

BIOINFORMATICS
INSTITUTE

Исследование транскриптома человека в норме и при патологии

Исполнитель: Нестеренко Максим
(магистр, 1 курс, биологический факультет СПбГУ)

Руководитель проекта: Андрей Сергеевич Готов
(Ресурсный центр «Центр Биобанк», СПбГУ)

Куратор проекта: Александр Владимирович Предеус
(Институт биоинформатики)

Патология: *гестационный сахарный диабет*

□ **Сахарный диабет беременных** — состояние, характеризующееся гипергликемией, возникающей на фоне беременности у некоторых женщин и обычно спонтанно исчезающее после родов ¹.



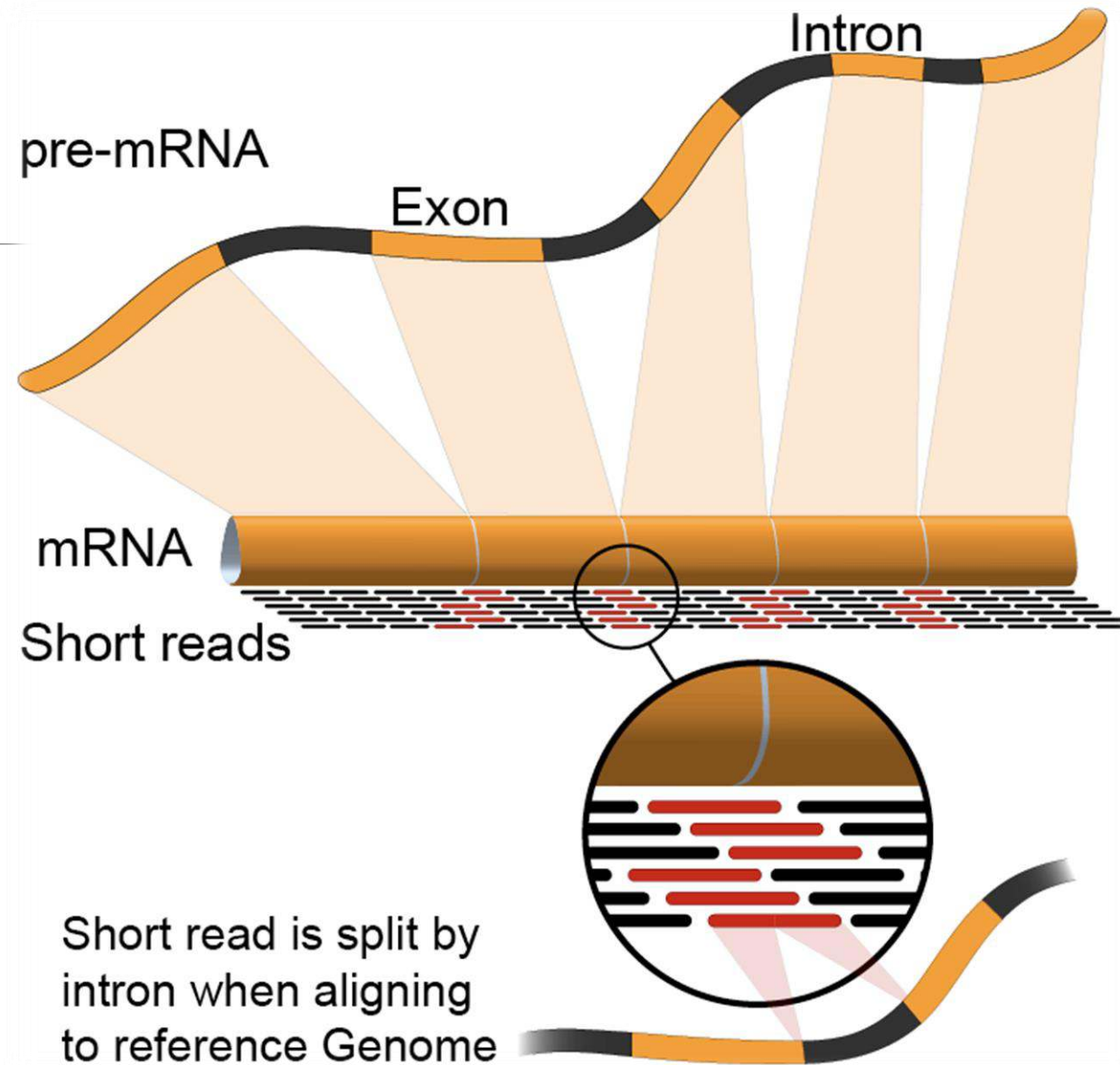
Цель и задачи

Цель проекта:

Сравнение профиля экспрессии генов при патологии и в норме

Задачи :

- Разобраться со стадиями контроля качества
- Проанализировать экспериментальные данные



FastQC: анализ качества, V010

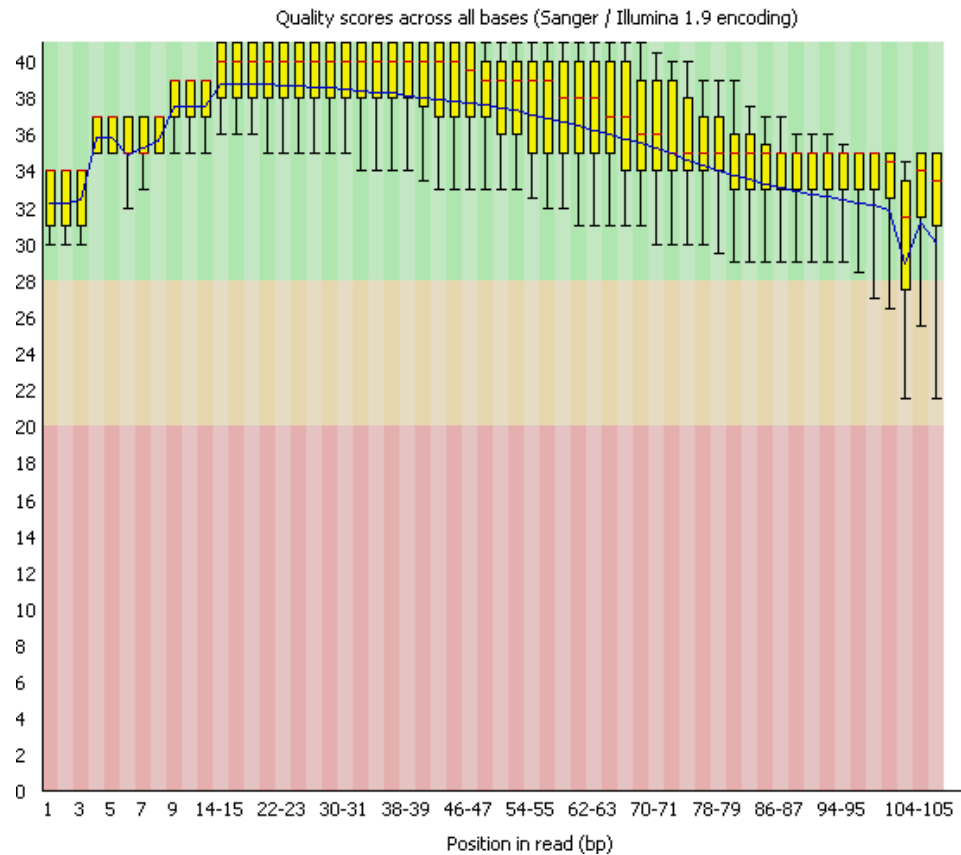


Рис.1. Качество определения нуклеотида на протяжении всей длины рида

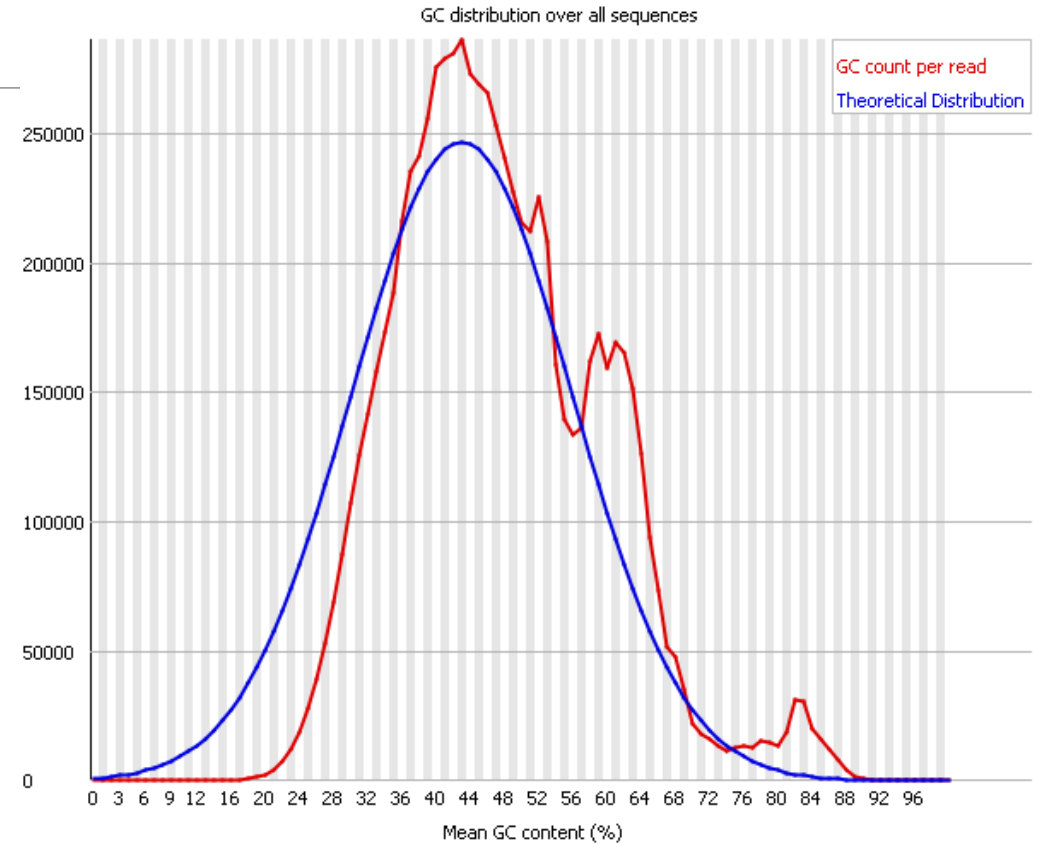


Рис.2. Распределение GC-состава в библиотеке

FastQC: анализ качества, B011

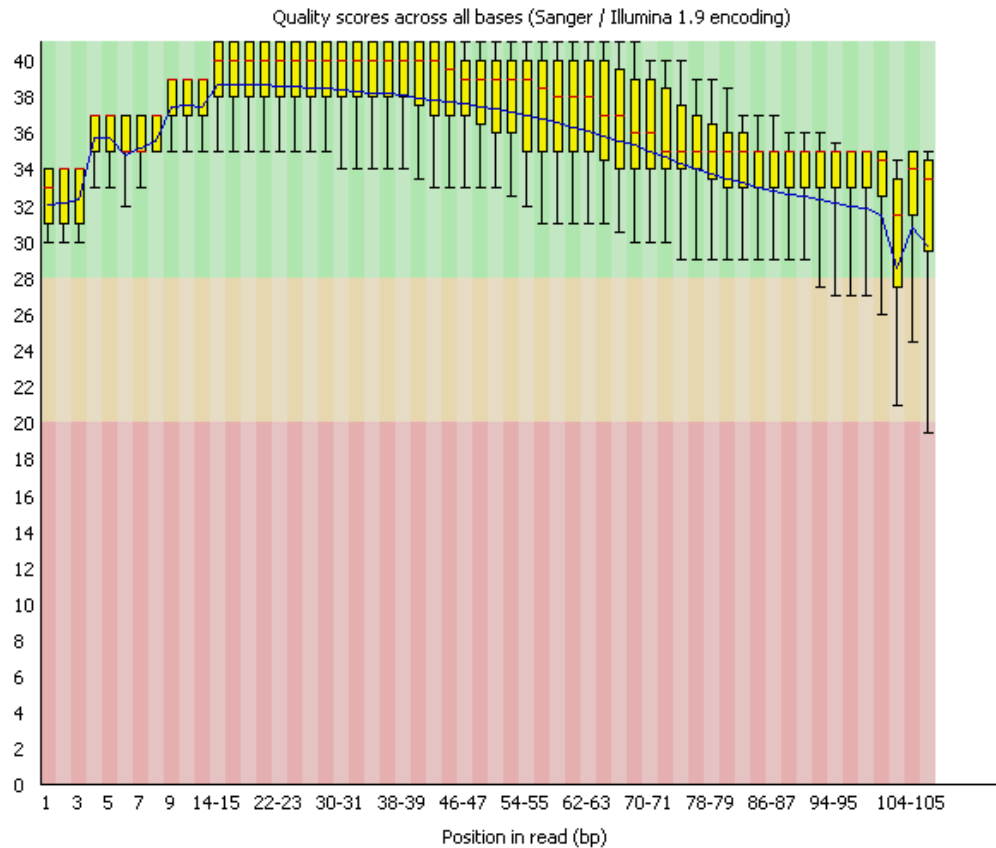


Рис.3. Качество определения нуклеотида на протяжении всей длины рида

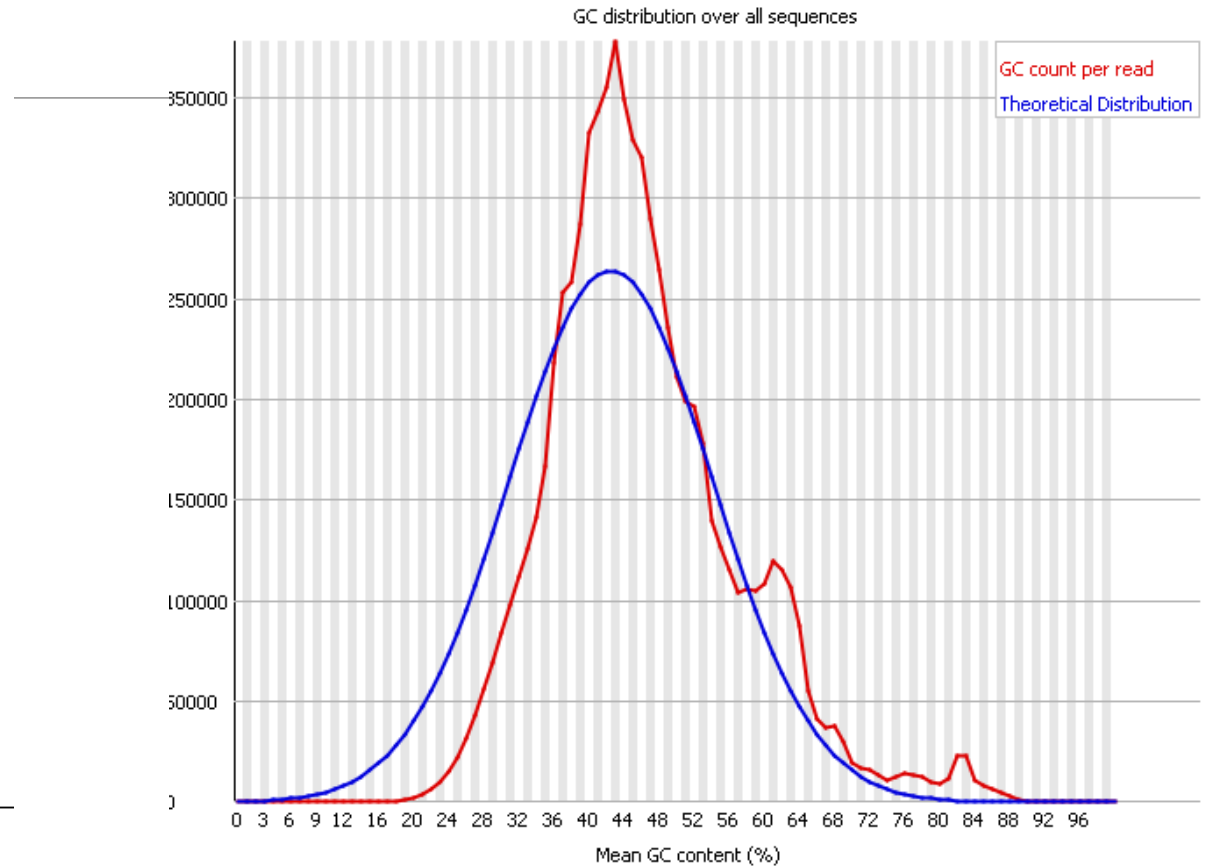


Рис.4. Распределение GC-состава в библиотеке

STAR: картирование библиотек на геном человека

Геном человека - **GRCh38.p7, primary assembly,**
версия аннотации – **25**

Образцы	Кол-во ридов	Общий % карт. ридов	% уник. карт.ридов	% некартир. ридов	% картир. на rRNA	% карт. на кодир.обл	% картир. на UTR	% картир. на интроны	% карт. на межген. обл
V010 (патология)	7976315	92,8	83,02	7,2	12,87	26,21	24,38	30,96	5,78
V011 (контроль)	7675812	88,81	81,83	11,19	29,99	17,67	17,71	29,71	5,43
V012	7277067	93,15	82,64	6,86	14,43	25,5	27,08	26,93	6,22

Kallisto: квантификация и нормализация данных

Фильтр:
в анализ – гены с не
менее 50
«нормализованными
единицами» в двух
образцах



Гены с «повышенным уровнем экспрессии» в образце В10 (патология)

Обозначение гена	Описание	B10_TPM	B11_TPM	B10/B11	Гены-маркеры ²
RPL9P9	ribosomal protein L9 pseudogene 9	287	2	96	
CSH1	chorionic somatomammotropin hormone 1 (placental lactogen)	67610	2469	27,37287449	
CCK	cholecystokinin	65	2	22	
CRH	corticotropin releasing hormone	270	13	19,35714286	*
CGB8	chorionic gonadotropin, beta polypeptide 8	57	2	19,33333333	
HTRA1	HtrA serine peptidase 1	397	21	18,09090909	*
TFPI2	tissue factor pathway inhibitor 2	4088	231	17,625	
LPL	lipoprotein lipase	86	4	17,4	
PSG9	pregnancy specific beta-1-glycoprotein 9	1491	88	16,76404494	
LGALS14	lectin, galactoside-binding, soluble, 14	279	16	16,47058824	
GPNMB	glycoprotein (transmembrane) nmb	147	10	13,45454545	
KISS1	KiSS-1 metastasis-suppressor	1850	142	12,94405594	
CGA	glycoprotein hormones, alpha polypeptide	2910	259	11,19615385	

* - потенциальные гены-маркеры патологий беременности

Гены-маркеры ГСД (литературные данные)

□ Дифференциальная экспрессия 66 генов у рожениц, страдающих гестационным сахарным диабетом.

TABLE 3
DAVID mapping of genes differentially expressed in GDM placenta^a

Gene list	Enrichment score	Cluster
CD63, GBA, HLA-DRA, IFI30, SMPD1	2.86	Lysosome/vacuole
INHA, INHBA, MIF, PLAB	2.02	Transforming growth factor- β and cytokine activity
INHA, INHBA, MIF, SRI	1.81	Regulation of cell activation, physiologic processes, and immune response
CEBPA, INHA, INHBA	1.75	Myeloid differentiation, hemopoiesis, and hemopoietic or lymphoid organ development
CALM1, CSEN, FLJ13612, SRI	1.68	Calcium binding EF-hand
C1QR1, INHA, INHBA, MIF, POLG2, PROCR	1.32	Immune cell activation and response to wounding, external stimulus, pathogen, or stress
ANXA4, CSEN, FOSL2, INHA, INHBA, MIF	1.09	Regulation of cell death
ADAM12, ADD2, ADFP, AGPAT4, AQP3, C1QR1, C2orf28, C22orf5, CALM1, CD63, CLDN7, COL17A1, FLJ22649, FLT1, FXYD5, GALNT2, GBA, HLA-DRA, HRASLS3, PROCR, RFNG, SC5DL, STEAP4, TMEM4, TUSC3	1.07	Transmembrane

^a Genbank accession numbers were mapped using functional annotation clustering in the DAVID 2007 pathway analysis tool. For each group, the processes or functions are tabulated with the gene list and enrichment score. Enrichment score is calculated as the geometric mean (in log scale) of members' *P* values in a corresponding annotation cluster. Clusters shown here are those with enrichment scores > 1.0.

Enquobahrie. Global placental gene expression in gestational diabetes mellitus. *Am J Obstet Gynecol* 2009.

Гены-маркеры ГСД в анализируемых данных

Обозначение гена	Описание	B10_TPM	B11_TPM	B10/B11
STEAP4	Six transmembrane protein of prostate 2	74	14	5
TUSC3	Tumor Suppressor Proteins	281	58	4,779661017
IFI30	Gamma-interferon-inducible lysosomal thiol reductase	102	27	3,678571429
CD63	Antigen CD63	246	97	2,520408163
PPIB	Cyclophilin B	177	84	2,094117647
ADAM12	ADAM (A Disintegrin and Metalloprotease) Proteins	1973	1016	1,94100295
FLT1	Fms-related tyrosine kinase 1	127	75	1,684210526
CLDN7	Claudin 7	32	19	1,65
SRI	Sorcin	36	22	1,608695652
UCK2	Uridine-cytidine kinase 2	89	62	1,428571429
ANXA4	Annexin A4	62	48	1,285714286
HDGF	Hepatoma-derived growth factor	40	31	1,28125
DYNLL1	Dynein, light chain, LC8-type 1	50	40	1,243902439
MIF	Macrophage migration inhibitory factor	92	90	1,021978022
CALM1	Calmodulin	181	205	0,883495146
NDRG1	N-myc downstream regulated gene 1	35	42	0,837209302
RPS29	Ribosomal protein S29	339	425	0,798122066
RPS7	Ribosomal protein S7	219	293	0,74829932

Голубым цветом обозначены гены, «уровень экспрессии» которых в образце с патологией (B10) понижен, по сравнению с контролем (B11)

В заключение

- ❑ Проведена оценка качества полученных библиотек
- ❑ Сравнение нормализованных данных позволило предположить наличие различий между образцами «патология» и «контроль»

