECTION -				*		
	+++	**	*			
					*	++
			+	*	*	*
				++	^	
			~~	~ ^ ^		مادياد
					L L	**
					*	**
	*	*				
EPTOR INTERACTION -	***	***				
GAP JUNCTION -	***	***		*		
	*		*	*		
NE THYROID DISEASE -	*	*				
SIGNALING PATHWAY				*		
IIONINE METABOLISM -					*	**
ENDOCYTOSIS -			**	*	*	**
DIATED PROTEOLYSIS -			**	*	**	**
VIRAL MYOCARDITIS -			*	*		
DHERENS JUNCTION -	**	**	*	**	**	**
IC ACID METABOLISM -	*			**		*
XYLATE METABOLISM -			*			
XYLATE METABOLISM – THYROID CANCER –			*			
XYLATE METABOLISM THYROID CANCER IMIDINE METABOLISM	*		* * **	**	**	**
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION	*		* * **	** **	** **	** **
XYLATE METABOLISM - THYROID CANCER - IMIDINE METABOLISM - OTHELIAL MIGRATION - TIGHT JUNCTION -	*		* * ** **	** ** **	** ** **	** ** **
XYLATE METABOLISM - THYROID CANCER - IMIDINE METABOLISM - DTHELIAL MIGRATION - TIGHT JUNCTION - CELL CYCLE -	*		* * ** ** **	** ** **	** ** **	** ** **
XYLATE METABOLISM – THYROID CANCER – MIDINE METABOLISM – DTHELIAL MIGRATION – TIGHT JUNCTION – CELL CYCLE – ASTHMA –	* * *	**	* * ** ** **	** ** *	** ** **	** ** *
XYLATE METABOLISM – THYROID CANCER – IMIDINE METABOLISM – DTHELIAL MIGRATION – TIGHT JUNCTION – CELL CYCLE – ASTHMA – SIGNALING PATHWAY –	* * **	**	* * ** ** **	** ** * *	** ** **	** ** *
XYLATE METABOLISM - THYROID CANCER - IMIDINE METABOLISM - DTHELIAL MIGRATION - TIGHT JUNCTION - CELL CYCLE - ASTHMA - SIGNALING PATHWAY - CTIN CYTOSKELETON -	* * ** **	** * ***	* * * ** ** ** ***	** ** * ***	** ** **	** ** *
XYLATE METABOLISM - THYROID CANCER - MIDINE METABOLISM - TIGHT JUNCTION - CELL CYCLE - ASTHMA - SIGNALING PATHWAY - CTIN CYTOSKELETON - OOCYTE MEIOSIS -	* * ** ** **	** * ***	* * * ** ** ** ***	** ** * ***	** ** **	** ** * *
XYLATE METABOLISM – THYROID CANCER – IMIDINE METABOLISM – DTHELIAL MIGRATION – TIGHT JUNCTION – CELL CYCLE – ASTHMA – SIGNALING PATHWAY – CTIN CYTOSKELETON – OOCYTE MEIOSIS – ETES OF THE YOUNG –	* * ** ** ** *	** * ***	* * * * * * * * * * * * * * * * * * *	** ** * *** ***	** ** **	** ** * *
XYLATE METABOLISM – THYROID CANCER – IMIDINE METABOLISM – DTHELIAL MIGRATION – TIGHT JUNCTION – CELL CYCLE – ASTHMA – SIGNALING PATHWAY – CTIN CYTOSKELETON – OOCYTE MEIOSIS – OOCYTE MEIOSIS – ETES OF THE YOUNG – BLADDER CANCER –	* ** ** ** *	** * ***	* * * * * * * * * * * * * * *	** ** ** *** ***	** ** **	** ** * *
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER SCLE CONTRACTION	* * ** ** * * * *	** * *** *	* * * * * * * * * * * * * * * * * * *	** ** ** *** *** **	** ** **	** ** * ** *
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM TIGHT JUNCTION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER FOCAL ADHESION	* ** ** ** * * * * *	** * *** *	* * * * * * * * * * * * * * * * * * *	** ** ** ** ** **	** ** ** **	** ** * * * *
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER SIGCAL ADHESION PROTEIN EXPORT	* * ** ** * * * *	** * * * **	* * * * * * * * * * * * * * * * * * *	** ** ** ** ** ** **	** ** ** **	** ** * ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER SCLE CONTRACTION FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE	* * ** ** * * * * *	** * ** * **	* * * * * * * * * * * * * * * * * * *	** ** ** ** ** ** ** **	** ** ** **	** ** * ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM OTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER ISCLE CONTRACTION FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE TTER REABSORPTION	* ** ** ** ** ** **	** * * * **	* * * * * * * * * * * * * * * * * * *	** ** ** *** ** **	** ** ** ** **	** ** * ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION TIGHT JUNCTION CELL CYCLE ASTHMA SIGNALING PATHWAY OCYTYE MEIOSIS ETES OF THE YOUNG BLADDER CANCER FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE TTER REABSORPTION CHIA COLI INFECTION	* ** ** ** ** **	** * ** * **	* * * * * * * * * * * * * * * * * * *	** ** * *** *** *** *** *** ***	** ** ** **	** ** * ** ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA CELL CYCLE ASTHMA CIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER BLADDER CANCER PROTEIN EXPORT RNA POLYMERASE TER REABSORPTION CHIA COLI INFECTION POOSHORYLATION	* ** ** ** ** ** **	** * ** ** **	* * * * * * * * * * * * * * * * * * *	** ** ** ** ** ** ** ** ** **	** ** ** ** ** **	** ** ** ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM OTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER SCLE CONTRACTION PROTEIN EXPORT RNA POLYMERASE TER REABSORPTION CHIA COLI INFECTION SPLICEOSOME	* ** ** ** ** ** **	** * * * ***	* * * * * * * * * * * * * * * * * * *	** ** ** ** ** ** ** ** ** ** ** ** **	** ** ** ** ** **	** ** ** ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER SCLE CONTRACTION FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE TER REABSORPTION CHIA COLI INFECTION E PHOSPHORYLATION SPLICEOSOME NTINGTONS DISEASE	* * ** ** ** ** **	** * * * ** ***	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** ***	** ** ** ** ** **	** ** ** ** ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER BLADDER CANCER FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE CHIA COLI INFECTION SPLICEOSOME PROTEASOME PROTEASOME	* ** ** ** ** ** **	** * * ** **	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** ** **	** ** ** ** ** *** ***
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM OTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER BLADDER CANCER BLADDER CANCER ROA POLYMERASE TER REABSORPTION PROTEIN EXPORT RNA POLYMERASE TER REABSORPTION CHIA COLI INFECTION SPLICEOSOME PROTEASOME PROTEASOME PROTEASOME LZHEIMERS DISEASE	* * ** ** ** ** *** ***	** *** * *** ***	* * ** ** ** ** ** ** ** *** *********	*** *** *** *** *** *** *** *** **** ****	** ** ** ** ** ** **	** ** ** ** ***
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM OTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY CTIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER ISCLE CONTRACTION FOCAL ADHESION PROTEIN EXPORT RNA POLYMERASE TITER REABSORPTION CHIA COLI INFECTION SPLICEOSOME PROTEASOME PROTEASOME PROTEASOME PROTEASOME LZHEIMERS DISEASE ARKINSONS DISEASE	* * * * * * * * * * * * * * * * * * *	** * * * ***	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** ** ** ** **	** ** ** ** ** **
XYLATE METABOLISM – THYROID CANCER – MIDINE METABOLISM – DTHELIAL MIGRATION – CELL CYCLE – ASTHMA – CELL CYCLE – ASTHMA – CIN CYTOSKELETON – OOCYTE MEIOSIS – ETES OF THE YOUNG – BLADDER CANCER – BLADDER CANCER – ISCLE CONTRACTION – FOCAL ADHESION – PROTEIN EXPORT – RNA POLYMERASE – CHIA COLI INFECTION – SPLICEOSOME – PROTEASOME – PROTEASOME – PROTEASOME – LZHEIMERS DISEASE – ARKINSONS DISEASE – RIBOSOME –	* * * * * * * * * * * * * * * * * * *	** * ** ** ** **	* * ** ** ** ** ** ** ** ** ** ** ** **	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** * ** ***	** ** ** ** ** ** **
XYLATE METABOLISM THYROID CANCER MIDINE METABOLISM DTHELIAL MIGRATION CELL CYCLE ASTHMA SIGNALING PATHWAY SIGNALING PATHWAY CIN CYTOSKELETON OOCYTE MEIOSIS ETES OF THE YOUNG BLADDER CANCER BLADDER CANCER PROTEIN EXPORT PHOSPHORYLATION SPLICEOSOME PROTEASOME PROTEASOME PROTEASOME ARKINSONS DISEASE RIBOSOME	* ** ** ** ** *** *** ***	** *** ** *** ***	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** ** *** ***	*** ** ** ** *** *** **** ****
XYLATE METABOLISM - THYROID CANCER - IMIDINE METABOLISM - CELL CYCLE - ASTHMA - SIGNALING PATHWAY - CELL CYCLE - ASTHMA - SIGNALING PATHWAY - CTIN CYTOSKELETON - OOCYTE MEIOSIS - BLADDER CANCER - ISCLE CONTRACTION - FOCAL ADHESION - PROTEIN EXPORT - RNA POLYMERASE - RNA POLYMERASE - NTER REABSORPTION - SPLICEOSOME - SPLICEOSOME - PROTEASOME - PROTEASOME - RIBOSOME - RIBOSOME -	* * ** ** ** ** *** *** *** *** *** **	** *** ** *** *** *** *** *** *** ***	* * *** *** *** *** *** *** *** *** ***	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** ** ** *** *** *** *** *	*** ** *** *** *** *** *** *** *** ***
XYLATE METABOLISM - THYROID CANCER - IMIDINE METABOLISM - OTHELIAL MIGRATION - TIGHT JUNCTION - CELL CYCLE - ASTHMA - SIGNALING PATHWAY - CTIN CYTOSKELETON - OOCYTE MEIOSIS - BLADDER CANCER - ISCLE CONTRACTION - FOCAL ADHESION - PROTEIN EXPORT - RNA POLYMERASE - RNA POLYMERASE - TIER REABSORPTION - CHIA COLI INFECTION - SPLICEOSOME - SPLICEOSOME - PROTEASOME - PROTEASOME - ARKINSONS DISEASE - RIBOSOME -	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** *** **	* * * * * * * * * * * * * * * * * * *	*** *** *** *** *** *** *** *** *** **	*** *** *** *** *** *** *** *** *** **	** ** ** ** ** *** *** *** *** *** ***

Paths with genes with more zeros than expected



BUTANOATE METABOLISM **OLFACTORY TRANSDUCTION** JAK STAT SIGNALING PATHWAY CELL ADHESION MOLECULES CAMS O GLYCAN BIOSYNTHESIS TYPE II DIABETES MELLITUS PHOSPHATIDYLINOSITOL SIGNALING SYSTEM ABC TRANSPORTERS CALCIUM SIGNALING PATHWAY NITROGEN METABOLISM TRYPTOPHAN METABOLISM NEUROACTIVE LIGAND RECEPTOR INTERACTION DORSO VENTRAL AXIS FORMATION LYSINE DEGRADATION INOSITOL PHOSPHATE METABOLISM T CELL RECEPTOR SIGNALING PATHWAY ALDOSTERONE REGULATED SODIUM REABSORPTIO MELANOGENESIS CHEMOKINE SIGNALING PATHWAY SMALL CELL LUNG CANCER TGF BETA SIGNALING PATHWAY MTOR SIGNALING PATHWAY TYPE I DIABETES MELLITUS TASTE TRANSDUCTION ADIPOCYTOKINE SIGNALING PATHWAY STARCH AND SUCROSE METABOLISM VASCULAR SMOOTH MUSCLE CONTRACTION **B CELL RECEPTOR SIGNALING PATHWAY** PROSTATE CANCER PATHWAYS IN CANCER AXON GUIDANCE LYSOSOME ENDOMETRIAL CANCER WNT SIGNALING PATHWAY **MELANOMA** LONG TERM POTENTIATION HYPERTROPHIC CARDIOMYOPATHY HCM NON SMALL CELL LUNG CANCER PROGESTERONE MEDIATED OOCYTE MATURATION COLORECTAL CANCER INSULIN SIGNALING PATHWAY DILATED CARDIOMYOPATHY ERBB SIGNALING PATHWAY RENAL CELL CARCINOMA ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOM GLIOMA **GNRH SIGNALING PATHWAY** PANCREATIC CANCER CHRONIC MYELOID LEUKEMIA ECM RECEPTOR INTERACTION GAP JUNCTION FC GAMMA R MEDIATED PHAGOCYTOSIS ADHERENS JUNCTION LINOLEIC ACID METABOLISM CELL CYCLE NEUROTROPHIN SIGNALING PATHWAY REGULATION OF ACTIN CYTOSKELETON **OOCYTE MEIOSIS** FOCAL ADHESION

Functional analysis of 5PL model residuals

- Genes included in top3 up- and downregulated KEGG pathways.
- The mechanism of reduced expression of some genes on single-cell level might be different to what was expected: some groups of genes, like ribosomal genes, have higher expression on a single-cell level than bulk level, so low expression genes could not be captured due to limited total number of sequencing read counts that are measured in one experiment.



Potential technical sources of "zeros"

- Biased reads coverage at 3' end of transcripts (TIN score).
- Difference in %GC content of the gene sequence.
- Length of the gene.
- No. of transcripts per gene.
- Lower mappability in a gene region.

A low gene expression level in the analysed sample is the main contributor for zeros, **but not only.**

Sample 1a Sample 1b Factor Reduced Full Reduced Full 1.011 1.014 GC (0.988; 1.034)(0.991; 1.037)Gene_length 0.999 (0.998;1) 0.999 (0.998;1) -1.029 1.082 1.067 1.11 N_transcripts (0.979; 1.078) $(1.071; 1.149)^*$ (1.046;1.119)* (1.009; 1.124)0.978 1.018 1.006 1.016 Mappability (0.906; 1.05) $(1.013; 1.022)^*$ $(0.972; 1.04)^*$ $(1.012; 1.021)^*$ 1.039 1.011 1.062 1.017 TIN (1.006;1.017)* (1.046;1.078)* (1.012;1.023)* $(1.025; 1.05)^*$ 1.304 1.325 1.107 1.243 Expression (1.168;1.440)* (1.269;1.382)* (0.978; 1.237) $(1.186;1.3)^*$ 1.119 1.554 Chromatin (0.866; 1.371) $(1.224; 1.884)^*$ 1.001 Methylation 1(0.994; 1.006)(0.996; 1.008)

Odds ratio with 95% CI

Conclusions

- The differences between biological / technical replicates are minimal on all platforms.
- The differences in gene expression between platforms are greater than when comparing replicates on the same platform.
- There are some potential biological factors that may indicate why some genes have a higher dropout rate than others.
- There are various technical factors that can potentially affect the dropout rate, the most important of which are read coverage across the transcript and mappability in the gene region.

Acknowledgments

POLSL:

- Joanna Polańska
- Joanna Żyła
- Anna Papież
- Hanna Słowik



Yale cancer center

A Comprehensive Cancer Center Designated by the National Cancer Institute

Yale:

- Jun Zhao
- Rihao Qu
- Xiaotong Li
- Yuval Kluger
- Christos Hatzis