

# NGS w transkryptomice – Raport

A. Varanko, J. Czołowski, F. Hajdyła

## 10.11.22

- Przygotowanie materiału biologicznego
- Izolacja i oczyszczanie RNA
- Elektroforeza RNA
- Kontrola jakości RNA

Celem ćwiczenia było przeprowadzenia izolacji RNA z przygotowanego materiału biologicznego.

*Wszystkie czynności wykonano zgodnie z protokołem.*

Zmierzono OD<sub>600</sub> po około 2 godzinach od założenia hodowli logarytmicznej. Bakterie w fazie logarytmicznej są preferencyjnie wykorzystywane, ponieważ komórki w tej fazie najbardziej się dzielą, mają zwiększoną ekspresję różnych białek i najbardziej podatne na działanie różnych środków ( np. antybiotyków, które wpływają na syntezę białek itd). Hodowle logarytmiczne zaindukowano chlorem kadmu poprzez dodanie tego związku do hodowli. Do kontroli jakości RNA użyto zestawu RNA 6000 Pico kit, w tym przypadku próbki RNA rozcieńczono do 1 ng, do analizy użyto 2 µl próbki RNA. Zmierzone wartości RIN zamieszczono w kolejnym sprawozdaniu.

## Izolacja i oczyszczanie RNA

Na ćwiczeniach użyto szczepów *S. aureus*:

- RN4220 (WT) pCN51- $\emptyset$  – szczep dziki / wild type (AWT2)
- RN4220  $\Delta\Delta$ saoBC pCN51- $\emptyset$  – mutant I (BBC2)
- RN4220  $\Delta\Delta$ saoBC pCN51-saoC – mutant II (CBC2)

Na koniec tego etapu zmierzono stężenia RNA aparatem DenNovix (podczas pomiaru przechowywano RNA na lodzie).

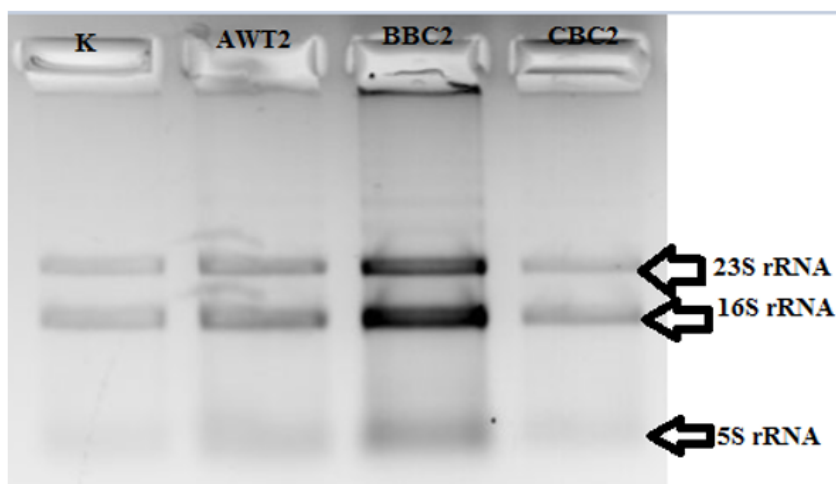
Uzyskane preparaty RNA cechują się stosunkowo wysoką czystością, bo stosunek A260/A280  $\approx$  2,0, czyli nie są zanieczyszczone białkiem, a otrzymany stosunek A260/A230 = 2,2 świadczy o braku kontaminacji jonami tiocyjanowymi, fenolanowymi oraz związkami organicznymi.

Tabela 1: Dane z NanoDrop 1000.

Szczep	Stężenie [ng/μl]	A260/A280	A260/A230
AWT2	86	2,08	2,22
BBC2	62	2,03	2,29
CBC2	12	2,05	2,27

## Elektroforetyczny rozdział RNA w warunkach denaturujących

Uzyskane na żelu prążki po rozdzieleniu RNA (Rys. 1) są zgodne z oczekiwaniami, porównując z markerem z pierwszej studzienki, natomiast w 3 kolumnie zauważono kontaminację genomowym DNA, co może wynikać z małej objętości zastosowanej DNAazy lub nieodpowiedniego czasu inkubacji po jej dodaniu.



Rysunek 1: Żel po elektroforezie.

### 17.11.22

- Kontrola ilości RNA
- Deplecja rRNA

*Ćwiczenie wykonano zgodnie z protokołem.*

Celem ćwiczenia jest zadbanie o wysoką jakość i ilość RNA w celu przygotowania bibliotek, ponieważ powodzenie eksperymentu zaczyna się od wysokiej jakości całkowitego RNA.

Do precyzyjnego pomiaru stężenia RNA zastosowano metodę fluorymetryczną, ponieważ fluorometr wykorzystuje barwniki specyficzne dla danych cząsteczek (można odróżnić RNA od DNA, ponieważ barwniki fluorescencyjne emitują światło tylko po związaniu z cząsteczkami docelowymi).

W wyniku tego, że rybosomalny rRNA stanowi 90 % prokariotycznego RNA, zastosowano deplecję rRNA. Ta metoda stosuje połączenia zmodyfikowanego oligonukleotydu komplementarnego do różnych regionów rRNA i kulek magnetycznych streptawidyny w celu wyeliminowania rRNA z próbek.

W pierwszej kolejności przygotowano próbki do pomiaru fluorymetrycznego za pomocą Qbit.

$$c_{\text{RNA}} = 824 \text{ ng}/\mu\text{l} \quad (1)$$

Na podstawie otrzymanego wyniku stężenia obliczono objętość jaką należało dodać do mieszaniny do hybrydyzacji aby uzyskać całkowitą masę RNA w próbce  $m = 2.5 \mu\text{g}$

$$V = \frac{2500 \text{ ng}}{824 \text{ ng}} \approx 3 \mu\text{l} \quad (2)$$

Ostateczny skład mieszaniny hybrydyzacyjnej:

- 50  $\mu\text{l}$  2x Hybridisation Buffer
- 3  $\mu\text{l}$  RNA
- 44  $\mu\text{l}$  H<sub>2</sub>O wolnej od nukleaz
- 3  $\mu\text{l}$  RiboMinus Pan-Prokaryote Probe Mix

Następnie dokonano deplecji RNA celem uzyskania frakcji mRNA. Przed i po deplecji zmierzono stężenie mRNA przy pomocy urządzeń Qbit i Nanodrop oraz wykonano elektroforezę kapilarną przy pomocy urządzenia Bioanalyzer. Wyniki przedstawiono poniżej.

Tabela 2: Wyniki dla RNA przed deplecją.

Próbka	RIN (RNA Integrity Number)
CBC2	7.9
BBC2	9.0
AWT2	8.6

Tabela 3: Wyniki dla RNA po deplecji (oczyszczonego mRNA).

Próbka	$c_{\text{RNA}}$ [pg/ $\mu\text{l}$ ]	względne zanieczyszczenie próbki [%]
CBC2	465	0,0
BBC2	922	1,2
AWT2	716	0,6

## 24.11.22

- Przygotowanie bibliotek cDNA z wcześniej otrzymanego mRNA:

- precyzyjny pomiar stężenia RNA (fluorymetria, Qubit) i wyrównanie stężenia RNA we wszystkich próbkach
- synteza 1, 2 nici cDNA
- Oczyszczenie cDNA z użyciem kulek magnetycznych

Celem ćwiczenia było przygotowanie bibliotek cDNA z wcześniej otrzymanego mRNA. Przeprowadzono odwrotną transkrypcję, gdzie dodanie aktynomycyny D do mieszaniny zapobiega fałszywej, zależnej od DNA syntezie, jednocześnie umożliwiając syntezę zależną od RNA i poprawiając specyficzność nici. Następnie nić jest usuwana i syntetyzowana jest nić zastępcza; włączenie dUTP wygasza nam drugą nić podczas amplifikacji; następnie kulki magnetyczne oddzielają ds cDNA od mieszaniny reakcyjnej. Rezultatem jest cDNA z tępymi końcami.

*Wszystkie czynności wykonano zgodnie z protokołem.*

Na początku wykonano precyzyjny fluorymetryczny pomiar stężenia mRNA za pomocą Qubit. Stężenie mRNA zmierzone na poprzednich ćwiczeniach wynosiło 12,7 ng/μl. Stężenie na początku tych ćwiczeń, wyniosło 8,74 ng/μl. Widać więc degradację ok. 30% mRNA na przestrzeni tygodnia. Wyrównano stężenia mRNA do 30 ng w 5 μl wody w próbkę do PRC, na podstawie poniższych obliczeń.

$$V = \frac{30 \text{ ng} \times 1 \text{ } \mu\text{l}}{8,74 \text{ ng}} \quad (3)$$

$V = 3,43 \text{ } \mu\text{l}$  próbki mRNA i dopełnić do 5 μl (czyli dodać 1,57 μl) wodą. Następnie wszystkie procedury, do końca, wykonywano zgodnie z instrukcją. Próbki zostały zamrożone w -20°C do następnych zajęć.

### 01.12.22

- Adenylacja 3' końców
- Ligacja adapterów
- Oczyszczenie cDNA z użyciem kulek magnetycznych

*Wszystkie czynności wykonano zgodnie z protokołem.*

Jeden nukleotyd A jest dodawany do końców 3' fragmentu aby zapobiec ich wzajemnej ligacji podczas ligacji adaptera. Jeden nukleotyd na końcu 3' będzie zapewniał nam komplementarną końcówkę do ligacji adaptera z fragmentem (to zapewnia niski współczynnik tworzenia się chimer). Następnie zachodzi ligacja wielu adapterów z końcami fragmentów ds cDNA co przygotowuje je do hybrydyzacji w komorze przepływowej (flowcell). Ten krok umożliwia regulowanie łączenia wielu bibliotek cDNA w końcowym etapie. Konieczne jest zatem sprawdzenie kompatybilności indeksów aby zsekwencjonować je w jednej reakcji.

## 08.12.22

- Wzbogacanie fragmentów cDNA (amplifikacja PCR)
- Oczyszczenie cDNA z użyciem kulek magnetycznych
- Sprawdzenie stężenia bibliotek cDNA oraz ocena jakości z zastosowaniem elektroforezy kapilarnej
- Rozcieńczenie i denaturacja bibliotek cDNA na potrzeby reakcji sekwencjonowania

*Wszystkie czynności wykonano zgodnie z protokołem.*

Podczas wzbogacenia fragmentów DNA w celu przygotowania bibliotek starano się zminimalizować liczbę cykli PCR aby uniknąć zniekształcenia reprezentacji bibliotek finalnej próbki. Na koniec amplifikacji PCR uzyskano biblioteki LS z finalnym indeksem.

Kulki Ampure XP stosowano do oddzielenia ds cDNA od mieszaniny reakcyjnej drugiej nici, nadmiaru soli, enzymów, starterów i nukleotydów. Ten system wykorzystuje technologię odwracalnej immobilizacji z kulkami paramagnetycznymi do przepustowego oczyszczania DNA. W celu osiągnięcia najwyższej jakości na platformach Illumina stworzono optymalną gęstość na klastrach flowcell, dlatego zmierzono stężenia DNA utworzonych bibliotek i oceniono jakość za pomocą elektroforezy kapilarnej. Zindeksowane biblioteki musimy znormalizować i połączyć w równych objętościach. Należy zrobić pomiary na świeżo i nie przechowywać próbki w postaci puli.

## Analiza danych

Obliczenia prowadzono na serwerze `mol058.mol.uj.edu.pl` z wykorzystaniem pakietów dostępnych w środowisku `ngs`, zapewnionym przez prowadzącego w postaci pliku `yml`.

### Analiza liczby odczytów

W celu wyznaczenia liczby odczytów każdej z replik zsumowano (`DataFrame.sum()`) tabelę `NumReads` wygenerowaną przez program `salmon`. Wyniki przedstawiono w tab. 4.

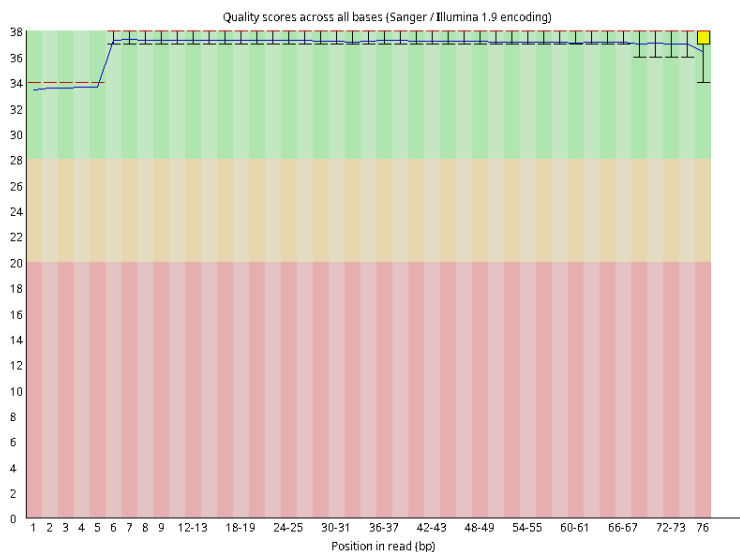
Tabela 4: Porównanie liczby odczytów w każdej z replik.

Nazwa repliki	liczba odczytów
wt51e_1	1707680
wt51e_2	1464820
wt51e_3	1361432
mt51e_1	1772327
mt51e_2	2149890
mt51e_3	1327580

## Analiza jakości odczytów

Analizę jakości odczytów wykonano przy pomocy **FastQC**. Poniżej przedstawiono wyniki analizy dla `rn_mt51e_lg_3_R2`.

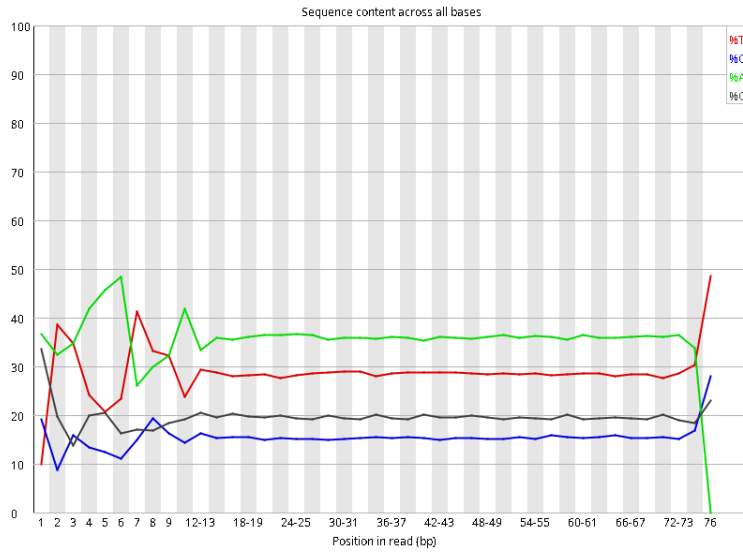
W przypadku uśrednionego Q score analiza nie wykazała istotnych spadków jakości dla żadnej z replik. Jest to spodziewany wynik, gdyż platforma Illumina, z której korzystano generuje odczyty o niskiej długości. Degradacja może występować głównie w przypadku długich odczytów co wynika z samej istoty procesu sekwencjonowania. Na rys. 2 widać, że mediana Q score dla żadnej z pozycji nie spada poniżej 34.



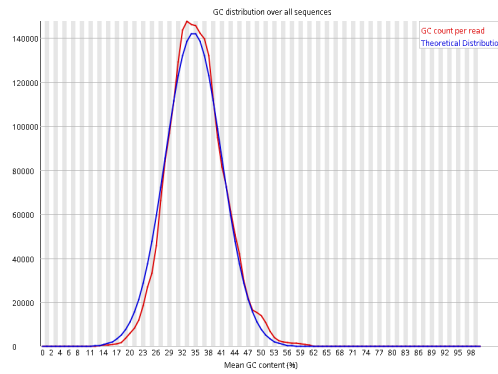
Rysunek 2: Uśredniony Q score względem pozycji dla wszystkich odczytów.

Analiza częstości występowania poszczególnych zasad na kolejnych pozycjach odczytów (Rys. 3) wykazała wysokie odstępstwo od „normalnej” biblioteki dla każdej z badanych replik. Jest to jednak spodziewany wynik dla bibliotek typu RNA-Seq. Biblioteki powstałe w wyniku primingu przy użyciu losowych heksamerów (w tym prawie wszystkie biblioteki RNA-Seq) oraz te, które zostały pofragmentowane przy użyciu transpozaz, mają skłonność do selekcji wybranych zasad w pozycjach, od których rozpoczynają się odczyty. Nie jest to coś, co można skorygować przez przycięcie i w większości przypadków nie wydaje się mieć negatywnego wpływu na dalszą analizę.

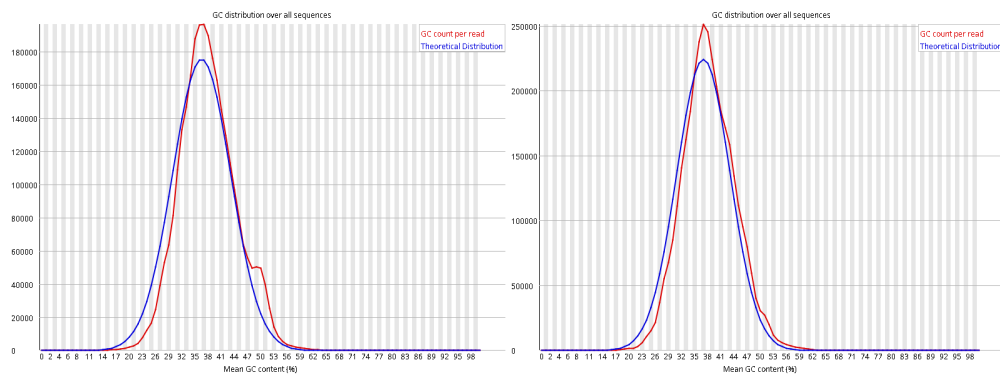
W przypadku analizy rozkładu częstości par GC w odczytach większość replik nie wykazała żadnego odstępstwa od rozkładu normalnego (Rys 4a). Świadczy to o prawdziwej losowości biblioteki. W przypadku replik `rn_mt51e_lg_1_R2` oraz `rn_mt51e_lg_2_R1` analiza wykazała jednak nieznaczące odstępstwo rozkładu od normalności (odpowiednio rys. 4b i 4c). Ostrzeżenia w tym module analizy zwykle wskazują na problem z biblioteką. Ostre piki są zwykle wynikiem specyficznego zanieczyszczenia (Rys. 4c), które może być wychwycone dalej przez analizę nadreprezentowanych sekwencji. Na rys. 4b widoczne jest nieznaczące poszerzenie krzywej rozkładu, co może wskazywać na zanieczyszczenie sekwencjami pochodzącymi z innego gatunku.



Rysunek 3: Uśredniona częstość występowania poszczególnych zasad na kolejnych pozycjach odczytów.



(a)

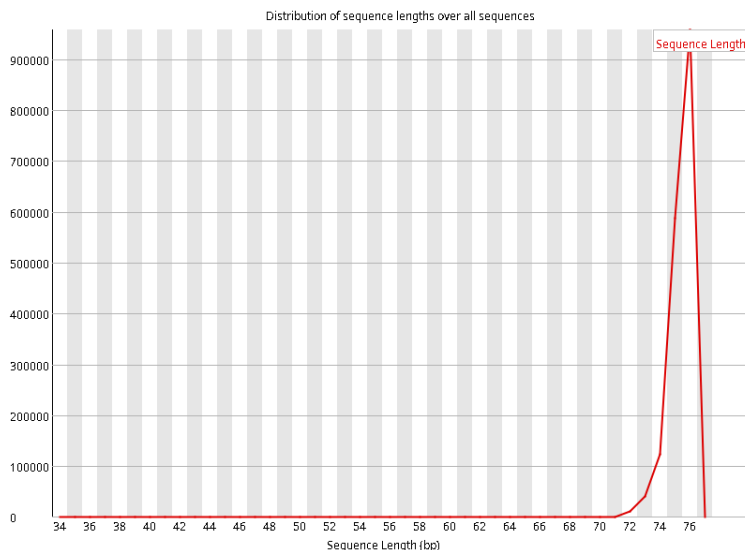


(b)

(c)

Rysunek 4: Rozkład częstości par GC w odczytach dla replik (a) rn\_mt51e\_lg\_3\_R2 (b) rn\_mt51e\_lg\_1\_R2 (c) rn\_mt51e\_lg\_2\_R1.

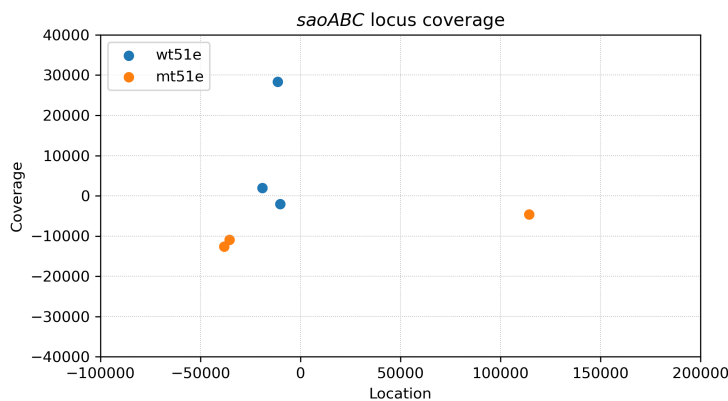
W przypadku analizy rozkładu długości odczytów (Rys. 5) odstępstwa od normalności wystąpiły w każdej z replik, jest to jednak związane z platformą jaka została użyta do sekwencjonowania. W naszym przypadku platforma Illumina produkuje odczyty o długości ok. 75 pz, co jest całkowicie normalne i powtarza się dla każdej z replik. Ostrzeżenia występujące w tym module analizy można zignorować.



Rysunek 5: Rozkład długości odczytów.

## Analiza rozrzutu replik

Na rys. 6 przedstawiono wynik analizy rozrzutu replik po zredukowaniu liczby wymiarów metodą PCA. Analizy dokonano w oparciu o ich wartości TPM (transcripts per million). Widoczna jest wyraźna separacja pomiędzy punktami pochodzącymi z próbki wt51e a tymi z mt51e. W przypadku próbki mt51e jeden punkt istotnie różni się od pozostałych.



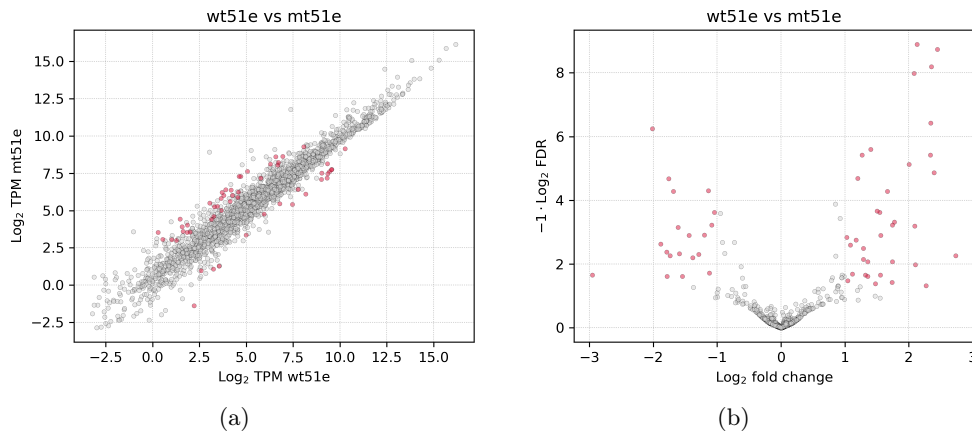
Rysunek 6: Analiza rozrzutu replik metodą PCA w oparciu o wartości TPM.



## Analiza DGE

Na rys. 7a cały zakres wartości TPM układa się dobrze na dopasowanej prostej, co wskazuje na brak lub niewielkie różnice w ekspresji genów między wt51e a mt51e. Kolorem czerwonym zaznaczono punkty istotnie różniące się między tymi próbkami ( $|\log_2 FC| > 1 \wedge p < 0.05$ ). To porównanie jest generalnie wizualizowane za pomocą wykresów punktowych, gdzie każdy punkt reprezentuje pojedynczy gen, a jego umiejscowienie wskazuje na jego średni poziom ekspresji odpowiednio u wt51e oraz mt51e. Ponieważ osie reprezentują poziomy ekspresji dla odpowiedniej próbki, punkty danych układające się wzdłuż przekątnej wskazują na podobne poziomy ekspresji w obu próbkach. Punkty danych powyżej lub poniżej przekątnej oznaczałyby odpowiednio wyższe lub niższe poziomy ekspresji w wt51e względem mt51e.

Wykres zależności logarytmicznej krotności zmiany od skorygowanej wartości  $p$  (Rys. 7b), nazywany również „volcano plot”, ze względu na podobieństwo do wybuchającego wulkanu, pokazuje statystyczną istotność różnicy w stosunku do wielkości różnicy dla każdego pojedynczego genu. Im wyżej na osi  $Y$  znajduje się punkt danych, tym mniejsza jest jego skorygowana wartość  $p$  (punkty istotnie różne między próbkami zaznaczono kolorem czerwonym). Oś  $X$  pokazuje różnice w wartościach krotności zmiany ekspresji. Punkty bliższe 0 reprezentują geny, które mają podobne poziomy ekspresji. Punkty rozrzucone daleko od 0 reprezentują geny, które różnicują między próbkami pod względem ekspresji. Na tym wykresie bardziej wyraźnie widać, że między próbkami występują różnice w ekspresji niektórych genów. Niektóre z nich wykazują wyższą ekspresję w mt51e względem wt51e ( $\log_2 FC > 1$ ), a inne niższą ( $\log_2 FC < -1$ ).



Rysunek 7: Wyniki analizy DGE. (a) wykres punktowy TPM vs TPM (b) volcano plot.

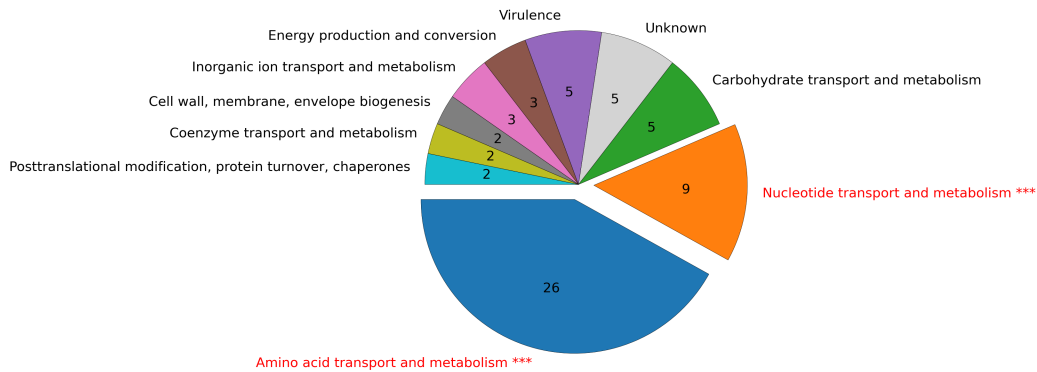
Poniżej zamieszczono tabelę zawierającą informacje o wszystkich różnicujących odczytach. Można z niej m.in. odczytać, że 35 odczytów to odczyty tzw. up-regulated, a więc takie, których ekspresja jest większa w mt51e względem wt51e oraz 19 odczytów to odczyty down-regulated, czyli takie, których ekspresja spada w mt51e względem wt51e. W tabeli znajdują się także informacje odnośnie średnich wartości TPM, wartości FDR (padj), typie sekwencji (type), produkcie (product) oraz ontologii (onts), której analiza zostanie przybliżona w kolejnym podrozdziale.

locus_tag	log2 Fold Change	padj	TPM	w51c	TPM	m51c	mean	type	product	onts
SAO1HSC_03010	2.790244119414	0.0054619536387036	1.224018	1.224018	11.048	3333.33333333334	CDS	imidazole glycerol phosphate synthase subunit HisH	Amino acid transport and metabolism	
SAO1HSC_03011	2.145719353333333	1.8311034321406-09	26.054133333333333	157.31739919999998	156.0031	CDS	flavinooxygenase	Vitamins		
SAO1HSC_02706	2.35058397150638	1.3671370000000001	156.0031	156.0031	156.0031	CDS	immunoglobulin G-binding protein 5b	Energy production and catabolism; Amino acid transport and metabolism		
SAO1HSC_02707	2.35058397150638	1.3671370000000001	156.0031	156.0031	156.0031	CDS	immunoglobulin G-binding protein 5b	Amino acid transport and metabolism		
SAO1HSC_02288	2.3431716081201	3.774841547601e-07	13.063401333333333	85.04546666666666	71.49286666666666	CDS	3-isoisopropylmalate dehydratase	Amino acid transport and metabolism		
SAO1HSC_00011	2.33568120917007	3.782946823285858e-06	33.527133333333334	199.729	199.729	CDS	hypodermal protein	Amino acid transport and metabolism		
SAO1HSC_02287	2.127070281109	1.72225530197096-09	13.819366666666667	21.588800000000003	21.588800000000003	CDS	3-isoisopropylmalate dehydratase, large subunit	Amino acid transport and metabolism		
SAO1HSC_00012	2.101703033333333	0.0104151028260013	1.4813316666666667	8.23785	8.23785	CDS	histidinol-phosphate aminotransferase	Amino acid transport and metabolism		
SAO1HSC_02285	2.095035710492	7.22558641000333e-06	12.629033333333334	56.7954	56.7954	CDS	2-isoisopropylmalate synthase	Amino acid transport and metabolism		
SAO1HSC_02284	2.081338748213	1.04030740833708-08	18.210800000000003	83.637333333333333	83.637333333333333	CDS	ketoacid reductoisomerase	Amino acid transport and metabolism		
SAO1HSC_03020	1.7753805160245	0.0084828707245154	95.293766666666666	380.238833333333333	380.238833333333333	CDS	homoserine dehydrogenase	Amino acid transport and metabolism		
SAO1HSC_03008	1.717503301848	0.0083335541681373	3.61100333333333333	16.225766666666666	16.225766666666666	CDS	imidazole glycerol phosphate synthase subunit HisF	Amino acid transport and metabolism		
SAO1HSC_03009	1.782005160928	0.03207472270933	2.6741166666666667	15.222176666666666	15.222176666666666	CDS	lysine dehydratase	Amino acid transport and metabolism		
SAO1HSC_03038	1.6620108412101	5.19347968646702e-05	1.0170366666666669	38.572333333333336	38.572333333333336	CDS	5-methyltetrahydropterin glutamate:homocysteine S-methyltransferase, Methionine synthase II (cobalamin-independent)	Amino acid transport and metabolism		
SAO1HSC_02844	1.55701807919369	0.022513231942816	77.8316	277.940333333333334	277.940333333333334	CDS	ABC-type metal ion transport system, periplasmic component/surface antigen	Amino acid transport and metabolism		
SAO1HSC_02717	1.5490792745148	0.002261662365173	19.21303333333333333	61.977066666666666	61.977066666666666	CDS	hypodermal protein	Vitamins		
SAO1HSC_02718	1.5490792745148	0.002261662365173	19.21303333333333333	61.977066666666666	61.977066666666666	CDS	hypodermal protein	Vitamins		
SAO1HSC_02240	1.4774540413178	0.041399658815205	3.0784866666666667	12.173129666666668	12.173129666666668	CDS	beta-chain forming cytochrome, aerotaxis/lambda-lysoin family protein	Amino acid transport and metabolism		
SAO1HSC_03031	1.4020930928615	2.50701453091052e-06	104.385533333333334	299.104	299.104	CDS	threonine synthase	Amino acid transport and metabolism		
SAO1HSC_02902	1.35782470989896	0.0083335541681373	3.79143333333333333	8.015716666666668	8.015716666666668	CDS	histidinol dehydrogenase	Amino acid transport and metabolism		
SAO1HSC_00026	1.31883108734188	0.021186827300157	12.614266666666668	38.61043333333333333	38.61043333333333333	CDS	ABC-type metal ion transport system, periplasmic component/surface antigen	Vitamins		
SAO1HSC_00435	1.289681729219719	0.007158382012105	4.0112066666666666	11.715886666666667	11.715886666666667	CDS	glutamate synthase, large subunit	Inorganic ion transport and metabolism		
SAO1HSC_00843	1.26728551678285	3.782946823285858e-06	54.523166666666667	144.05000000000002	144.05000000000002	CDS	ABC-type methionine transport system, periplasmic component	Amino acid transport and metabolism		
SAO1HSC_00316	1.200391659159616	2.083907089920298e-05	23.379566666666665	58.957066666666666	58.957066666666666	CDS	peptidoglycan N-acetylglucosylase, PgdA, CDAI family	Carbohydrate transport and metabolism		
SAO1HSC_00339	1.1754265010454	0.0017967192428538	9.967395	24.149200000000003	24.149200000000003	CDS	bifunctional homocysteine S-methyltransferase/5,10-methylenetetrahydrofolate reductase, Methionine synthase I (cobalamin-dependent)	Amino acid transport and metabolism		
SAO1HSC_00436	1.1882923088723	0.0029317022826705	102.098833333333336	21.289033333333333	21.289033333333333	CDS	glutamate synthase, small subunit, subunit beta	Amino acid transport and metabolism		
SAO1HSC_03019	1.0420568702844	0.0310210911402898	12.705833333333337	32.2293	32.2293	CDS	separate class	Amino acid transport and metabolism		
SAO1HSC_00309	1.02842110089747	0.0014588806963302	267.230333333333336	619.7306666666667	619.7306666666667	CDS	glutamine synthetase, type I	Amino acid transport and metabolism		
SAO1HSC_02857	1.0520260384892	0.00001571701086	62.110033333333334	566.397	566.397	CDS	adenosuccinate synthetase	Nucleotide transport and metabolism		
SAO1HSC_00019	1.0520260384892	0.00001571701086	62.110033333333334	566.397	566.397	CDS	adenosuccinate synthetase	Nucleotide transport and metabolism		
SAO1HSC_00145	1.1370167576766	4.97513308917954e-05	641.22889999999999	85.36286666666666	85.36286666666666	CDS	phosphoribosyl transferase system IIC components, glucose	Nucleotide transport and metabolism		
SAO1HSC_01818	1.119790065160283	0.0012211502634293	110.04666666666668	441.242333333333334	441.242333333333334	CDS	aldehyde dehydrogenase	Amino acid transport and metabolism		
SAO1HSC_01910	1.38410710388773	0.00252529138218	520.70966666666666	184.710033333333333	184.710033333333333	CDS	biomolecular pyrimidine regulatory protein PyrB, uracil phosphoribosyltransferase	Nucleotide transport and metabolism		
SAO1HSC_01696	1.4364017175751	0.0024221145442908	31.4671	15.657203333333333	15.657203333333333	CDS	nitrate reductase, alpha subunit	Energy production and catabolism		
SAO1HSC_01168	1.58882900680714	0.004880114918827	693.6046666666667	136.641166666666666	136.641166666666666	CDS	phosphotransferase, catalytic chain	Nucleotide transport and metabolism		
SAO1HSC_01170	1.616827540014	0.00077995187817	7538.14	218.0800000000004	218.0800000000004	CDS	carbamoyl-phosphate synthase, large subunit	Amino acid transport and metabolism		
SAO1HSC_00188	1.7920100096019	0.0045015026307036	176.285333333333335	214.388333333333333	214.388333333333333	CDS	ornithine 5-phosphate decarboxylase	Amino acid transport and metabolism		
SAO1HSC_00176	1.7828694420404	2.0024022145442908	11.637066666666668	42.920666666666666	42.920666666666666	CDS	pyruvate-formate lyase-activating enzyme	Posttranslational modification, protein turnover, chaperones		
SAO1HSC_01055	1.917978787857916	0.01458288390794	288.9016666666667	2.4249076666666665	2.4249076666666665	CDS	malate-binding periplasmic protein MalE	Carbohydrate transport and metabolism		
SAO1HSC_02413	1.8821991699759	0.0022920818018563	9.43036333333333337	2.10898333333333337	2.10898333333333337	CDS	tagatose 6-phosphate kinase	Nucleotide transport and metabolism		
SAO1HSC_01169	2.105208774273	5.721923920240357e-07	635.629	15.225766666666666	15.225766666666666	CDS	carbamoyl-phosphate synthase, small subunit	Amino acid transport and metabolism		
SAO1HSC_00295	2.195880830835838	0.021871826728504	4.6408733333333333	0.883197	0.883197	CDS	dihydrodipicolinate synthase/N-acetylglucosaminase lyase	Amino acid transport and metabolism; Cell wall, membrane, envelope biogenesis		

Tabela 5: Informacje o wszystkich różnicujących wynikach.

## Analiza Ontologii

Ontologia odczytów (genów) mówi o tym w jaki sposób dany odczyt jest połączony w rzeczywistości z otaczającym go światem oraz w jaki sposób funkcjonuje. Mówi nam ona więc o funkcji danych odczytów. Przypisano ontologie do locus poprzez dopasowanie danych znajdujących się w tabeli `rn_meta.tsv`. Poniższy wykres kołowy ilustruje jakie funkcje pełnią geny różnicujące między próbkami pod względem ekspresji. Widać, że największa grupa genów różnicujących związana jest z transportem oraz metabolizmem aminokwasów. Wyróżnia się także grupa związana z transportem i metabolizmem nukleotydów. Widać, że pozostałe geny różnicujące mają różne funkcje w komórce *S. aureus*: wirulencja, biogeneza ściany komórkowej, błony oraz otoczki, transport i metabolizm koenzymów itd. Spośród wszystkich genów różnicujących funkcje 5 z nich pozostają nieznane.



Rysunek 8: Wykres kołowy przedstawiający wyniki analizy ontologii.