

MEP-LINCs PC3 SS2 Pilot Analysis

2015-07-28

Summary

The MEP-LINCs PC3 SS2 datasets include four levels of high content imaging data from on Microenvironment Microarrays (MEMAs). After QA filtering, there are 1748 Microenvironment Perturbations (MEPs) that are pairwise combinations of 46 printed ECM proteins and 38 ligands or growth factors. MEPS that have extreme proliferation, cell count or H3K9me3 signals are identified. All quantitated data, merging and analysis code is stored at Synapse and GitHub and is available at <https://www.synapse.org/#!Synapse:syn4624330>.

Introduction

The LINCs Pilot PC3 SS2 experiment was performed with PC3 cells grown in eight 8-well plates. The SS2 staining set includes, DAPI, H3K9me3 (488nm), Fibrillarin (555nm) and EdU (647nm). Four color images of the cells at each spot were gathered on an Olympus ScanR automated microscope. All data for this staining set comes from the nuclei as defined by the DAPI staining.

Intensity, position and a limited set of morphology data are gathered for each cell, merged with the experiment metadata, normalized and aggregated. The dataset is organized to the four LINCs imaging categories as follows:

Level 1 - Raw data

Level 2 - Normalized data

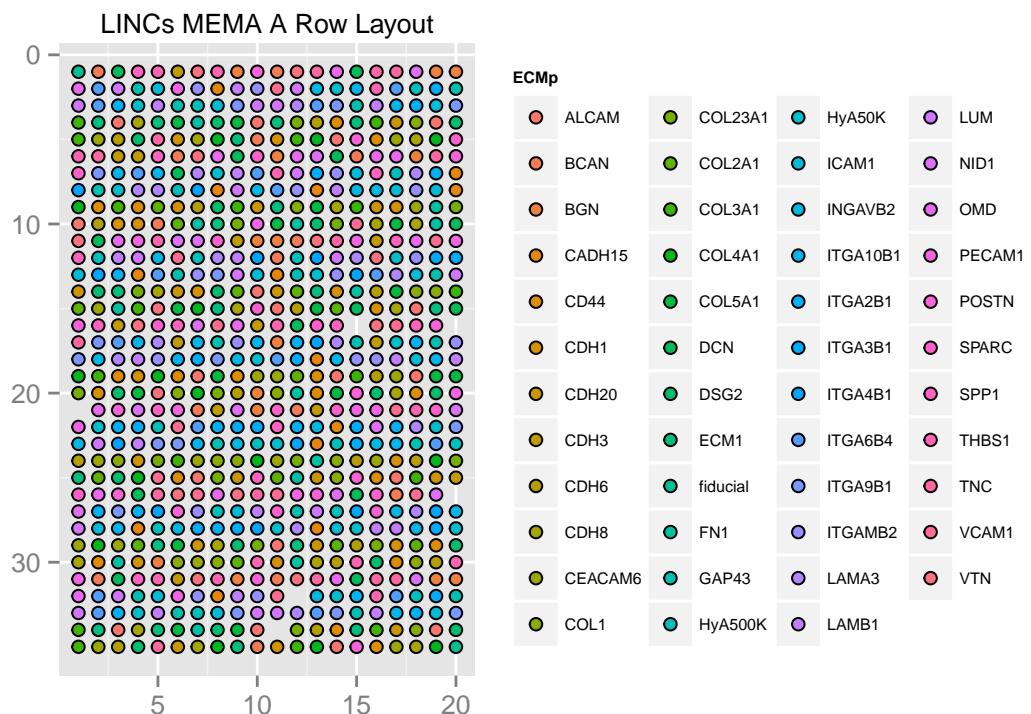
Level 3 - Normalized data aggregated to the spot level

Level 4 - Normalized data aggregated to the replicate (MEP) level

The data merging and analysis is done in R using open source software.

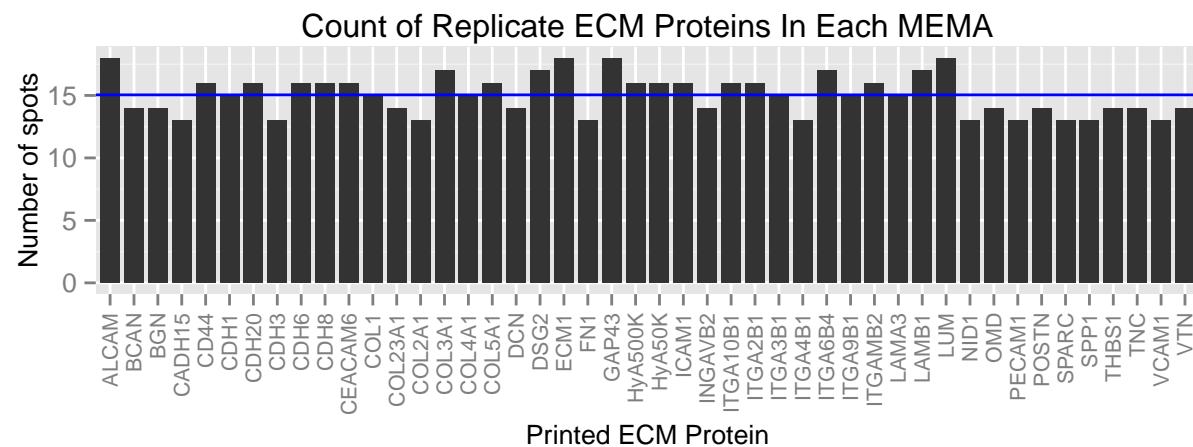
MEMA Layout

All MEMAs in the experiment are in separate wells and have the same design of 46 ECM proteins spotted in 35 rows and 20 columns. The proteins are randomly assigned to spots in the top 30 rows. Rows 31-35 are replicates of rows 1-5. The upper left and bottom right corners of each MEMA are image fiducials in the 488nm channel and there are four blank spots for checking orientation in all channels.



Replicate Count

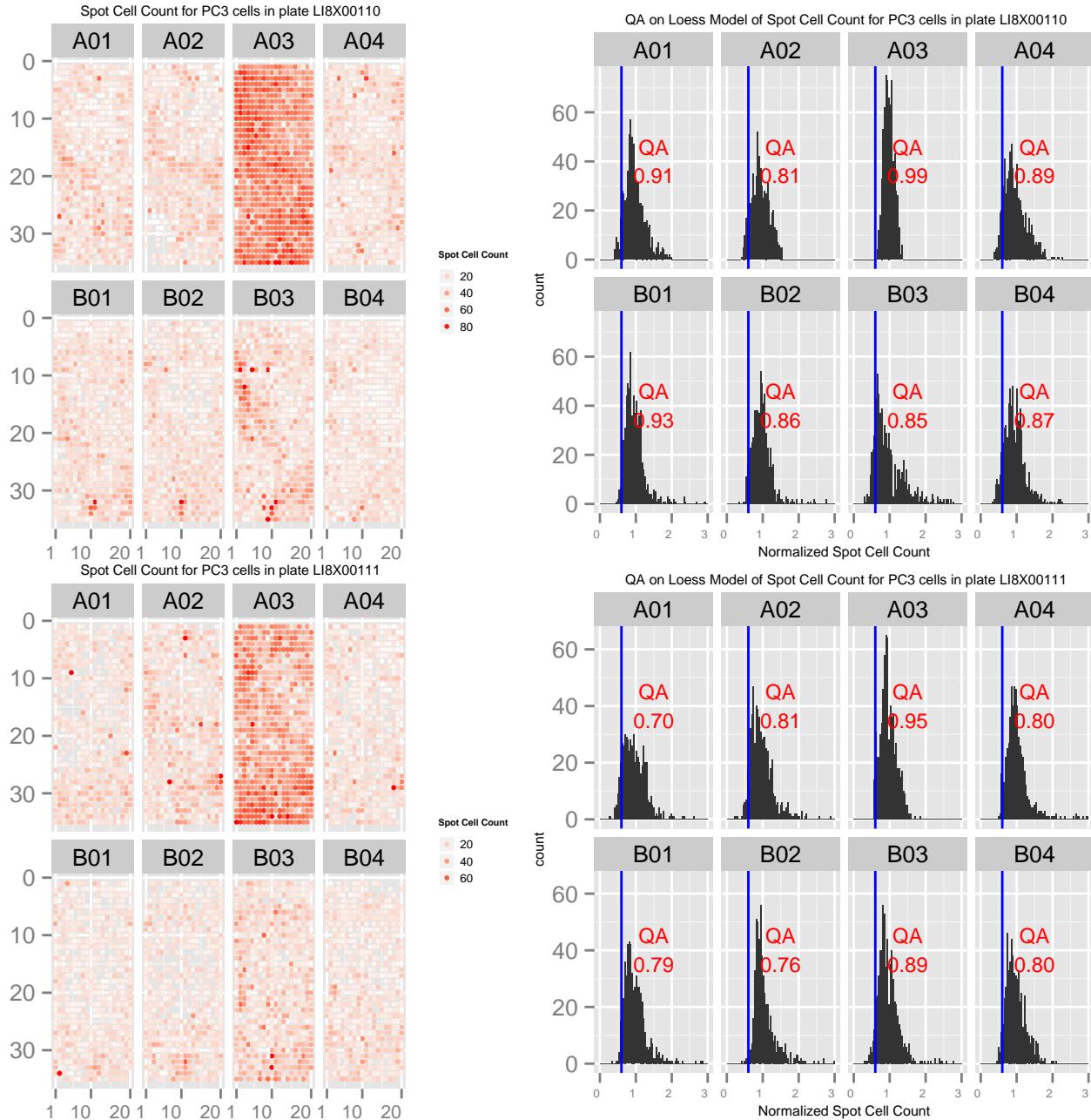
The MEMAs have an average of 15 replicates with a range from 13 to 19.

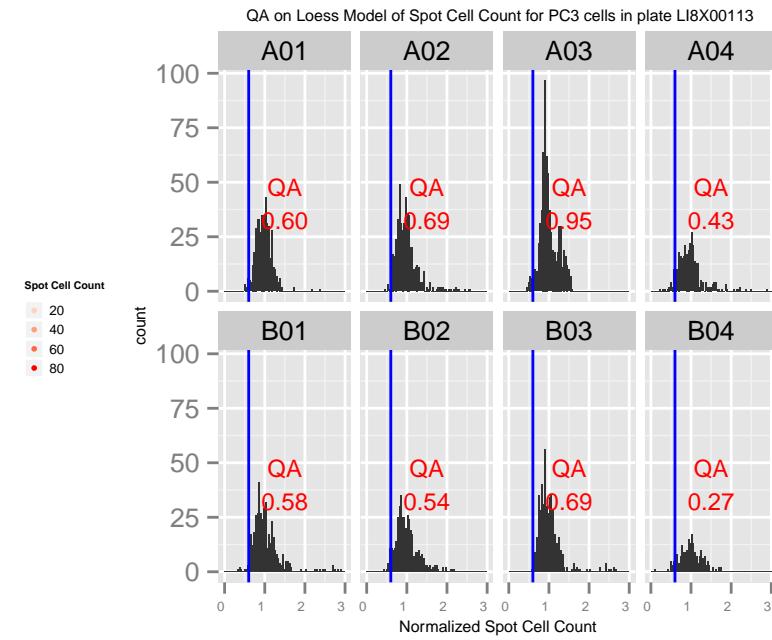
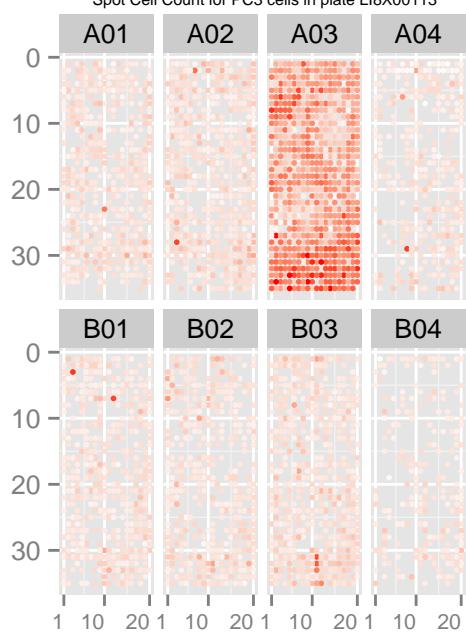
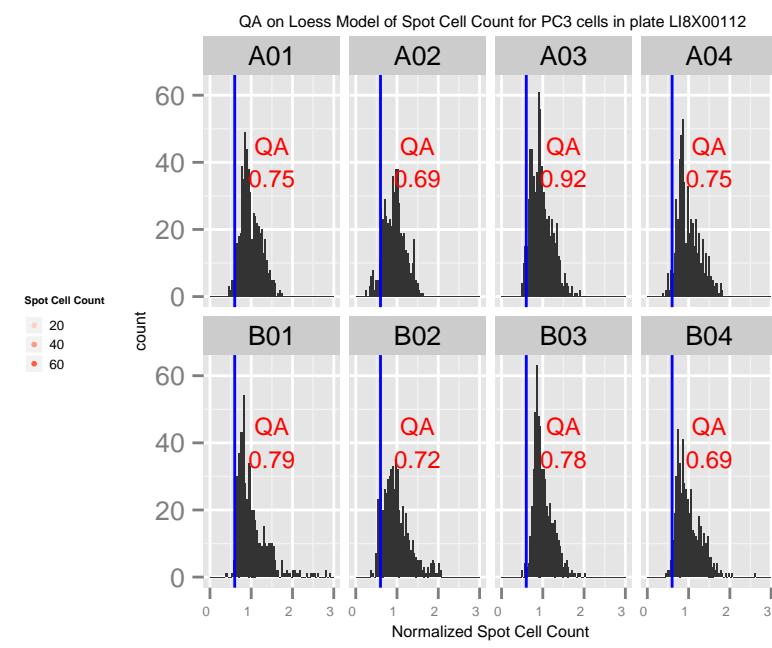
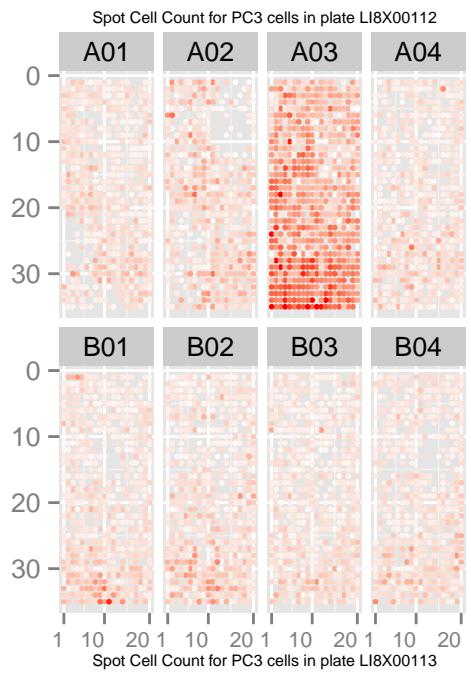


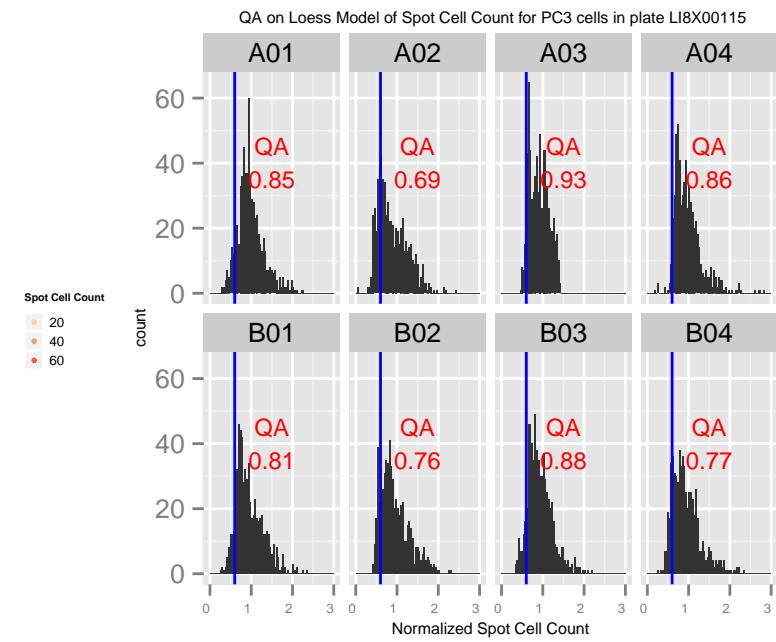
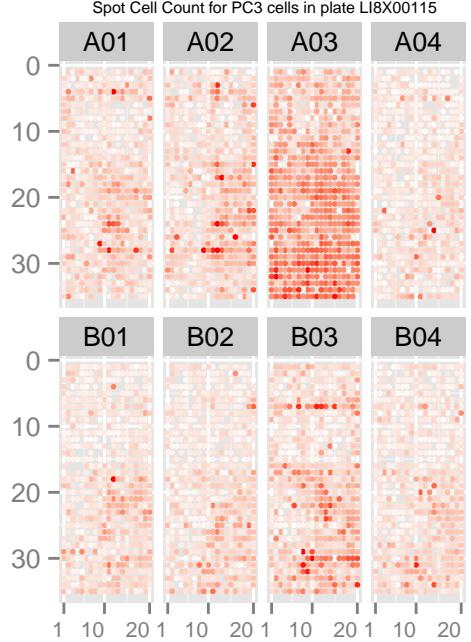
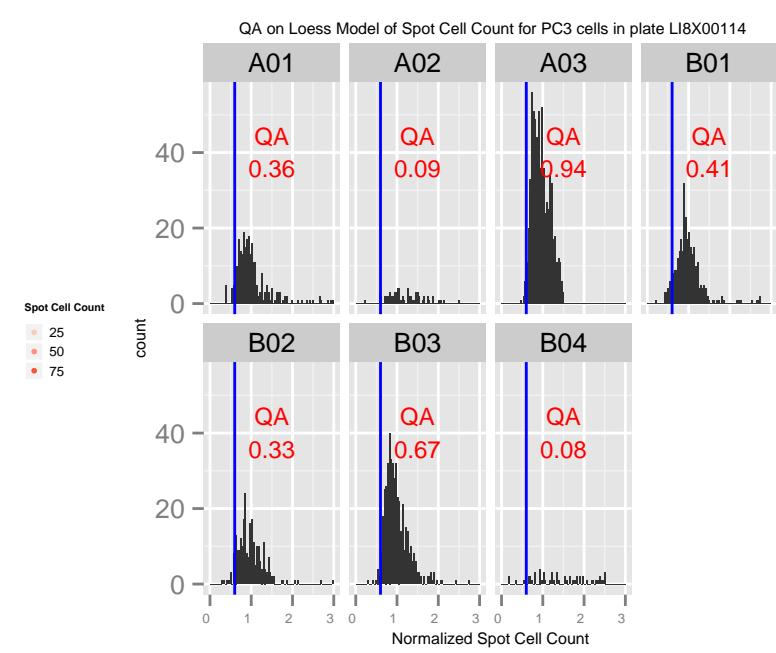
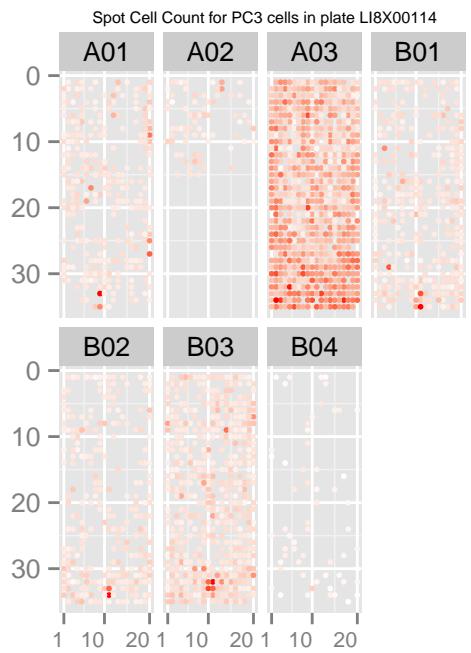
QA Scoring of the dataset

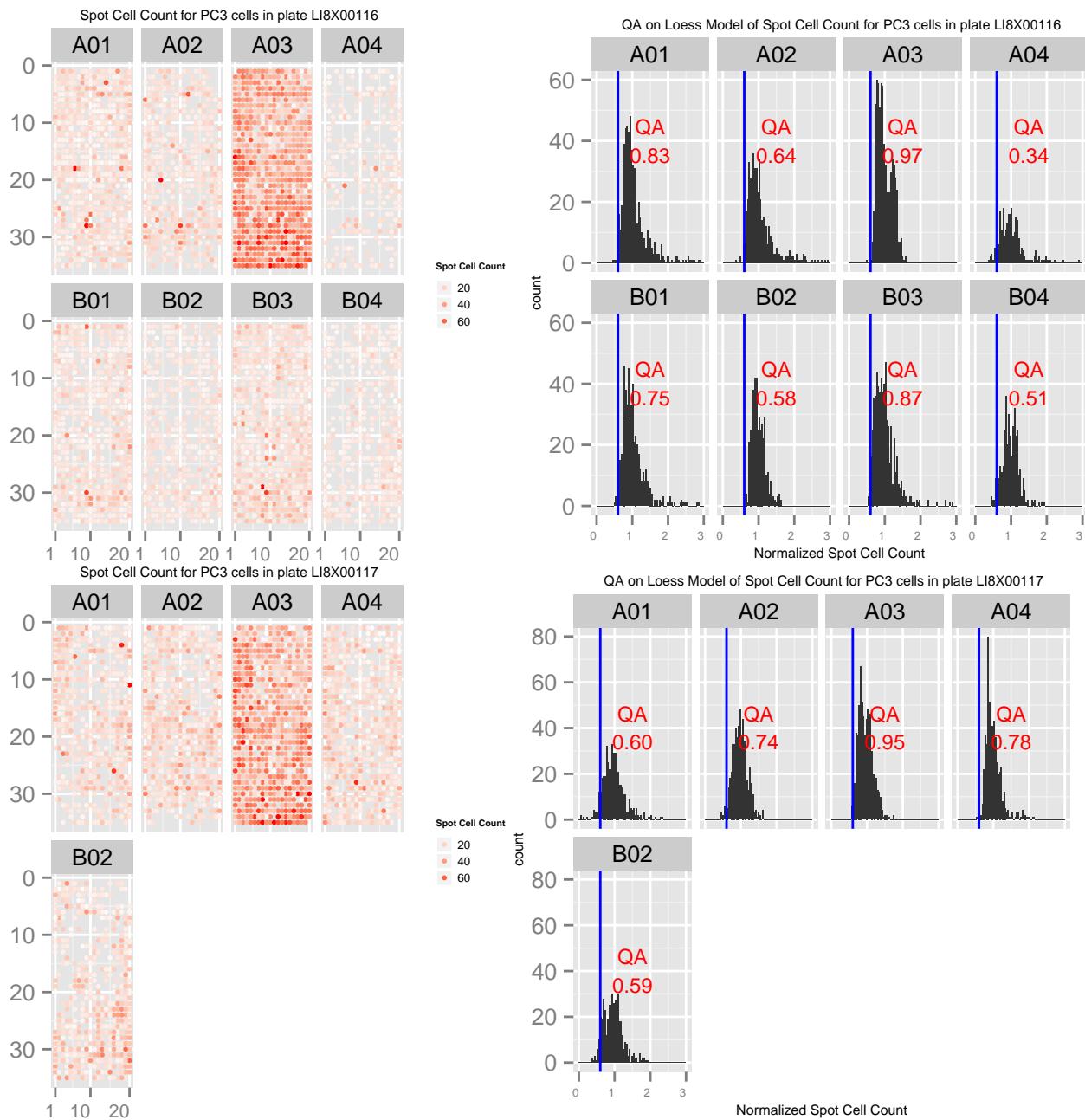
Each well is scored for even cell seeding according to the count of the DAPI-stained nuclei. A detailed explanation of the QA method is in the supplemental material. In brief, the level 2 and 3 data have cell counts at the spot level and locally-averaged cell counts at the neighborhood level. Both of these parameters are used to score the wells and filter the dataset. QA Scores range from 0 to 1 and represent the proportion of the spots that have at least one cell and are not in low cell count neighborhoods.

The following plots are pseudoimages each MEMA's spot cell count and a histogram of the loess model used for QA scoring.





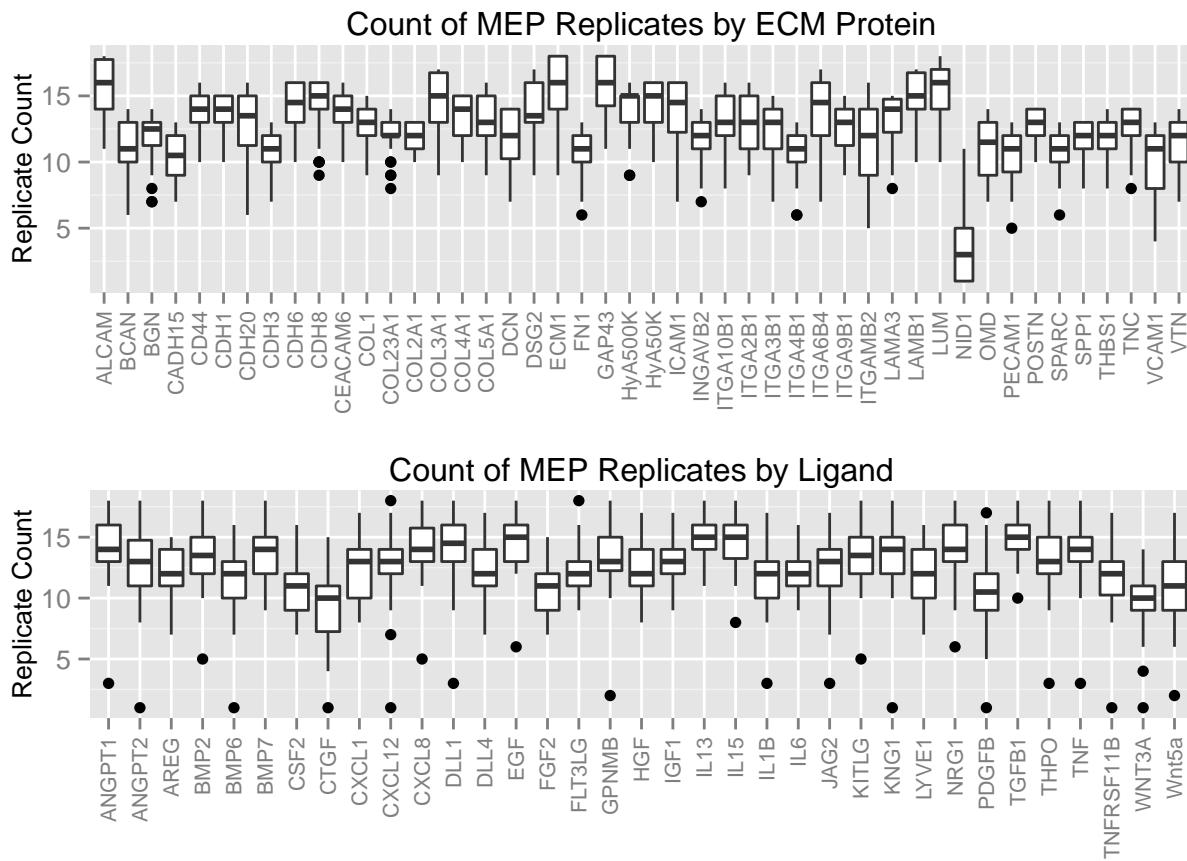




Filtering

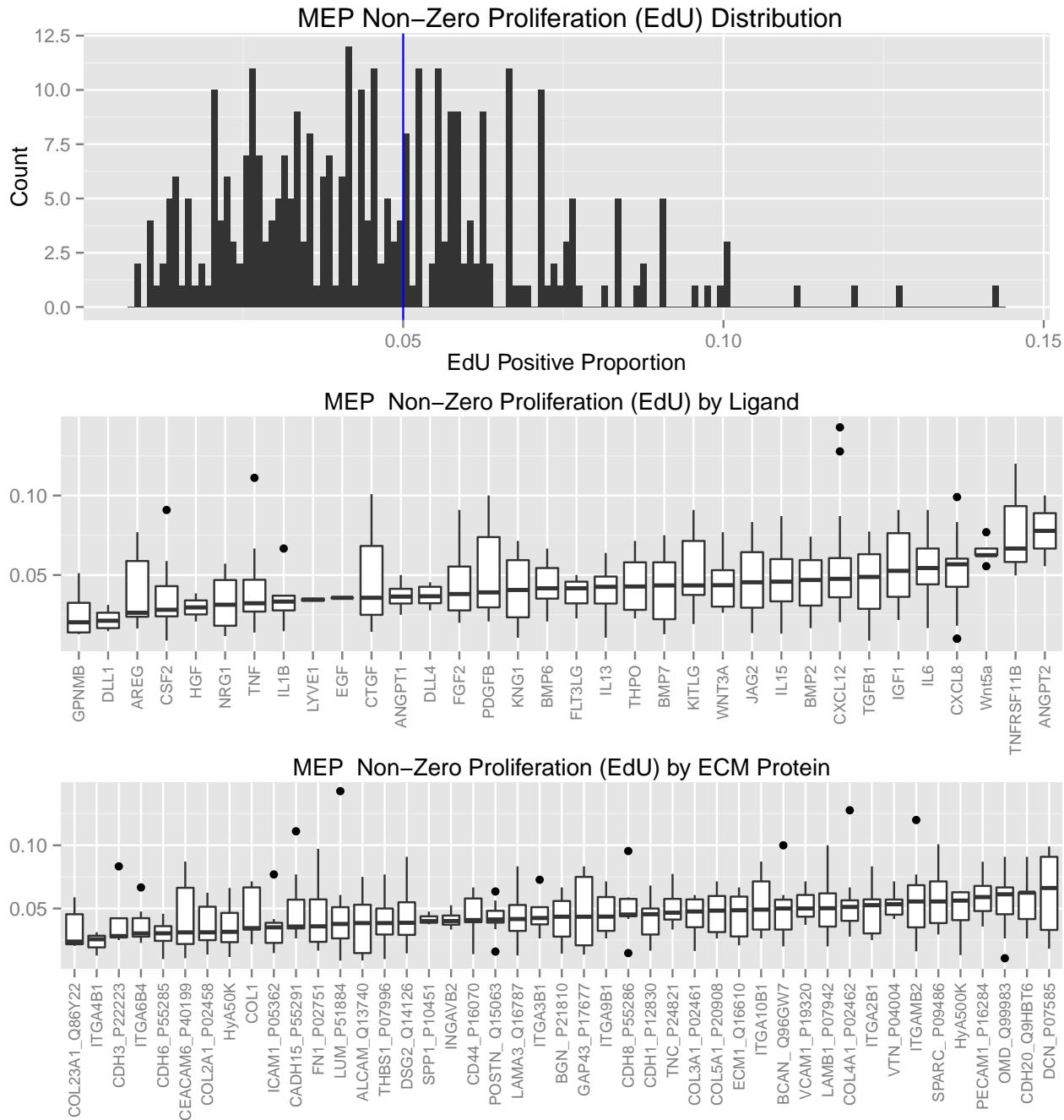
Wells with QA scores below 0.6 and the HighSerum control wells are removed from further analysis of the dataset. After filtering on the well QA score there are 38 ligands in the dataset.

Each spot represents a MEP that is a pairwise combination of the ECM protein printed at a spot and the ligand in the solution of the well. The number of replicate MEPs after removing low-quality wells are shown below.



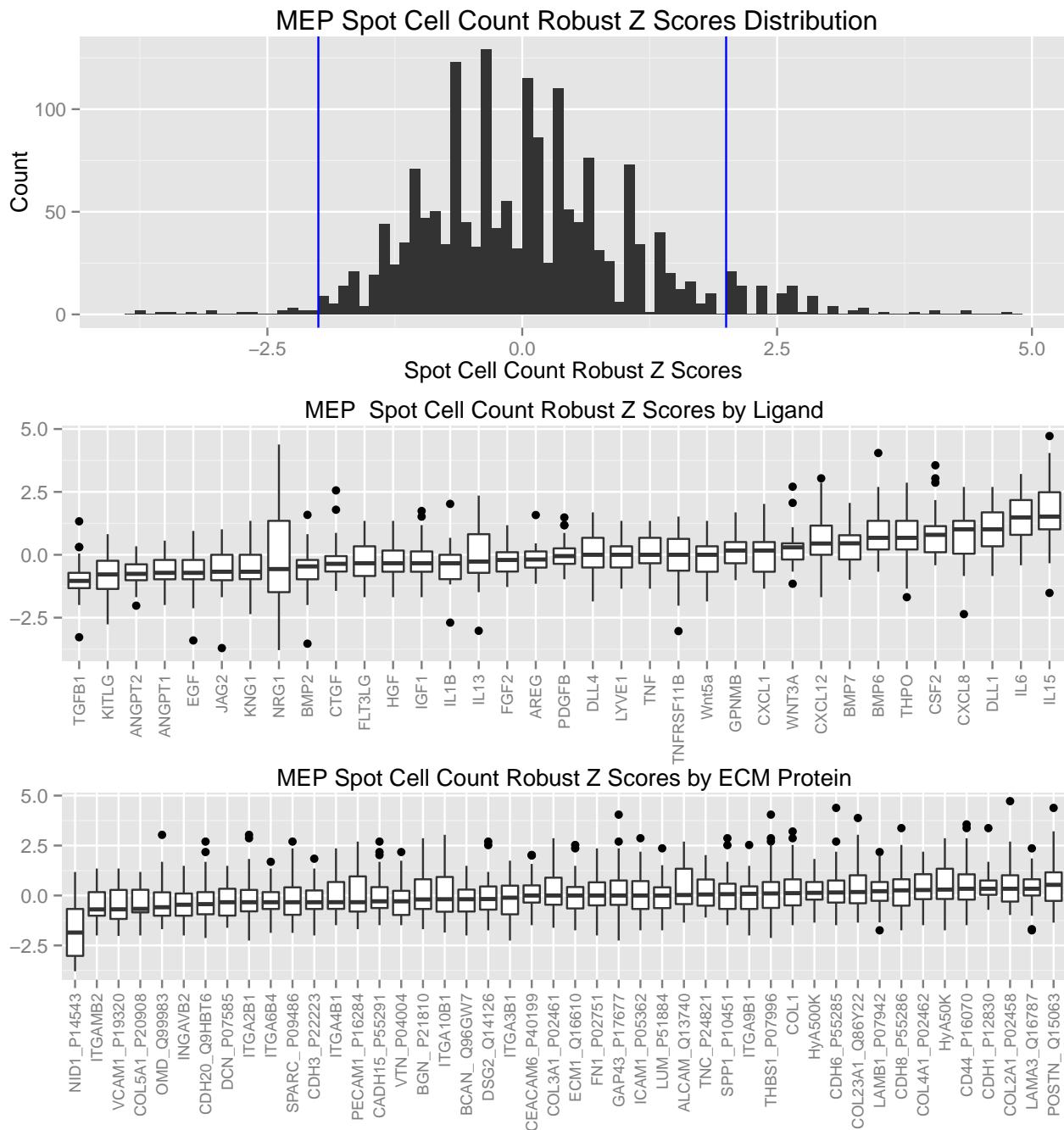
EdU-based Proliferation

Proliferation can be measured as the proportion of cells in S Phase by introducing EdU during the last hour of the assay. The plot below shows the distribution of the proportion of EdU positive cells that are non-zero stratified by MEP. The blue line shows an EdU positive proportion of 0.05. The proliferation of the MEPs is also shown by ligand and ECM protein. A listing of the MEPs with proliferation proportions above 0.05 is in the supplemental material. The dataset for these analyses has been filtered for MEPS with at least six replicates.



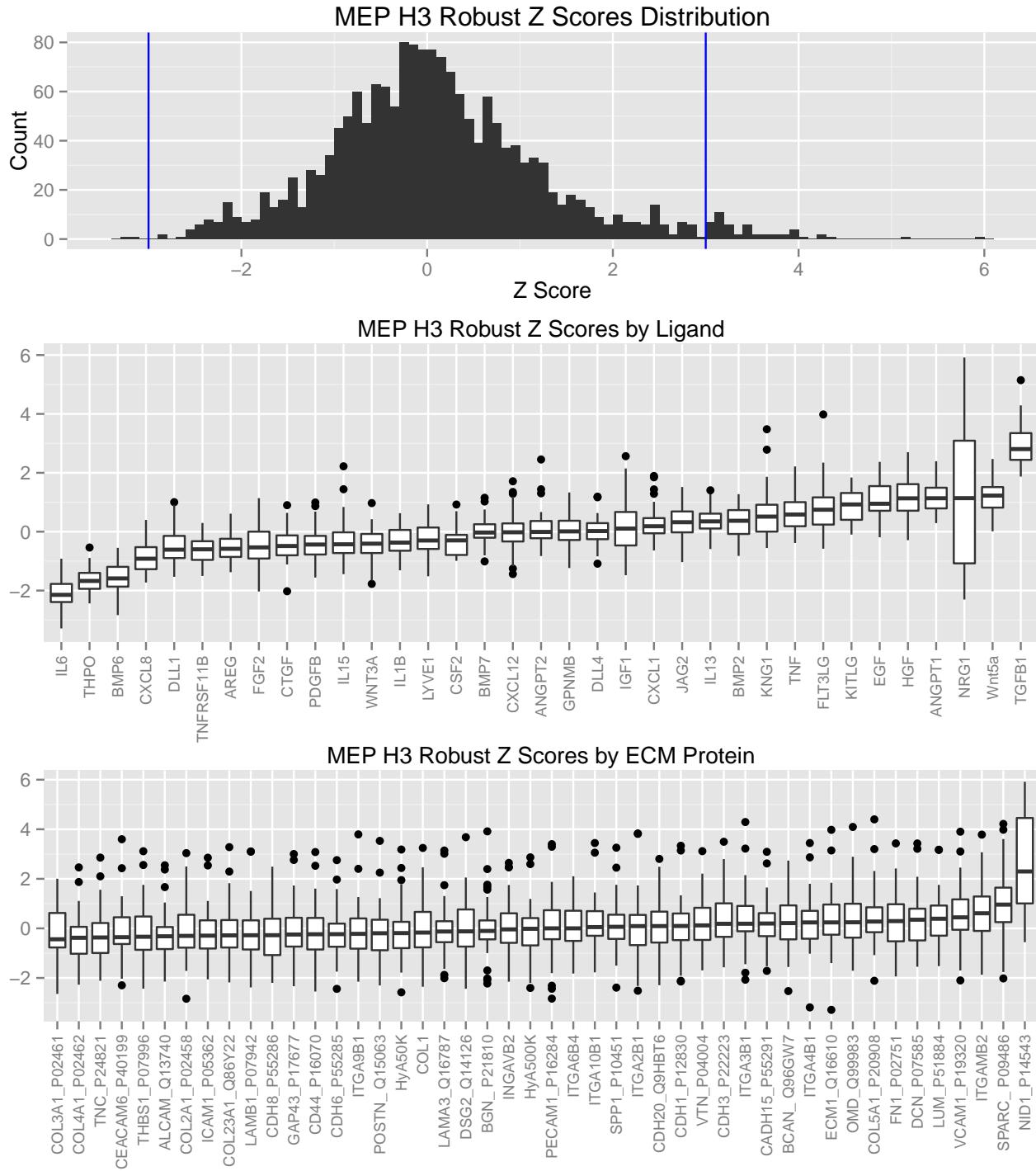
Spot Cell Count Analysis

The spot cell count analysis uses robust Z scores to identify MEPs with extreme population sizes. First, the count of cells at each spot is normalized by dividing it by the median cell count of all spots in the same plate's high serum well. The normalized spot cell counts are then summarized by the median of their replicates. The median and mad of the distribution of normalized and summarized values are used to convert to robust Z scores and are shown below. The blue lines at +/- 2 show thresholds for selecting MEPs of interest. Below the distribution plot are plots with Z scores stratified by ligand and ECM protein. A listing of the MEPs outside of the blue lines is in the supplemental material.



H3K9me3 Response

The following plots look at the plate normalized responses of the H3K9me3 signal stratified by ligand, ECM protein and MEP. The distribution of the Z scores is shown with blue lines highlighting the +/-3 robust z score positions. The MEPs outside of the blue lines are listed in the supplemental material.



Supplemental Material

Quality Analysis

The variance of the signal in MEMA data comes from biological and technical factors. The technical factors create regions of low cell counts per spot and uneven staining across the array. The goal of the QA pipeline is to quantify the technical factors to identify wells or plates that need to be removed from downstream processing and/or be replaced by wells from a new experiment.

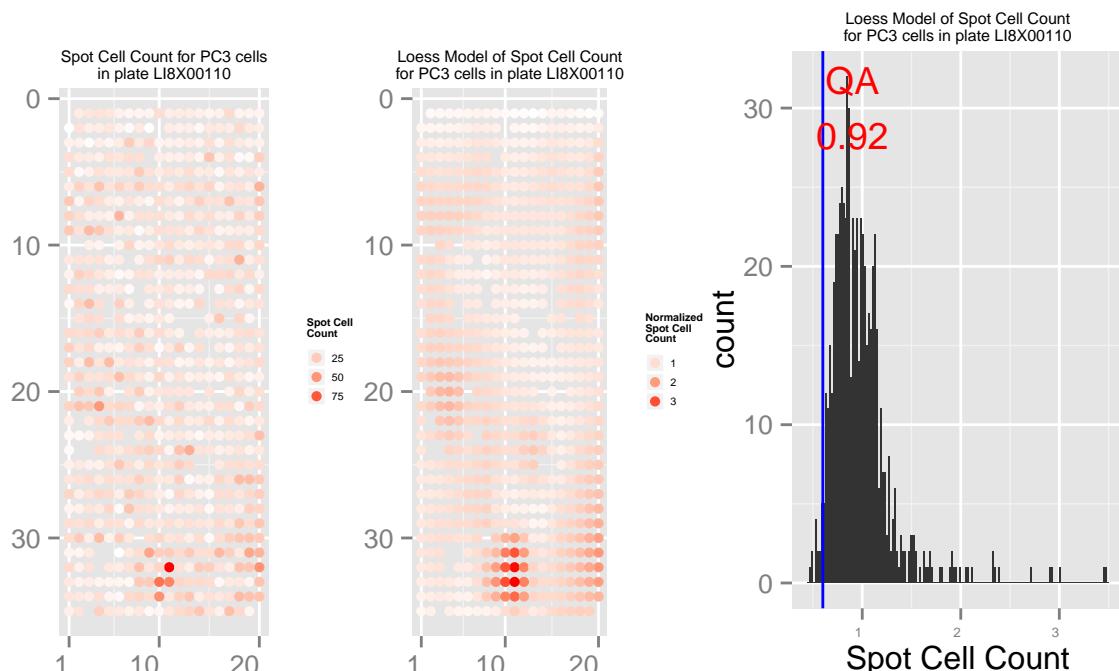
The hypothesis for the MEMA QA process is that the biological signal comes from individual spots while the technical variations come from regions of low signal. A bivariate loess model can be used to quantify the number of spots in low signal regions, leading to a MEMA QA score.

Loess Model Explanation

The loess model of a MEMA is the mean value of a weighted version of each spot's region or neighborhood. In a 700 spot array, a loess span value of 0.1 sets the size of the neighborhood to be the nearest 70 points (within approximately 5 spots in all directions). The weights are a tricubic function of the euclidean distance between the spot being modeled and the neighborhood spots. These weights vary from 1 to 0 as distances increase from the nearest to the farthest neighbor. In other words, each spot in the model takes on the mean value of its 70 nearest neighbors with the closest neighbors having the largest impact. Therefore, the loess model is dominated by the technical regional factors as opposed to individual biological responses.

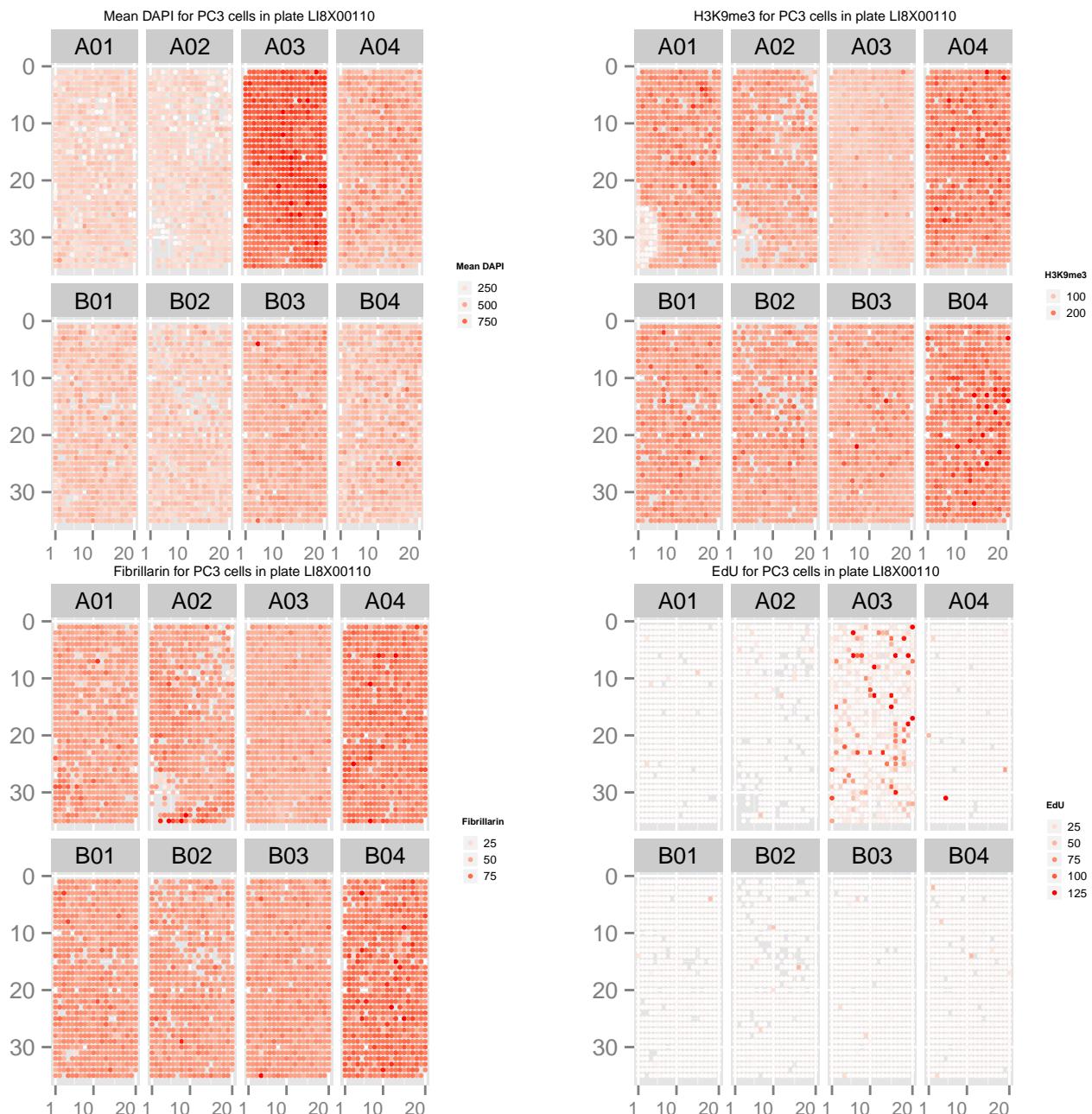
A MEMA's QA score is derived from the loess model of the control-well-normalized values by calculating the proportion of spots in low signal regions(LSR). A threshold for classifying spots as LSR is based on the median of each plate's control well. To have higher scores reflect increasing quality, the MEMA QA score is defined as the proportion of non-LSR spots to total spots. This value will be 1 for MEMAs with no low signal regions and approach 0 as the number of LSR spots increases.

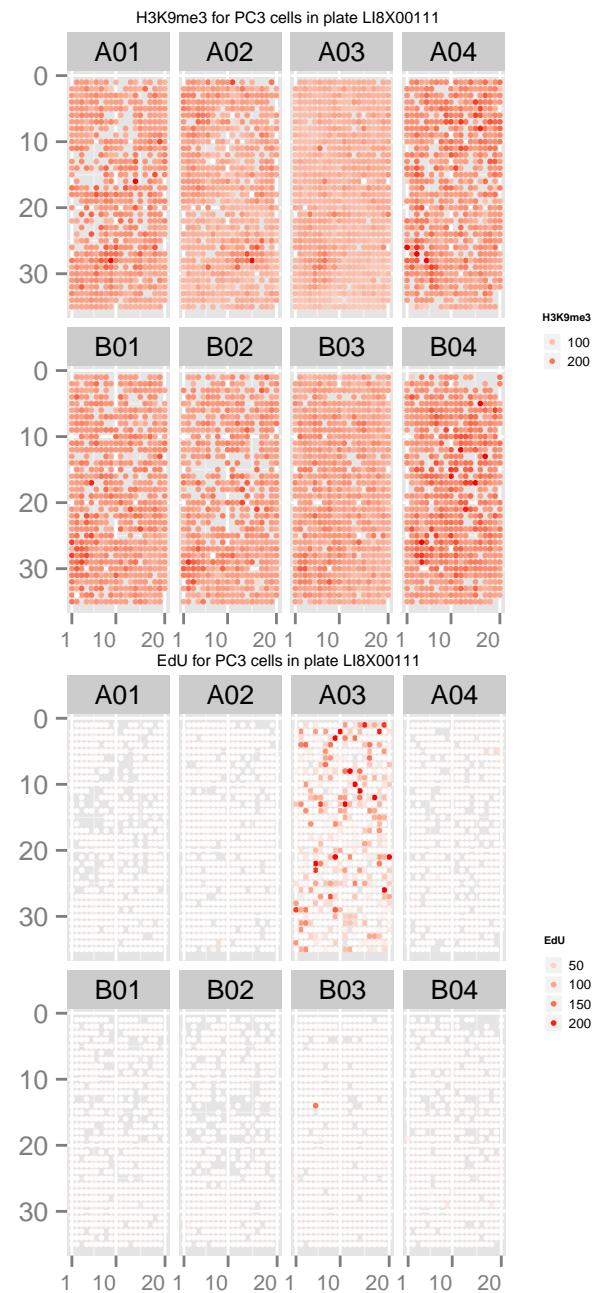
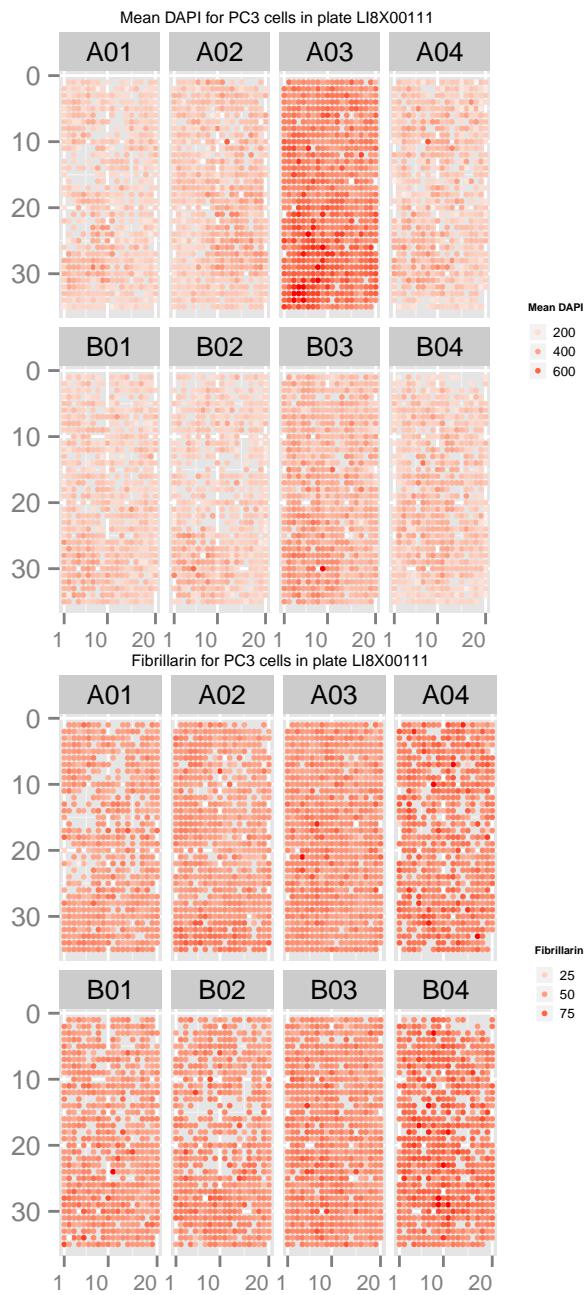
Below are plots showing data from well A02 from plate LI8X00101 from LINCs staining set 1 (DAPI, phalloidin, CellMask and MitoTracker). The LSR spots are those to the left of the blue vertical line at the threshold value of 0.6 in the histogram.

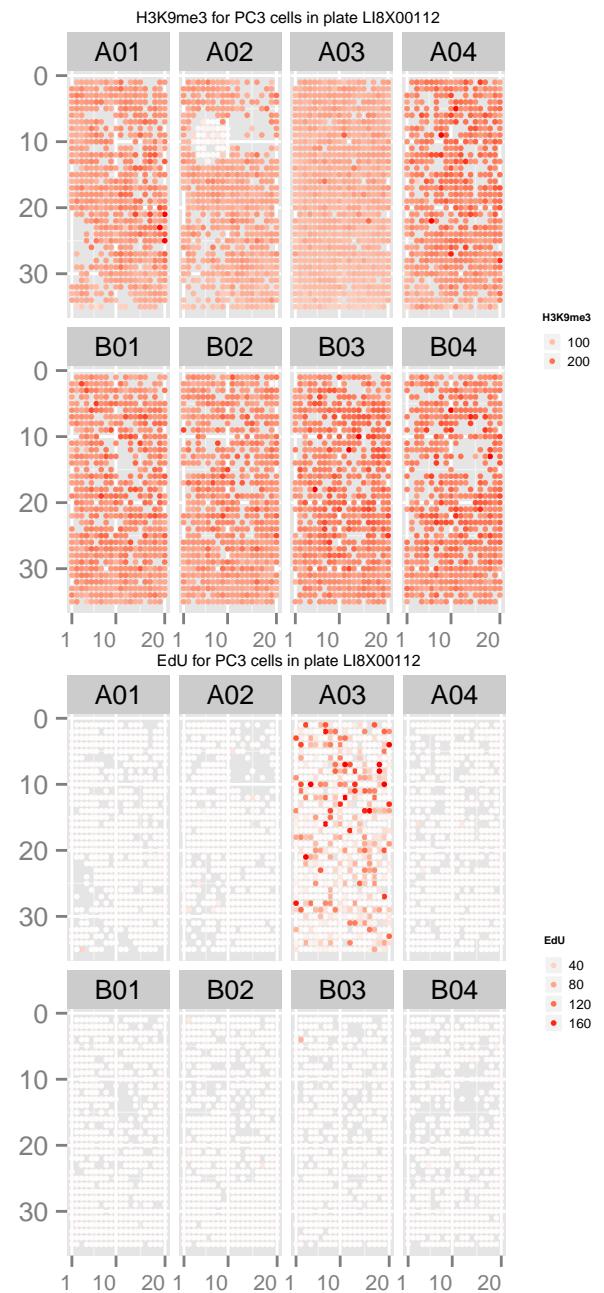
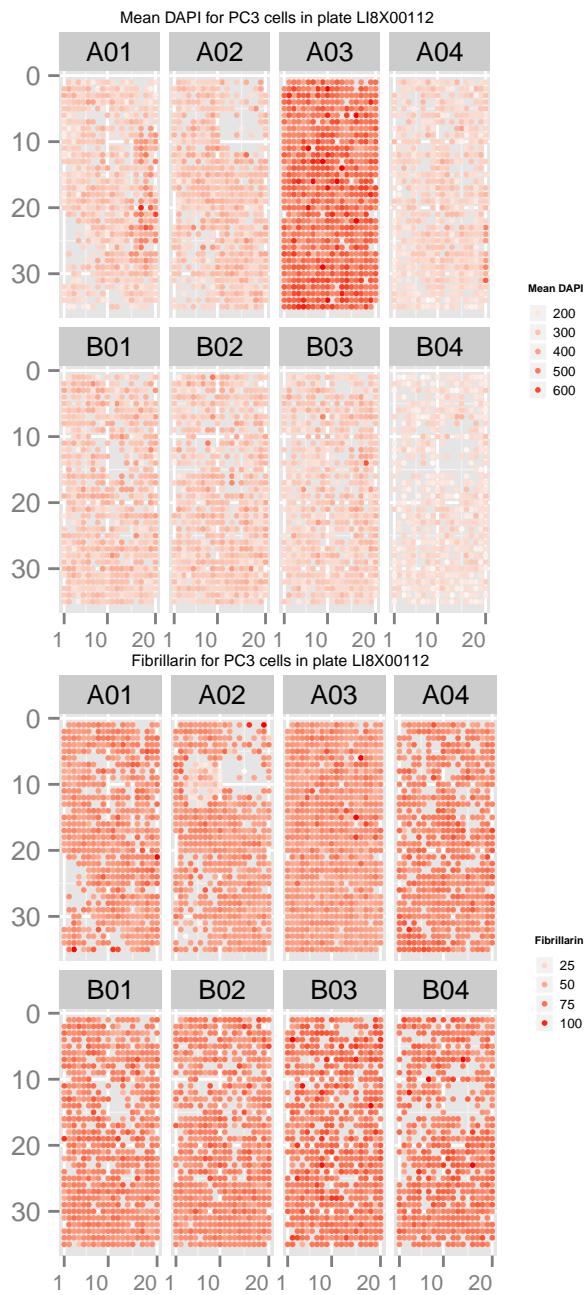


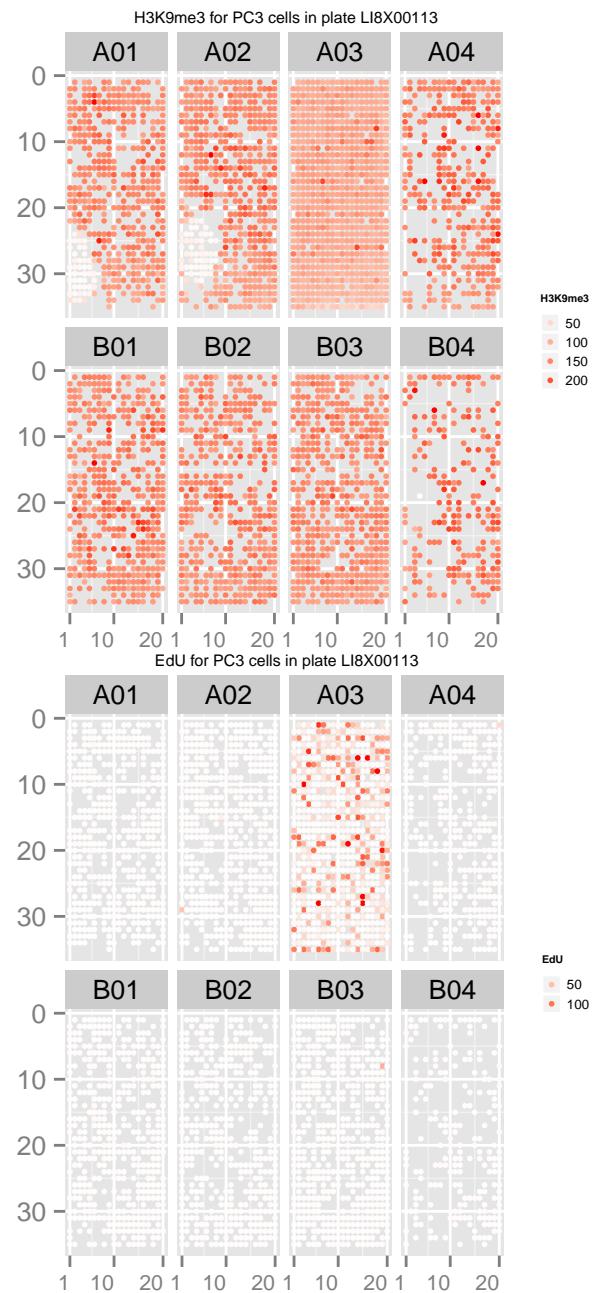
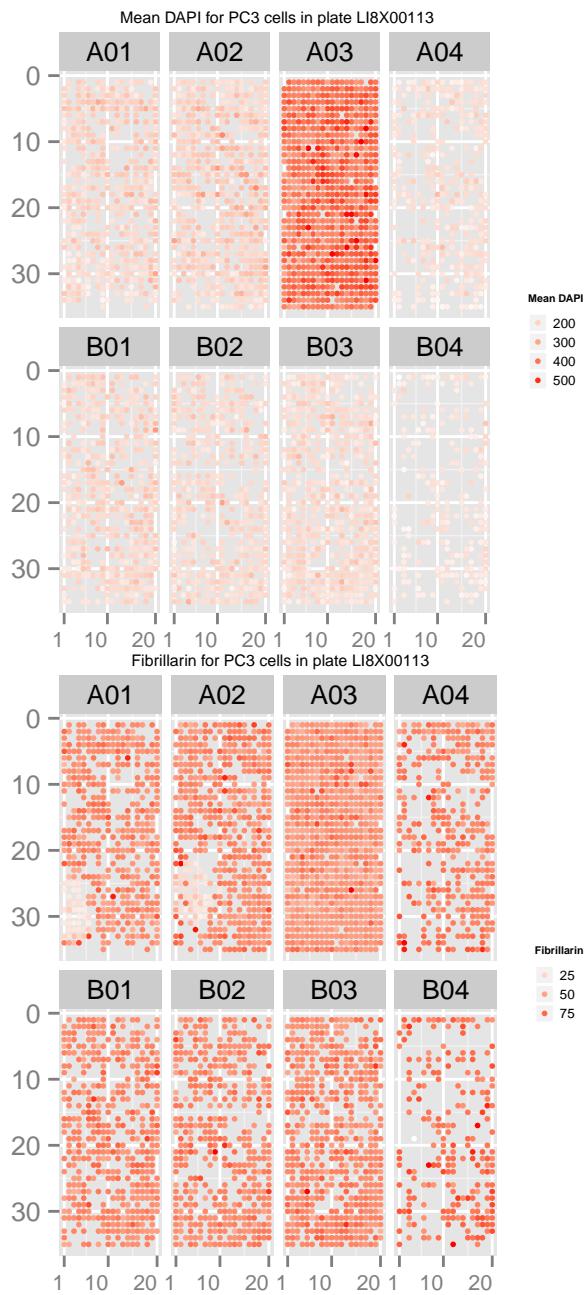
Stain Pseudoimages

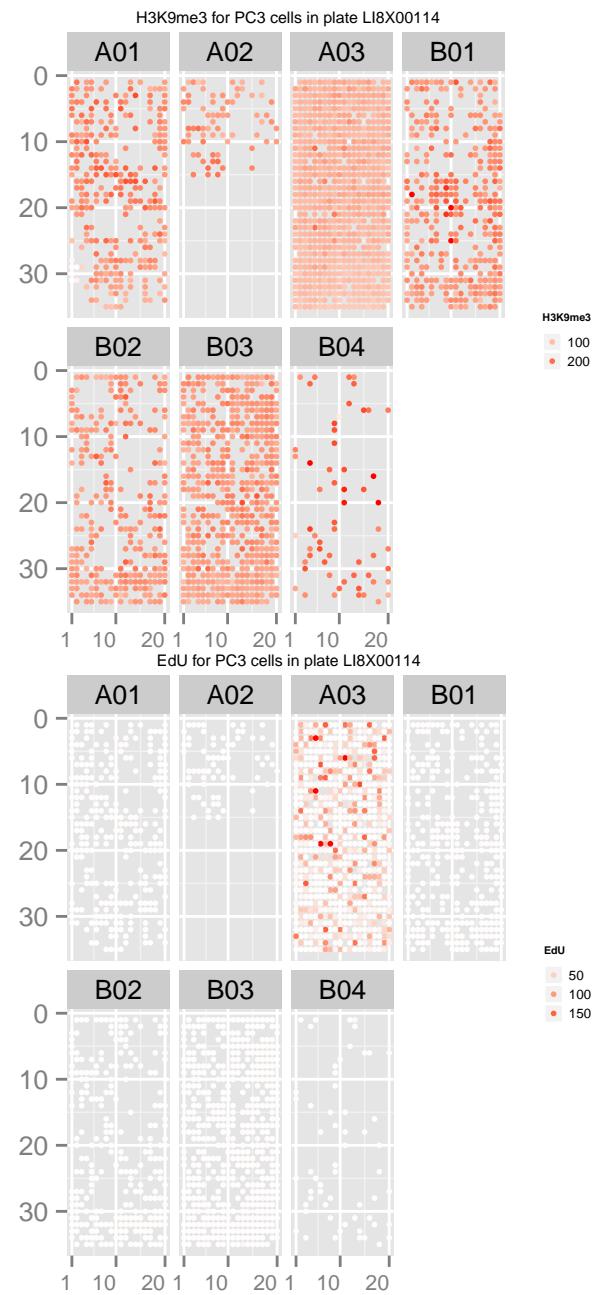
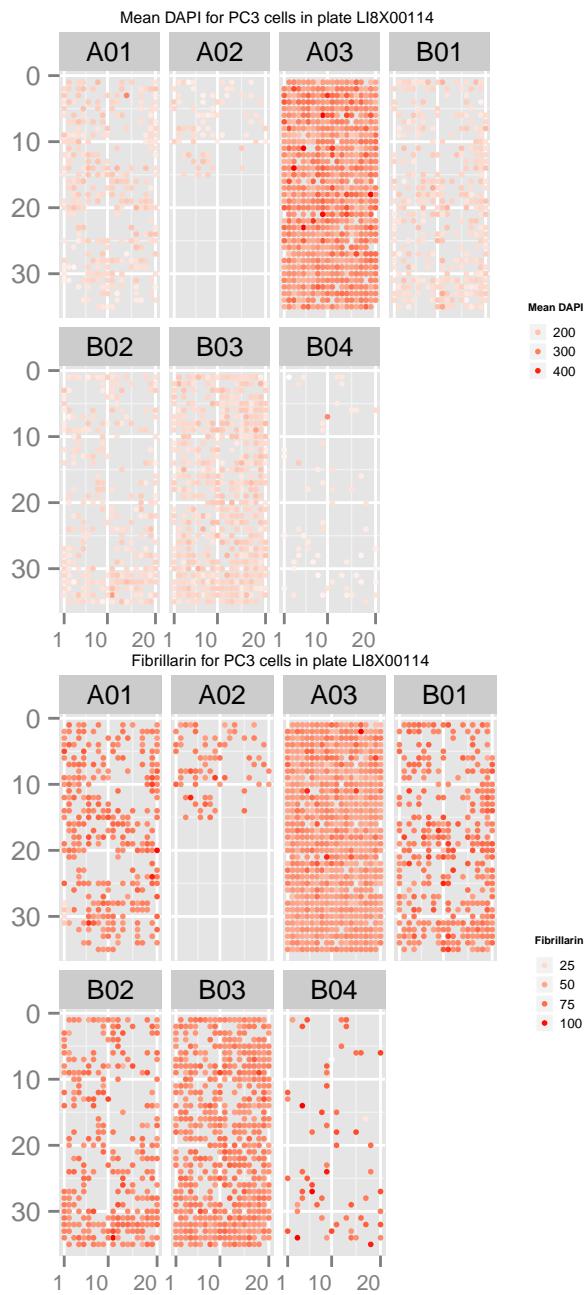
The pseudoimages of each well's raw signals are shown in the plots below. Wells that could not be successfully imaged due to focus issues are missing from the pseudoimages.

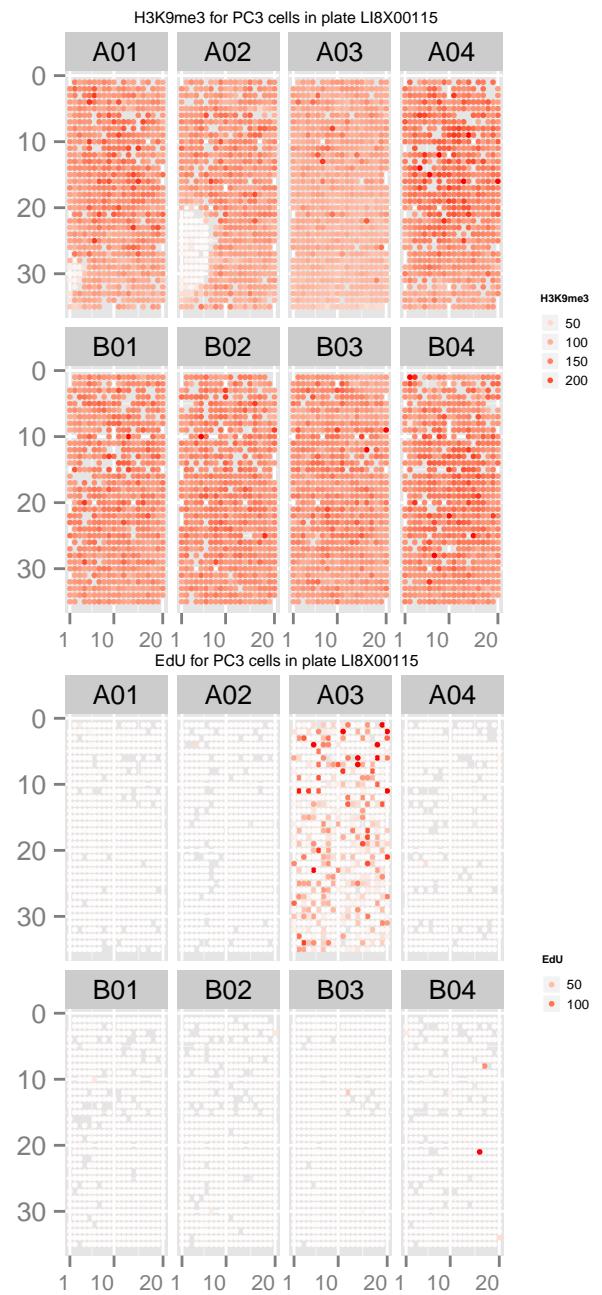
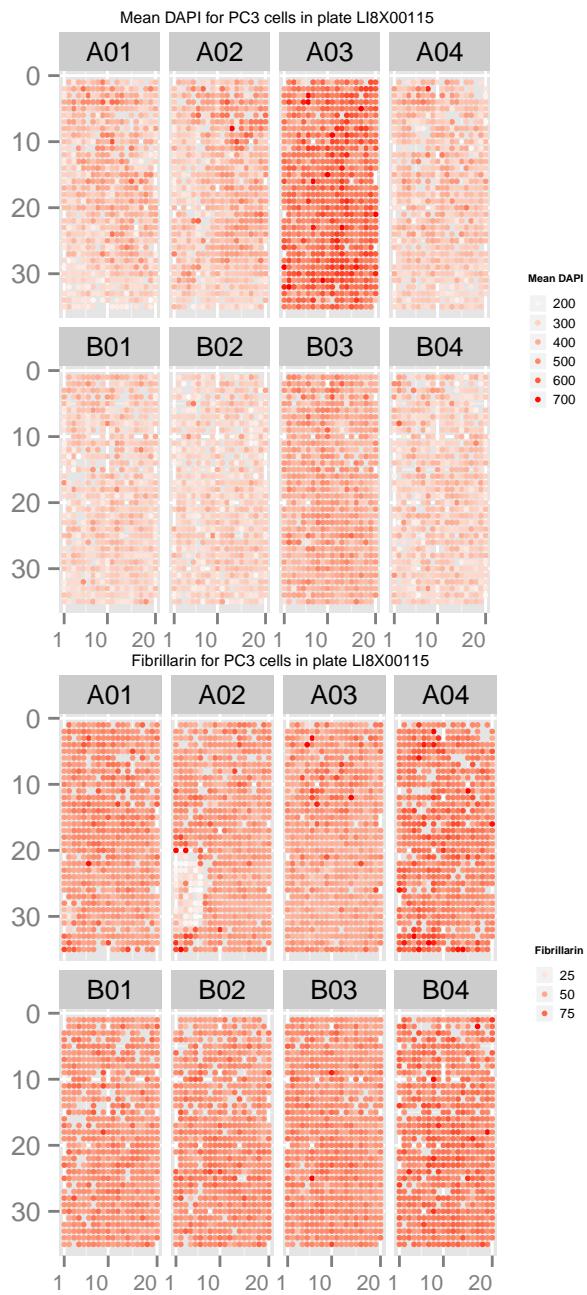


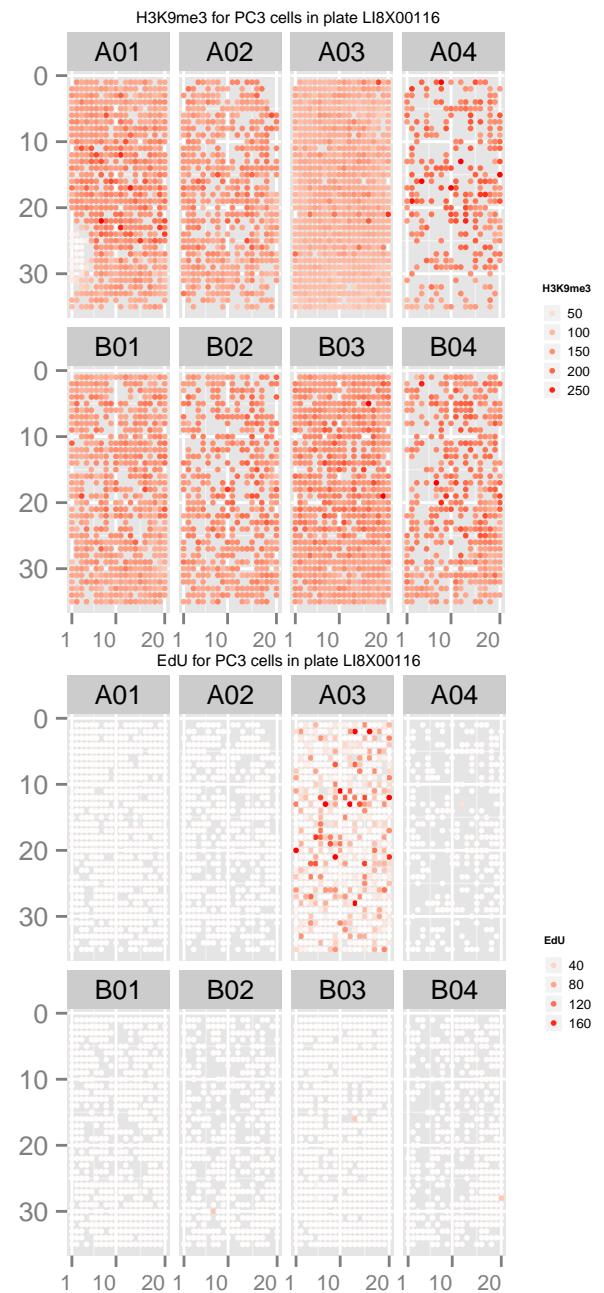
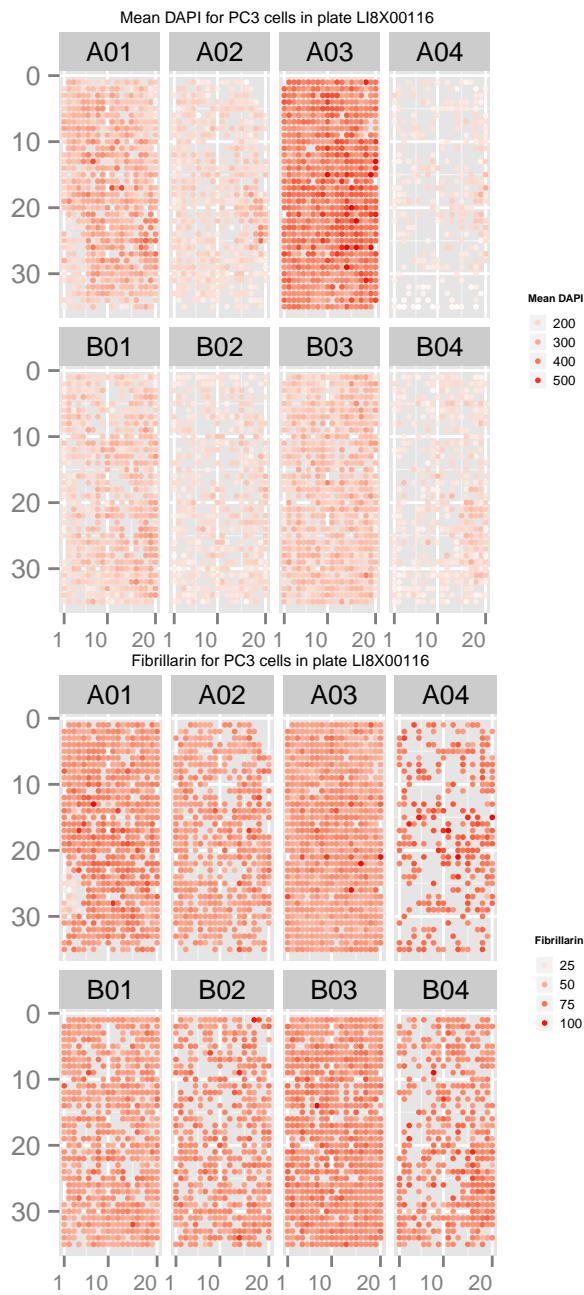


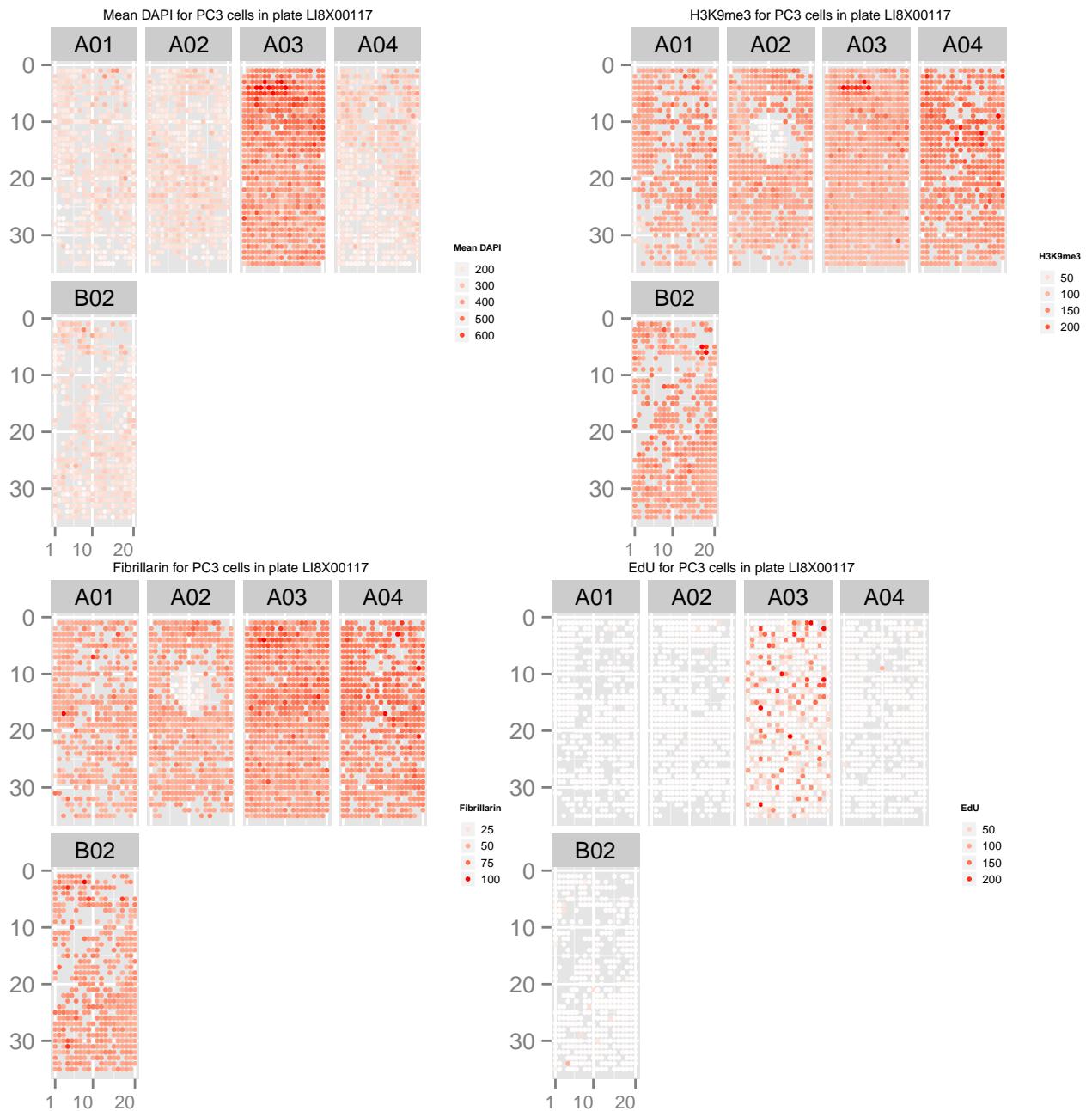












Extreme EdU Proliferation MEPs

Ligand	ECMP	EdUPositiveProportion
ANGPT1	CDH1	0.05
BMP2	BCAN	0.05
CXCL12	ALCAM	0.05
CXCL12	COL2A1	0.05
FGF2	VCAM1	0.05
FLT3LG	THBS1	0.05
KNG1	HyA500K	0.05
NRG1	VTN	0.05
GPNMB	VTN	0.05
CXCL8	PECAM1	0.05
BMP6	ITGAMB2	0.05
CXCL12	LUM	0.05
CXCL8	LAMA3	0.05
IGF1	BGN	0.05
IL6	BCAN	0.05
IL6	INGAVB2	0.05
KITLG	COL4A1	0.05
PDGFB	ITGA2B1	0.05
CXCL12	LAMA3	0.05
NRG1	ALCAM	0.05
CXCL8	HyA50K	0.05
IL6	COL4A1	0.05
ANGPT2	VTN	0.06
BMP2	BGN	0.06
BMP2	COL3A1	0.06
BMP7	COL2A1	0.06
CXCL12	VTN	0.06
CXCL8	ALCAM	0.06
CXCL8	LAMB1	0.06
IL6	ECM1	0.06
IL6	SPARC	0.06
WNT3A	ITGA9B1	0.06
Wnt5a	PECAM1	0.06
CXCL8	POSTN	0.06
BMP6	THBS1	0.06
IL15	BCAN	0.06
IL15	CDH8	0.06
IL15	BGN	0.06
NRG1	COL4A1	0.06
CXCL12	COL5A1	0.06
CXCL8	ITGA2B1	0.06
CXCL8	TNC	0.06
IL13	CD44	0.06
IL15	VTN	0.06
CXCL8	VTN	0.06
AREG	DCN	0.06
BMP7	COL23A1	0.06
CSF2	ITGAMB2	0.06
CXCL12	TNC	0.06
CXCL8	CDH1	0.06

Ligand	ECMP	EdUPositiveProportion
CXCL8	CDH8	0.06
IL15	COL3A1	0.06
IL6	LAMB1	0.06
IL6	OMD	0.06
CXCL8	ITGA10B1	0.06
CXCL8	CD44	0.06
BMP2	ECM1	0.06
CXCL12	COL3A1	0.06
CXCL8	BCAN	0.06
BMP2	LUM	0.06
CXCL12	FN1	0.06
IL15	OMD	0.06
BMP6	CDH20	0.06
BMP6	OMD	0.06
IL15	CDH20	0.06
IL6	BGN	0.06
IL6	COL2A1	0.06
KNG1	PECAM1	0.06
THPO	SPARC	0.06
Wnt5a	HyA500K	0.06
Wnt5a	ITGA9B1	0.06
TNF	POSTN	0.06
IL13	HyA500K	0.06
IL15	HyA50K	0.07
IL1B	ITGAMB2	0.07
BMP6	PECAM1	0.07
CXCL12	CD44	0.07
CXCL12	OMD	0.07
CXCL8	COL4A1	0.07
IL6	GAP43	0.07
PDGFB	COL1	0.07
TNF	ECM1	0.07
TNFRSF11B	ITGA6B4	0.07
Wnt5a	BGN	0.07
IL6	COL5A1	0.07
BMP2	CDH1	0.07
CXCL8	ITGAMB2	0.07
BMP7	PECAM1	0.07
CXCL8	COL1	0.07
CXCL8	SPARC	0.07
FGF2	COL5A1	0.07
IGF1	VTN	0.07
IL15	LAMB1	0.07
IL6	ITGA9B1	0.07
KITLG	ITGA10B1	0.07
KNG1	VCAM1	0.07
THPO	LAMA3	0.07
IL6	ITGA3B1	0.07
CXCL12	SPARC	0.07
IL15	DCN	0.07
BMP2	OMD	0.07
BMP7	ALCAM	0.08

Ligand	ECMP	EdUPositiveProportion
IL15	GAP43	0.08
CXCL8	CEACAM6	0.08
AREG	CADH15	0.08
CXCL12	THBS1	0.08
IL6	ITGA10B1	0.08
Wnt5a	ICAM1	0.08
WNT3A	ITGAMB2	0.08
TGFB1	TNC	0.08
IGF1	GAP43	0.08
CXCL12	CDH3	0.08
CXCL8	GAP43	0.08
IL6	ITGA2B1	0.08
IL6	LAMA3	0.08
JAG2	DSG2	0.08
KITLG	PECAM1	0.09
CXCL12	ITGA10B1	0.09
IL15	CEACAM6	0.09
CSF2	DCN	0.09
FGF2	OMD	0.09
IGF1	DCN	0.09
IL6	CDH20	0.09
KITLG	DSG2	0.09
PDGFB	CDH8	0.10
PDGFB	FN1	0.10
CXCL8	DCN	0.10
ANGPT2	LAMB1	0.10
PDGFB	BCAN	0.10
CTGF	SPARC	0.10
TNF	CADH15	0.11
TNFRSF11B	ITGAMB2	0.12
CXCL12	COL4A1	0.13
CXCL12	LUM	0.14

Extreme Spot Cell Count MEPs

Ligand	ECMP	SpotCellCount	MedNorm	RobustZ	SpotCellCount
NRG1	NID1			-3.79	1.0
JAG2	NID1			-3.71	1.0
BMP2	NID1			-3.53	2.0
EGF	NID1			-3.41	2.5
TGFB1	NID1			-3.28	3.0
TNFRSF11B	NID1			-3.04	3.0
IL13	NID1			-3.02	4.0
KITLG	NID1			-2.77	5.0
IL1B	NID1			-2.70	4.0
CXCL8	NID1			-2.36	5.0
KNG1	NID1			-2.36	5.0
NRG1	GAP43			-2.25	7.0
NRG1	ITGA2B1			-2.25	7.0
NRG1	ITGA3B1			-2.25	7.0
EGF	THBS1			-2.13	7.5
NRG1	CDH20			-2.13	7.5
ANGPT2	NID1			-2.02	6.0
TNFRSF11B	VCAM1			-2.02	6.0
CXCL12	COL2A1			2.00	17.5
IL6	CEACAM6			2.00	17.5
CXCL1	COL4A1			2.02	18.0
CXCL8	COL2A1			2.02	18.0
CXCL8	ITGA10B1			2.02	18.0
CXCL8	TNC			2.02	18.0
DLL1	CEACAM6			2.02	18.0
DLL1	SPARC			2.02	18.0
IL15	CEACAM6			2.02	18.0
IL15	COL4A1			2.02	18.0
IL15	HyA50K			2.02	18.0
IL1B	HyA50K			2.02	18.0
NRG1	CADH15			2.02	18.0
NRG1	CDH8			2.02	18.0
NRG1	FN1			2.02	18.0
NRG1	GAP43			2.02	18.0
NRG1	LAMB1			2.02	18.0
THPO	HyA50K			2.02	18.0
THPO	POSTN			2.02	18.0
BMP7	CD44			2.06	19.0
WNT3A	ITGA4B1			2.06	19.0
CSF2	COL1			2.18	18.0
CSF2	VTN			2.18	18.0
CXCL12	THBS1			2.18	18.0
IL6	CADH15			2.18	18.0
IL6	CDH20			2.18	18.0
IL6	CDH8			2.18	18.0
IL6	GAP43			2.18	18.0
IL6	ITGA4B1			2.18	18.0
IL6	LAMB1			2.18	18.0
CXCL8	BGN			2.19	18.5
DLL1	CDH6			2.19	18.5

Ligand	ECMP	SpotCellCount	MedNormRobustZ	SpotCellCount
NRG1	ICAM1		2.19	18.5
THPO	ALCAM		2.19	18.5
NRG1	COL4A1		2.19	18.5
CXCL12	POSTN		2.35	18.5
IL13	ITGA10B1		2.35	25.0
CXCL8	ALCAM		2.36	19.0
CXCL8	COL1		2.36	19.0
DLL1	ECM1		2.36	19.0
DLL1	FN1		2.36	19.0
IL15	COL23A1		2.36	19.0
IL15	LUM		2.36	19.0
IL15	SPARC		2.36	19.0
NRG1	CD44		2.36	19.0
NRG1	COL3A1		2.36	19.0
NRG1	ITGA4B1		2.36	19.0
NRG1	LAMA3		2.36	19.0
THPO	GAP43		2.36	19.0
IL6	DSG2		2.52	19.0
IL6	ITGA10B1		2.52	19.0
IL6	SPP1		2.52	19.0
BMP6	ECM1		2.53	19.5
DLL1	CDH8		2.53	19.5
IL15	ITGA9B1		2.53	19.5
NRG1	ALCAM		2.53	19.5
BMP6	COL1		2.53	19.5
CXCL8	POSTN		2.53	19.5
CTGF	THBS1		2.56	21.5
IL6	COL23A1		2.69	19.5
BMP6	BGN		2.70	20.0
CXCL8	PECAM1		2.70	20.0
DLL1	POSTN		2.70	20.0
IL15	ALCAM		2.70	20.0
IL15	CDH6		2.70	20.0
IL15	DSG2		2.70	20.0
NRG1	HyA50K		2.70	20.0
NRG1	SPARC		2.70	20.0
THPO	CDH20		2.70	20.0
DLL1	CADH15		2.70	20.0
DLL1	GAP43		2.70	20.0
NRG1	COL2A1		2.70	20.0
THPO	COL23A1		2.70	20.0
WNT3A	THBS1		2.71	21.0
CSF2	BGN		2.87	20.0
CXCL12	COL1		2.87	20.0
CXCL12	COL3A1		2.87	20.0
CXCL12	HyA50K		2.87	20.0
CXCL12	SPP1		2.87	20.0
IL15	ICAM1		2.87	20.5
IL15	THBS1		2.87	20.5
IL6	ITGA2B1		2.87	20.0
THPO	CD44		2.87	20.5
IL15	OMD		3.04	21.0

Ligand	ECMP	SpotCellCountMedNormRobustZ	SpotCellCount
NRG1	ITGA2B1	3.04	21.0
CSF2	ITGA10B1	3.04	20.5
CXCL12	COL23A1	3.04	20.5
NRG1	COL1	3.20	21.5
IL6	POSTN	3.21	21.0
IL15	CD44	3.37	22.0
IL15	CDH1	3.37	22.0
IL15	CDH8	3.37	22.0
CSF2	CD44	3.56	22.0
NRG1	COL23A1	3.88	23.5
BMP6	THBS1	4.05	24.0
IL15	GAP43	4.05	24.0
NRG1	CDH6	4.38	25.0
NRG1	POSTN	4.38	25.0
IL15	COL2A1	4.72	26.0

Extreme H3 MEPs

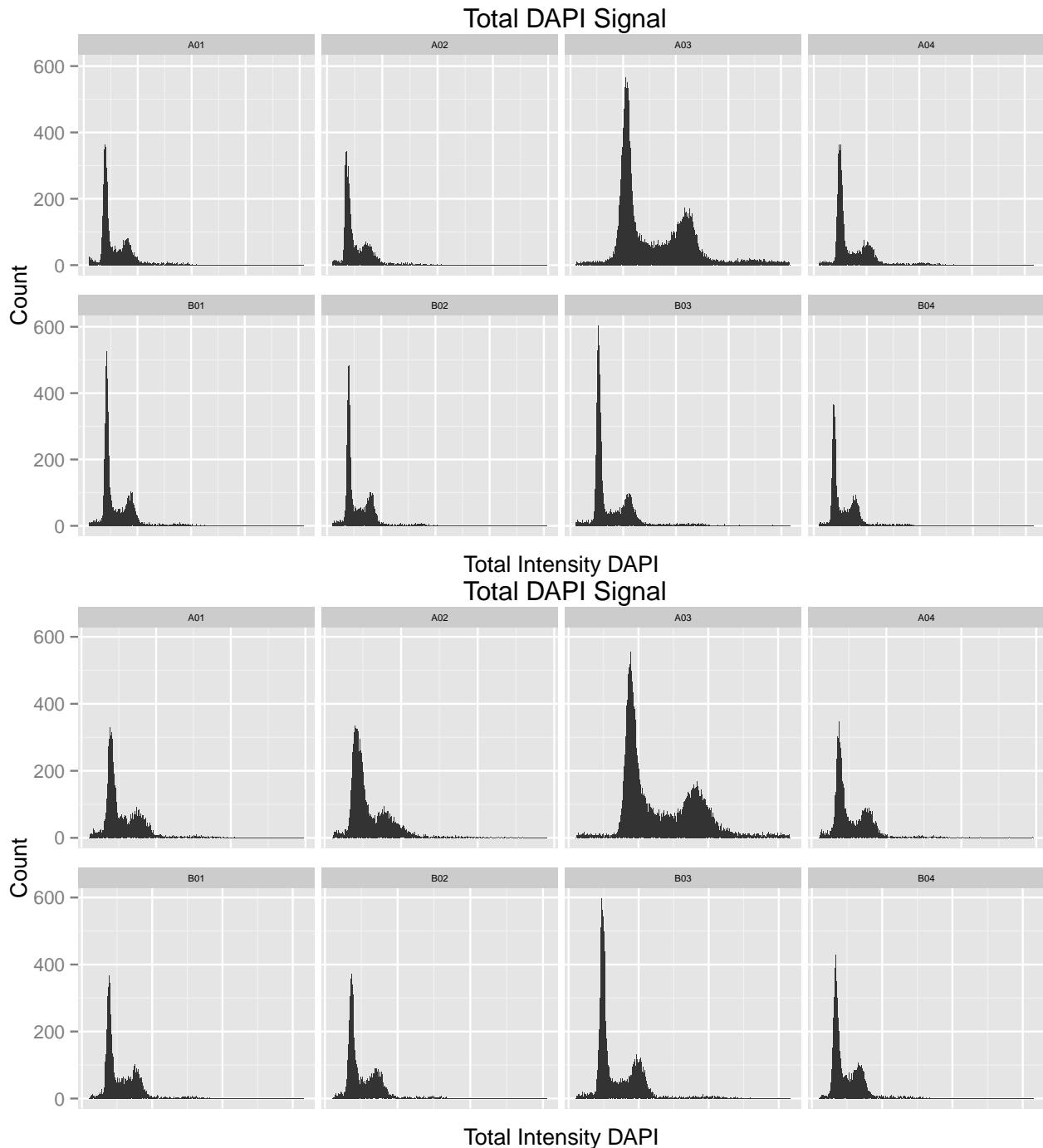
The following lists are the MEPs with H3K9me3 robust Z scores less than -3 or greater than 3.

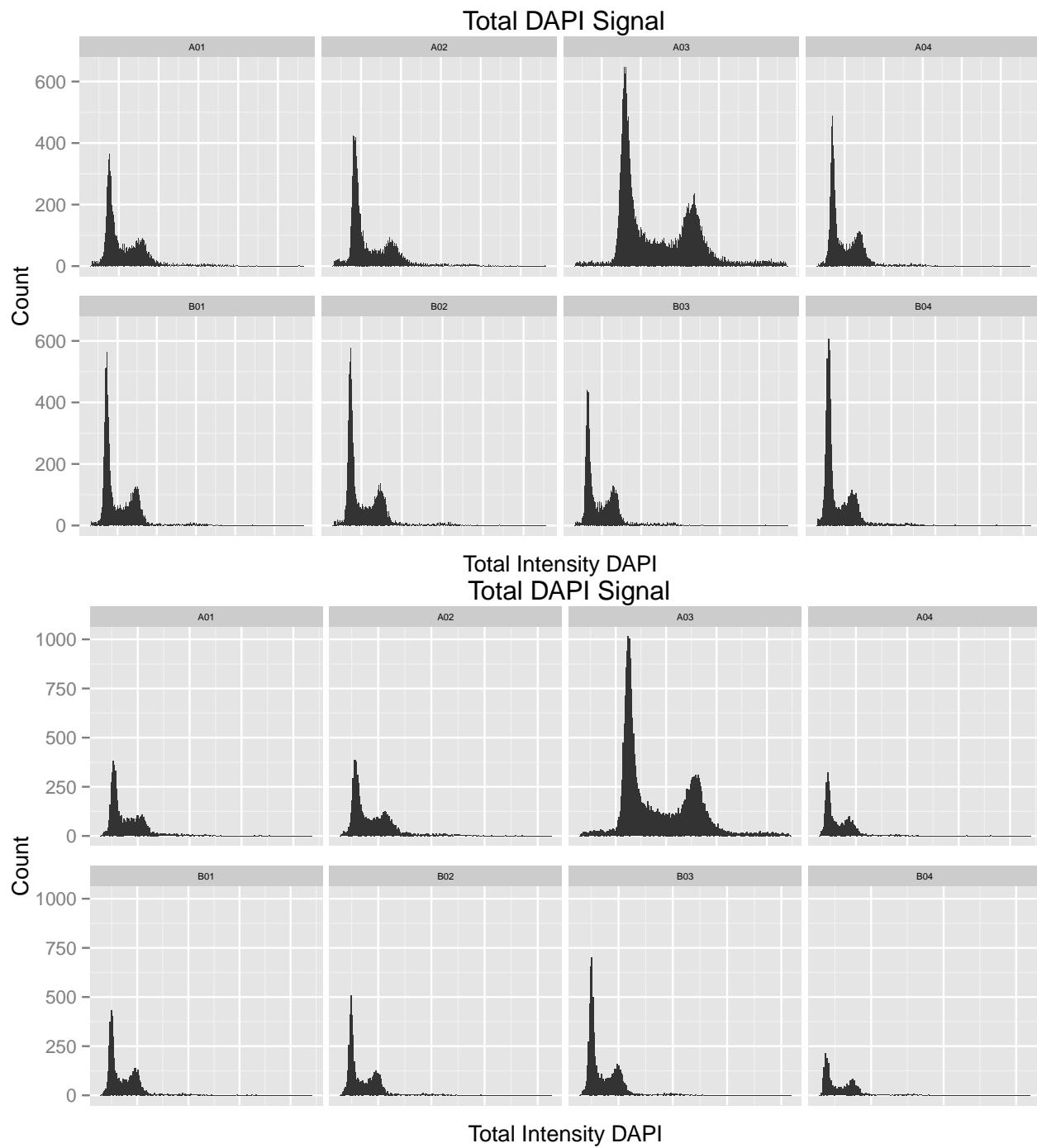
Ligand	ECMP	H3MedNormRobustZ
IL6	ECM1	-3.29
IL6	ITGA4B1	-3.19
TGFB1	LAMA3	3.01
NRG1	COL2A1	3.04
NRG1	ITGA10B1	3.05
TGFB1	ITGAMB2	3.07
NRG1	CD44	3.08
NRG1	CADH15	3.09
NRG1	LAMB1	3.10
TGFB1	VCAM1	3.10
TGFB1	LAMB1	3.10
NRG1	THBS1	3.11
TGFB1	VTN	3.12
NRG1	LAMA3	3.14
TGFB1	CDH1	3.14
TGFB1	ECM1	3.14
TGFB1	LUM	3.17
NRG1	LUM	3.17
NRG1	HyA50K	3.18
TGFB1	COL5A1	3.20
NRG1	DCN	3.21
NRG1	ITGA3B1	3.22
NRG1	COL1	3.25
NRG1	SPP1	3.26
NRG1	COL23A1	3.28
NRG1	PECAM1	3.29
NRG1	CDH1	3.33
TGFB1	PECAM1	3.40
TGFB1	DCN	3.43
TGFB1	FN1	3.43
TGFB1	ITGA4B1	3.44
TGFB1	ITGA10B1	3.44
KNG1	SPARC	3.48
TGFB1	CDH3	3.50
NRG1	POSTN	3.53
TGFB1	CEACAM6	3.60
NRG1	SPARC	3.61
NRG1	DSG2	3.68
NRG1	ITGAMB2	3.78
NRG1	ITGA9B1	3.79
TGFB1	ITGA2B1	3.81
NRG1	ITGA2B1	3.83
NRG1	VCAM1	3.90
NRG1	BGN	3.91
NRG1	ECM1	3.98
FLT3LG	SPARC	3.98
TGFB1	OMD	4.09
TGFB1	SPARC	4.21
TGFB1	ITGA3B1	4.29

Ligand	ECMp	H3MedNormRobustZ
NRG1	COL5A1	4.40
TGFB1	NID1	5.15
NRG1	NID1	5.92

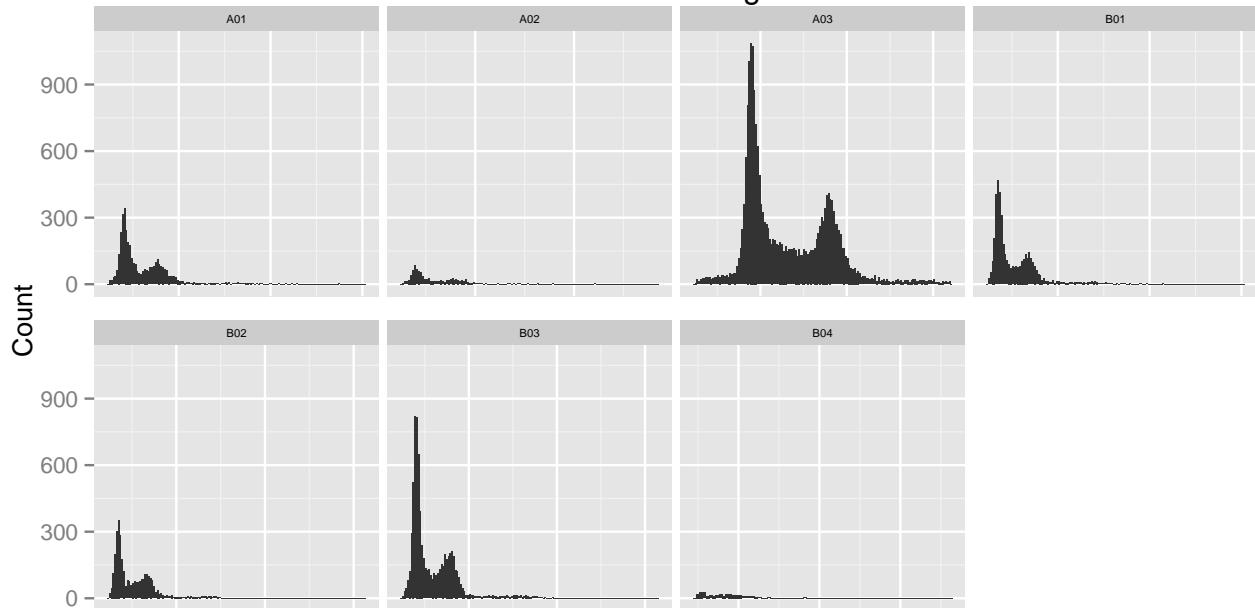
Cell Cycle Plots

Cell Cycle plots include univariate plots of the total DAPI signal and bivariate plots which add the mean EdU signal. The EdU signal has been gated to create EdU+ populations.

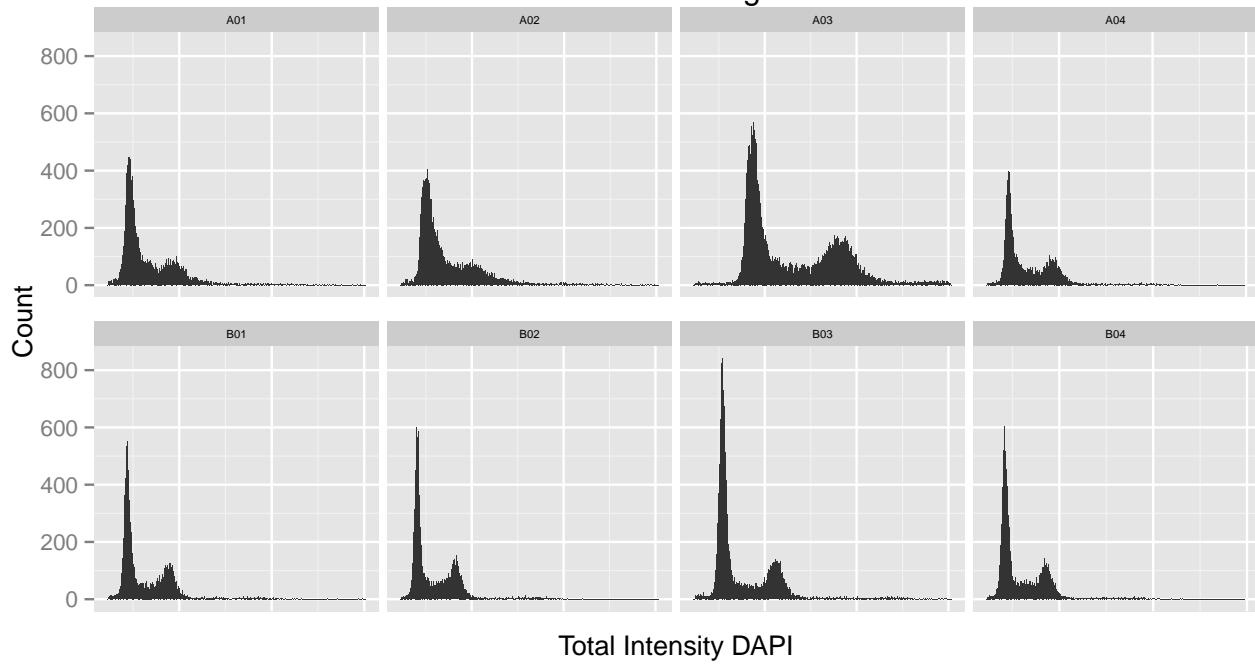




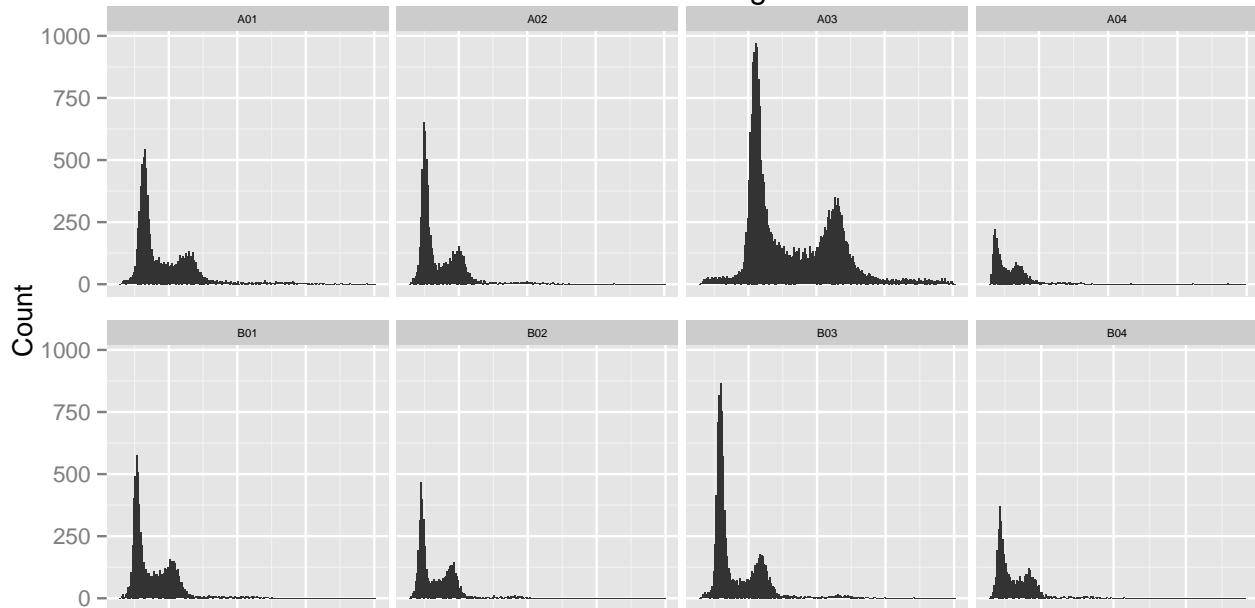
Total DAPI Signal



Total Intensity DAPI Total DAPI Signal



Total DAPI Signal



Total Intensity DAPI Total DAPI Signal

