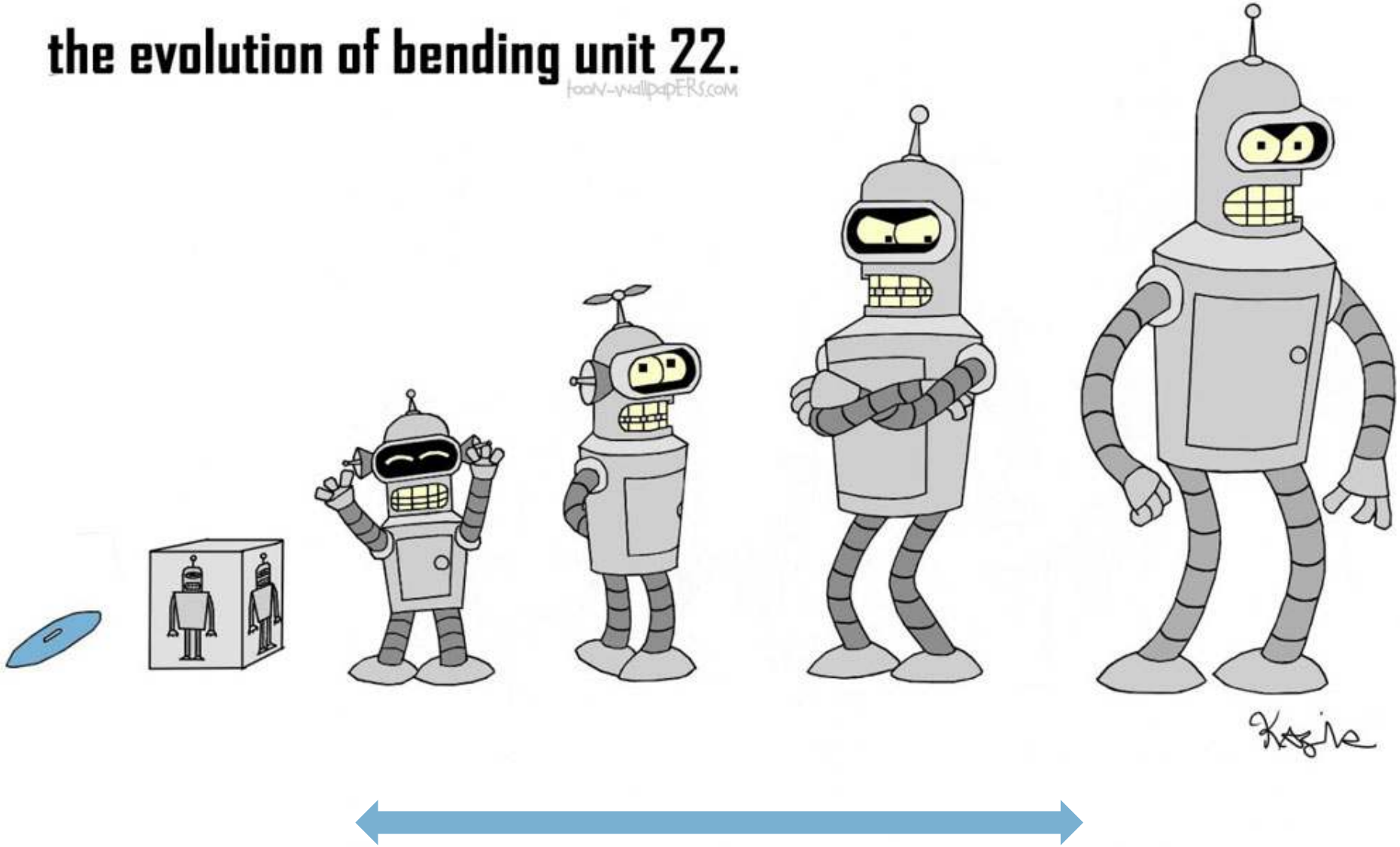
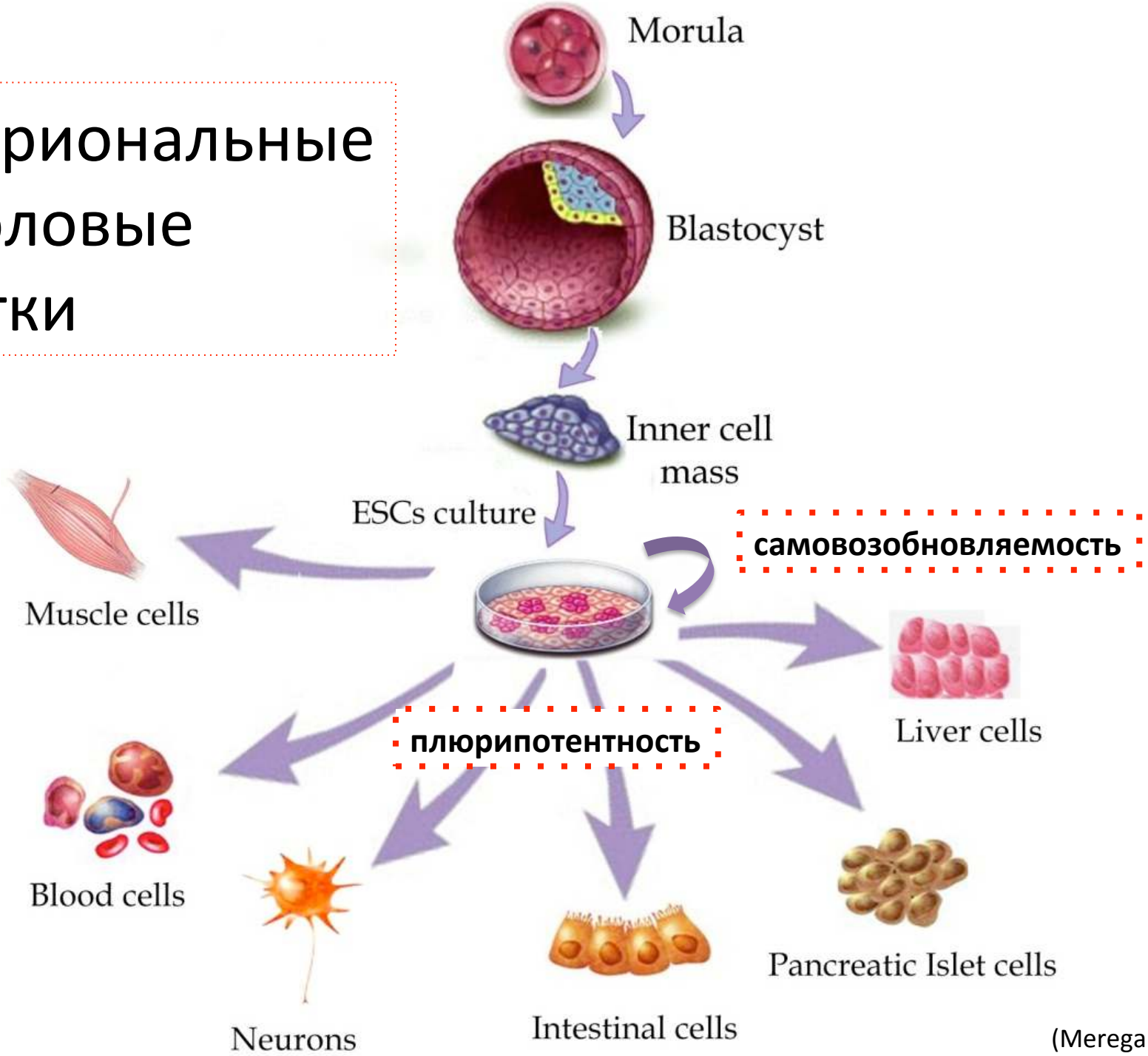
Биоинформационная оценка совершенства стволовых клеток человека

Мария Шутова, Сергей Киселев
ИОГен РАН
Владимир Наумов, Дмитрий Ищенко,
НИИФХМ РАН

the evolution of bending unit 22.

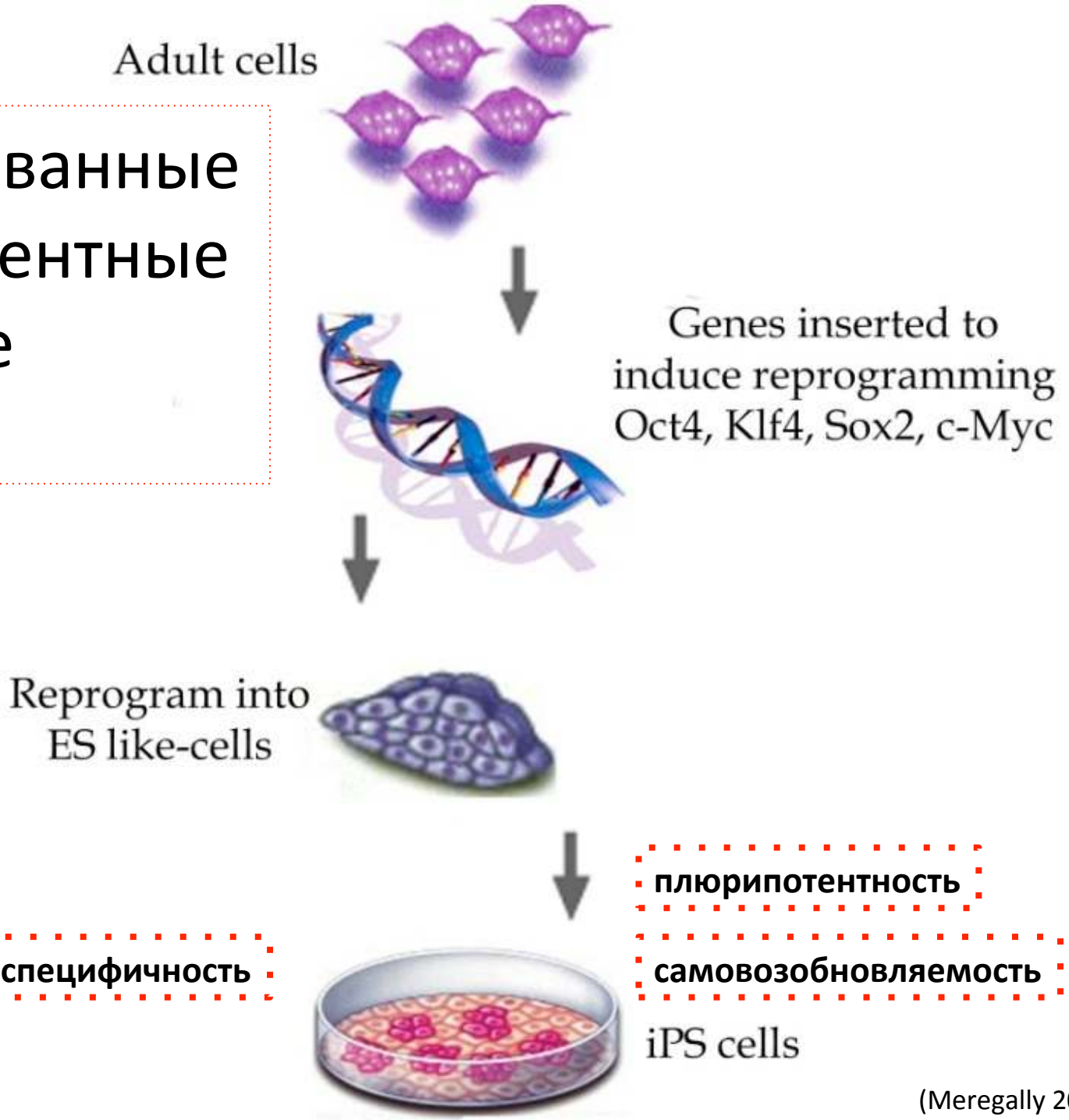


Эмбриональные стволовые клетки



(Meregally 2011)

Индуцированные плюрипотентные стволовые клетки



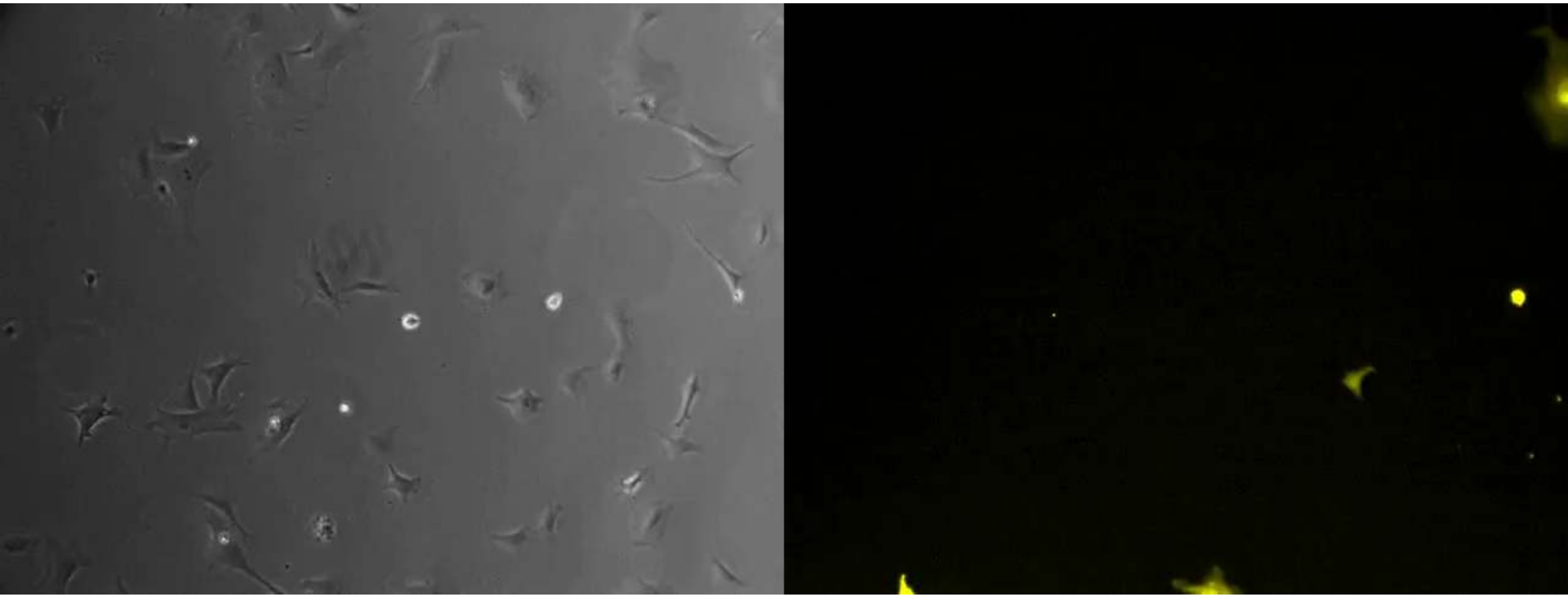
плюрипотентность

пациент-специфичность

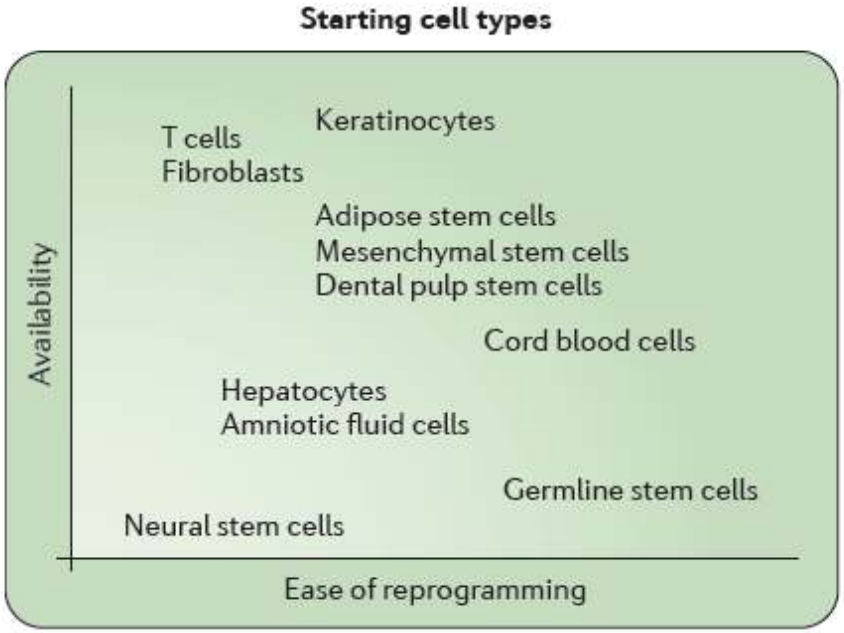
самовозобновляемость

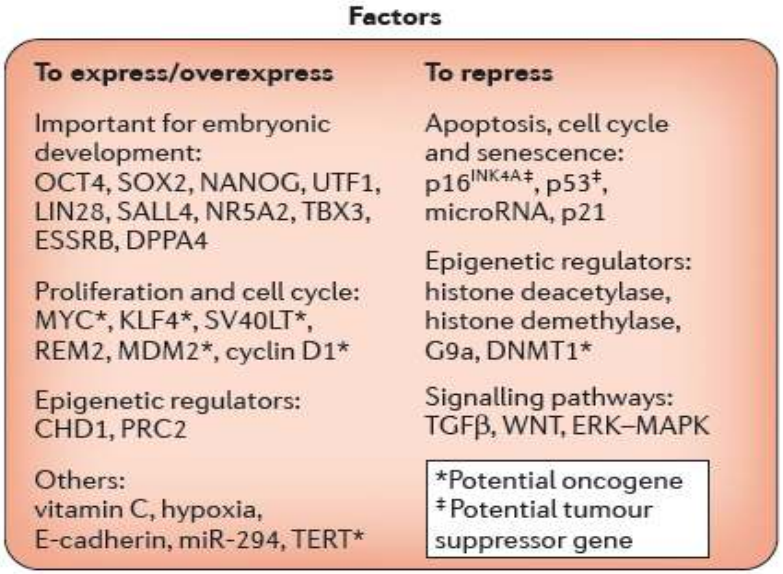
(Meregally 2011)

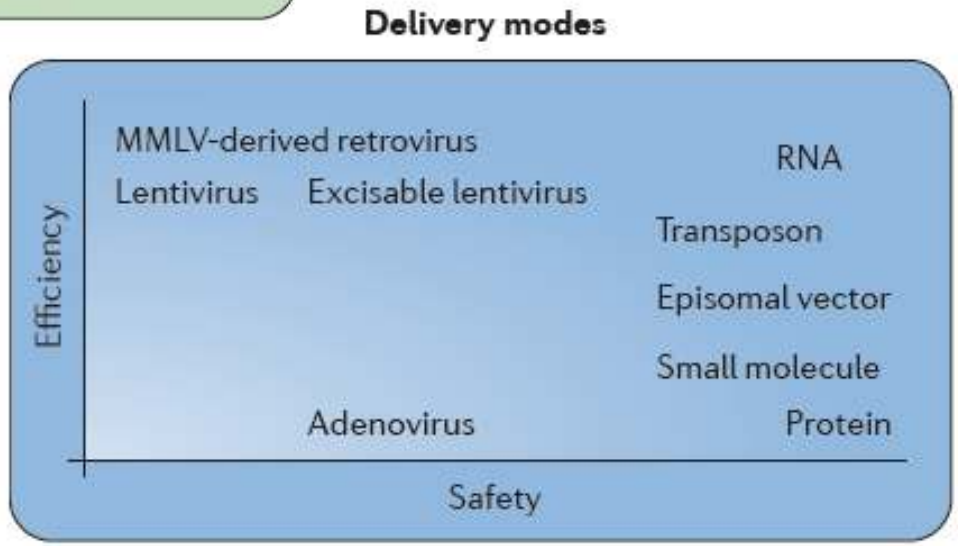
как?



как?

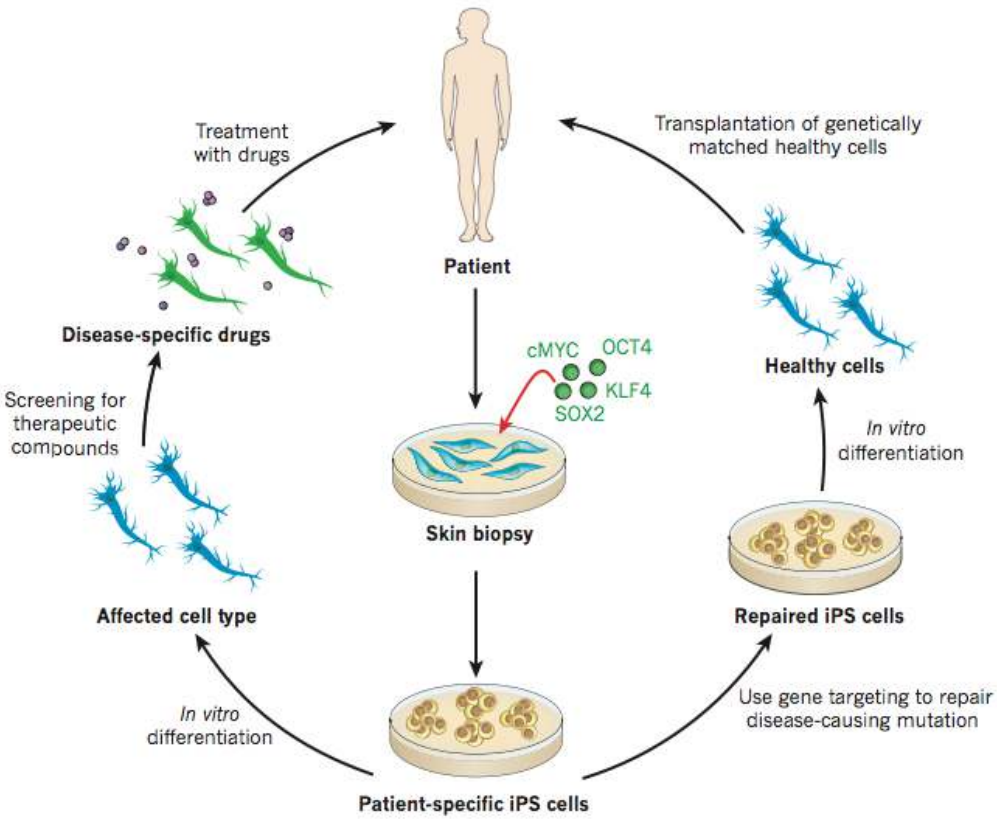






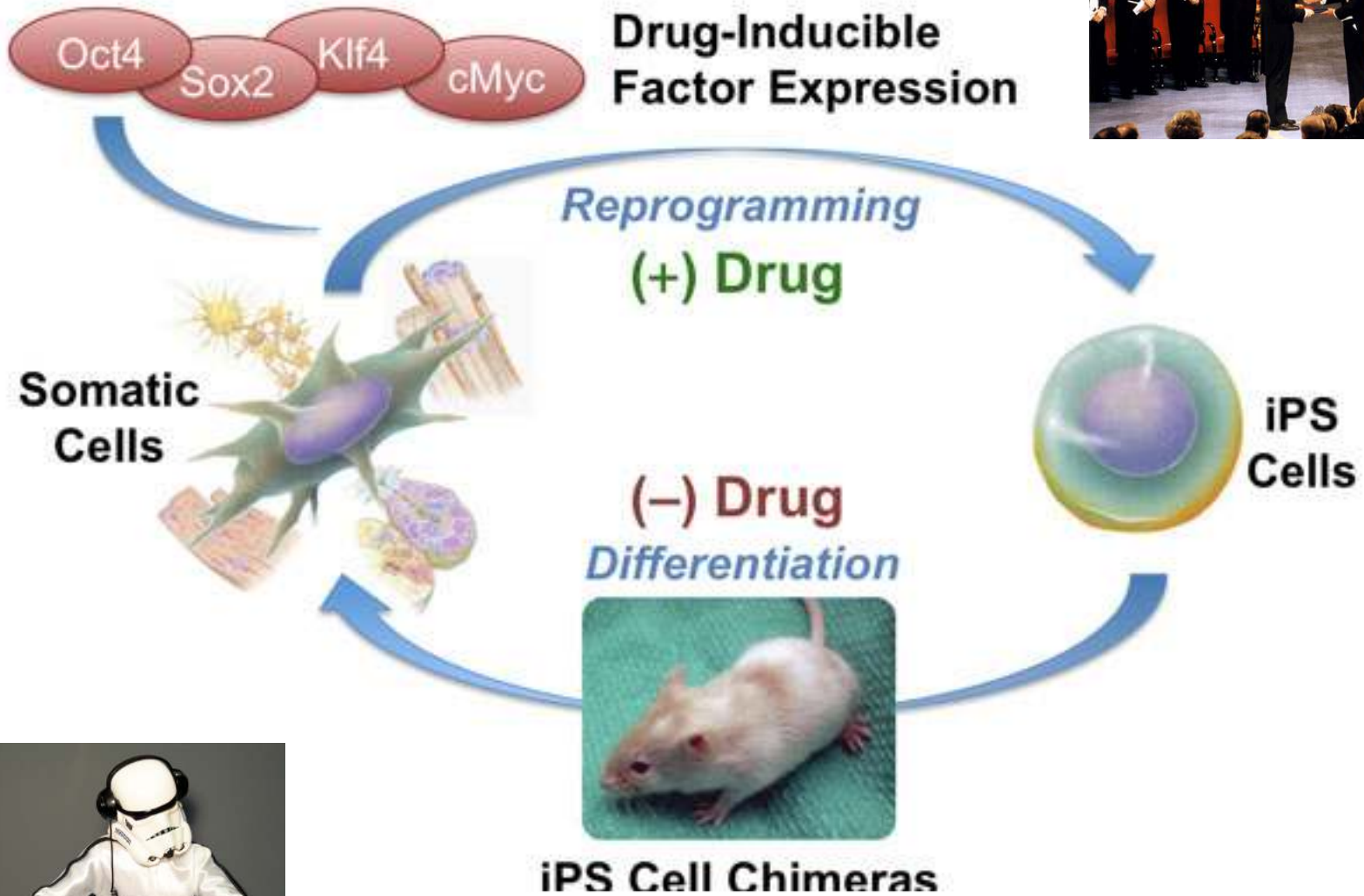
(Belmonte 2012)

зачем?



Disease	Molecular defect of donor cell	Cell type differentiated from iPS cells	Disease phenocopied in differentiated cells	Drug or functional tests
Neurological				
Amyotrophic lateral sclerosis (ALS)	Heterozygous Leu144Phe mutation in *SOD1*	Motor neurons and glial cells	ND	No
Spinal muscular atrophy (SMA)	Mutations in *SMN1*	Neurons and astrocytes, and mature motor neurons	Yes	Yes
Parkinson's disease	Multifactorial; mutations in *LRRK2* and/or *SNCA*	Dopaminergic neurons	No	Yes
Huntington's disease	72 CAG repeats in the huntingtin gene	None	NA	No
Down's syndrome	Trisomy 21	Teratoma with tissue from each of the three germ layers	Yes	No
Fragile X syndrome	CGG triplet repeat expansion resulting in the silencing of *FMR1*	None	NA	No
Familial dysautonomia	Mutation in *IKBKAP*	Central nervous-system lineage, peripheral neurons, haematopoietic cells, endothelial cells and endodermal cells	Yes	Yes
Rett's syndrome	Heterozygous mutation in *MECP2*	Neural progenitor cells	Yes	Yes
Mucopolysaccharidosis type IIIB (MPS IIIB)	Homozygous mutation in *NAGLU*	Neural stem cells and differentiated neurons	Partially	Yes
Schizophrenia	Complex trait	Neurons	Yes	Yes
X-linked adrenoleukodystrophy (X-ALD), childhood cerebral ALD (CCALD) and adrenomyeloneuropathy (AMN)	Mutation in *ABCD1*	Oligodendrocytes and neurons	Partially	Yes
Haematological				
ADA SCID	Mutation or deletion in *ADA*	None	ND	No
Fanconi's anaemia	*FAA* and *FAD2* corrected	Haematopoietic cells	No (corrected)	No
Schwachman–Bodian–Diamond syndrome	Multifactorial	None	NA	No
Sickle-cell anaemia	Homozygous HbS mutation	None	NA	No
β-Thalassaemia	Homozygous deletion in the β-globin gene	Haematopoietic cells	ND	No
Polycythaemia vera	Heterozygous Val617Phe mutation in *JAK2*	Haematopoietic progenitors ($CD34^+CD35^-$)	Partially	No
Primary myelofibrosis	Heterozygous mutation in *JAK2*	None	NA	No
Metabolic				
Lesch–Nyhan syndrome (carrier)	Heterozygous mutation in *HPRT1*	None	NA	No
Type 1 diabetes	Multifactorial; unknown	β-Cell-like cells (express somatostatin, glucagon and insulin; glucose-responsive)	ND	No
Gaucher's disease, type III	Mutation in *GBA*	None	NA	No
α1-Antitrypsin deficiency (A1ATD)	Homozygous mutation in the α1-antitrypsin gene	Hepatocyte-like cells (fetal)	Yes	No
Glycogen storage disease Ia (GSD1a)	Defect in glucose-6-phosphate gene	Hepatocyte-like cells (fetal)	Yes	No
Familial hypercholesterolaemia	Autosomal dominant mutation in *LDLR*	Hepatocyte-like cells (fetal)	Yes	No
Crigler–Najjar syndrome	Deletion in *UGT1A1*	Hepatocyte-like cells (fetal)	ND	No
Hereditary tyrosinaemia, type 1	Mutation in *FAHD1*	Hepatocyte-like cells (fetal)	ND	No
Pompe disease	Knockout of *GAA*	Skeletal muscle cells	Yes	No
Progressive familial cholestasis	Multifactorial	Hepatocyte-like cells (fetal)	ND	No
Hurler syndrome (MPS IH)	Genetic defect in *IDUA*	Haematopoietic cells	No	No
Cardiovascular				
LEOPARD syndrome	Heterozygous mutation in *PTPN11*	Cardiomyocytes	Yes	No
Type 1 long QT syndrome	Dominant mutation in *KCNQ1*	Cardiomyocytes	Yes	No
Type 2 long QT syndrome	Missense mutation in *KCNH2*	Cardiomyocytes	Yes	Yes
Primary immunodeficiency				
SCID or leaky SCID	Mutation in *RAG1*	None	NA	No
Omenn syndrome (OS)	Mutation in *RAG1*	None	NA	No
Cartilage-hair hypoplasia (CHH)	Mutation in *RMRP*	None	NA	No
Herpes simplex encephalitis (HSE)	Mutation in *STAT1* or *TLR3*	Mature cell types of the central nervous system	No	No
Other category				
Duchenne muscular dystrophy	Deletion in the dystrophin gene	None	NA	No
Becker muscular dystrophy	Unidentified mutation in dystrophin	None	NA	No
Dyskeratosis congenita (DC)	Deletion in *DKC1*	None	NA	No
Cystic fibrosis	Homozygous deletion in *CFTR*	None	NA	No
Friedreich's ataxia (FRDA)	Trinucleotide GAA repeat expansion in *FXN*	Sensory and peripheral neurons, and cardiomyocytes	Partially	No
Retinitis pigmentosa	Heterogeneity in causative genes and mutations: mutations in *RP9*, *RP1*, *PRPH2* or *RHO*	Retinal progenitors, photoreceptor precursors, retinal-pigment epithelial cells and rod photoreceptor cells	Yes	Yes
Recessive dystrophic epidermolysis bullosa (RDEB)	Mutation in *COL7A1*	Haematopoietic cells, and epidermis-like keratinocytes that differentiate into cells of all three germ layers in vivo	Partially	Yes
Scleroderma	Unknown	None	NA	No
Osteogenesis imperfecta	Mutation in *COL1A2*	None	NA	No

(Robinton 2012)


Oct4
Sox2
Klf4
cMyc
Drug-Inducible
Factor Expression
Reprogramming
(+) Drug
Somatic
Cells
iPS
Cells
(–) Drug
Differentiation
iPS Cell Chimeras

идентичность ЭСК и iPS?

- одинаковые свойства? “нормальность”
- одинаковая эффективность дифференцировки? “удобство”
- безопасность использования в медицине?
- “адекватность” использования iPS клеток как модели болезней?

идентичность ЭСК и iPS?

Reference	Cell lines	Level of analysis	Key findings
Marchetto *et al.* (2009)[17]	2 hiPSCs + 2 hESCs	Transcriptome via microarray	Distinct gene expression signature of hiPSCs
Chin *et al.* (2009)[13]	4 hESCs + 5 hiPSCs + 3 somatic	Transcriptome via microarray MicroRNA-ome via microarray Histone methylation via ChIP	Distinct gene expression signature of hiPSCs
Ghosh *et al.* (2010)[20]	4 hiPSCs + 4 somatic	Transcriptome via microarray	Transcriptional memory of somatic cell of origin
Guenther *et al.* (2010)[12]	6 hiPSCs + 6 hESCs	Histone methylation via ChIP-seq Transcriptome via microarray	Inconsistent hiPSC vs. hESC differences Lab-specific gene expression differences
Newman *et al.* (2010)[27]	17 hESCs + 67 hiPSCs	Meta-analysis of microarray data	Lab-specific gene expression signatures
Feng *et al.* (2010)[3]	6 hiPSCs + 14 hESCs	Hemangioblastic differentiation propensity Endothelial cell differentiation propensity	Early senescence of hiPSC progeny
Hu *et al.* (2010)[1]	5 hESCs + 12 hiPSCs	Neural differentiation propensity	Variable yield of neural progeny
Polo *et al.* (2010)[22]	12 Murine iPSCs	mRNA transcripts via qPCR DNA methylome via HELP Histone modification via ChIP	Epigenetic memory abrogated by extended passaging
Kim *et al.* (2010)[21]	31 Murine iPSCs + 14 murine ESCs + somatic	DNA methylome via CHARM Hematopoietic differentiation potential Osteogenic differentiation potential	Epigenetic memory of somatic cell of origin
Narsinh *et al.* (2011)[2]	3 hESCs + 4 hiPSCs	mRNA transcripts via single-cell qPCR Cardiovascular differentiation propensity	Single-cell heterogeneity of hiPSCs Variable yield of cardiovascular progeny
Lister *et al.* (2011)[23]	2 hESCs + 5 hiPSCs	DNA methylome via methylC-seq Histone methylation via ChIP-seq Transcriptome via RNA-Seq	Hot spots of aberrant methylation
Bock *et al.* (2011)[4]	20 hESCs + 12 hiPSCs	Transcriptome via microarray DNA methylome via RRBS mRNA transcripts via fluorescent counting	Bioinformatic analysis predicts differentiation propensity
Laurent *et al.* (2011)[28]	69 hESCs + 37 hiPSCs + somatic	Genomic stability via SNP genotyping	CNV in hiPSCs and hESCs

(Narsinh 2011)

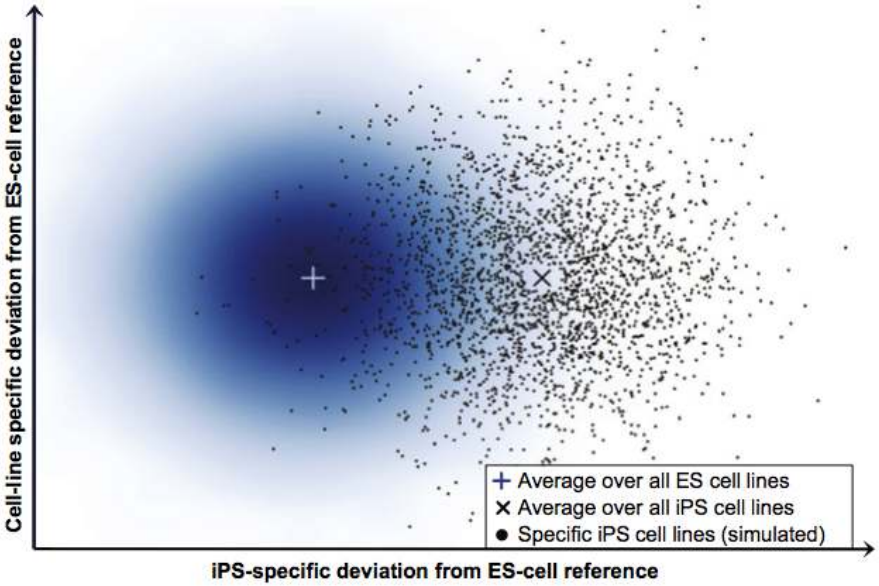
идентичность ЭСК и iPS?

Conclusion about the Relationship between ESCs and iPSCs	First Author	Year	Clone Numbers ESC	Clone Numbers iPSC
It is difficult to distinguish between them	A.M. Newman	2010	23	68
	M.G. Guenther	2010	36	54
	C. Bock	2011	20	12
There are notable differences	M. Chin	2009	3	5
	C.M. Marchetto	2009	2	2
	J. Deng	2009	3	4
	Z. Ghosh	2010	6	4
	A. Doi	2011	3	9
	Y. Ohi	2011	3	9
	K. Kim	2011	6	12
	R. Lister	2011	2	5

D

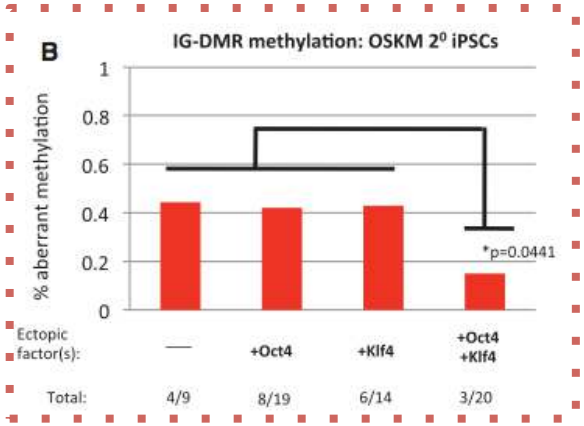
	iPS Cell Classifier	Accuracy	Sensitivity	Specificity	TN (ES)	FN	FP	TP
Published iPS cell signatures	Chin2009 gene expression signature	63%	0%	100%	20	12	0	0
	Doi2009 DNA methylation signature	63%	0%	100%	20	12	0	0
	Stadtfeld2010 single-gene signature (MEG3)	72%	100%	55%	11	0	9	12
Cross-validated ES vs iPS classifier	ES cell lines vs. iPS cell lines (DNA methylation)	78%	67%	86%	17	4	3	8
	ES cell lines vs. iPS cell lines (gene expression)	68%	48%	81%	16	6	4	6
	ES cell lines vs. iPS cell lines (both data types)	81%	64%	91%	18	4	2	8
Positive controls (ES vs fib. classifier)	ES cell lines vs. fibroblasts (DNA methylation)	100%	100%	100%	20	0	0	6
	ES cell lines vs. fibroblasts (gene expression)	94%	73%	100%	20	2	0	4
	ES cell lines vs. fibroblasts (both data types)	99%	95%	100%	20	0	0	6
Negative controls (trivial classifiers)	Predict everything as ES cell line	63%	0%	100%	20	12	0	0
	Predict everything as iPS cell line	38%	100%	0%	0	0	20	12
	Flip a coin (50% chance to predict as ES cell line)	50%	50%	50%	10	6	10	6

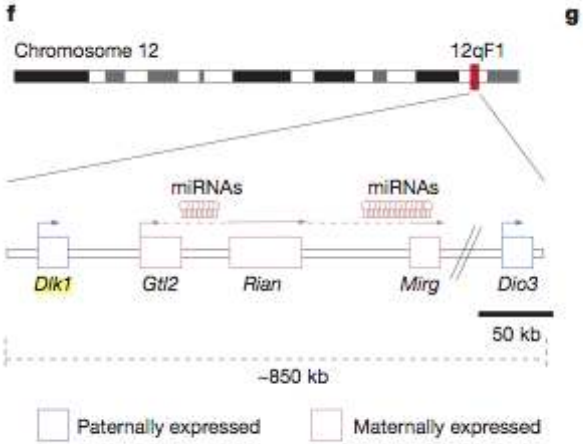
(Yamanaka 2012)

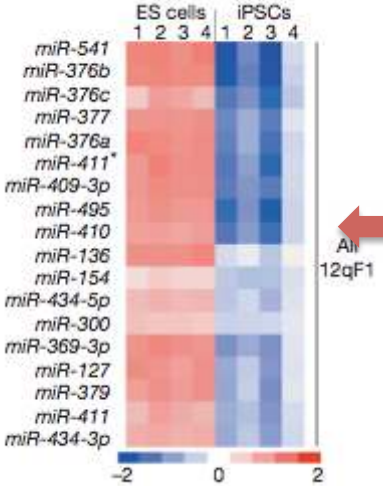


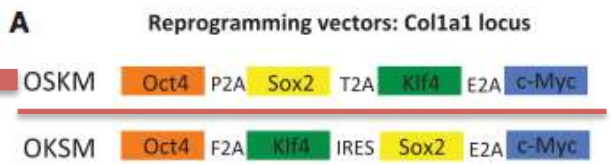
(Bock 2011)

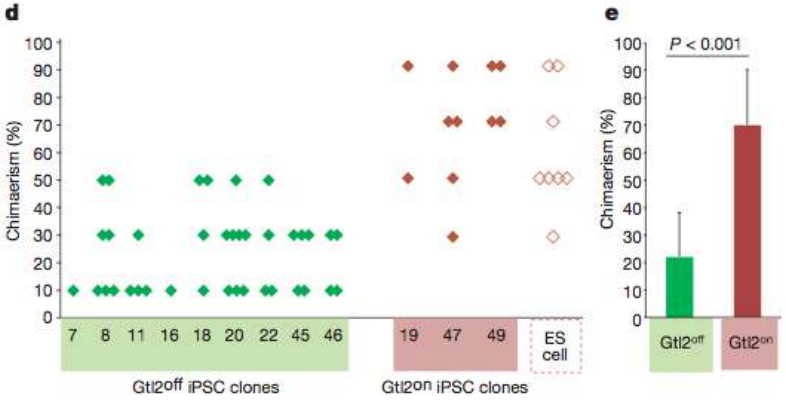
Dlk1-Dio3 locus?

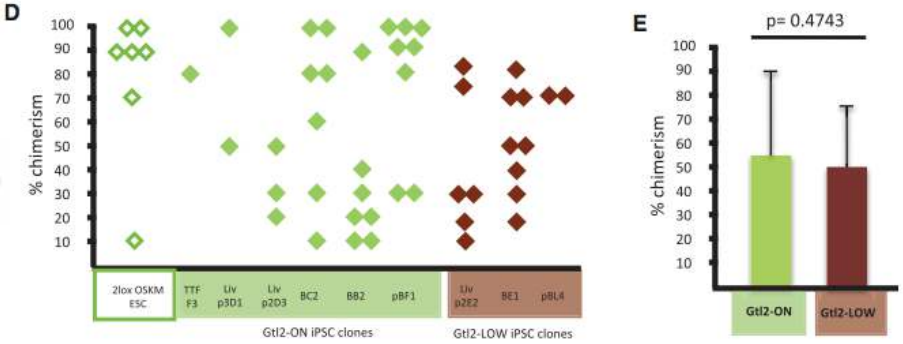








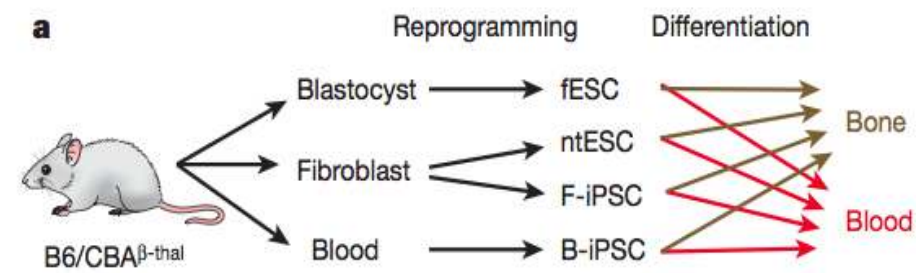


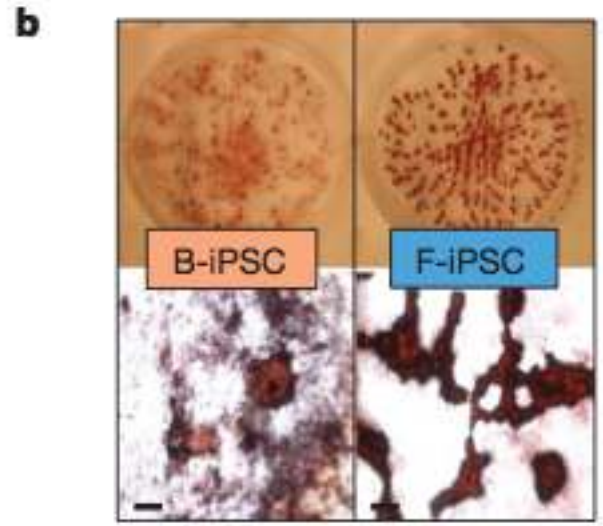


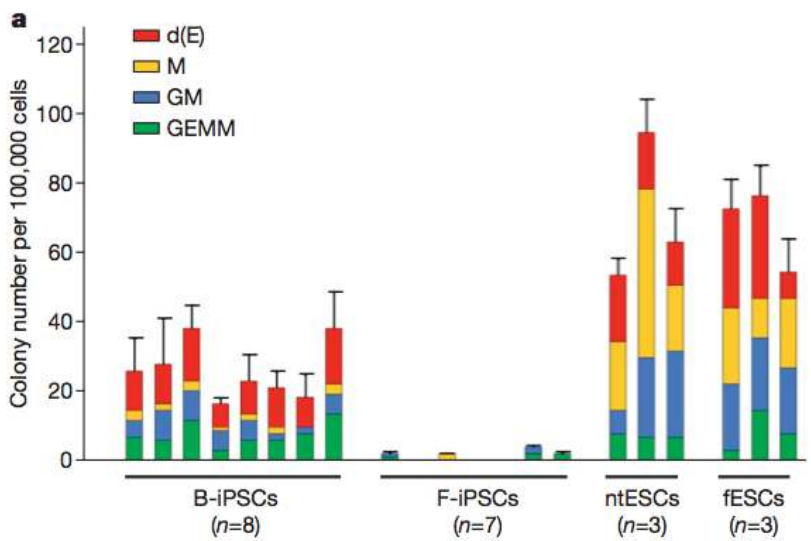
(Stadtfeld 2010)

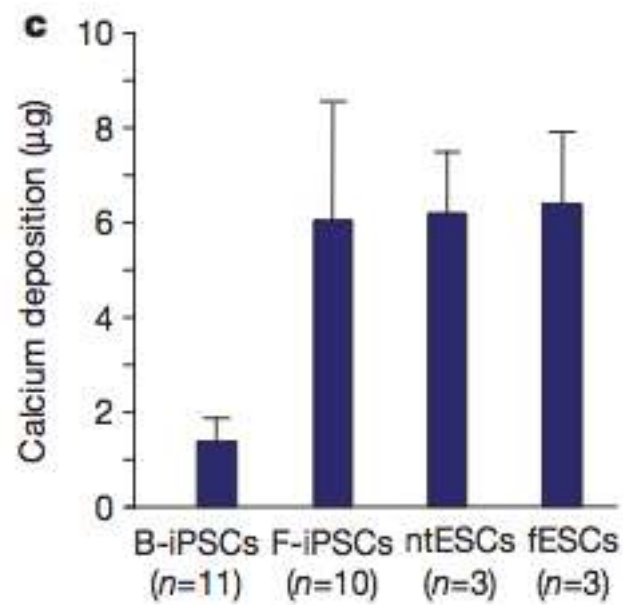
(Carey 2011)

somatic memory?







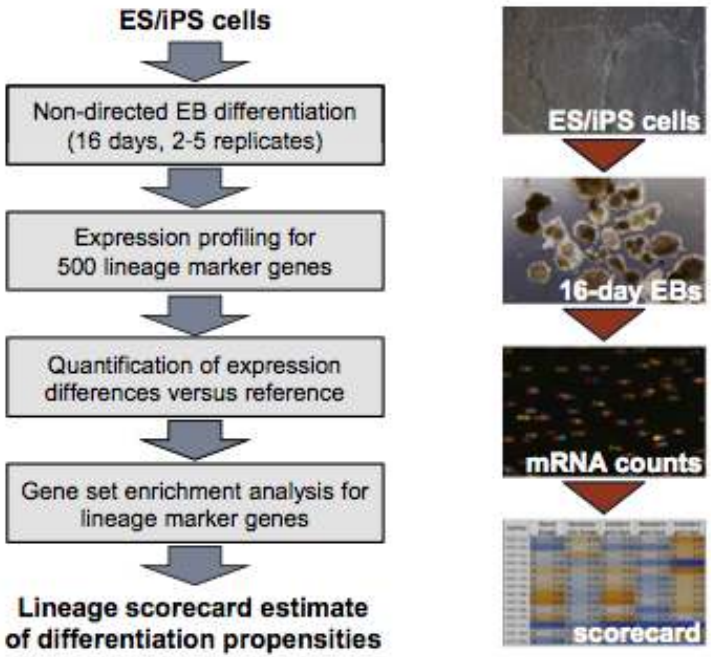


(K Kim, A Doi 2010)

(Bock 2011)

commitment?

A

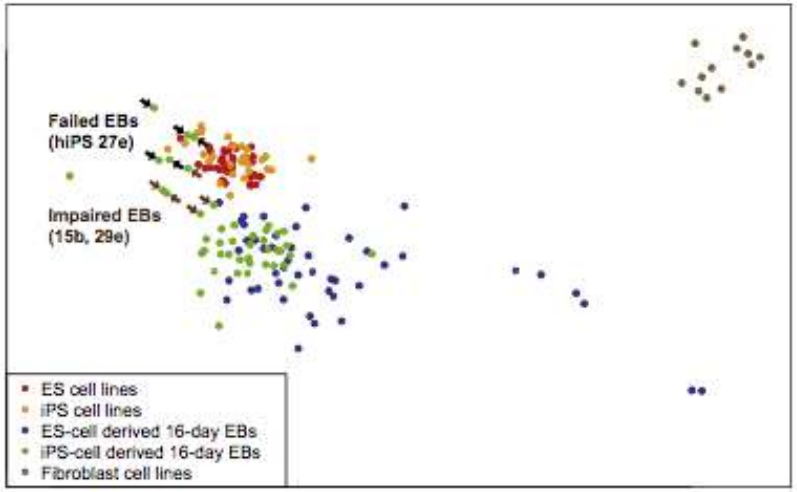


B

Cell line	Neural lineage	Hematopo-ietic lineage	Ectoderm germ layer	Mesoderm germ layer	Endoderm germ layer
HUES1	-1.84	-0.30	-1.56	0.06	-0.59
HUES3	-0.29	-0.01	-0.23	-0.07	0.08
HUES6	-0.78	-0.26	-0.51	-0.05	-0.47
HUES8	-0.15	0.69	-0.17	0.68	1.45
HUES9	-0.89	0.31	-0.75	0.51	0.37
HUES28	-1.33	-0.11	-0.91	1.03	-0.07
HUES44	0.70	-0.27	0.52	-0.48	-0.45
HUES45	-0.46	-0.26	-0.49	-0.02	0.65
HUES48	0.83	0.18	0.70	0.24	0.55
HUES49	0.19	0.07	0.03	-0.66	-0.26
HUES53	-0.95	0.65	-1.19	-0.22	-0.20
HUES62	0.25	-0.15	0.15	-0.60	0.24
HUES63	0.62	0.39	0.72	0.34	0.61
HUES64	1.45	-0.07	1.44	-0.56	-0.61
HUES65	0.19	0.02	0.22	0.19	-0.15
HUES66	0.59	-0.67	0.36	-1.22	-0.37
H1	1.54	-0.29	1.21	0.07	-0.56
H9	1.08	0.01	1.10	0.55	-0.16

Differentiation propensity: high medium low

C



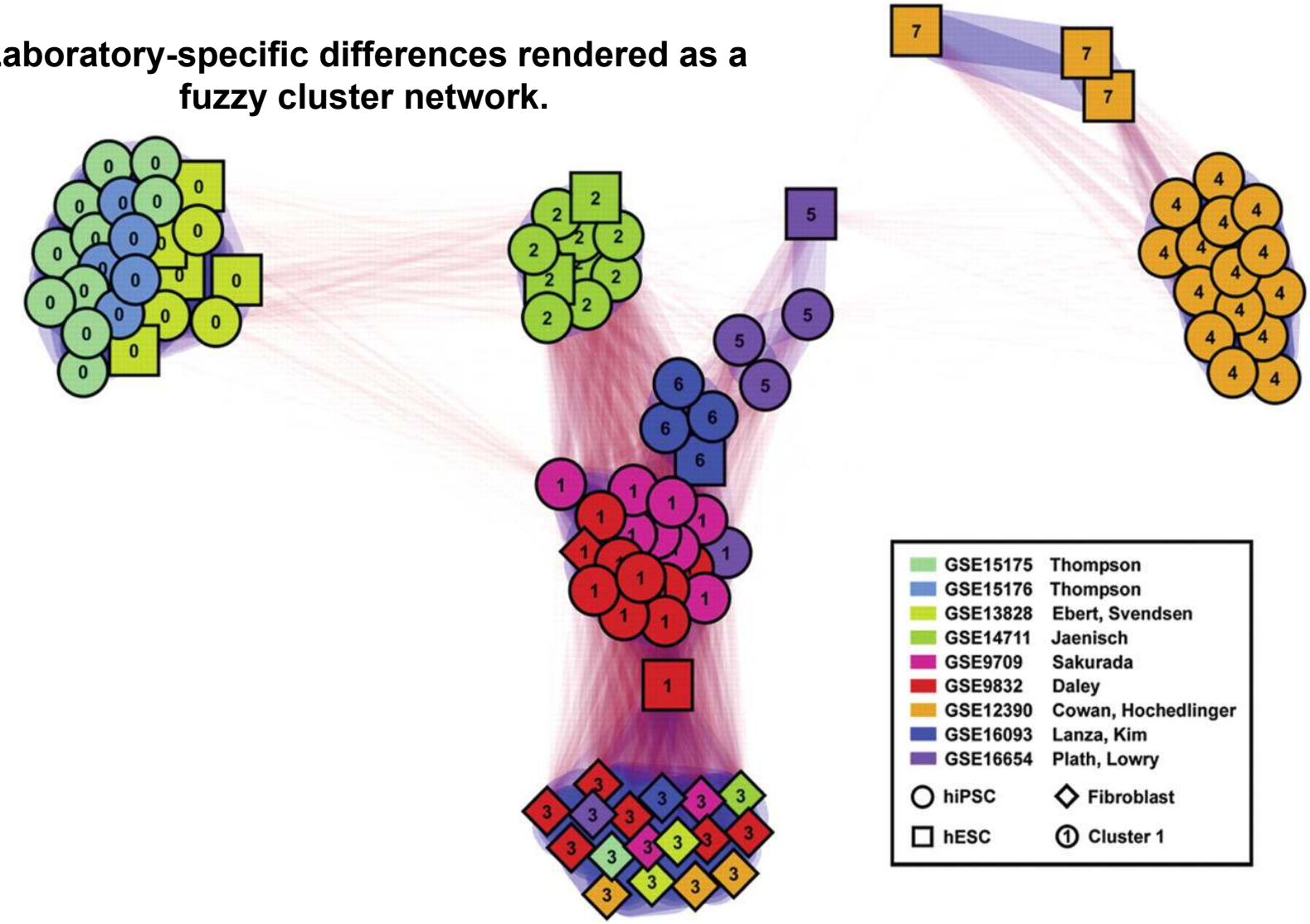
D

Cell line	Neural lineage	Hematopo-ietic lineage	Ectoderm germ layer	Mesoderm germ layer	Endoderm germ layer
hiPS 11a	-0.69	0.18	-0.37	-0.23	0.83
hiPS 11b	-1.17	-0.23	-0.96	-1.03	0.47
hiPS 11c	-0.22	0.40	-0.03	-0.16	0.37
hiPS 15b	-0.48	-0.78	-0.63	-1.11	-2.49
hiPS 17a	0.19	0.05	0.33	0.00	1.16
hiPS 17b	-0.07	-0.48	-0.02	-0.83	0.20
hiPS 18a	0.28	-0.52	0.31	-0.67	0.20
hiPS 18b	0.80	-0.72	0.84	-0.62	0.15
hiPS 18c	0.93	-0.65	1.05	-0.41	0.10
hiPS 20b	-0.37	-0.47	-0.30	-1.16	0.56
hiPS 27b	0.52	-0.50	0.68	-0.71	-0.42
hiPS 27e	-1.61	-1.04	-2.12	-1.82	-3.27
hiPS 29d	-0.25	-0.04	0.00	-0.11	0.83
hiPS 29e	-0.99	-0.60	-1.15	-1.14	-1.08

Differentiation propensity: high medium low

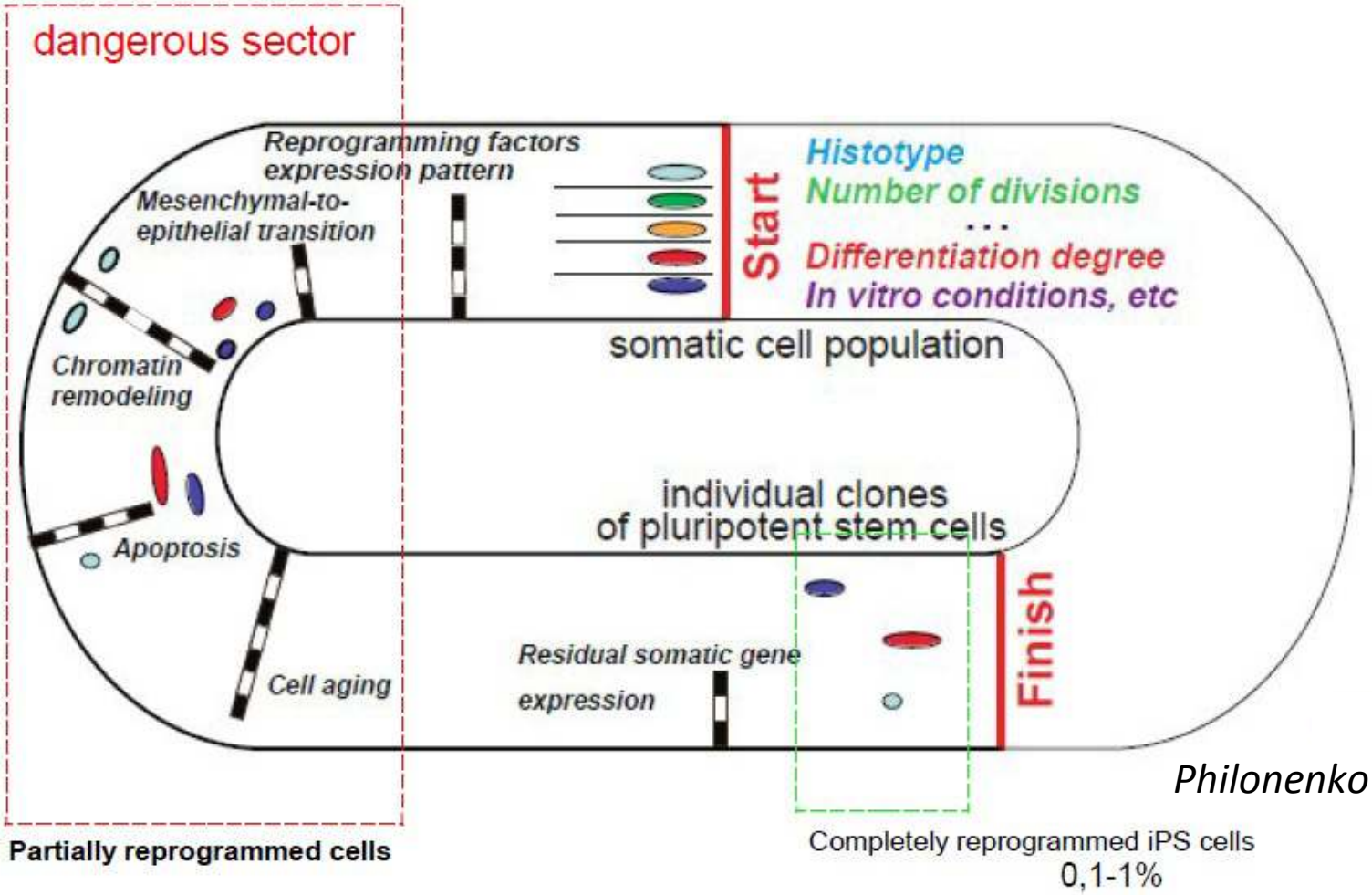
lab-specific stochasticity?

Laboratory-specific differences rendered as a fuzzy cluster network.



Sullivan G et al. Mol. Hum. Reprod. 2010;16:880-885

OLYMPIC REPROGRAMMING ARENA



Philonenko et al, 2011

Does the reprogramming process retrace the developmental pathway?
Does incomplete reprogramming influence tumorigenic potential of iPS cells?

проверка точности репрограммирования

- изогенная система (ES=somatic=iPS)
- одна лаборатория, одни руки, одни среды
- учет состояния соматических клеток
- три типа соматических/iPS клеток в четырех/шести повторностях каждый
- параллельный анализ метилирования (450k) и экспрессии (HT12v4) всех клеток

ES05

изогенная система

new5

фибробласты кожи

нейроны

клетки глаза

iPS7

iPS47

iPS27

iPS29

iPS22

iPS14

methylome

transcriptome

ES05

transgene silencing

new5

pluripotency tests

MACS Sorting

ICC

CD31-CD105+

NCAM+

RPE65+

FACS

iPS7 iPS47 iPS27 iPS29 iPS22 iPS14

pluripotency tests *RT-PCR* *karyotype* *ICC*

methylome

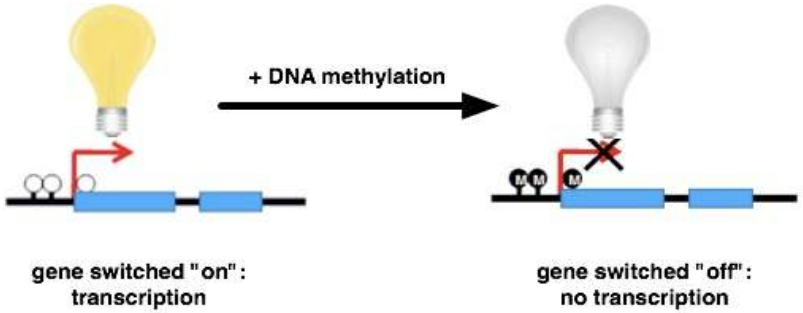
transcriptome

стандартная характеристика ES/iPSCs

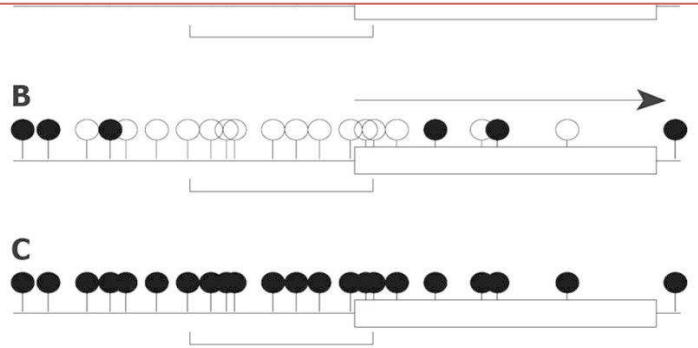
- морфология
- пролиферативная активность, самоподдержание
- нормальный кариотип
- дифференцировка in vitro
- дифференцировка in vivo

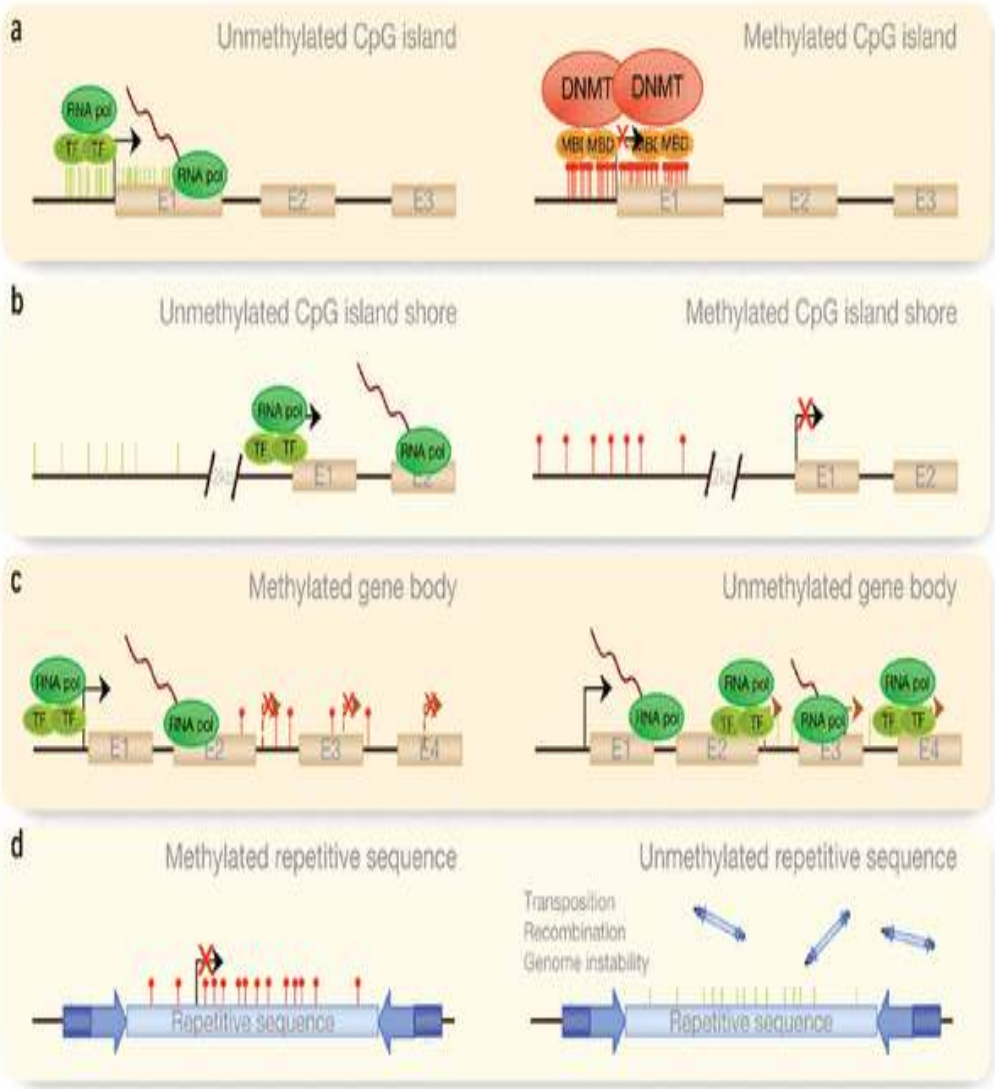
этого достаточно, или нужен полногеномный анализ?

Эпигенетическая регуляция экспрессии генов



CpG island is a region with at least 200 bp, and a GC percentage that is greater than 50%, and with an observed-to-expected CpG ratio that is greater than 60%.

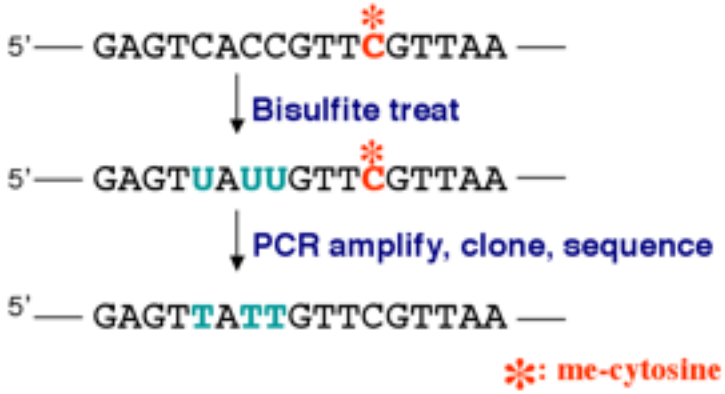




Анализ метилирования

illumina Infinium HumanMethylation450 BeadChip

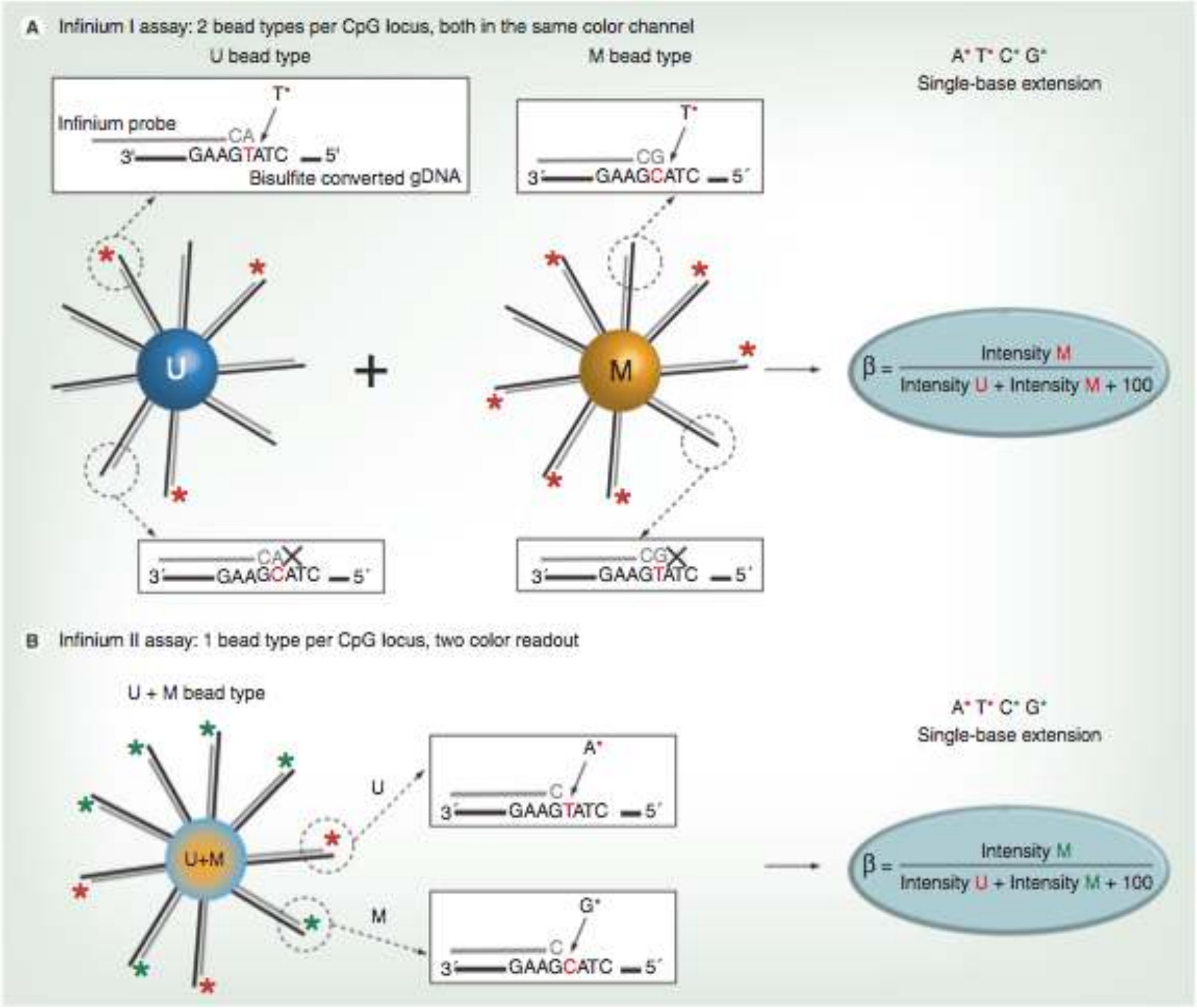
- показания уровня метилирования 485,512 CpG динуклеотидов (1,5% из всех которые есть в геноме)
- покрытие – 99% RefSeq аннотированных генов
- воспроизводимость для технических повторностей больше 99%
- 600 отрицательных контролей



бисульфитная конверсия

нанесение на чип (microarray)

цветовая детекция

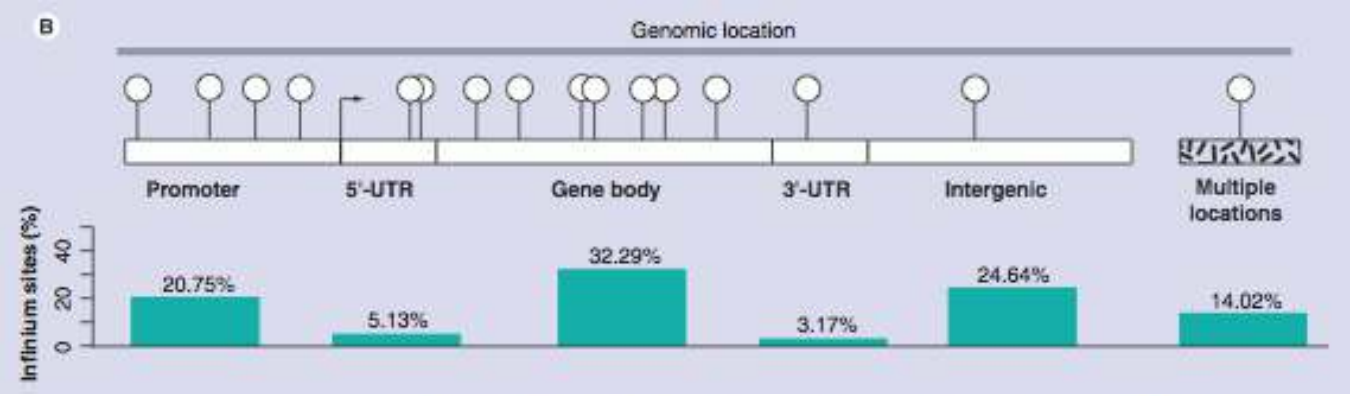

A Infinium I assay: 2 bead types per CpG locus, both in the same color channel
U bead type
M bead type
A* T* C* G*
Single-base extension
Infinium probe
T*
CA
3' GAAGTATC 5'
Bisulfite converted gDNA
T*
CG
3' GAAGCATC 5'
U
M
CA
3' GAAGCATC 5'
CG
3' GAAGTATC 5'
β = Intensity M / Intensity U + Intensity M + 100
B Infinium II assay: 1 bead type per CpG locus, two color readout
U + M bead type
A* T* C* G*
Single-base extension
U+M
U
A*
C
3' GAAGTATC 5'
M
G*
C
3' GAAGCATC 5'
β = Intensity M / Intensity U + Intensity M + 100

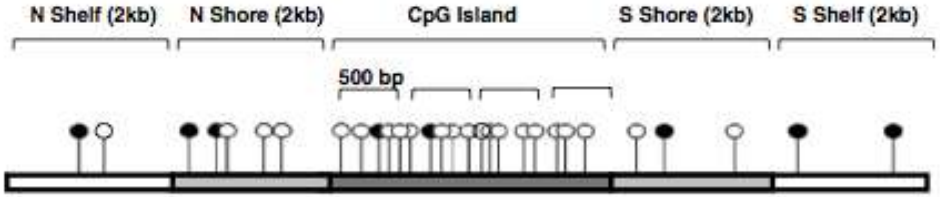
Beta value

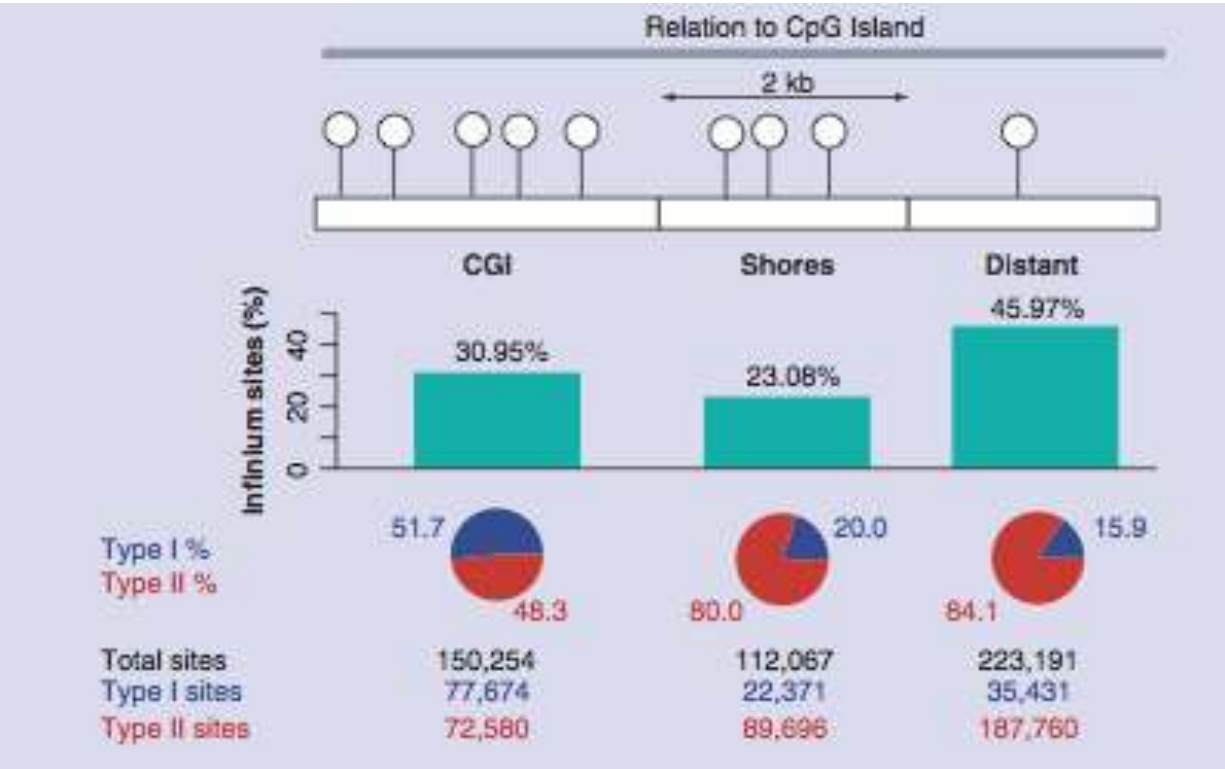
Index	TargetID	CHR	COLOR	INFINIUM_DESIGN_TYPE	PROBE_SNPS	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_	MAPINFO
1	cg00000029	16		II		RBL2	TSS1500	N_Shore	53468112
2	cg00000108	3		II	rs9857774	C3orf35;C3orf35	Body;3'UTR		37459206
3	cg00000109	3		II	rs9864492	FNDC3B;FNDC3B	Body;Body		171916037
4	cg00000165	1		II				S_Shore	91194674
5	cg00000236	8		II		VDAC3;VDAC3	3'UTR;3'UTR		42263294
6	cg00000289	14		II		ACTN1;ACTN1;ACTN1	3'UTR;3'UTR;3'UTR	N_Shore	69341139
7	cg00000292	16		II	rs62037371	ATP2A1;ATP2A1	1stExon;1stExon	N_Shore	28890100
8	cg00000321	8		II		SFRP1	TSS1500	S_Shore	41167802
9	cg00000363	1		II				N_Shore	230560793
10	cg00000622	15	Red	I		NIPA2;NIPA2;NIPA2;NIPA2	TSS200;TSS200;TSS200;T	Island	23034447
11	cg00000658	9		II		MAN1B1	Body	Island	139997924
12	cg00000714	19		II		TSEN34;TSEN34	Body;Body	S_Shore	54695678
13	cg00000721	6		II		LRRC16A	Body	S_Shelf	25282779
14	cg00000734	3		II		CNBP;CNBP;CNBP;CNBP;CN	5'UTR;5'UTR;5'UTR;5'UTI	Island	128902377
15	cg00000769	12	Grn	I		DDX55	TSS200	Island	124086477

- для единичного CpG значения от 0 до 1 (от 0% до 100%)
- число = степень метилирования, учитываются значения для CpG, которые детектировали больше трех бидов
- максимальная разница в значениях 0.15 для технических и 0.24 для биологических повторностей

- 40% проб находятся вне известных регулятроных регионов
- большинство из них покрыты с помощью Inf2 типа бидов
- Inf1 имеет 28% покрытия CpG, Inf2 – 72%

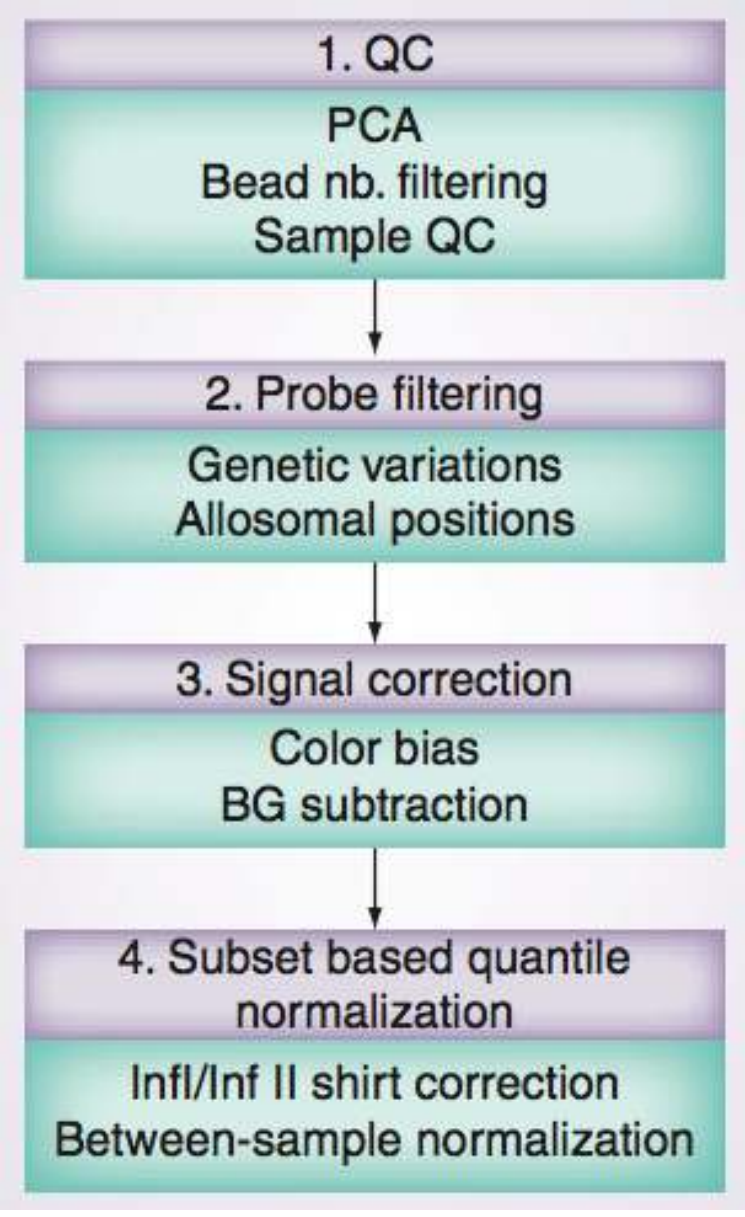






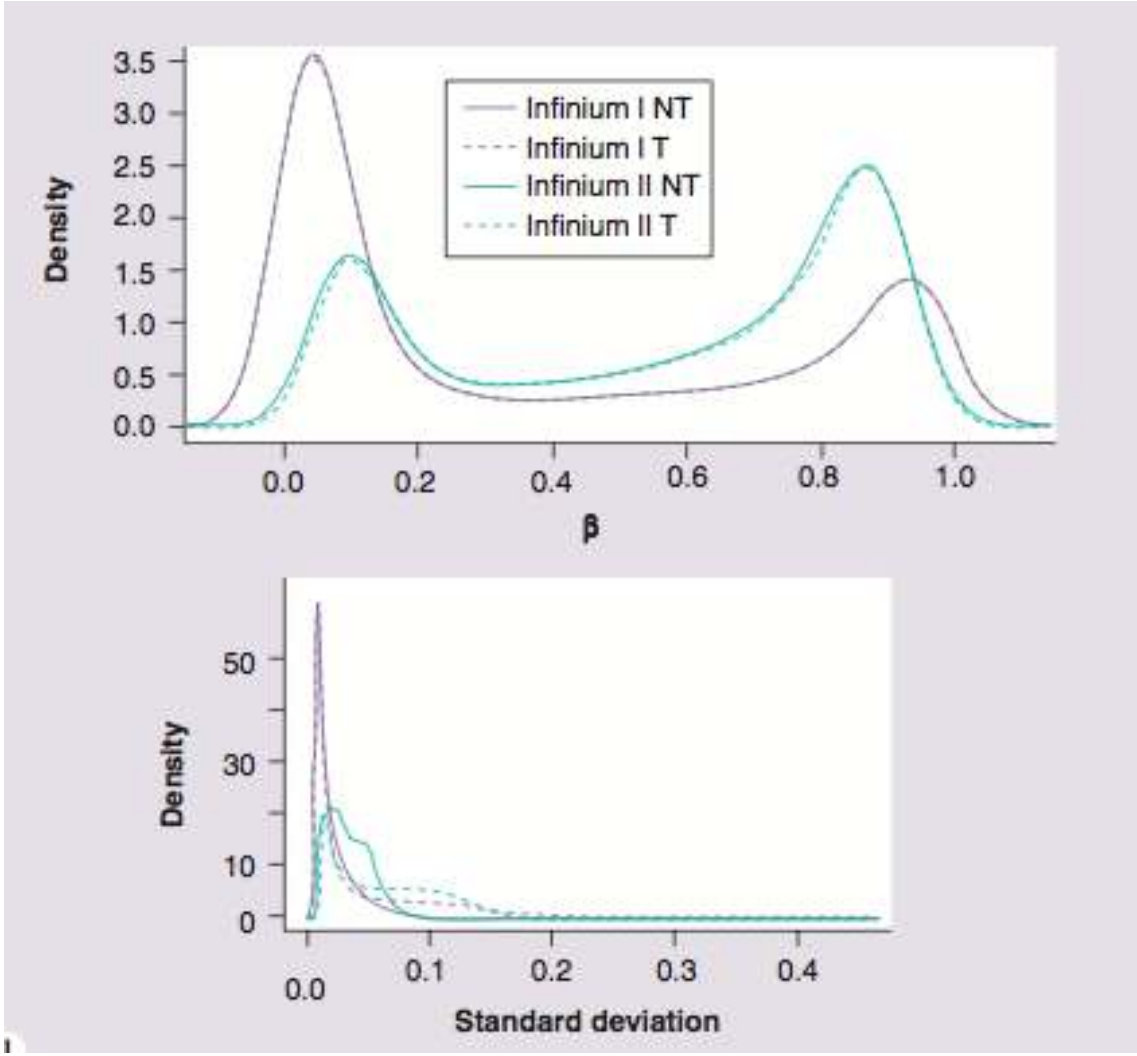
Preprocessing

- коррекция batch effect на основе PCA
- фильтрация сэмплов с <90% high-confidence значений (pval>0.01)
- фильтрация проб, соответствующих часто встречающимся SNP
- коррекция разницы в эффективности мечения/ свойствах сканирования/ интенсивности двух цветов на основе значений отрицательных контролей



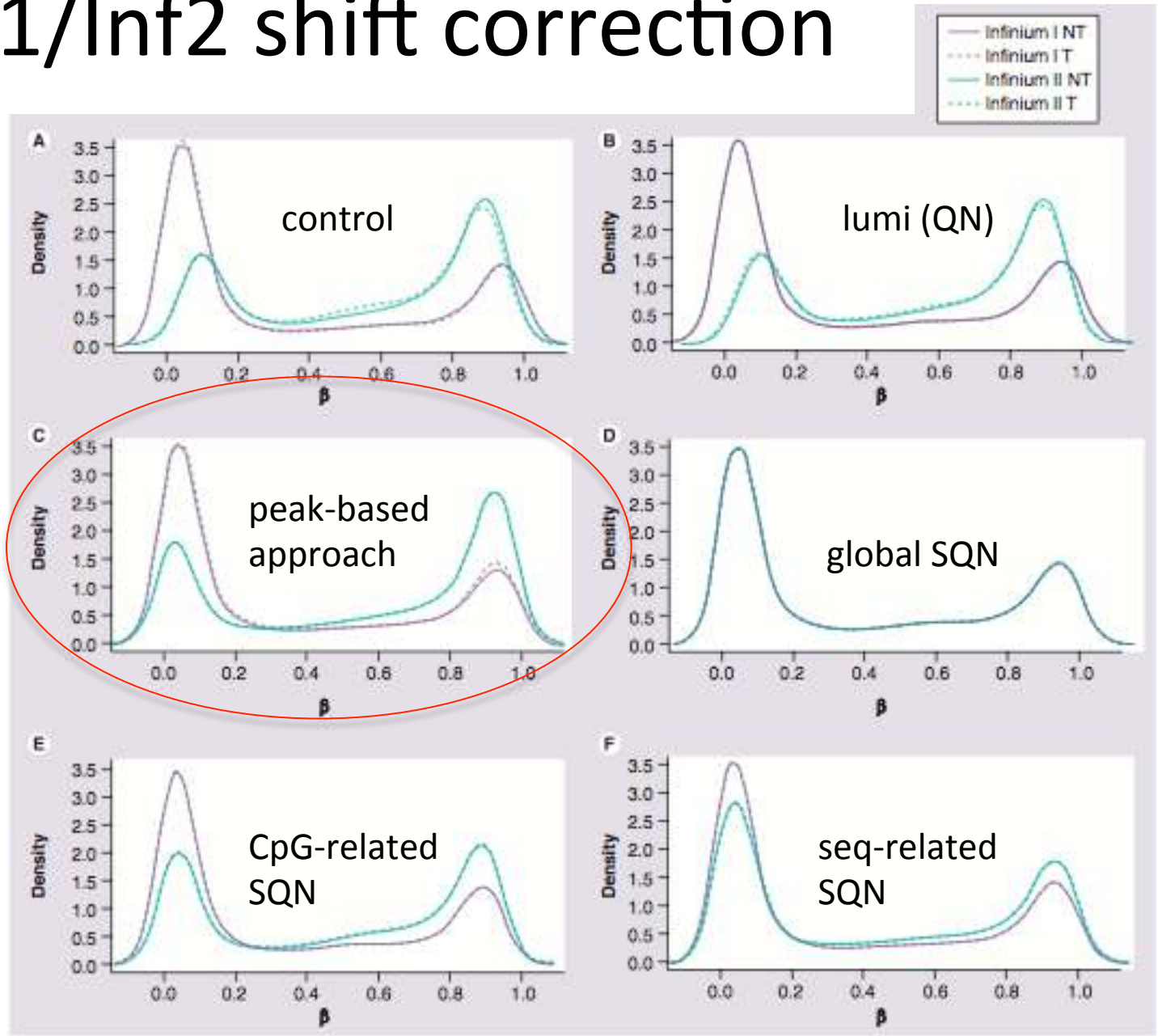
Inf1 vs Inf2 probes

- Inf1 сигналы более стабильны
- Inf1 сигналы имеют расширенный динамический диапазон значений
- Inf1 лучше детектирует близкие к 0 значения, Inf2 – наоборот
- значения Inf1 ближе к «идеальным» результатам
- SD выше у Inf2



Inf1/Inf2 shift correction

- use InfI signals as the anchors to estimate a reference distribution of quantiles
- correct the data so that non_anchor and anchor probes of the same percentiles will have the same value
- “relation to CpG” means S shore/shelf et cet

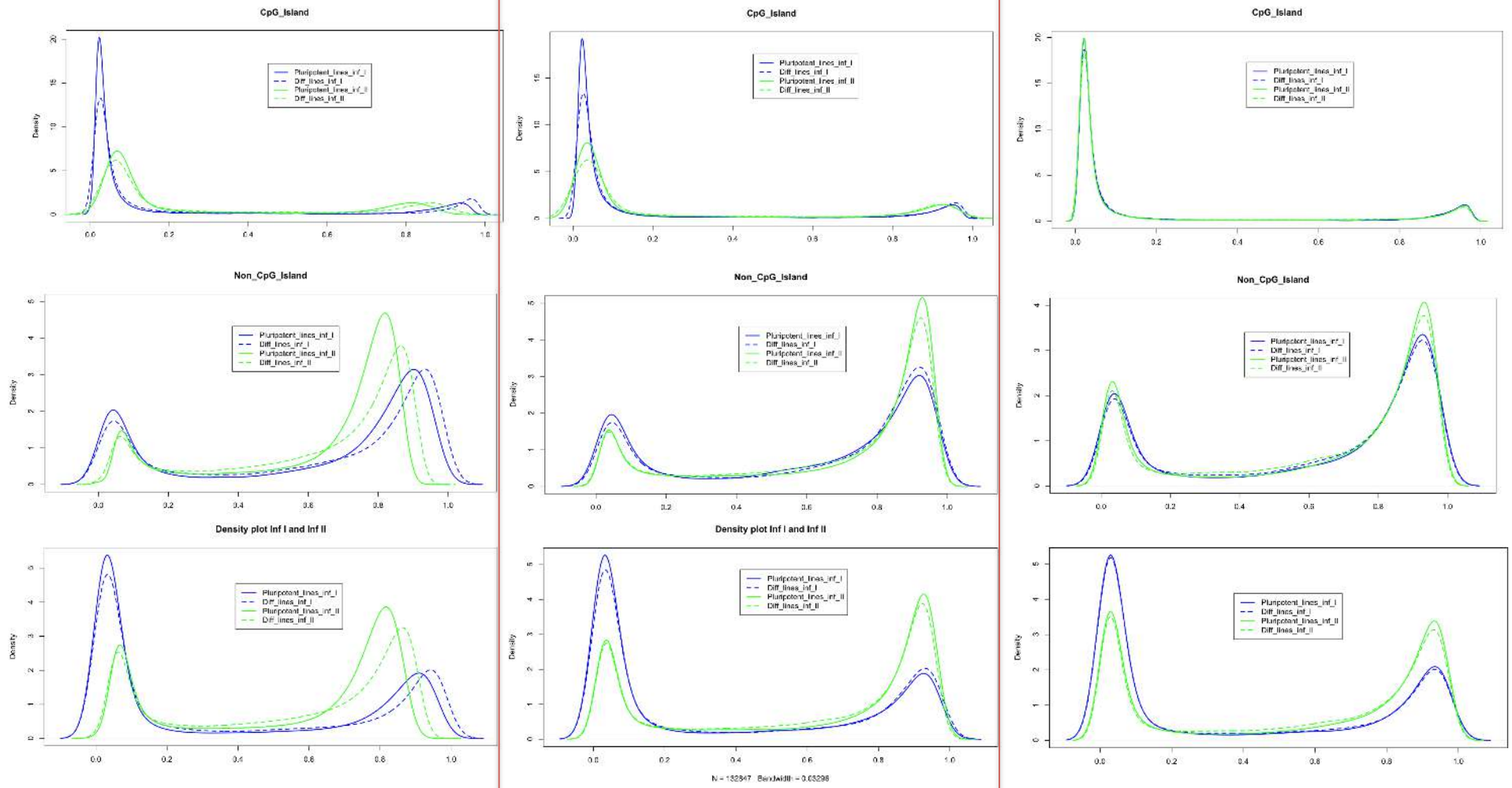


нормализация

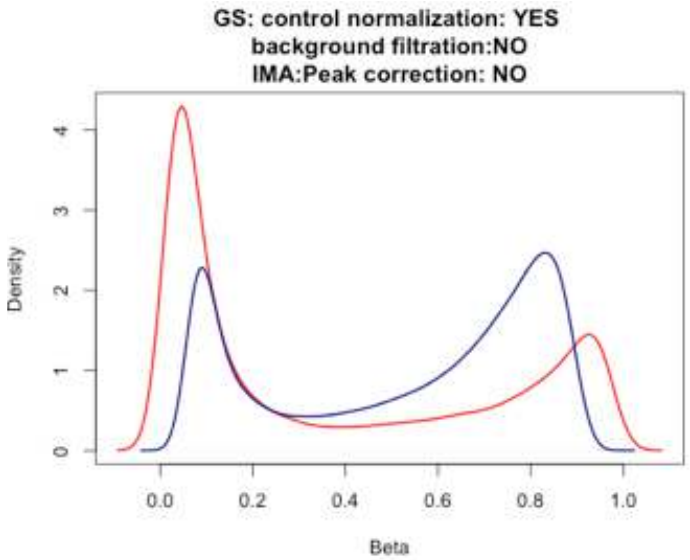
raw data

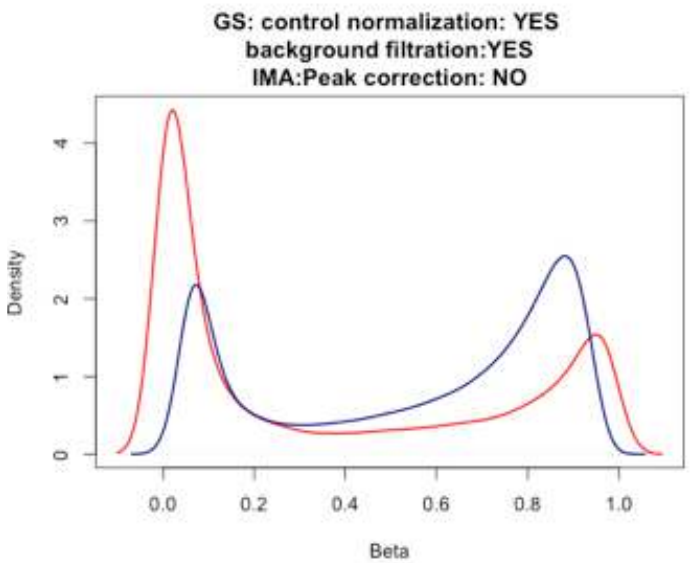
peak correction

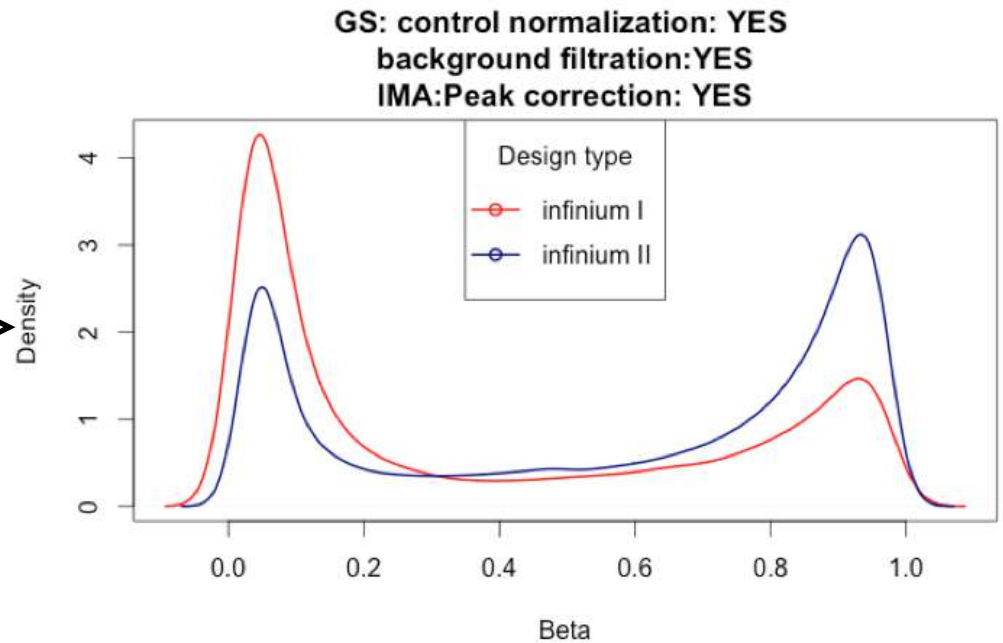
(CpG-related SQN)
Tost pipeline



IMA pipeline + ComBat for batch correction

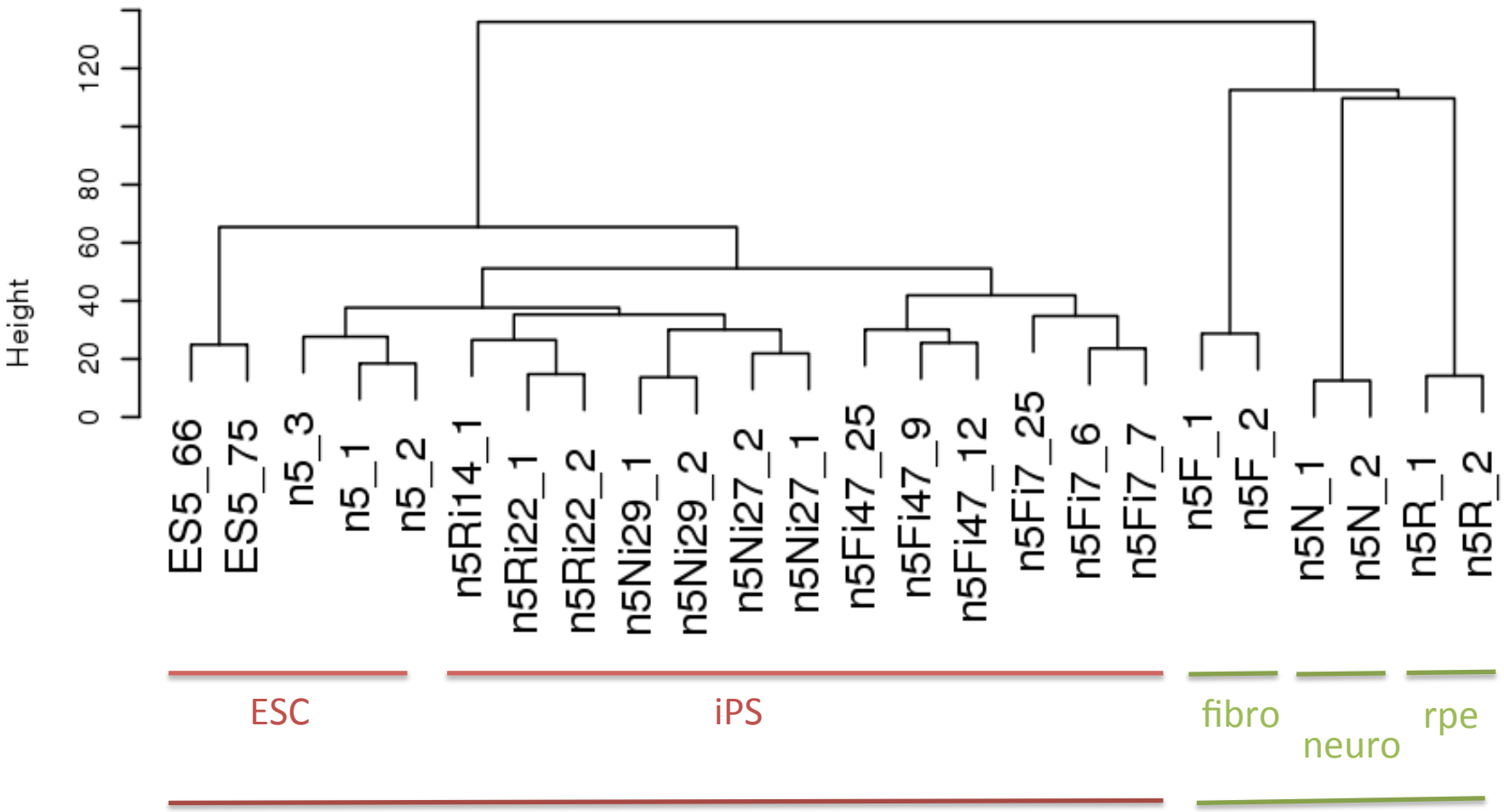




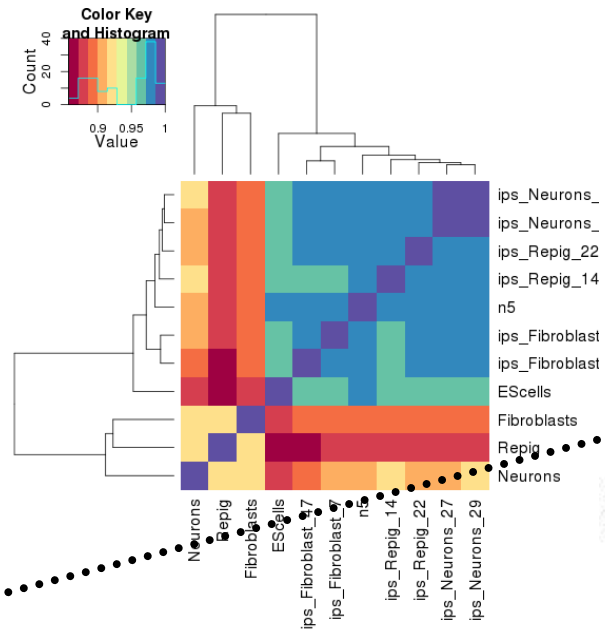


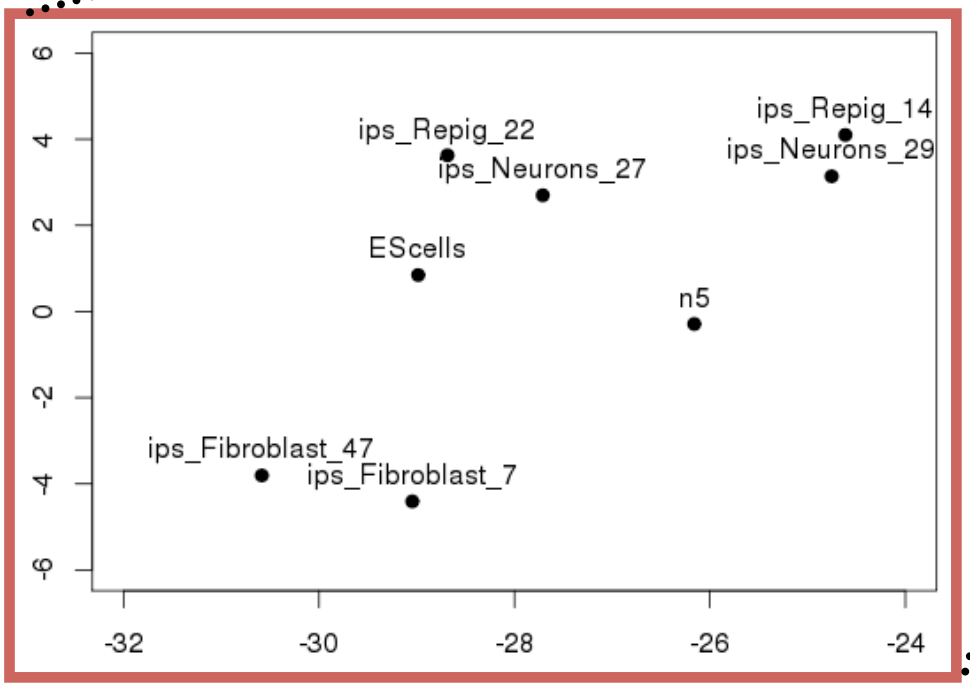
filterdetectP, whether or not to filter probes based on detection P-value;
Xchrom, whether or not to remove the loci from the X,Y chromosome or both;
peakcorrection, whether or not to perform peak correction;
transfm, whether or not to transfer the raw beta value using arcsine square root or logit;
normalization, whether or not to perform quantile normalization;
na.omit, whether or not to remove the loci containing missing beta values;
snpfilter, whether or not to filter out loci whose methylation level are measured by probes containing SNP(s) at/near the targeted CpG site

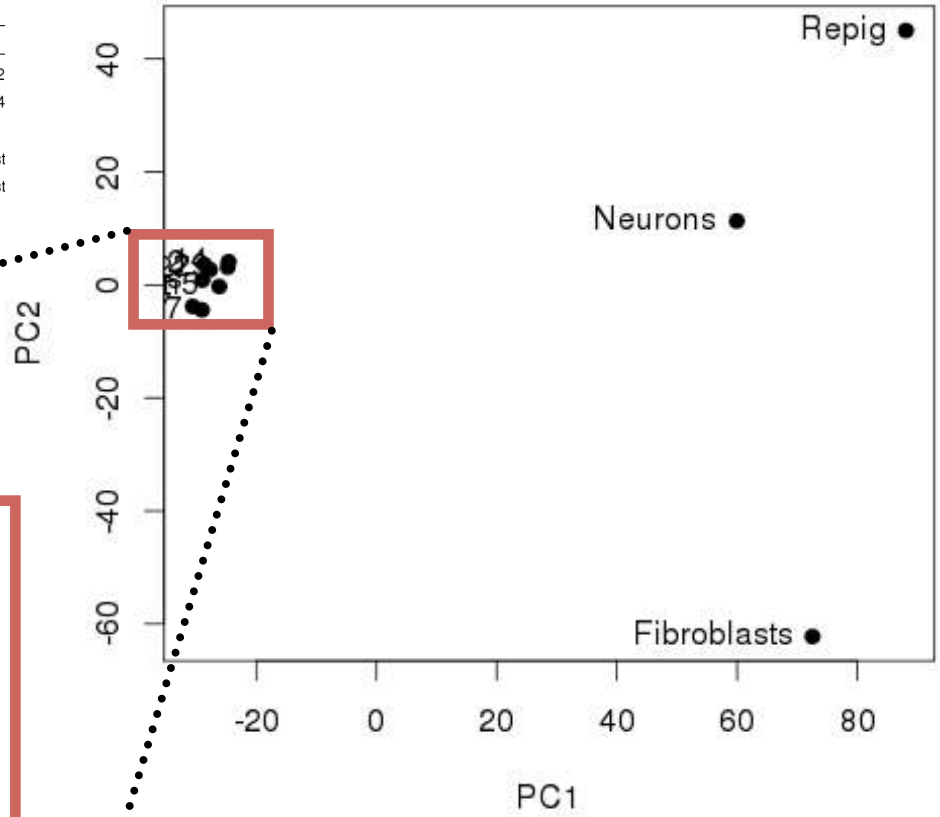
данные после нормализации



корреляция Спирмана с кластеризацией



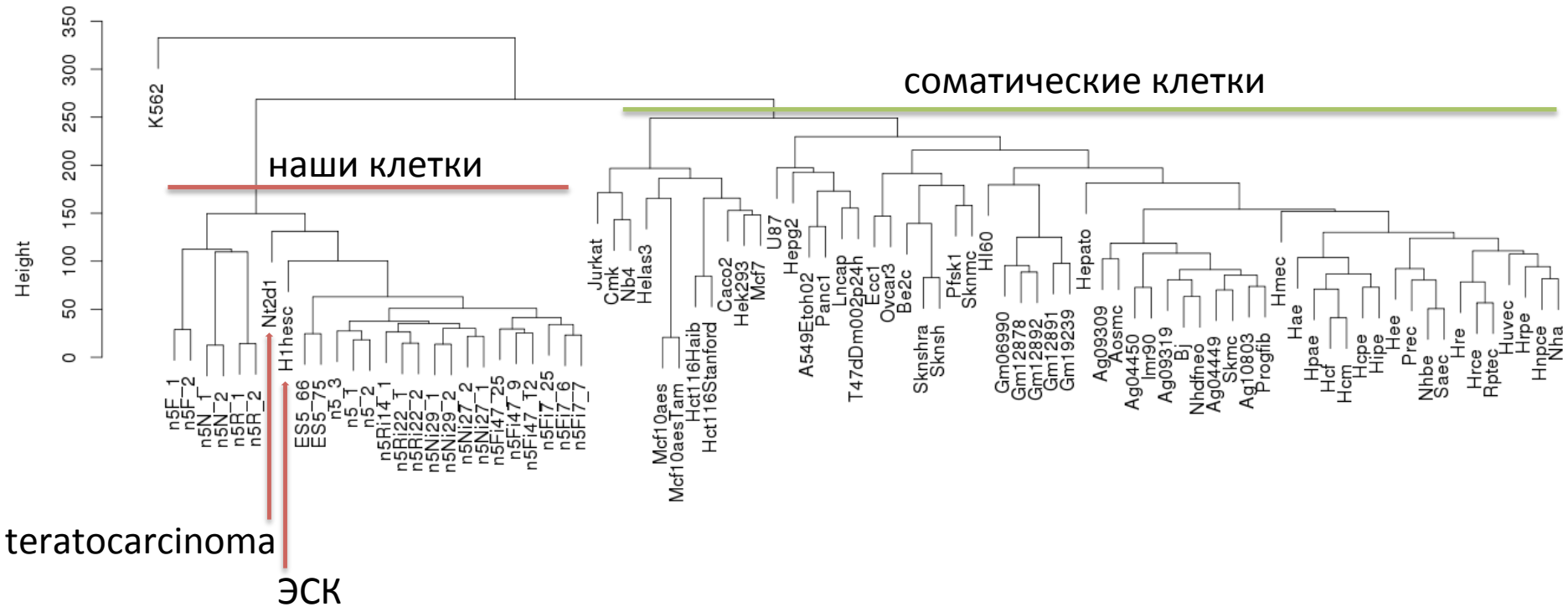




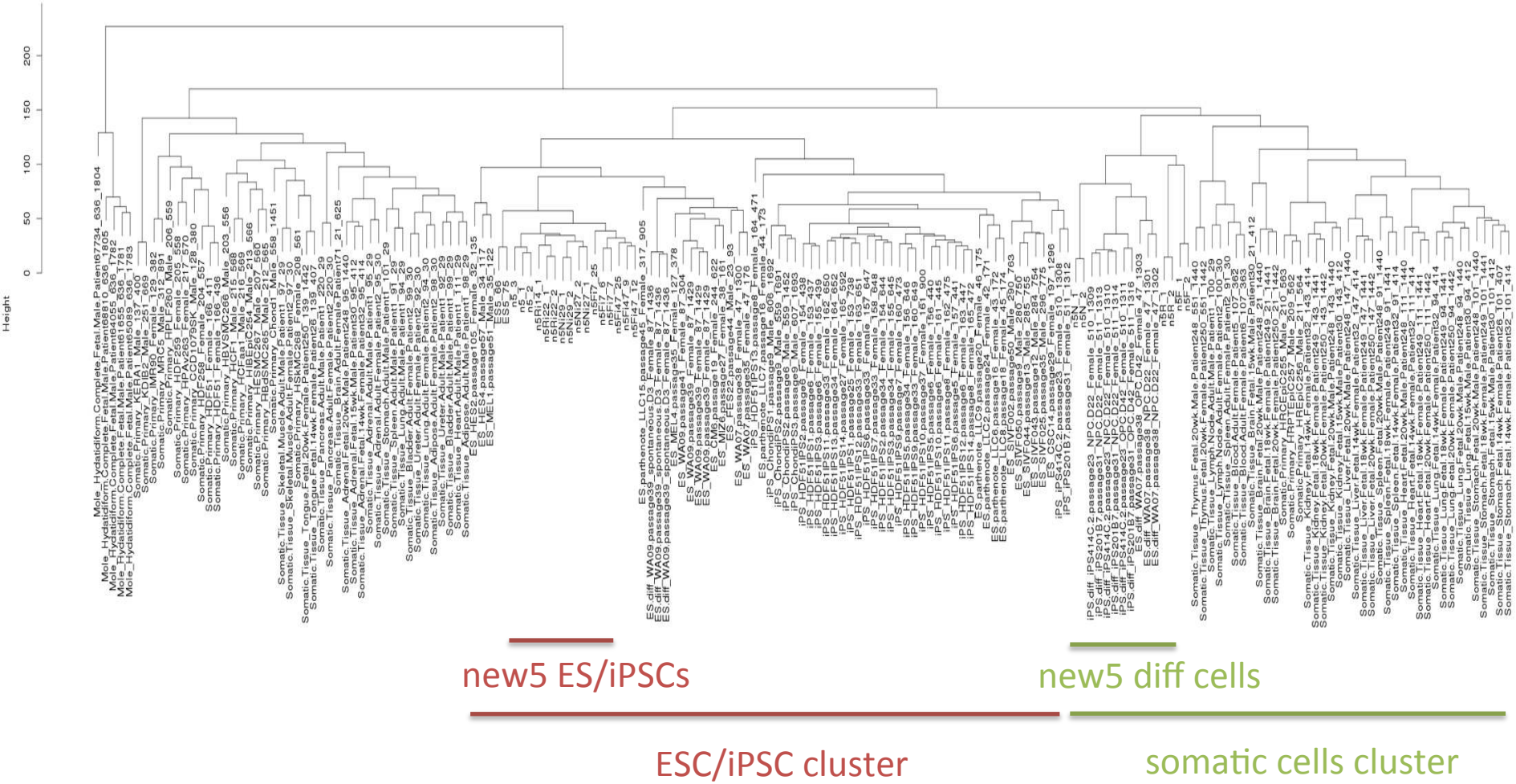

```
Importance of components:
                          PC1     PC2
Standard deviation     47.6691 24.7187
Proportion of Variance  0.5723  0.1539
Cumulative Proportion   0.5723  0.7261
```

сравнение с данными из открытых источников

Encode (63 типа клеток)



сравнение с данными из открытых источников

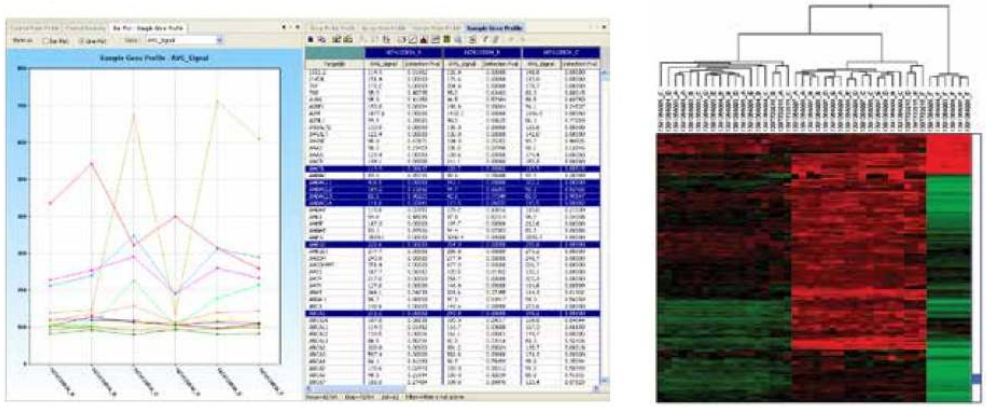


GSE31848 dataset (Nazor 2012)

Illumina HT12 v.4

Probes	Description	Human HT-12 v4.0* 12-sample
RefSeq Content		
NM	Coding transcript, well-established annotation	28,688
XM	Coding transcript, provisional annotation	11,121
NR	Non-coding transcript, well-established annotation	1,752
XR	Non-coding transcript, provisional annotation	2,209
Source	RefSeq source release	Human RefSeq Rel 38

Figure 4: GenomeStudio Gene Expression Module



The GenomeStudio software interface (left) provides a flexible graphical interface for data and controls display. GenomeStudio software contains powerful built-in data display tools, such as line graphs, tables, and heat maps (right) for expression analysis.

Figure 2: Direct Hyb Gene Expression Profiling Bead Design

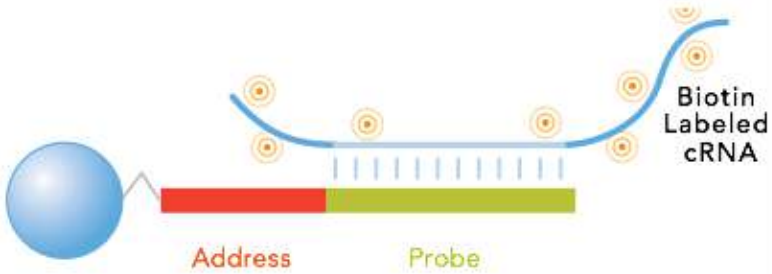


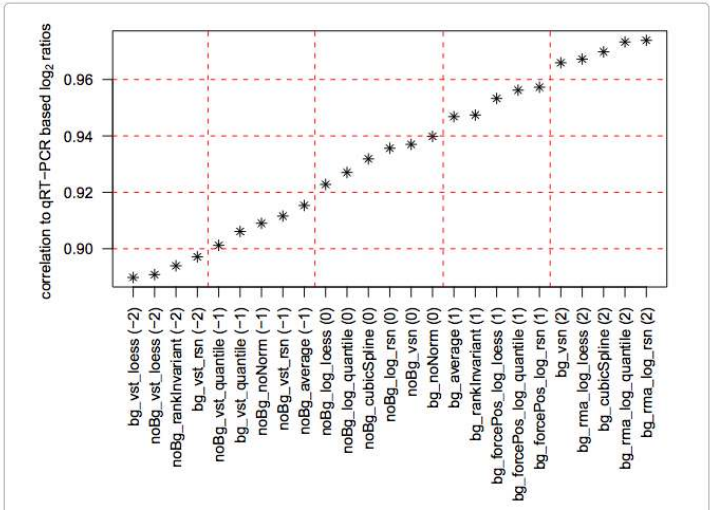
Table 2: Direct Hybridization Gene Expression Assay Product Specifications

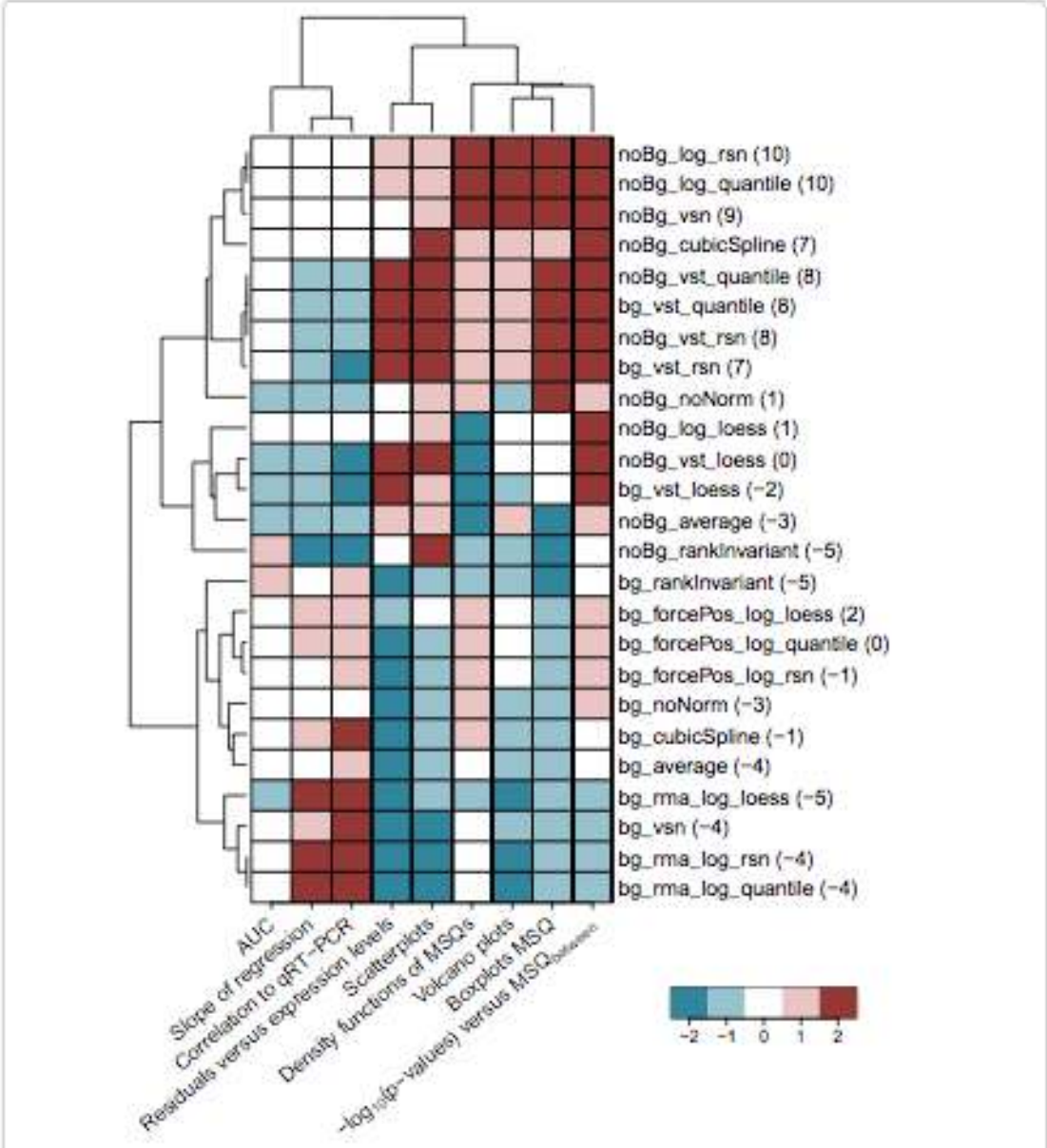
Parameter	Specification
Probe-Length	50-mer gene-specific probe, plus 29-mer address sequence
Sensitivity	≤ 1:250,000
Dynamic Range	≥ 3 logs
Detectable Fold Change	≤ 1.35 fold
Reproducibility CV	< 10%
Input RNA Required	50–500 ng

Normalization methods

TGF-β stimulated and control samples measured at three time points (2 h, 4 h, 12 h) in four replicates

(Schmid 2010)





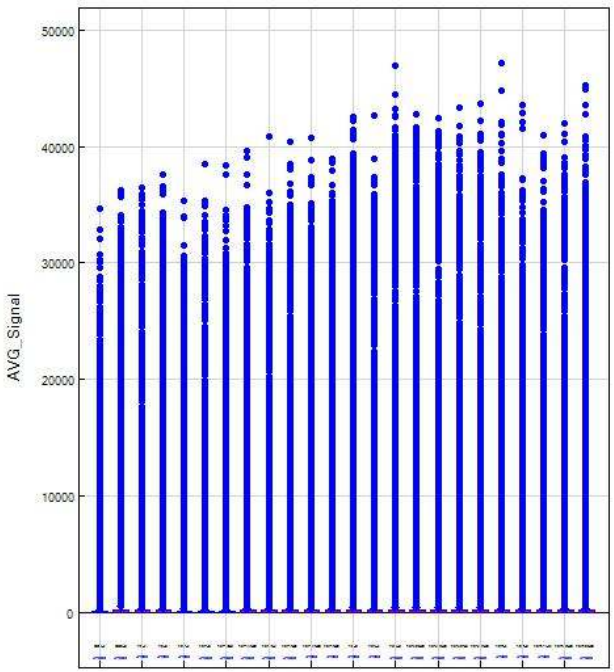
Genome Studio normalization methods

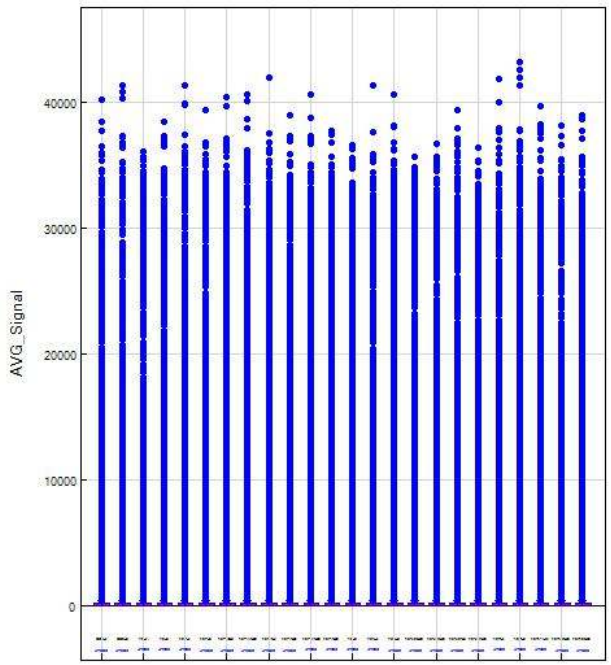
- **Average** - average signal of all samples becomes equal to the global average of all sample signals
- **Quantile** - is a method used to make the distribution, median, and mean of probe intensities the same for every sample. The normalization distribution is chosen by averaging each quantile across samples. Like cubic spline, this method assumes that all samples have similar distributions of transcript abundance.
- **Cubic spline** - For each sample, the vector of quantile intensities is computed. Similarly, quantiles for the "virtual" averaged sample after background subtraction are computed. Cubic B-spline is computed and used for interpolation.
- **Rank invariant** - Rank invariant normalization uses a set of probes that is rank invariant between a given sample and a virtual sample. Rank invariant normalization operates under the assumption that probes with similar ranking between samples have similar expression levels.

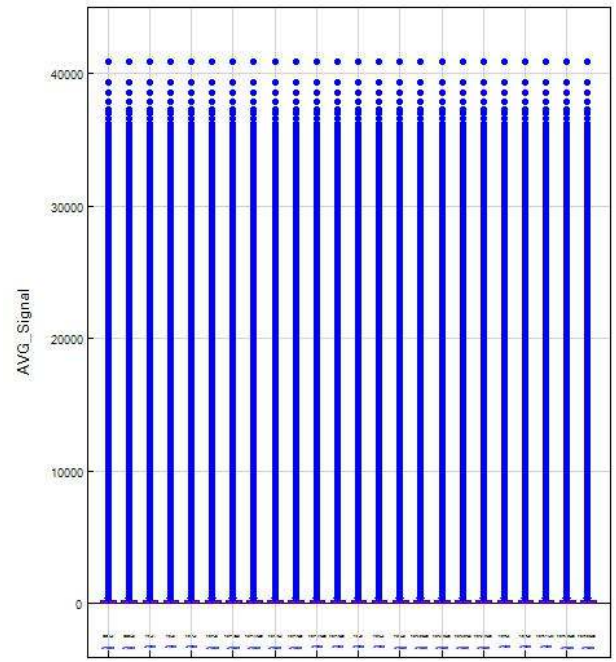
No normalization

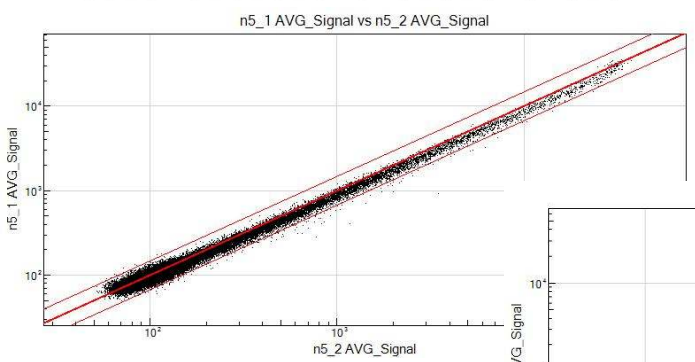
cubic

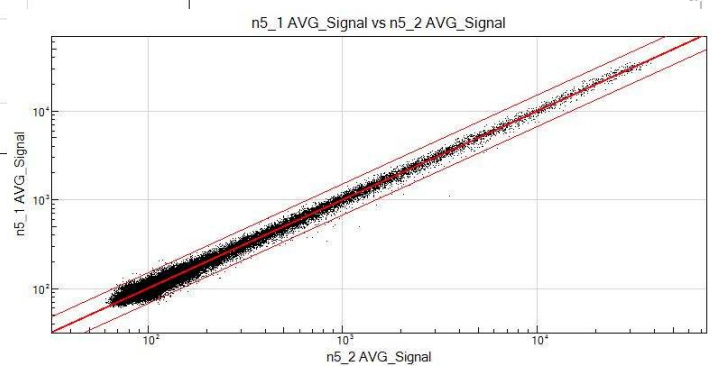
quantile

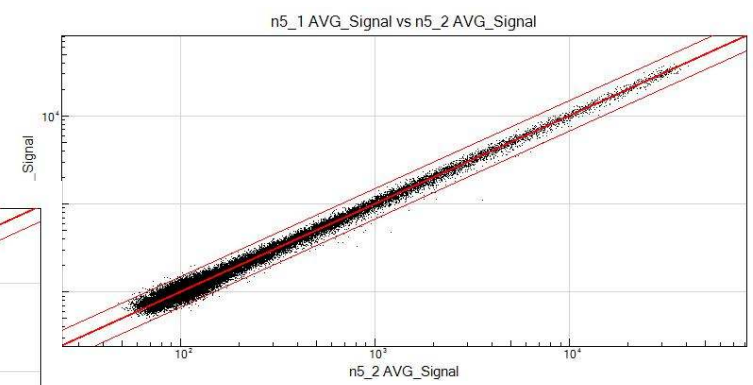


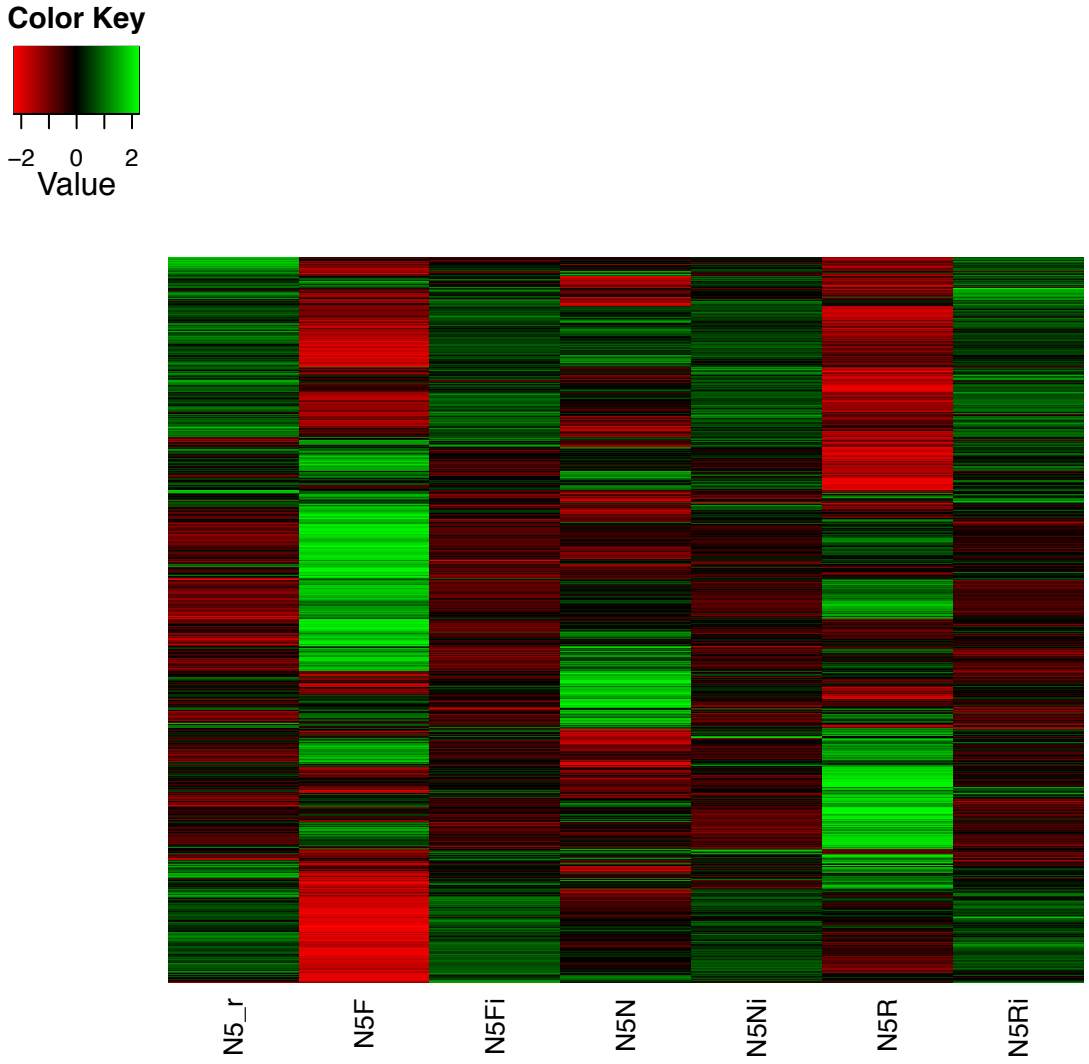




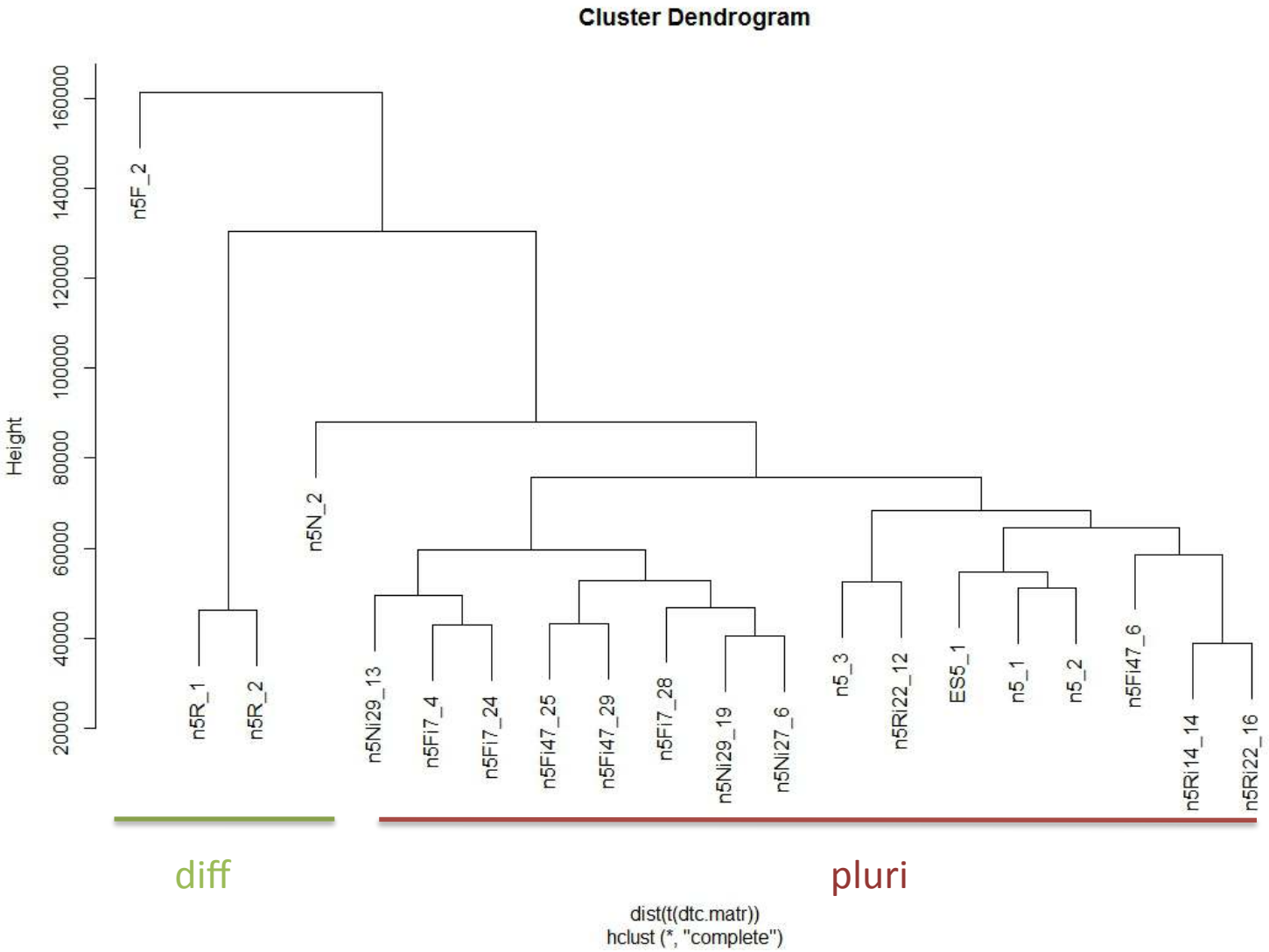


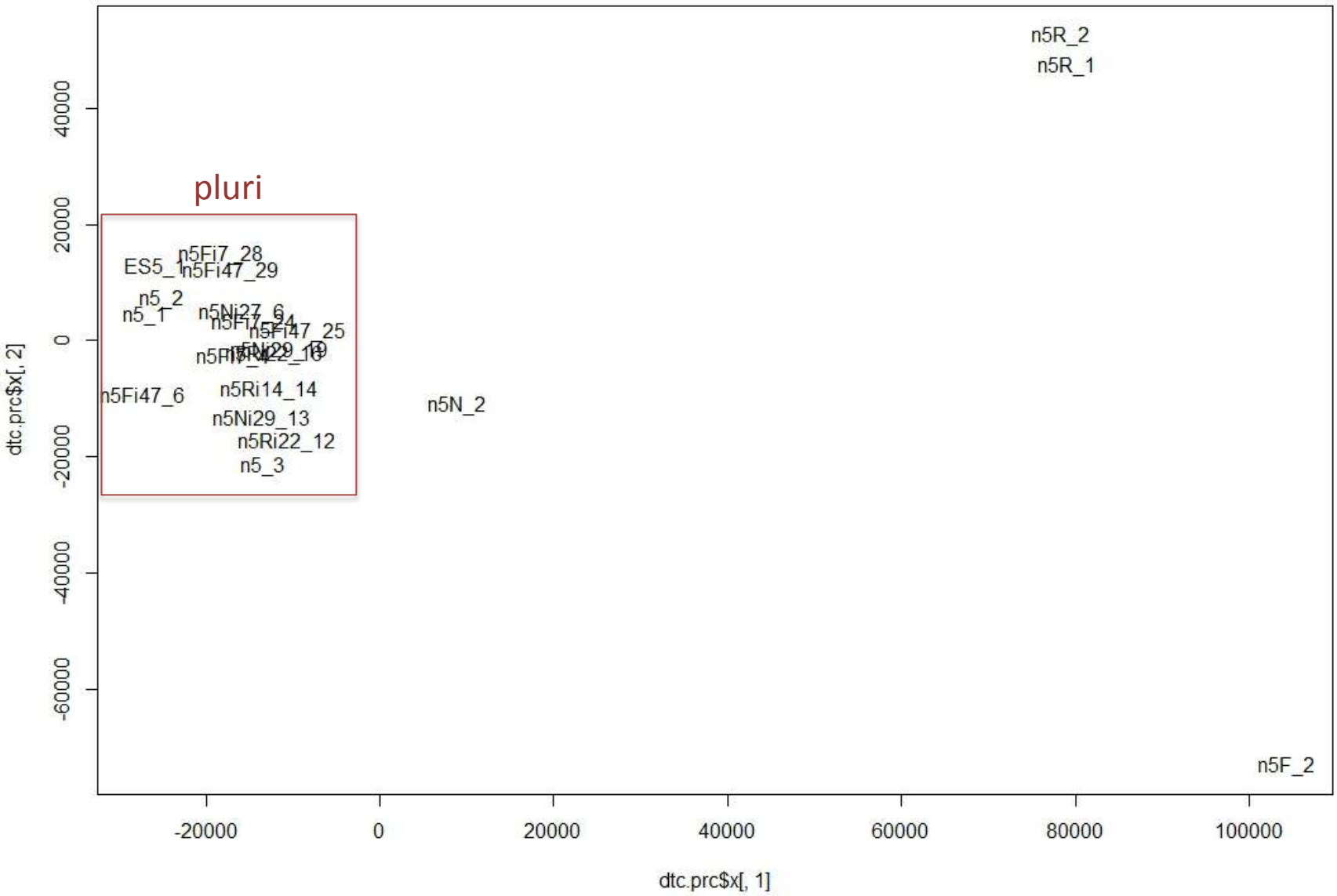





Color Key
−2
0
2
Value
N5_r
N5F
N5Fi
N5N
N5Ni
N5R
N5Ri

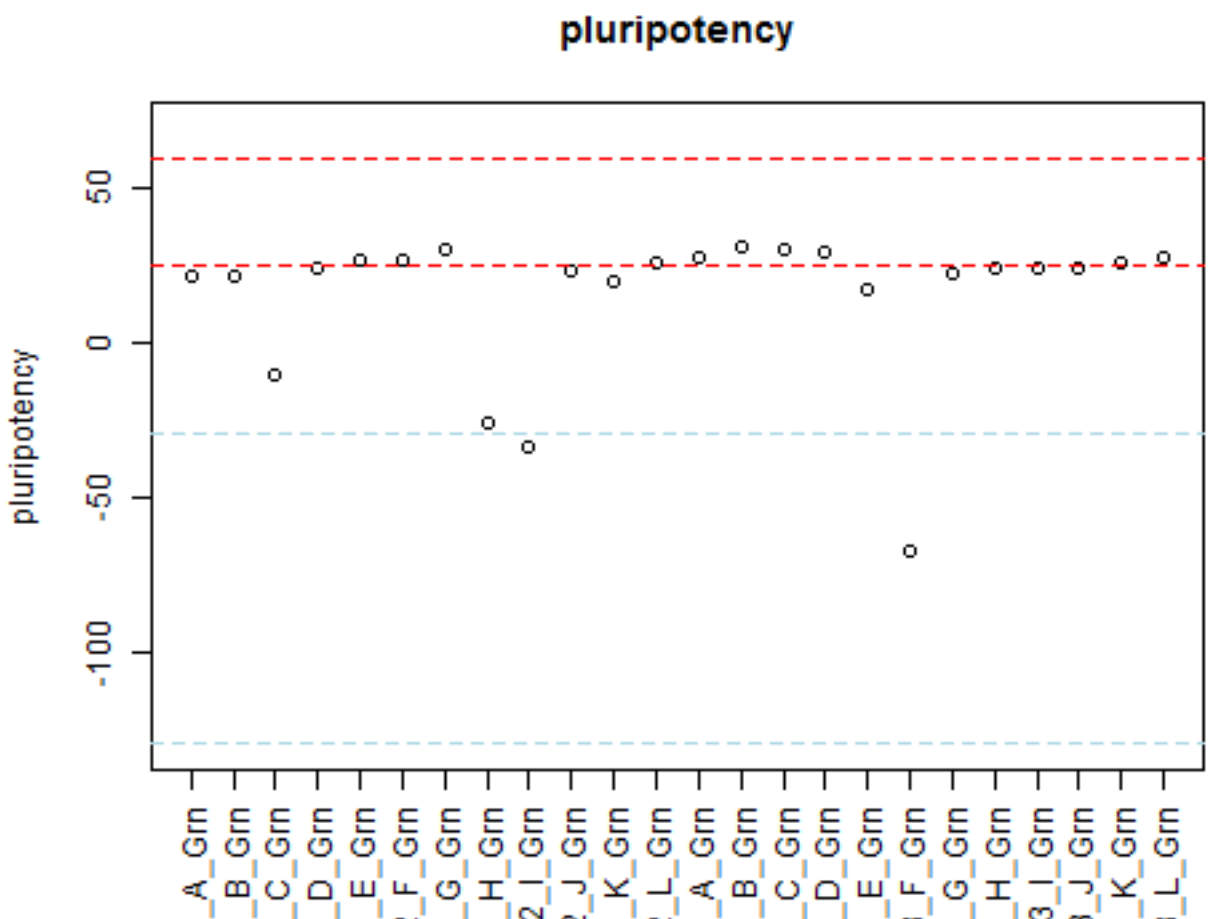
данные после нормализации




pluri
n5R_2
n5R_1
n5Fi7_28
ES5_1
n5Fi47_29
n5_2
n5_1
n5Ni27_6
n5Fi47_25
n5Fi47_6
n5Ri14_14
n5Ni29_13
n5Ri22_12
n5_3
n5N_2
n5F_2
40000
20000
0
-20000
-40000
-60000
-20000
0
20000
40000
60000
80000
100000
dtc.prc$x[, 1]
dtc.prc$x[, 2]

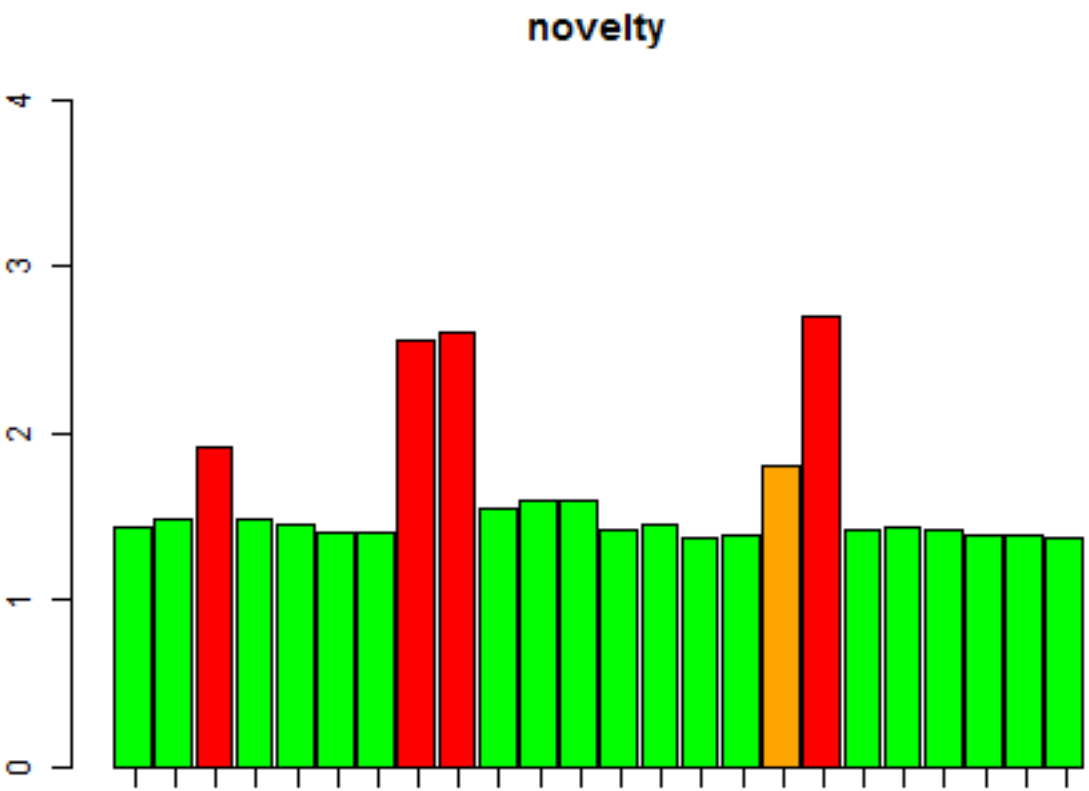
“pluritest” by Alex Meissner

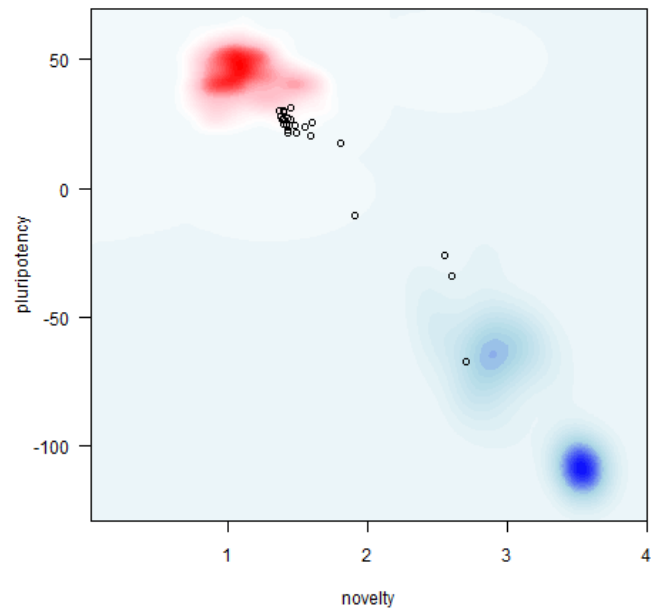
- *Pluripotency Score*: A score that is based on all samples (pluripotent cells, somatic cells and tissues) in the stem cell model matrix (red - 95% pluri samples, blue – 95% somatic)



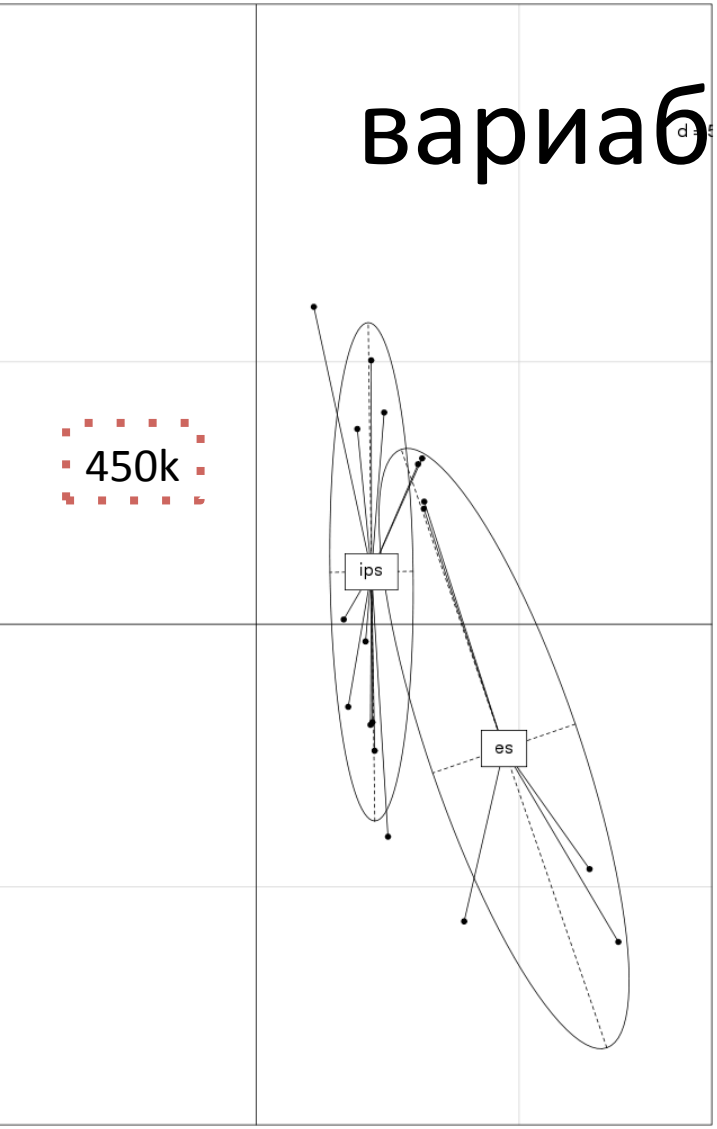
“pluritest” by Alex Meissner

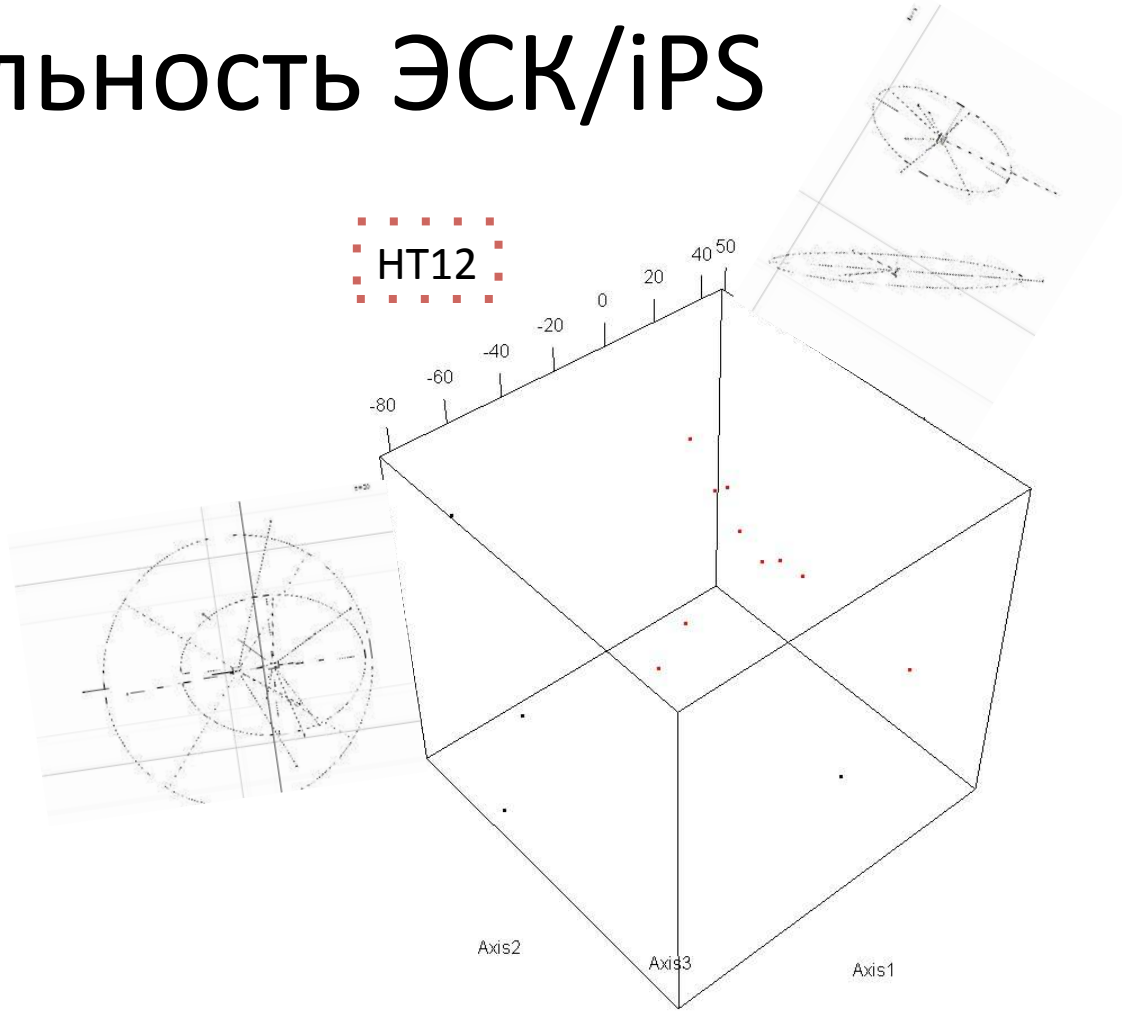
- *Novelty Score*: A score that is based on well-characterized pluripotent samples in the stem cell model matrix. A low Novelty Score indicates that the test sample can be well reconstructed based on existing data from other well-characterized iPSC and ESC lines. A high Novelty Score indicates that there are patterns in the tested sample that cannot be explained by the currently existing data from well-characterized, karyotypic normal pluripotent stem cells.





вариабельность ЭСК/iPS

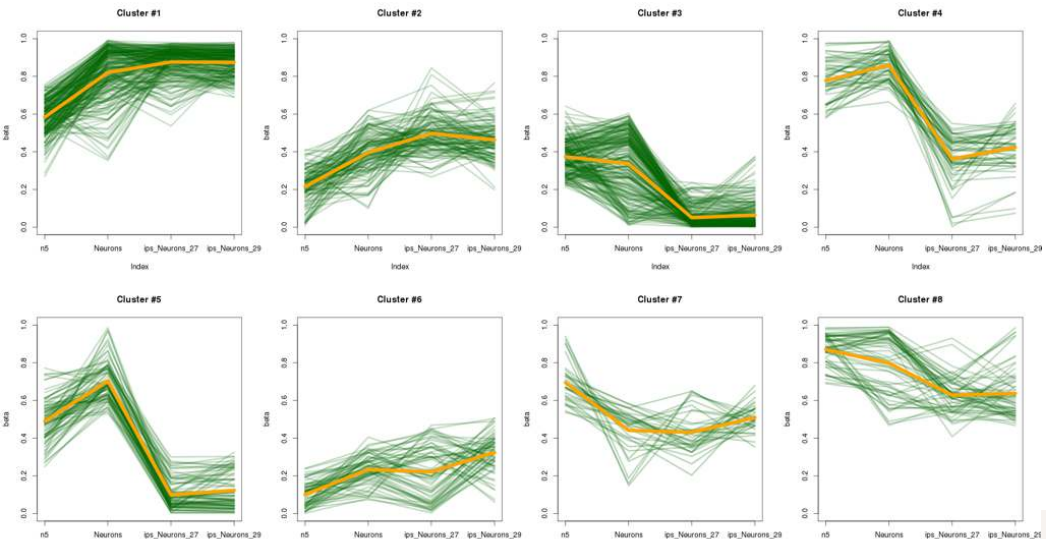


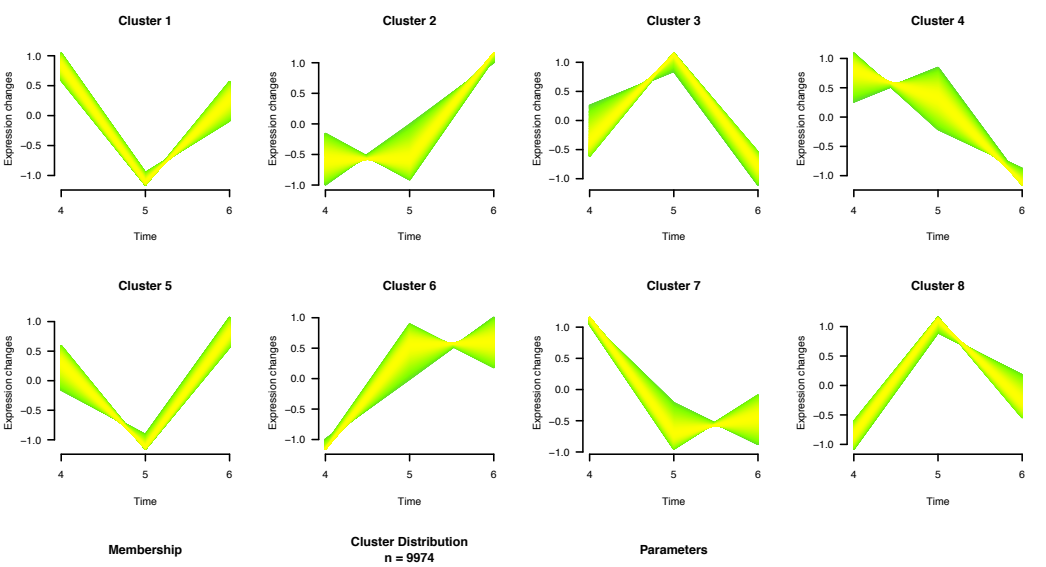


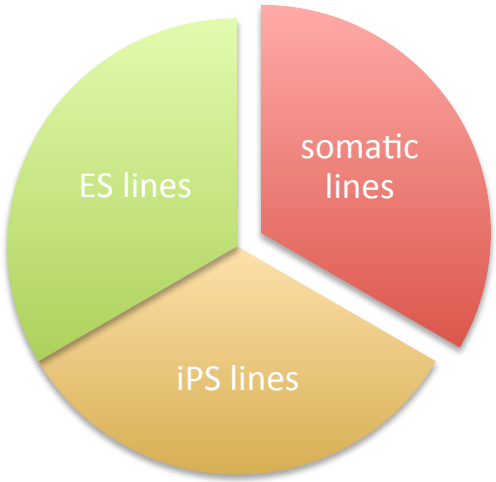
INFO_PERCENT	AXIS
27.356%	1
19.195%	2

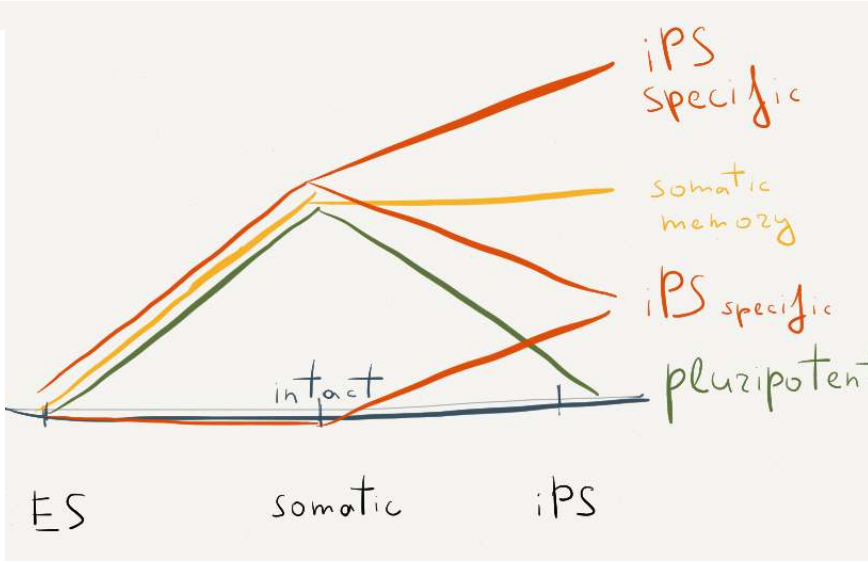
INFO_PERCENT	AXIS
20.068%	1
15.751%	2
11.547%	3

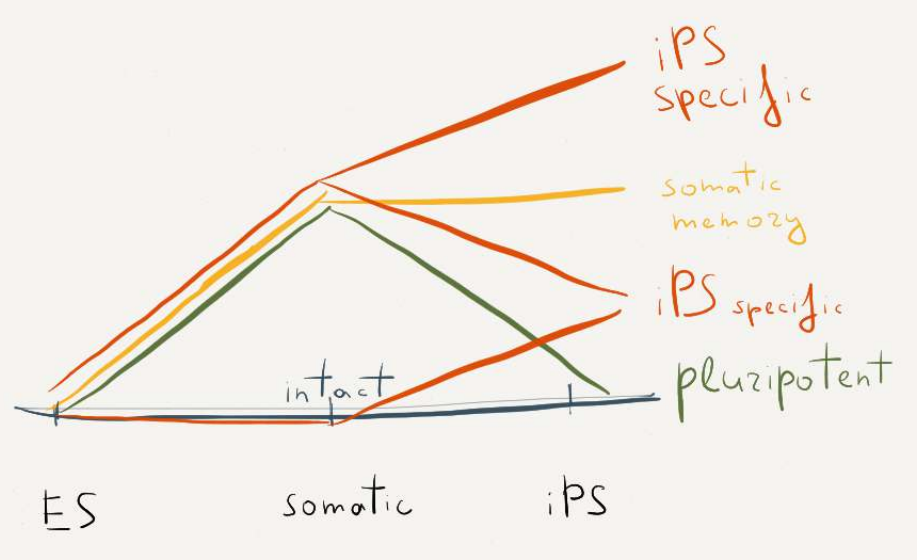
принцип анализа – кластеры?

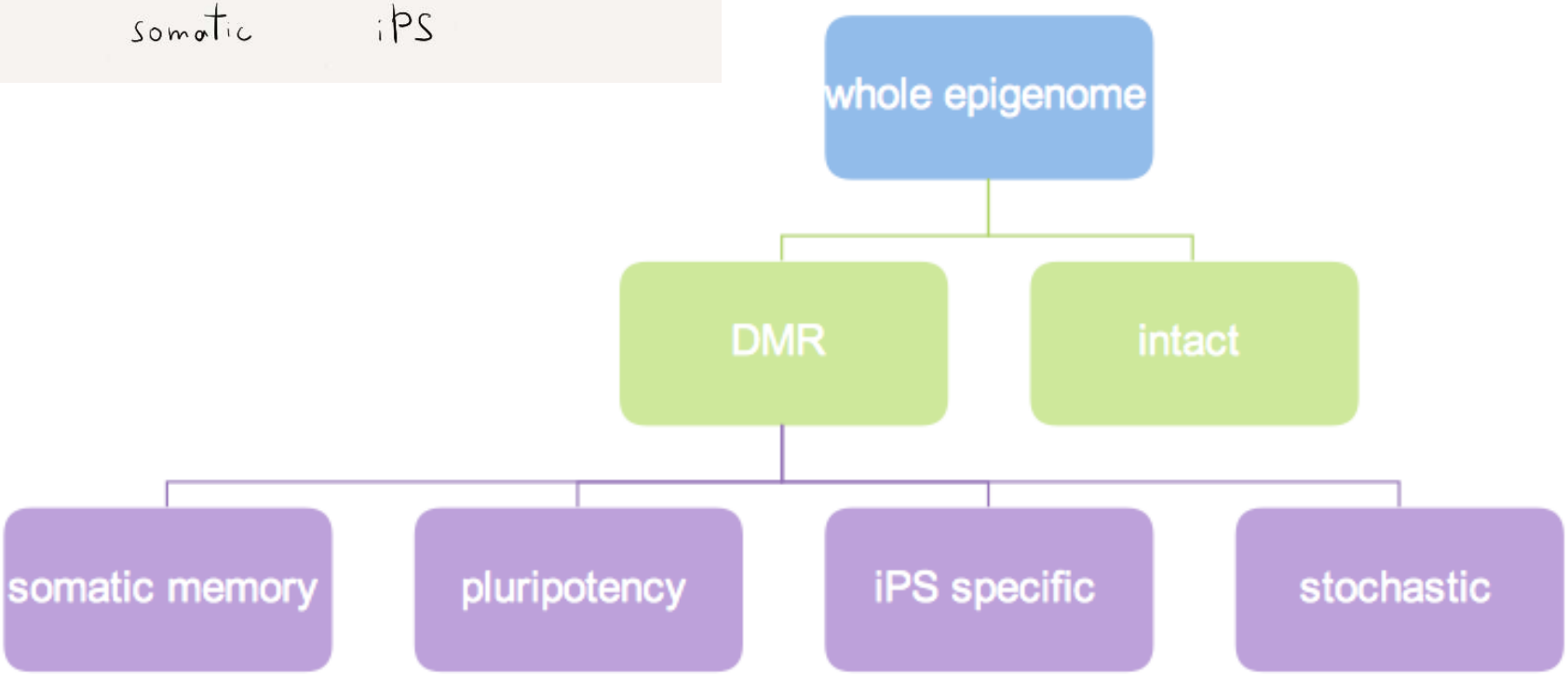




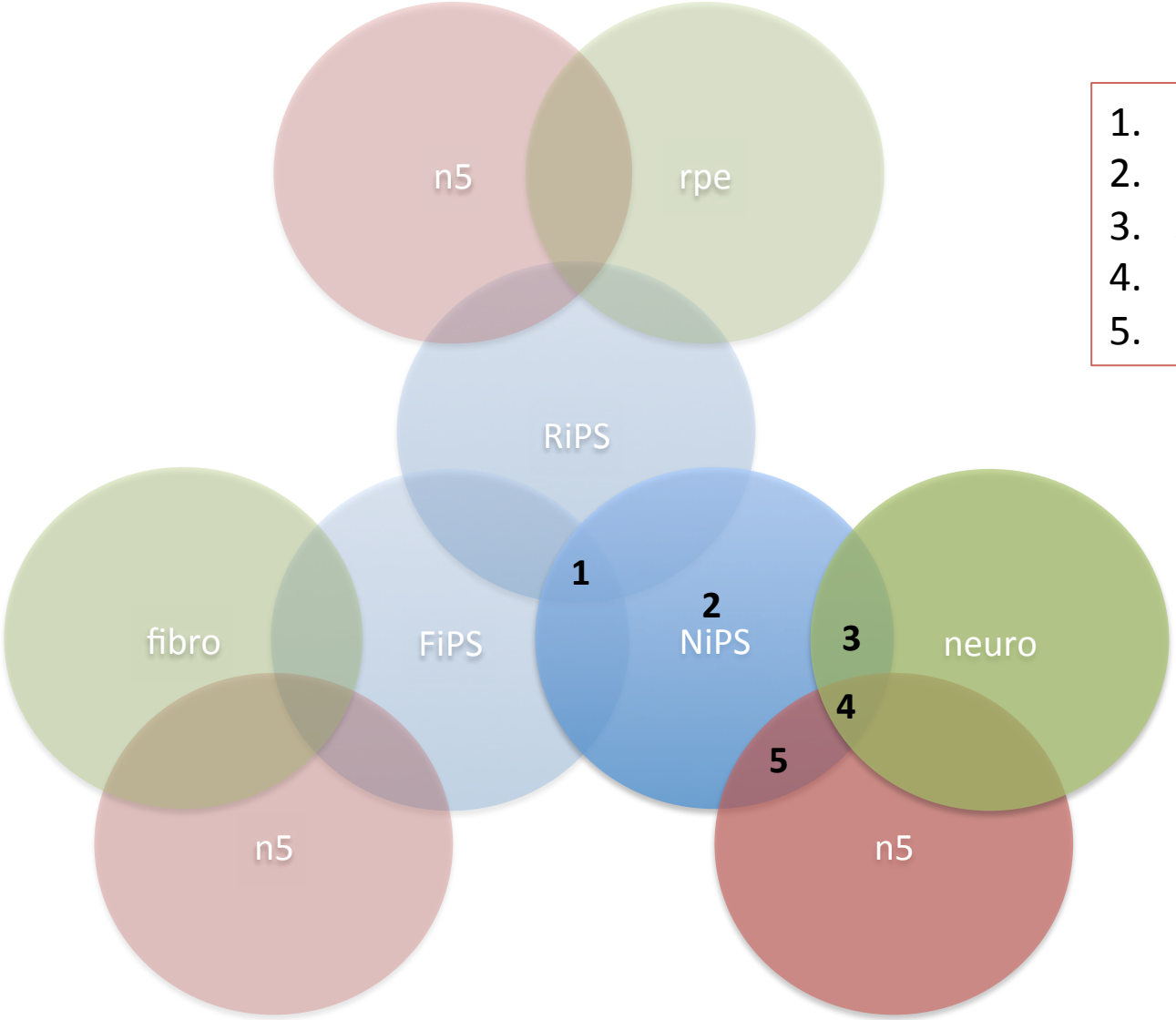





iPS specific
somatic memory
iPS specific
pluripotent
intact
ES
somatic
iPS

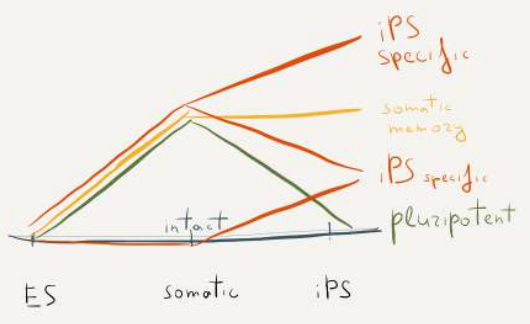

whole epigenome
DMR
intact
somatic memory
pluripotency
iPS specific
stochastic

- 450k – IMA (Wilcoxon rank-sum test), pval 0.01, delta 0.2
- HT12 – t-test, pval 0.01/0.05, foldchange 1.5, qvalue 0.05

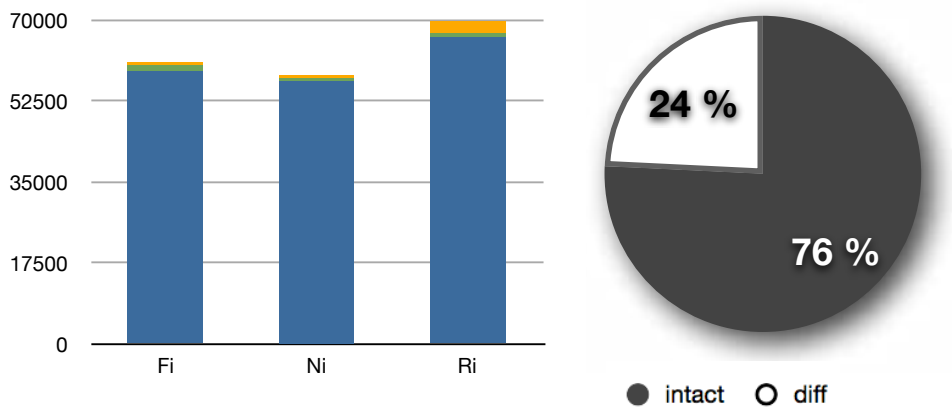


1. reprogramming trace
2. iPS-specific
3. somatic memory
4. intact
5. pluripotent

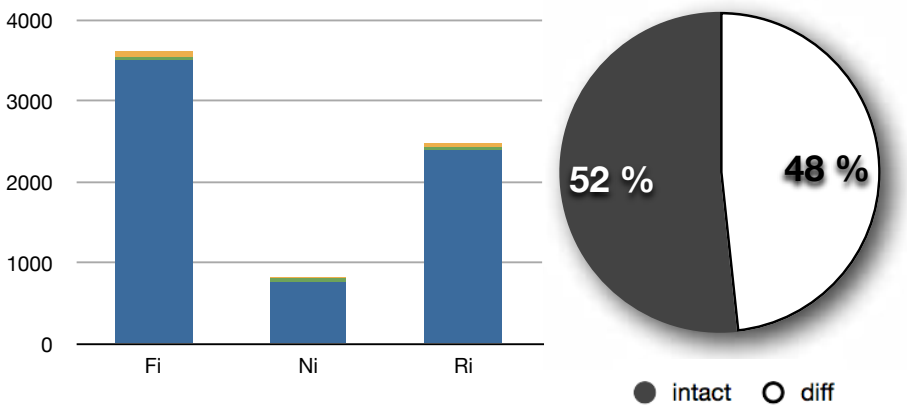
соотношение кластеров

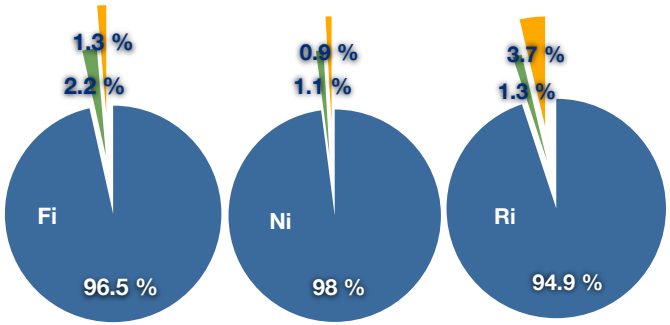


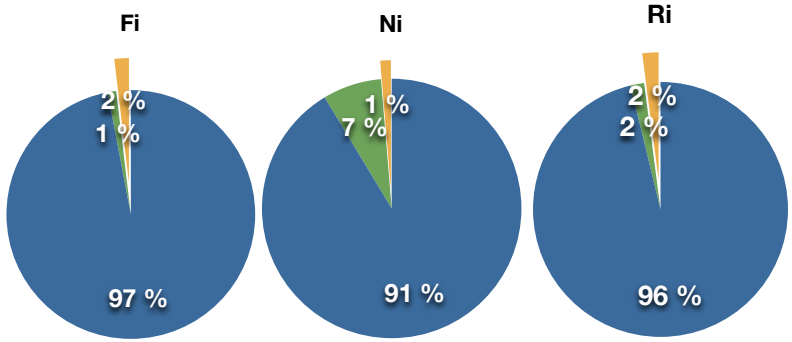
450k



HT12

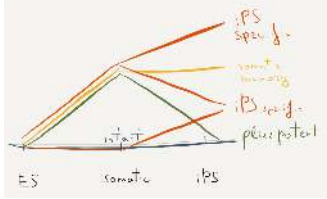


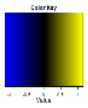




pluri somatic memory specific

функциональность?

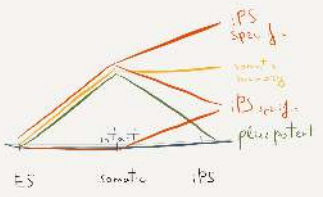


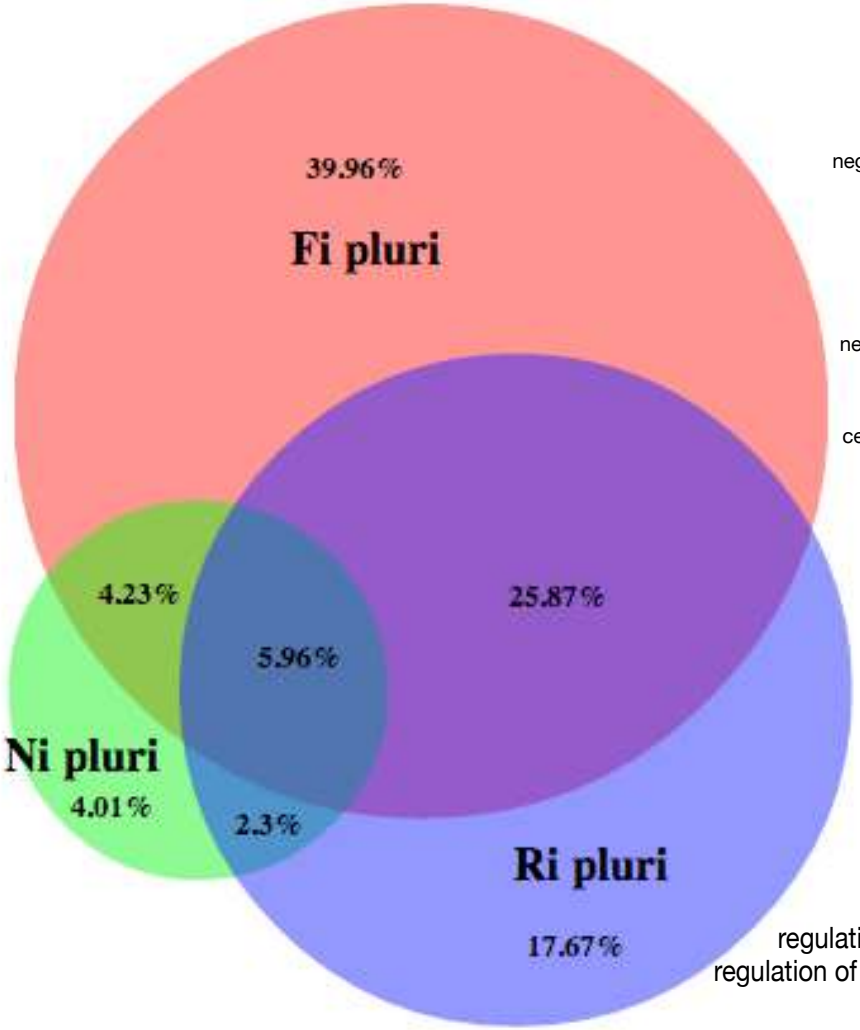


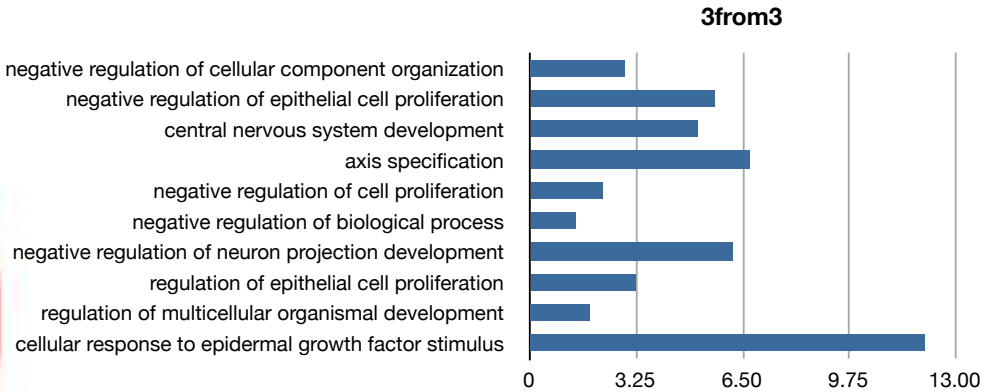
(gprox.sourceforge.net)

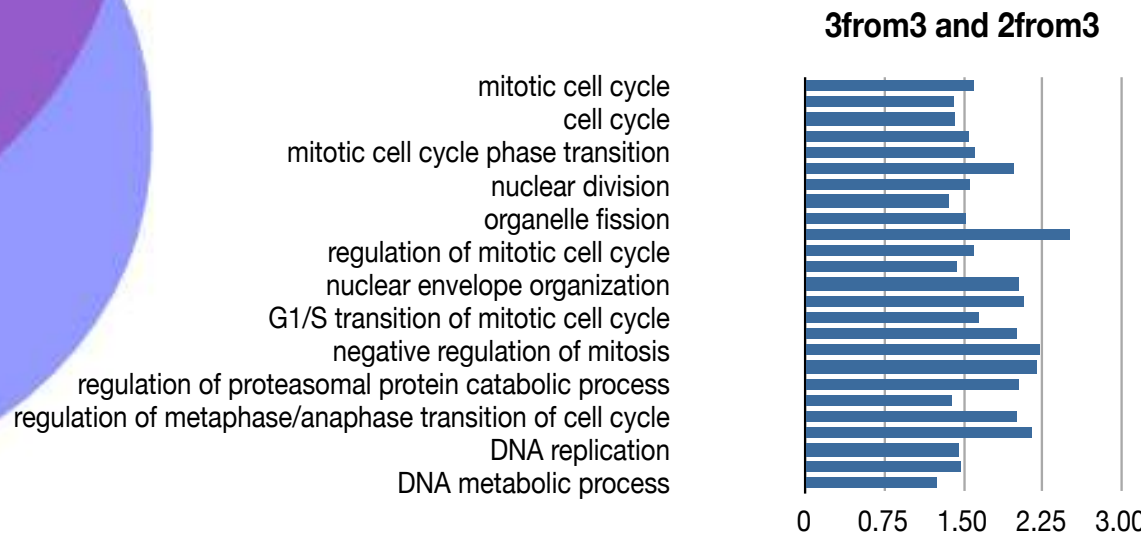
GO enrichment	1. pluripotent	2. somatic memory	3. iPS specific
common processes for all groups	microtubule motor activity, M phase of **mitotic cell cycle**, mitotic sister chromatid segregation, calcium ion binding, nuclear pore, chromosome, centromeric region, sulfuric ester hydrolase activity	**negative regulation of transcription from RNA polymerase II promoter**, neural tube closure, patterning of blood vessels, NAD(P)H dehydrogenase (quinone) activity	metal ion binding, nuclear chromatin, fibroblast growth factor−activated receptor activity, muscle organ development, calcium−mediated signaling

pluripotency-related part

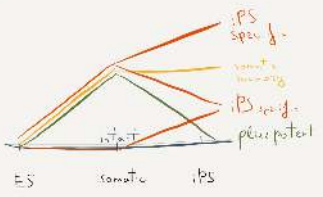


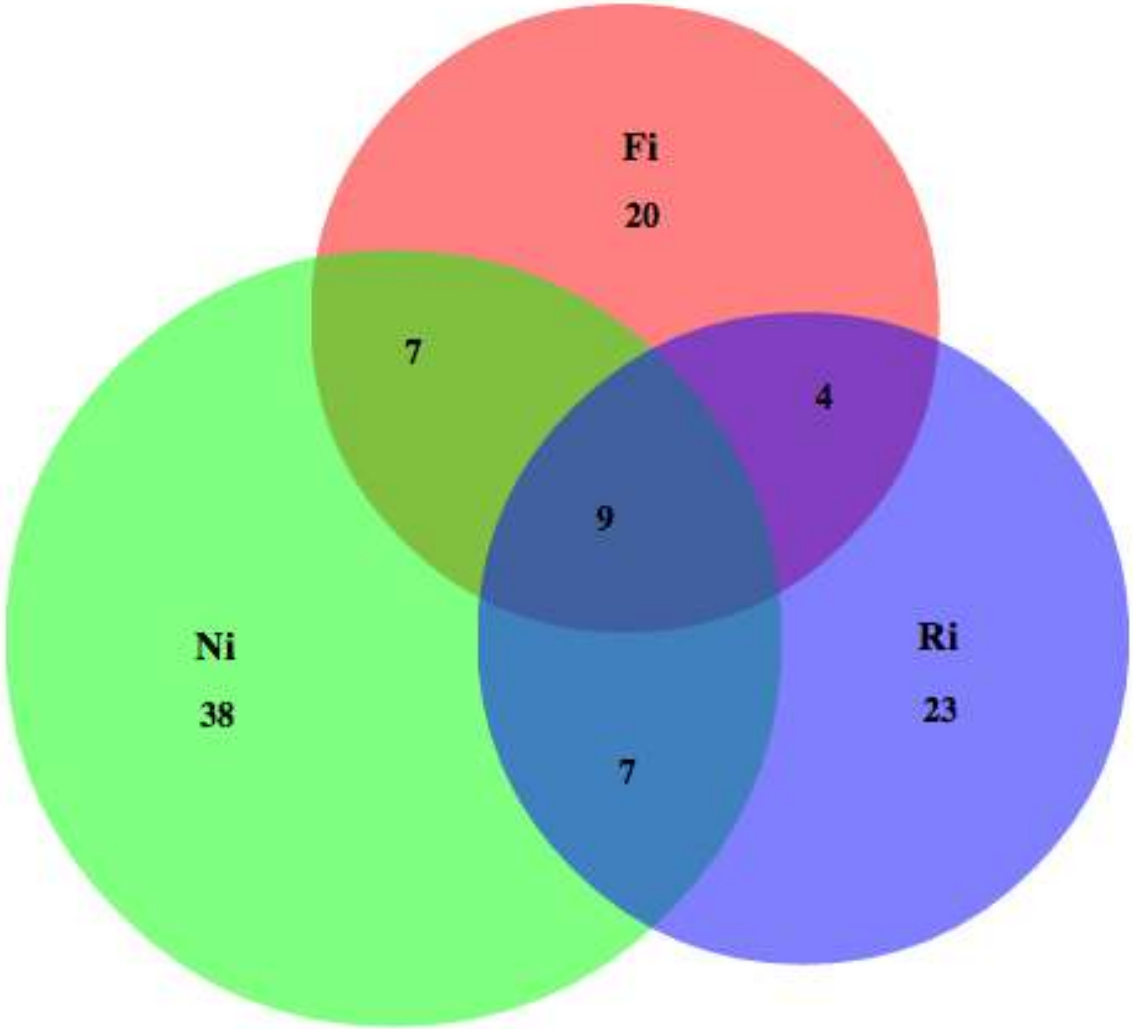






somatic memory part

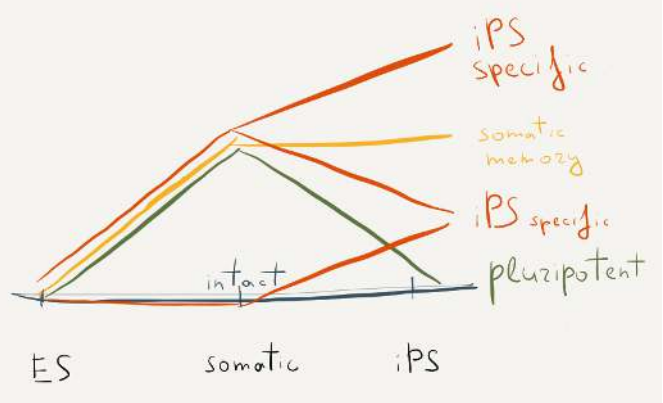




old somatic memory vs young

HT12	intact_old	sm_old	pluri_old	specific_old	diff sum (young only)
intact_young	6250	35	2338	38	2411
sm_young	15	18	10	19	44
pluri_young	107	0	1117	9	116
specific_young	1	1	6	10	8
diff sum (old only)	123	36	2354	66	

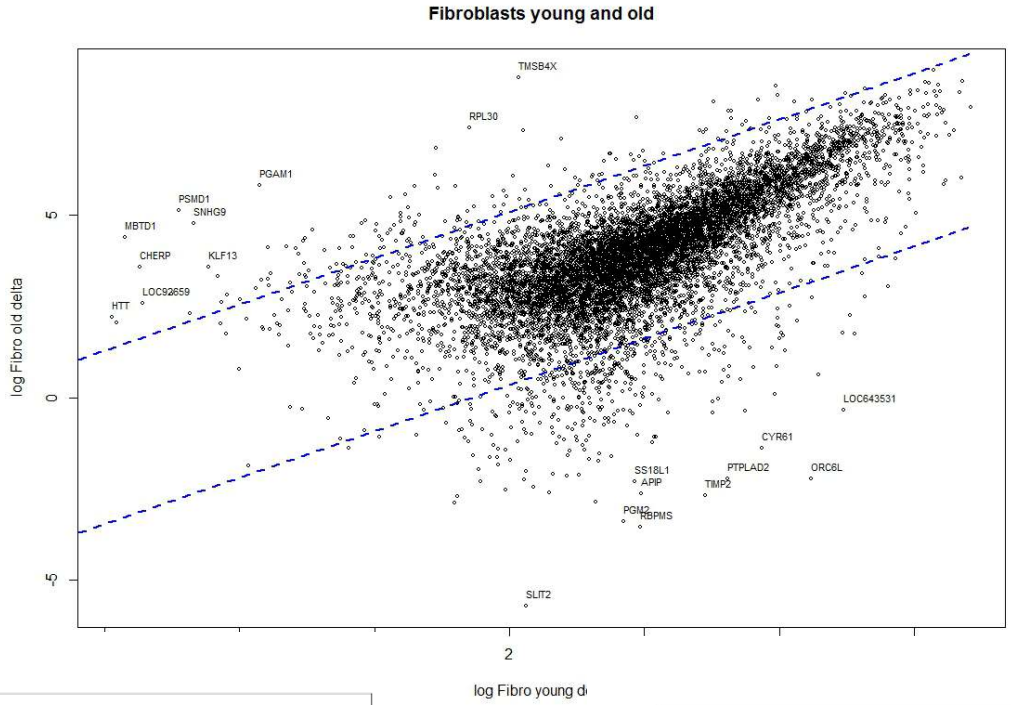
450k	intact_old	sm_old	pluri_old	specific_old	diff sum (young only)
intact_young	413275	156	3461	88	3705
sm_young	426	317	26	33	485
pluri_young	7376	21	49369	92	7489
specific_young	319	28	211	425	558
diff sum (old only)	8121	205	3698	213	

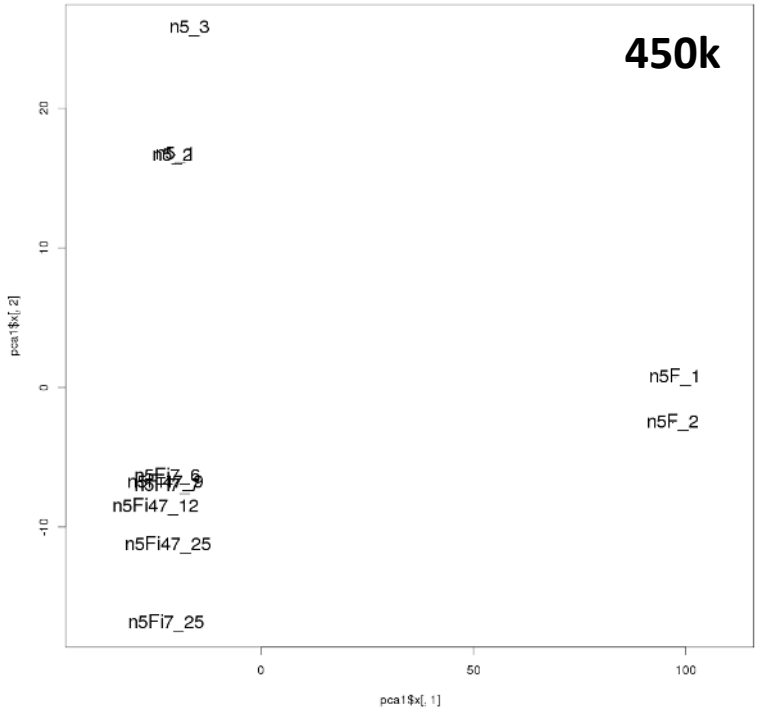


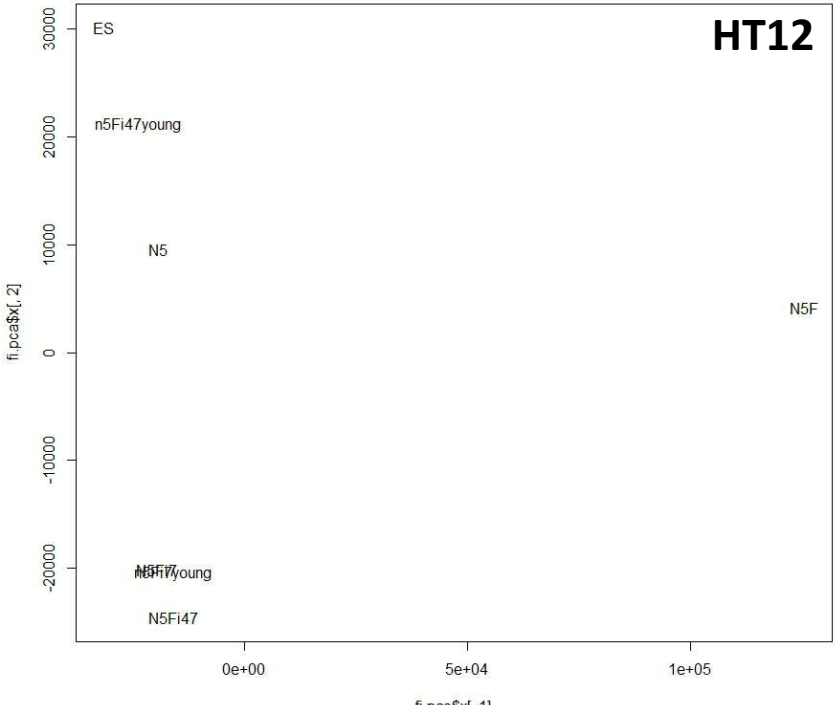
only in young_group	chip	stayed	max changed	old_group with higher impact
▼ somatic memory				
somatic memory	450k	40 %	54 %	intact
somatic memory	HT12	29 %	31%, 24%,16%	spec, intact, pluri
▼ specific				
specific	450k	43 %	32%, 21%	intact, pluri
specific	HT12	56 %	33 %	pluri
▼ pluri				
pluri	450k	86 %	12 %	intact
pluri	HT12	91 %	9 %	intact
▼ inact				
inact	450k	99 %	1 %	pluri
inact	HT12	72 %	27 %	pluri

HT12: iPS specific old = 25%sm+ 50%intact+ 12%pluri

450k: iPS specific old = 14%intact+ 14%pluri

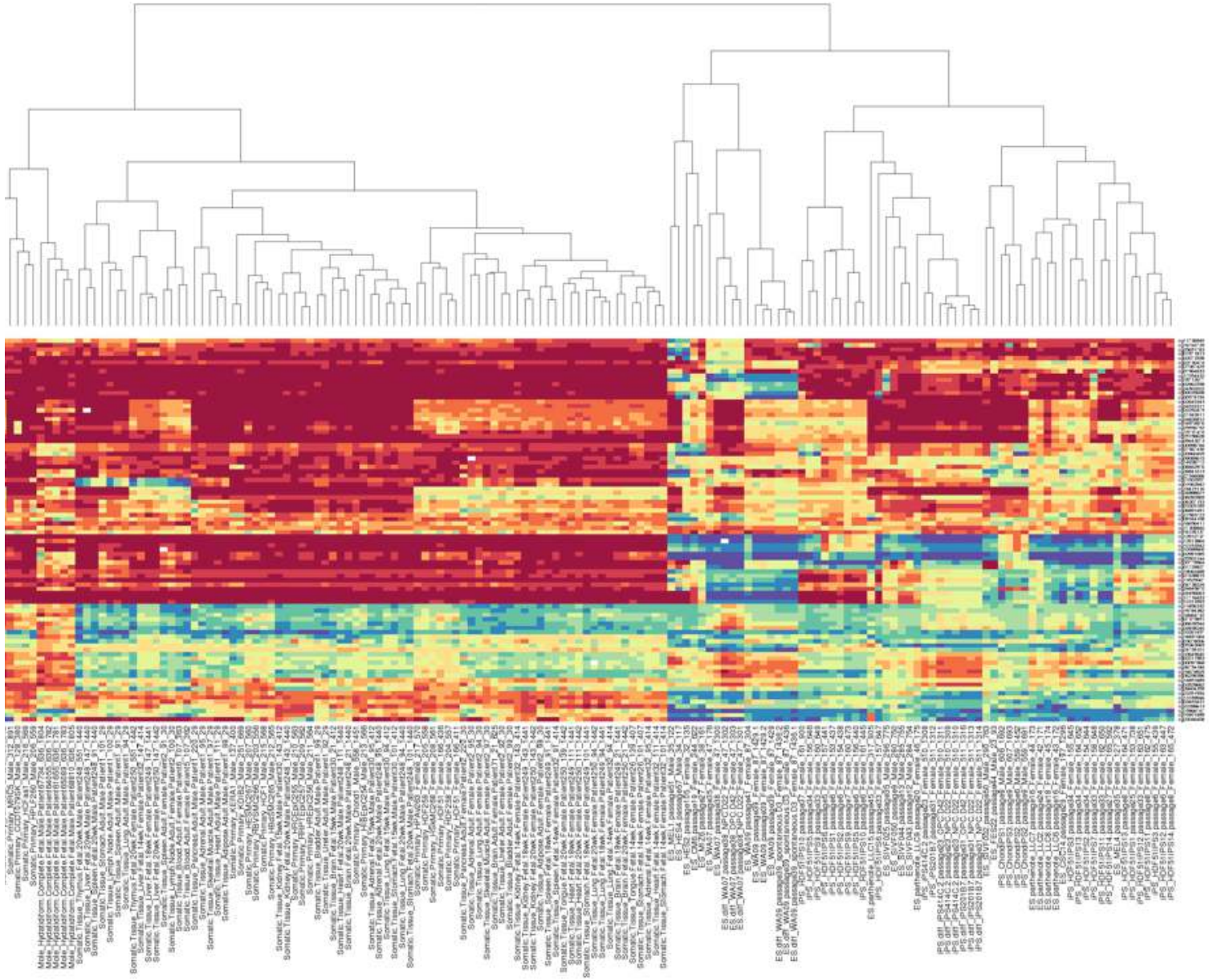

Fibroblasts young and old
TMSB4X
RPL30
PGAM1
PSMD1
SNHG9
MBTD1
CHERP
KLF13
LOC92659
HTT
LOC643531
CYR61
SS18L1
APIP
PTPLAD2
TIMP2
ORC6L
PGM2
RBPMS
SLIT2
log Fibro old delta
log Fibro young d
5
0
-5
2


450k
n5_3
n5F_1
n5F_2
n5Fi47_12
n5Fi47_25
n5Fi7_25
pca1$x[, 1]
pca1$x[, 2]
0
50
100
20
10
0
-10


HT12
ES
n5Fi47young
N5
N5F
N5Fi47
fi.pca$x[, 2]
30000
20000
10000
0
-10000
-20000
0e+00
5e+04
1e+05

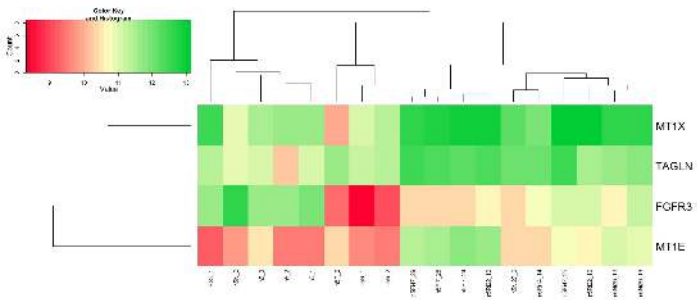
“markers” on Nazor’ data?

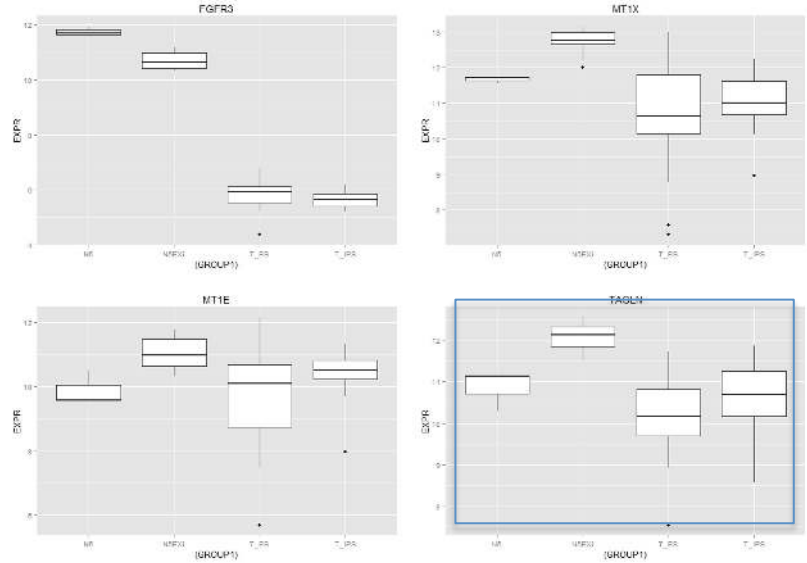
GSE31848



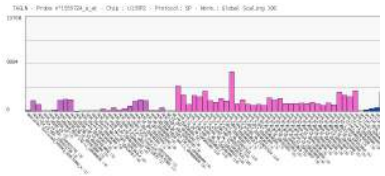
reprogramming-specifc locuses
4 genes - expression

new5 data vs GSE25970



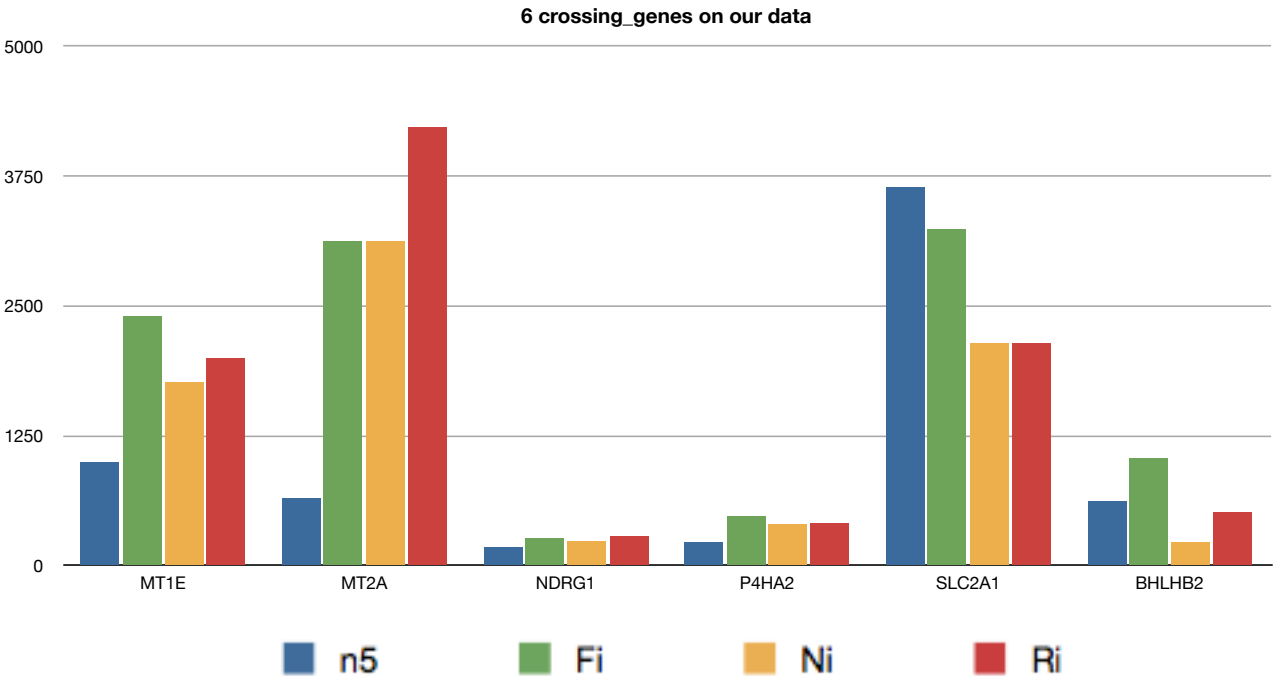


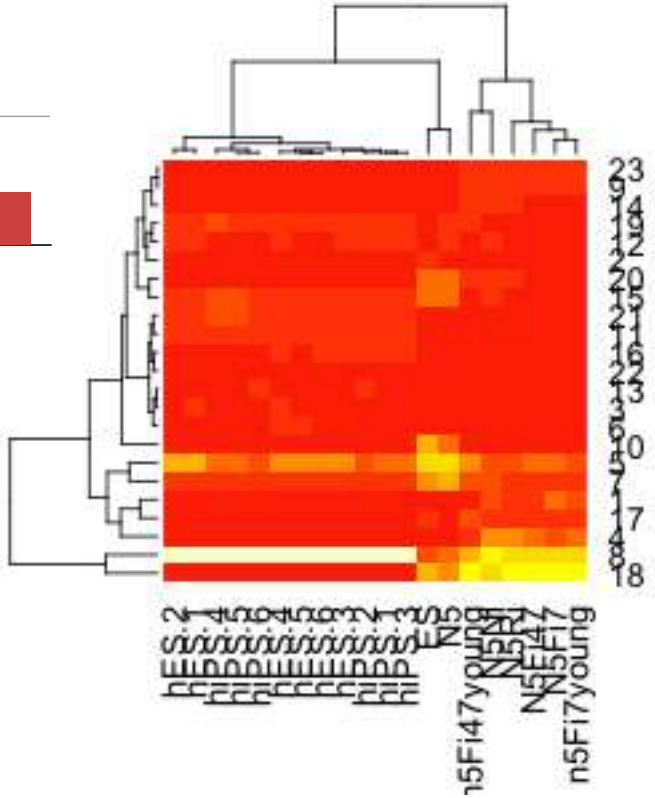
http://amazonia.transcriptome.eu/



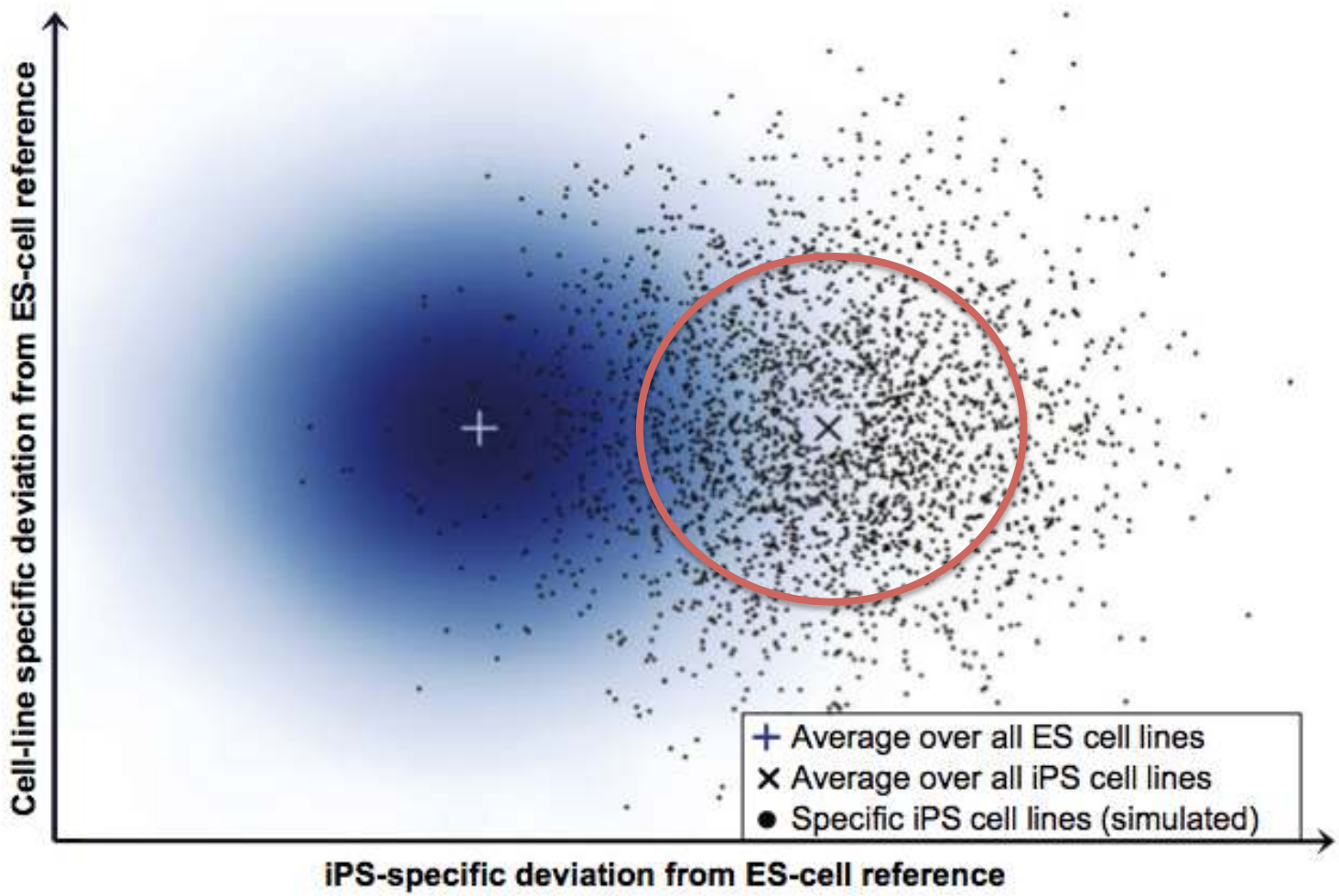
iPS Cell Classifier (Bock2011)	Accuracy	Sensitivity	Specificity	TN (ES=ES)	FN (iPS=ES)	FP (ES=iPS)	TP (iPS=iPS)
Chin2009 gene expression signature	63 %	0 %	100 %	20	12	0	0
Stadtfeld2010 MEG3	72 %	100 %	55 %	11	0	9	12
Shutova, unpublished, one-gene signature	69 %	90 %	33 %	18	8	2	4

“markers” on Hochedlinger’ data?





“markers” = stochastic + somatic memory



(Bock 2011)