

納品レポートの見方

RNA-Seq解析



CAUTION

納品物につきまして

弊社pdf形式のレポートファイルの、Data Download Informationの頁にRaw dataおよび解析結果のダウンロードリンクがあります。

Raw dataはデータベース登録時や、弊社でのデータ保管期間経過後に、追加データ解析をご希望の場合に必要なファイルとなります。

ダウンロードできる期間は約2週間となっておりますので、**必ず**全てのファイルをダウンロード頂きますよう、宜しくお願い致します。

なお、2週間経過後も3カ月はデータを保管しておりますので、再度ダウンロードが必要な場合はngs@macrogen-japan.co.jpまでご連絡ください。

CAUTION

md5sum値の確認

ダウンロードしたFastq.gzファイルもしくはHDD内のFastq.gzファイルは、ファイルの解凍前に”md5sum値”のご確認ください。

”QuickHash-GUI”という、フリーのアプリケーションもございます。

<https://www.quickhash-gui.org/downloads/>

もしお手持ちのソフトウェアが無ければこちらをご取得下さい。

QuickHash-GUIでのmd5sum値の確認方法は下記となります。
 “QuickHash-GUI.exe”アプリケーションを起動します(例1)。

①“FileS”タブをクリックし、

②“Algorithm”、“MD5”を選択してください。

md5sum値を確認したいfastqファイルが入っているフォルダを

③“Select Directory” から選択していただきますと、自動的に解析が進行致します。

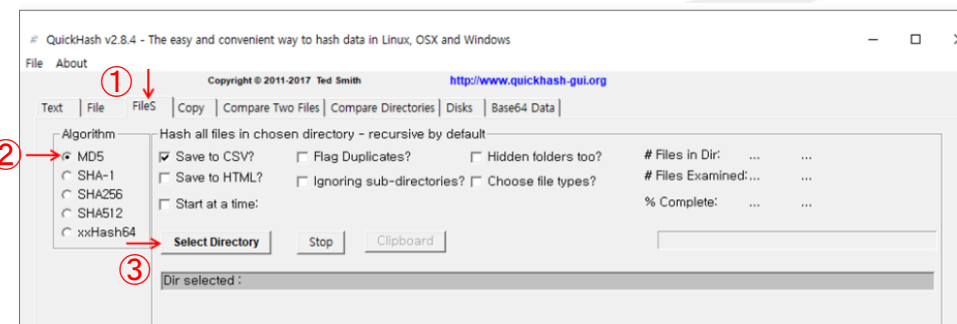
出力は、csvファイルまたはtxtファイルとして保存することができます。

使用するシステムの性能により、処理に時間がかかる場合がございます。

表示された数字と、レポートに記載のmd5sum値の一致を確認できましたら作業完了となります。

あわせてQuickHash-GUIのユーザーマニュアルもご確認ください(例2)。

例1 QuickHash-GUI 起動画面



例2 QuickHash-GUI file 中身

Quickhash-GUI.exe	2,090,414	6,505,472
sqlite3-win32.dll	429,646	852,754
sqlite3-win64.dll	717,149	1,742,848
UserManual.pdf	512,697	576,987

Contents

本”納品レポートの見方”では納品しておりますレポートファイルに沿って解説しております。
レポートファイルと併せてご確認ください。

1. Experimental Methods and Workflow
2. Analysis Methods and Workflow
3. Summary of Data Production
4. Reference Mapping and Assembly Results
5. Differentially Expressed Gene Analysis Results
※5.5. KEGG Enrichment Analysisは解析希望時のみ、有償オプション。
6. SNP and Indel Discovery
※解析希望時のみ、有償オプション
7. Fusion Gene Prediction Results
※解析希望時のみ、有償オプション
8. Data Download Information
9. Appendix

※本”納品レポートの見方”中のデータは一般的な納品例となっております。
お手元のレポートと一致しない場合もございますので、ご了承ください。

1. Experimental/ 2. Analysis Methods and Workflow

こちらの項目では、ライブラリ作製からシーケンス解析までを”Experimental”、得られたリードデータを基に実施したデータ解析を”Analysis”と分類し、各々のワークフローを簡単に記載しています。

例1 RNA Sequencing Experiment Workflow

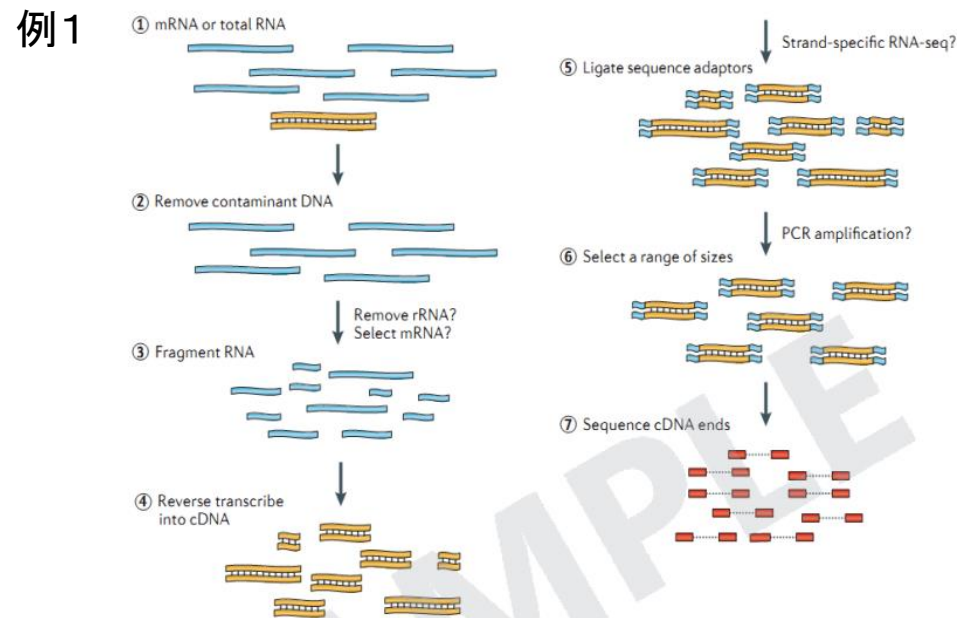


Figure 8. RNA Sequencing Experiment Workflow

例2 Analysis Workflow

例2



Figure 9. Analysis Workflow

3. Summary of Data Production

こちらの項目では得られたリードデータのリード数、GC%、Qualityの確認結果を記載しています。

例1 Raw data stats

・ Phredというプログラムで算出したQuality Score(QS)
= Phredクオリティスコアベーススコールにおけるエラー率の予測指標。

Q20: Phred QSが20以上の塩基の割合

Q30: Phred QSが30以上の塩基の割合

※QSの詳細は9. Appendixにも記載があります。

例1

Table 1. Raw data stats

Index	Sample id	Total read bases*	Total reads	GC (%)	Q20 (%)	Q30 (%)
1	MG_control_1	6,122,979,156	60,623,556	51.08	96.45	94.11
2	MG_control_2	6,550,687,896	64,858,296	52.73	96.84	94.70
3	MG_test_1	6,406,095,690	63,426,690	51.18	96.50	94.21
4	MG_test_2	6,806,696,838	67,393,038	51.28	96.50	94.21

(* Total read bases = Total reads x Read length)

- Total read bases: Total number of bases sequenced
- Total reads: Total number of reads
- GC (%): GC content
- Q20 (%): Ratio of bases that have phred quality score greater than or equal to 20
- Q30 (%): Ratio of bases that have phred quality score greater than or equal to 30

例2 Read quality at each cycle of sample

FASTQCで算出したQSを基に、Forward(read1)およびReverse(read2)について、リードの位置ごとのQSを図示しています。

・縦軸: QS、横軸: リード上での位置
緑色領域: Good Quality、黄色領域: Acceptable Quality、赤色領域: Bad Quality を示しており、得られたリードを平均して評価した際に、どのQualityにあたるのか確認できます。

例2

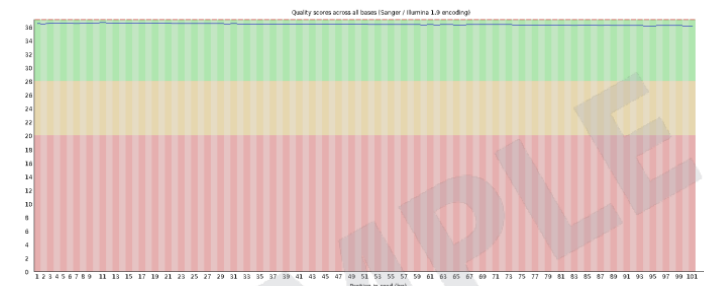


Figure 10. Read quality at each cycle of MG_CTRL_1 (read1)

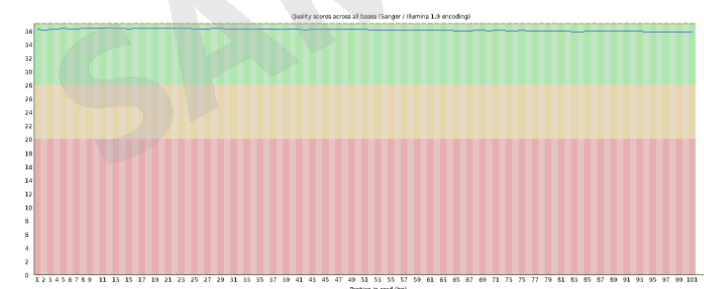


Figure 11. Read quality at each cycle of MG_CTRL_1 (read2)

- Yellow box: Interquartile range (25-75%) of phred score at each cycle
- Red line: Median of phred score at each cycle
- Blue line: Average of phred score at each cycle
- Green background: Good quality
- Orange background: Acceptable quality
- Red background: Bad quality

4. Reference Mapping and Assembly Results

こちらの項目では得られたリードデータをリファレンス情報にマッピングし、マッピングの状況から算出した、既知転写産物および遺伝子の発現量のリストおよび新規転写産物として予測した転写産物の発現量算出結果リストを確認できます。

例1 Known transcripts Expression Level

リードカウント、Transcript Lengthおよび発現量など、得られた情報をまとめた一覧をExcel形式でまとめています。

こちらのリストはResult_RNASeq_Excelフォルダ内のExpression_profileに以下名称にて保管されています。

- Expression_Profile.GRCh38.transcript.xlsx
- Expression_Profile.GRCh38.genes.xlsx

例1

Table 4. Known transcripts Expression Level (example)

Transcript_ID	Gene_ID	Gene Symbol	Description	Transcript_Locus	Transcript Length	AM Read_Count	BM Read_Count	AM_FPKM	BM_FPKM	AM_TPM	BM_TPM
NM_130786	11	ITIBG	alpha-1-B glycoprotein	chr19:58345183-58353492	3382	88	163	0.432306	0.678319	0.847053	1.504474
NR_048112	3	A2M	alpha-2-macroglobulin pseudogene	chr12:9229533-9234207	1201	0	0	0	0	0	0
XM_017013947	9	NAT1	N-acetyltransferase 1, transcript 8	chr8:18170419-18223859	2784	0	211	0	0.108737	0	0.241173
NM_001291962	9	NAT1	N-acetyltransferase 1, transcript 9	chr8:18170467-18223859	2122	0	0	0	0	0	0
NM_000015	10	NAT2	N-acetyltransferase 2	chr8:18391282-18491218	1285	0	0	0	0	0	0
NM_001085	12	SERPINA3	serpin family A member 3	chr14:94612377-94624063	1560	8	75	0.084216	0.694787	0.184454	1.474401
XM_005247104	13	AACD	arylacetamide deacetylase, transcript 1	chr3:151814308-151828488	1620	0	12	0	0.162886	0	0.228152
NM_001086	13	AACD	arylacetamide deacetylase, transcript 2	chr3:151814116-151828488	1663	108	106	1.152576	0.971041	2.524427	2.153215
XM_024452712	14	AAMP	angio associated migratory cell	chr2:218264127-218270181	2002	106	101	0.879142	0.710738	1.925533	1.576378
NM_001302545	14	AAMP	angio associated migratory cell	chr2:218264129-218270137	1763	1621	1787	15.408488	14.424821	33.74835	31.99344
NM_001087	14	AAMP	angio associated migratory cell	chr2:218264129-218270137	1760	9332	10212	88.854179	82.119453	194.6122	182.1363
NM_001166579	15	AANAT	aralkylamine N-acetyltransferase	chr17:76467646-76470117	1913	2	8	0.019876	0.053728	0.023387	0.116948
XM_017024259	15	AANAT	aralkylamine N-acetyltransferase	chr17:76467646-76470117	4252	4	11	0.013221	0.03452	0.028658	0.076564
NR_110548	15	AANAT	aralkylamine N-acetyltransferase	chr17:76467646-76470117	1082	0	0	0	0	0	0
NM_001088	15	AANAT	aralkylamine N-acetyltransferase	chr17:76467603-76470117	971	0	0	0	0	0	0
NR_933220	16	AARS	alanine tRNA synthetase, transcript 1	chr16:70252295-70289509	3258	90	160	0.481517	0.694592	1.01834	1.409466
NM_001695	16	AARS	alanine tRNA synthetase, transcript 2	chr16:70252384-70289509	3344	22367	68204	112.089745	288.699189	245.5037	640.2521

- Transcript_ID: Splicing variant (isoform/transcript)
- Gene_ID: Gene ID
- Gene_Symbol: Symbol of gene
- Gene_Description: Description of gene
- Transcript_Locus: Transcript locus
- Transcript_Length: Transcript length
- [Sample Name]_Read_Count: Read count of a sample
- [Sample Name]_FPKM: FPKM normalized value of a sample
- [Sample Name]_TPM: TPM normalized value of a sample

4. Reference Mapping and Assembly Results

例2 新規転写産物の分類

例2で図示されておりますSplicingパターンの取り方を基に、計13パターンに分けて評価分類しています。

例2

Table 6. Description of class code for various splicing alternative transcript type

=	complete match of intron chain	s	intron match on the opposite strand (likely a mapping error)	e	single exon, overlapping intron, possibly pre-mRNA fragment (unspliced intron)
c	contained in reference (and intron isoform compatible)	x	exonic overlap on the opposite strand (like 'o' or 'e' but on the opposite strand)	o	other same strand overlap with reference exons
k	containment of reference (reverse containment)	i	fully contained in a reference intron	p	possible polymerase run-on (no actual overlap)
j	at least one junction match	y	contains a reference within its intron(s)	r	repeat (at least 50% bases soft-masked)
				u	none of the above (unknown, intergenic)

例3 Novel Transcript list

例2での分類結果、リードカウント、Exon数および発現量など、本解析で得られた情報をExcel形式にまとめています。

例3

Table 9. Novel transcript list (Example)

Transcript_ID	MSTRG_291.1	MSTRG_299.4	MSTRG_1212.1	MSTRG_1249.1	MSTRG_1322.1
Gene_ID	MSTRG_291	MSTRG_299	MSTRG_1212	MSTRG_1249	MSTRG_1322
Transcript_Locus	chr1:16739133-16740150	chr1:17439826-17441498	chr1:108272698-108337462	chr1:109787692-109792410	chr1:115179809-115223491
Strand	-	+	+	-	+
Exon_Count	1	2	6	3	5
Exon_Start	16739133	17439826,17441458	108272698,108284240,108331171,108333657,108335977,108337203	109787692,109789594,109792191	115179809,115217877,115218519,115221798,115222869
Exon_End	16740150	17441053,17441496	108273321,108284351,108331278,108333824,108336127,108337462	109789409,109789684,109792410	115180226,115217963,115218598,115222500,115223491
Class_Code	u	u	u	u	u
AM_Read_Count	71	2	101	203	0
BM_Read_Count	82	92	194	290	61
AM_FPKM	1.144176	0.024883	1.159767	3.22923	0
BM_FPKM	1.107626	1.006513	1.888131	3.912317	0.433582
AM_TPM	2.610777	0.056778	2.646351	7.368444	0
BM_TPM	2.522665	2.292375	4.300299	8.910469	0.987502

ファイルはResult_RNASeq_Excelフォルダ内のNovel Transcript Analysisに以下名称にて保管されています。

- Expression_Profile_with_Novel.GRCh38.transcript.xlsx
※こちらは転写産物のリストとなります。
- Expression_Profile_with_Novel.GRCh38.genes.xlsx
※こちらは遺伝子でのリストとなります。

- Transcript_ID: If there are detected novel transcripts with novel exons, StringTie assigns these transcripts to "MSTRG.xxxx.yy" of temporary transcript ID.
- Gene_ID: If there are detected novel genes in the intergenic region or unknown region, StringTie assigns these genes to "MSTRG.xxxx" of temporary gene ID.
- Transcript_Locus: Start and end position of transcript on genomic region
- Transcript_Length: Length of transcript
- Strand: Strand of transcript on genomic region
- Exon_Count: The number of exon in the transcript
- Exon_Start, End: The start and end position for each exon in the transcript
- Class_Code: Class code corresponding to transcript ID (Refer to Table 7)
- [Sample Name]_Read_Count: Read count of a sample
- [Sample Name]_FPKM: FPKM value for each sample (normalized value)
- [Sample Name]_TPM: TPM normalized value of a sample

5. Differentially Expressed Gene Analysis Results

こちらの項目の解析結果の内容は「Result_RNASeq_Excel」「DEG_Result」に格納されている、レポートファイル-“Analysis_Result.html”にまとめています。こちらのレポートファイルと併せてご確認ください(例1)。

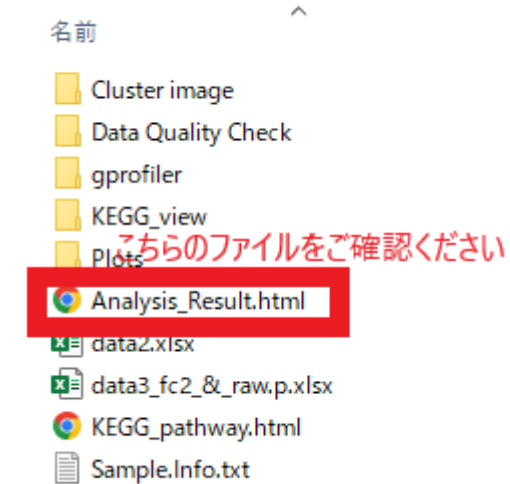
以下6項目に関してこれ以降のページでご案内します。

- ① 遺伝子のフィルタリング
- ② 各サンプルの相関関係の把握
- ③ 発現変動遺伝子の分布状況の確認
- ④ Heatmap Analysis
- ⑤ GO Enrichment Analysis
- ⑥ KEGG Enrichment Analysis ※解析希望時のみ、有償オプション。

データ解析で使用しております数値データのRaw dataは以下の名称でフォルダに格納されています。

- Raw data : Expression Profile.GRCh38.gene.xlsx
- ※ “result_RNAseq_excel”内の”Expression Profile”フォルダに格納。
- Normalization後のRaw data : data2.xlsx
- FC値、P値を加味した統計データ : data3_fc2_&_raw.p.xlsx

例1 Analysis_Result.html について



例2 Expression Profile.GRCh38.gene.xlsx

Table 5. Known genes Expression Level (example)

Gene_ID	Transcript_ID	Gene Symbol	Description	AM_Read_Count	BM_Read_Count	AM_FPKM	BM_FPKM	AM_TPM	BM_TPM
1	NM_130796	A1BG	alpha-1-B glycoprotein	88	163	0.432396	0.678319	0.947053	1.504474
2	NM_000014.NM_001347423	A2M	alpha-2-macroglobulin	0	0	0	0	0	0
3	NR_040112	A2MP1	alpha-2-macroglobulin pseudogene	0	0	0	0	0	0
9	NM_000662.NM_001160170	NAT1	N-acetyltransferase 1	288	217	2.411185	1.490984	5.281078	3.306918
10	NM_000015.NM_017012838	NAT2	N-acetyltransferase 2	10	6	0.097138	0.050729	0.212756	0.112513
12	NM_001085	SERPINA3	serpin family A member 3	8	75	0.084216	0.664787	0.184454	1.474461
13	NM_001086.XM_005247104	AADAC	arylacетamide deacetylase	108	120	1.152579	1.073907	2.524427	2.381867
14	NM_001087.NM_001302545	AAMP	ango associated migratory cell prot	11059	12110	105.141819	97.255012	230.2861	215.7062
15	NM_001088.NM_001166579	AANAT	aralkylamine N-acetyltransferase	6	19	0.023899	0.087248	0.052345	0.193512
16	NM_001605.XR_933220	JAARS	alanyl-tRNA synthetase	22457	68364	112.551262	289.363781	246.5145	641.7927
18	NM_000663.NM_001127448	ABAT	4-aminobutyrate aminotransferase	327	175	1.143824	0.441216	2.505251	0.978593
19	NM_005502.XM_005251773	ABCA1	ATP binding cassette subfamily A n	1496	2719	2.403716	3.695320	5.264719	8.196482
20	NM_001606.NM_212533.XM	ABCA2	ATP binding cassette subfamily A n	2500	3968	5.218521	6.986245	11.42962	15.4951
21	NM_001089	ABCA3	ATP binding cassette subfamily A n	2214	4876	5.619098	10.452255	12.30719	23.18251
22	NM_001271696.NM_0012714BCB7	ABCB7	ATP binding cassette subfamily B n	2618	1974	9.550061	6.097788	20.91695	13.52455
23	NM_001025091.NM_001090	ABCF1	ATP binding cassette subfamily F n	11449	11921	58.366045	49.563715	123.4553	109.9295
24	NM_000350	ABCA4	ATP binding cassette subfamily A n	62	139	0.140036	0.267738	0.306712	0.583827

- Gene_ID: Gene ID
- Transcript_ID: Splicing variant (isoform/transcript)
- Gene_Symbol: Symbol of gene
- Gene_Description: Description of gene
- [Sample Name]_Read_Count: Read count of a sample
- [Sample Name]_FPKM: FPKM normalized value of a sample

5. Differentially Expressed Gene Analysis Results

全比較内容のうち少なくとも1つにおいて、 $|fc| \geq 2$ 、 $raw.p < 0.05$ の条件を満たす遺伝子をフィルタリングした結果をまとめたものが、“data3_fc2_&_raw.p.xlsx”となります。

例3 data3_fc2_&_raw.p.xlsx

【カラム項目詳細】

- A列: Gene ID
- B列: Transcript ID
- C列: Gene Symbol
- D列: データベース記載の遺伝子説明
- E列: 遺伝子の分類
- F列: Protein ID
- G列: HGNC(HUGO Gene Nomenclature Committee) ID
- H列: MIM(Mendelian Inheritance in Man) ID
- I列: Ensembl ID
- J列: IMG/GENE-DB ID
- K列: Test/Control でのFold Change(FC)値
- L列: Test/Control でのlog Counts per Million reads (CPM)値
- M列: Test/Control でのP値
- N列: Test/Control でのFDR correction P値

	A	B	C	D	E	F	G	H	I	J	K	L
	Gene ID	Transcript ID	Gene Symbol	Description	gene biotype	Protein ID	HGNC	MIM	Ensembl	IMG/GENE-DB	S1/S2_fc	S1/S2_logCPM
1												
2		NM_000014	AK2M	alpha-2-macr protein_coding	NP_000006.3			7	103950		4.152384	2.108262
3		NM_001605	XFAARS1	alanyl-tRNA s protein_coding	NP_001596.2			20	601065		5.617145	7.015987
4		NM_005502	XNABCA1	ATP binding ci protein_coding	NP_005493.2	X		29	600046		-10.574885	8.461171
5		NM_001059	ABCA3	ATP binding ci protein_coding	NP_001080.2	X		33	601615		3.550961	2.428642
6		NM_000350	ABCA4	ATP binding ci protein_coding	NP_000341.2	X		34	601691		11.519017	1.339513
7		NM_001093	XNACCB	acetyl-CoA ci protein_coding	NP_001084.3	X		65	601557		4.640649	3.742363
8		NM_001141	94ACTA2	actin alpha 2, protein_coding	NP_001135417	X		130	102620		4.201726	1.071240
9		gene-AC	TBP6ACTBP6	ACTB pseudo;pseudogene				141			-4.283912	-0.573778
10		gene-AC	TG1FACTG1P1	actin gamma 1pseudogene				146			4.622086	0.690769
11		NM_004301	NKACTL6A	actin like 6A protein_coding	NP_004292.1	N		24124	604958		6.170424	4.819939
12		NM_001104	NKACTN3	actinin alpha 3 protein_coding	NP_001095.2	N		165	102574		-5.175477	0.446137
13		NM_001278	57ACVFR2A	activin A rece protein_coding	NP_001265505	X		173	102581		-4.265381	1.144517
14		NM_001106	XNACVFR2B	activin A rece protein_coding	NP_001097.2	X		174	602730		3.435856	6.370163
15		NM_001107	NKACYP1	acylphosphata protein_coding	NP_001098.1	N		179	600875		11.100987	3.189284
16		101	NM_001109	NKADAMB	ADAM metallo protein_coding	NP_001100.3	N		215	602267	-10.649603	3.659583
17		104	NM_001112	NKADARB1	adenosine dea protein_coding	NP_001103.1	N		226	601218	4.269774	2.778297

	M	N	O	P	Q	R	S	T	U	V	W	X
	S1/S2_raw.pvs	S1/S2_bh.pval	S1_mean	S2_mean	N S1	N S2	S1_Read_Cour	S2_Read_Cour	S1_FPKM	S2_FPKM	S1_TPM	S2_TPM
	0.02023408	0.114031759	2.977272	1.406124	1.408124	2.977272	84	395	0.255331	1.208290	0.568872	2.585372
	0.00462099	0.04207104	7.788569	5.328481	5.328481	7.788569	1990	12742	8.832669	56.215288	19.678998	119.353204
	0.00019328	0.00443287	5.951634	9.332989	9.332989	5.951634	32704	3525	51.523588	5.613255	114.793442	11.917753
	0.0359	0.184633756	3.221251	1.741165	1.741165	3.221251	119	482	0.271425	1.104783	0.604729	2.345614
	0.00023742	0.00504701	2.475155	0.479013	0.479013	2.475155	20	264	0.041292	0.543516	0.091975	1.153963
	0.01138005	0.078178854	4.519647	2.517397	2.517397	4.519647	240	1270	0.346556	2.040649	0.772118	4.332594
	0.02228167	0.122125486	2.109882	0.838008	0.838008	2.109882	40	192	0.340040	1.685961	0.757603	3.537078
	0.03319928	0.156998566	0.312471	1.031073	1.031073	0.312471	53	14	0.450983	0.118435	1.004791	0.251454
	0.01640391	0.098871588	1.830465	0.633483	0.633483	1.830465	28	149	0.212271	1.150434	0.472935	2.442538
	0.00305781	0.03165393	5.628990	3.146673	3.146673	5.628990	399	2807	3.281811	22.996632	7.311806	48.825198
	0.01036063	0.07336829	0.517874	1.699042	1.699042	0.517874	114	25	0.556068	0.121319	1.238908	0.257578
	0.0203832	0.114854426	0.871166	2.183998	2.183998	0.871166	180	48	0.520596	0.134568	1.159877	0.285708
	0.03794841	0.189696968	7.011993	5.258226	5.258226	7.011993	1893	7414	2.259149	9.157604	5.033337	19.442924
	0.00018166	0.00421239	4.140339	1.319550	1.319550	4.140339	76	963	1.857646	25.482089	4.138797	54.102172
	0.00021391	0.00474993	1.659568	4.585986	4.585986	1.659568	1169	125	5.563584	0.644729	12.385547	1.368854
	0.01702052	0.101558059	3.587444	1.839687	1.839687	3.587444	131	638	0.206225	1.646652	0.459466	3.498082

O-X列につきましては、各サンプルごとの平均値、raw signaのTMM補正值、リードカウント、FPKM値、TPM値を記載しています。

※詳細は“ Analysis_Result.html”“ III-5. Column Information”にございます。

こちらのExcelファイルにGO Enrichment Analysis、KEGG Enrichment Analysis※オプションの結果も別シートで記載していますので、あわせてご確認ください。

①解析対象遺伝子のフィルタリング

例1 Distribution of genes with various number of zero counts

いずれかのサンプルでリードカウントが0となった遺伝子を解析対象から外すゼロカウントフィルタリングを行います。

図ではサンプルデータごとにリードカウント=0の遺伝子をフィルタリングし、サンプル数を跨いで0となる遺伝子数を表示しています。

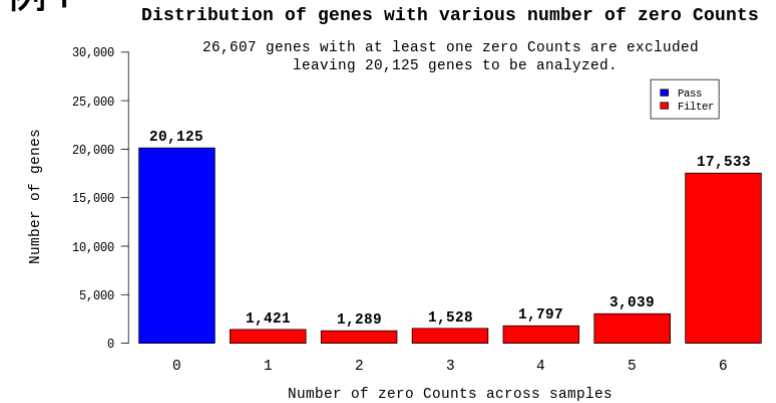
全てのサンプルで発現している遺伝子のみ(青Box)、以降の解析では使用しています。
※Raw dataのExcelにはリードカウントが0の遺伝子も記載しています。

ご希望いただいた場合、0カウントフィルタリングの条件を変更して、再解析も可能ですので、ご希望の場合は担当者にご連絡ください。

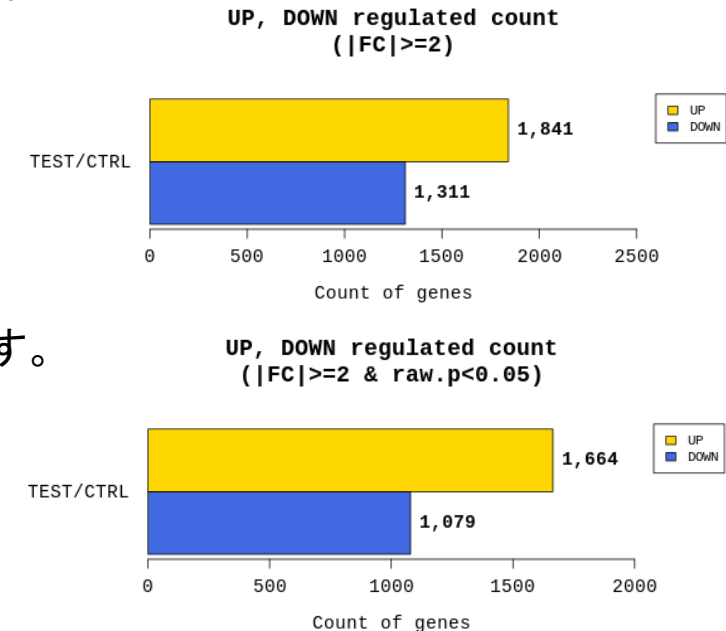
例2 Up, Down regulated count by Fold Change and P-value

リードカウントを基にTest/Controlで計算しFold change(FC)を算出。
あわせてp値も算出し、 $|FC| \geq 2$ かつp値 < 0.05 の遺伝子をフィルタリングしています。
(フィルタリングされる遺伝子数が極端に少ない場合は $|FC| \geq 1.5$ にて、フィルタリングを行う場合もあります。)

例1



例2



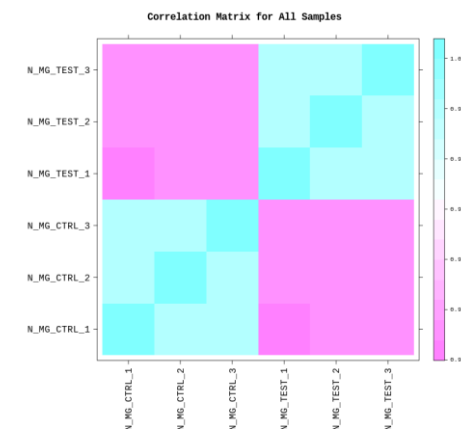
②各サンプルの相関関係の把握

各サンプルの相関関係を複数の図表で表しています。

例1 Correlation Matrix for All Samples

サンプルデータごとにピアソンの積率相関係数を取り、ヒートマップ化しています。

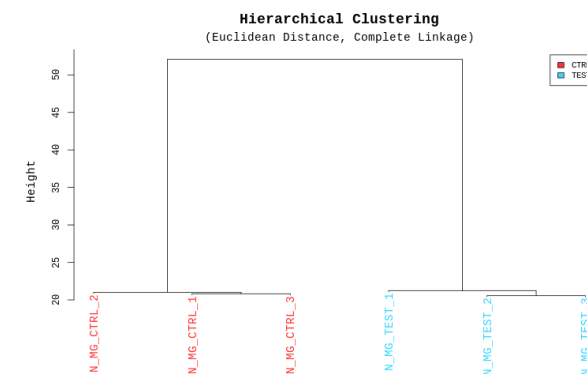
例1



例2 Hierarchical Clustering Analysis

サンプルデータごとのノーマライズ値を元に、発現の類似性を階層的に表しています。

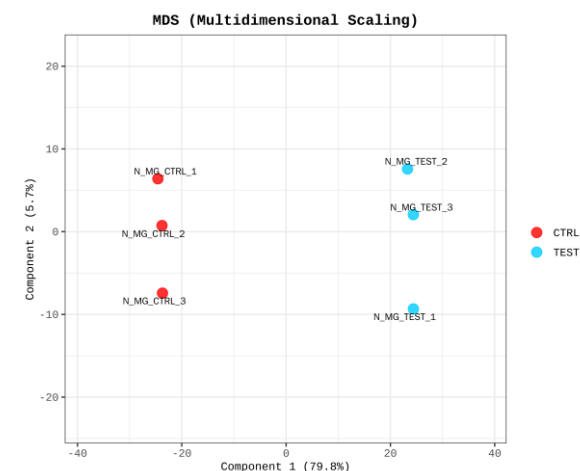
例2



例3 Multidimensional Scaling Analysis

二次元プロットにより、発現量のばらつきとサンプルとの相関を表しています。

例3



③発現変動遺伝子の分布状況の確認

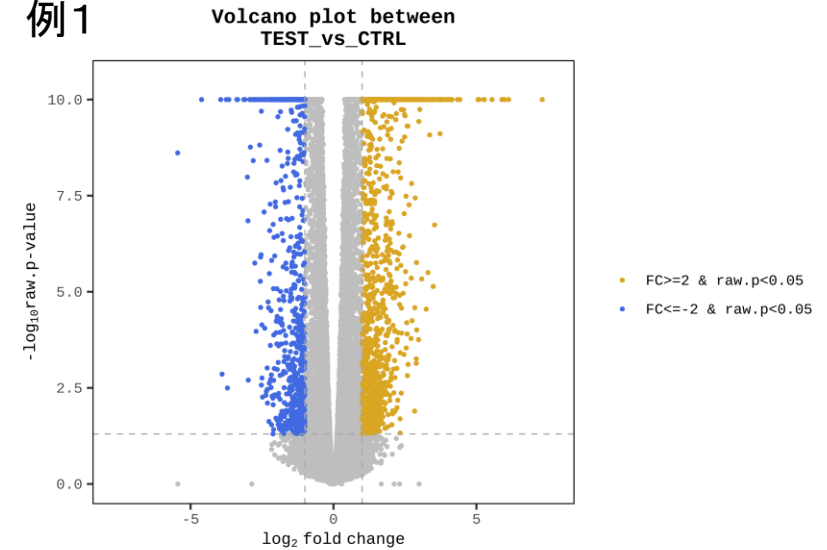
|FC| \geq 2かつp値<0.05の遺伝子に対して
各種Plotを作図しています。

例1 Volcano Plot between TEST vs CTRL

縦軸をp値の対数、横軸をFC値の対数として分布図化することで、
|FC| \geq 2かつp値<0.05の遺伝子の-(青色(-), 黄色(+))の分布を
表しています。

※n \geq 2の際に作成されます。

例1

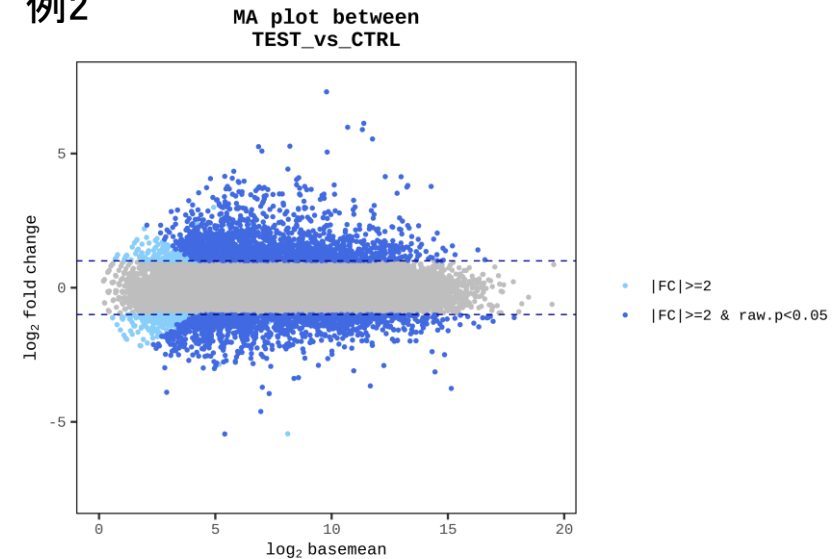


例2 MA Plot between TEST vs CTRL

縦軸をFC値の対数、横軸を発現量の相乗平均
として分布図化することで、
|FC| \geq 2かつp値<0.05の遺伝子-(水色、青色)の分布を表しています。

※解析結果によっては|FC| \geq 2が条件でない場合もございます。

例2

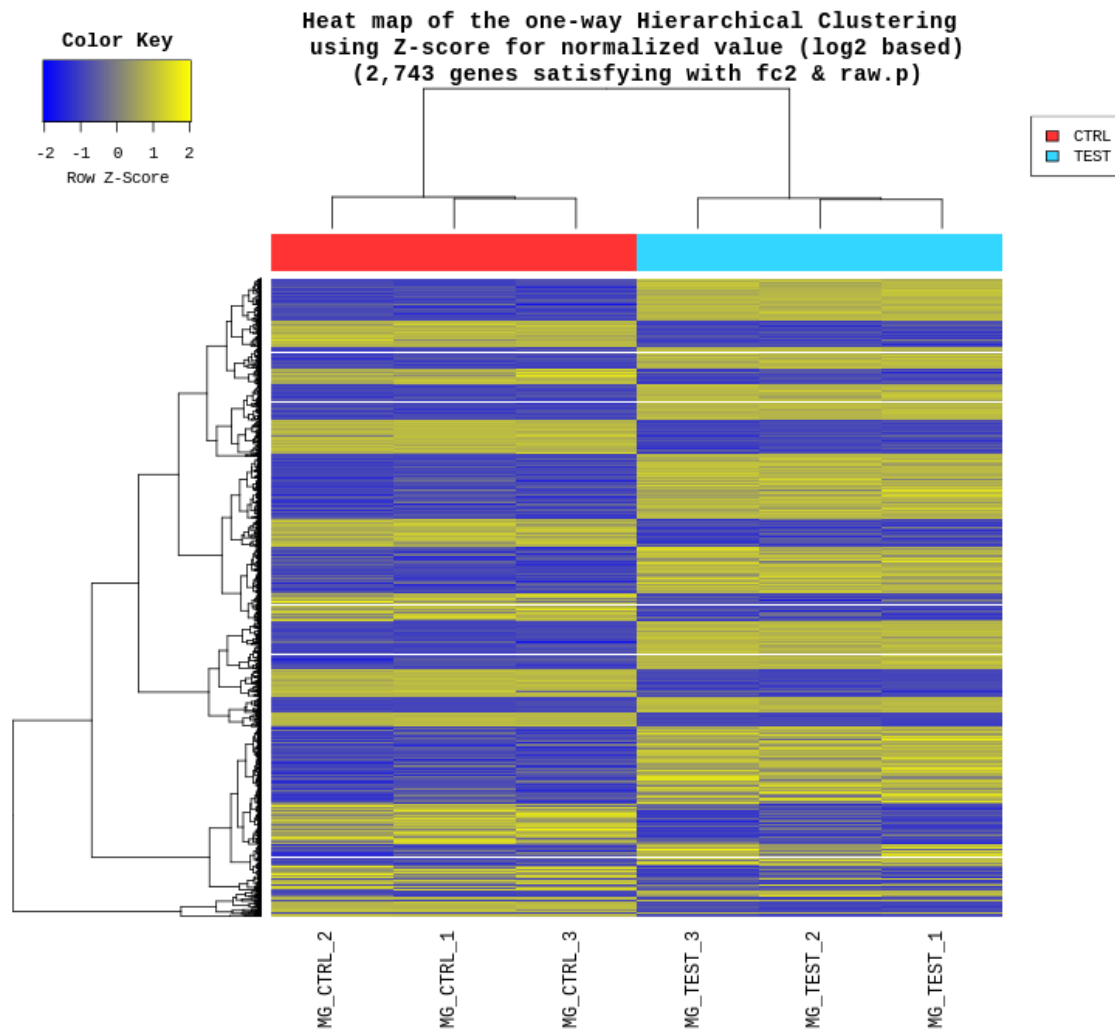


④ Heatmap Analysis

例1 Heatmap

$|FC| \geq 2$ かつ p 値 < 0.05 の遺伝子に対してあわせてノーマライズしたすべてのサンプル及び各比較パターンごとにHeatmapを作成しています。

例1

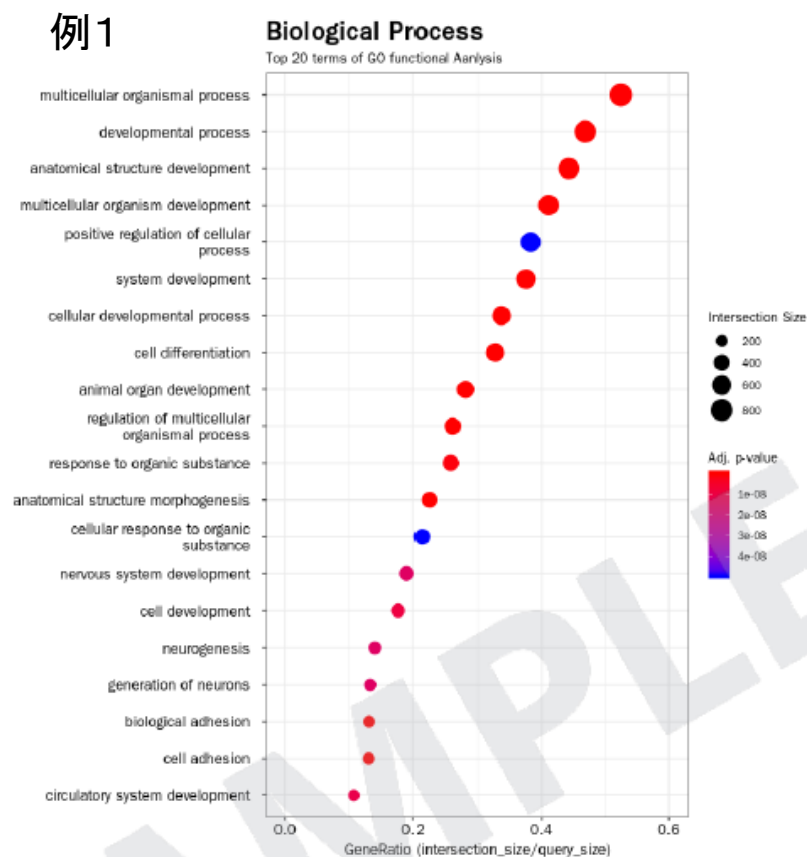


⑤ GO Enrichment Analysis

Gene Ontology(<http://geneontology.org/>) database登録情報より、発現量変動が大きかった遺伝子がBiological Process (BP), Cellular Component (CC), Molecular Function (MF)の3つの観点からどのような機能グループ(GO term)に属するかを分類し、発現量変動の大きかったGO termの一覧表及びPlotを作成しています。

例1 Biological ProcessにおけるGO解析結果
検定によるp値 < 0.05, 0.01, 0.001における、
上位20位までの遺伝子をグラフ化しています。

- ・縦軸: 遺伝子系統名、
- ・横軸: 全体の遺伝子数における
該当遺伝子数の割合
- ・●: 遺伝子数の割合
- ・カラーバー: p値



「result_RNAseq_excel」→「DEG_result」→「gprofiler」内に各種Plotが格納されています。
比較パターンごとのフォルダ内には“gprofiler.bar.png”および“gprofiler.png”がありますが、
同一内容の図表の表示の仕方を棒グラフ(.bar.png)とドットプロット(.png)で表しています。
例1はドットプロットでの図表となります。

⑤ GO Enrichment Analysis

【sizefilt.pngにつきまして】

格納されております”gprofiler.sizefilt.png”は、
p-valueが0.05未満かつterm_sizeが10～500のもののみから作成されたPlotになります。

GO termはterm size(そのtermに含まれる遺伝子数)が様々です。
非常に大きいまたは小さいGO termは、計算方法によっては、統計的有意性が大きく算出されることがあります。
そのため、通常の解析結果と合わせて、term sizeが10-500にフィルタリングして、別途作成したPlotも納品しています。

例2 GO_stat Sheet

データ解析で使用しております数値データのRaw dataは
以下の内容で”data3_fc2_&_raw.p.xlsx”に記載しています(例2)。

- GO_stats
- GO_genes

source	term_id	term_name	adjusted_p_value	term_size	query_size	intersection_size	effective_domain_size	intersections
GO:CC	GO:002626	cytosolic ribosome	2.77598E-17	115	1921	30	18797(6134_6206_6155_6204_6168_20091)	
GO:BP	GO:000614	GTP-dependent cotranslational protein targeting to membrane	1.60123E-13	96	1824	44	17816(6134_6206_6155_6204_6168_6747_4)	
GO:MF	GO:0001720	structural constituent of ribosome	1.82611E-14	170	1860	39	18098(6134_6206_6155_6204_6168_20091)	
GO:BP	GO:0006613	translational protein targeting to membrane	2.09918E-14	101	1824	44	17816(6134_6206_6155_6204_6168_6747_4)	
GO:MF	GO:0001518	structural molecule activity	4.45123E-14	739	1860	151	18098(6134_6206_127294_4586_301_1887_0)	
GO:BP	GO:0045047	protein targeting to ER	7.18109E-14	109	1824	45	17816(6134_6206_6155_6204_6168_6747_4)	
GO:CC	GO:0044191	ribosomal subunit	2.96614E-11	109	1921	61	18797(6134_6206_6155_6204_6168_20091)	
GO:BP	GO:007259	establishment of protein localization to endoplasmic reticulum	2.82077E-11	113	1824	45	17816(6134_6206_6155_6204_6168_6747_4)	
GO:BP	GO:007072	protein localization to endoplasmic reticulum	4.06110E-11	137	1824	47	17816(6134_6206_6155_6204_6168_6747_4)	
GO:CC	GO:0005840	ribosome	1.94090E-10	264	1921	62	18797(6134_6206_6155_6204_6168_20091)	
GO:CC	GO:002625	cytosolic large ribosomal subunit	1.69728E-10	64	1921	28	18797(6134_6155_6168_20091_6167_6161)	
GO:BP	GO:000134	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	7.11140E-10	122	1824	42	17816(6134_6206_6155_6204_6168_6747_4)	
GO:CC	GO:1904450	plasma membrane part	1.34640E-09	2979	1921	402	18797(6134_6206_6155_6204_6168_20091)	
GO:CC	GO:0071944	cell periphery	1.88910E-09	5662	1921	709	18797(6134_6206_6155_6204_6168_20091)	
GO:CC	GO:0005880	plasma membrane	5.37824E-09	5539	1921	692	18797(6134_6206_6155_6204_6168_20091)	
GO:CC	GO:0044444	cytoplasmic part	5.37824E-09	9685	1921	1225	18797(6134_6206_6155_6204_6168_20091)	
GO:CC	GO:0005717	cytoplasm	5.47210E-09	11334	1921	1509	18797(6134_6206_6155_6204_6168_20091)	
GO:BP	GO:0009880	tissue development	5.79946E-09	2068	1824	305	17816(6405_5054_8322_5743_144195_1272)	
GO:BP	GO:0006613	translational protein targeting to membrane	5.99600E-09	109	1824	54	17816(6134_6206_6155_6204_6168_6747_4)	
GO:BP	GO:0051178	localization	1.25007E-08	6751	1824	824	17816(6134_6206_10226_10794_21256_6405)	
GO:CC	GO:1903561	intracellular vesicle	1.65130E-08	2165	1921	309	18797(629_5054_10103_2068_9518_4151_4)	
GO:CC	GO:0001230	intracellular organelle	1.66690E-08	2167	1921	309	18797(629_5054_10103_2068_9518_4151_4)	
GO:CC	GO:0044448	cytosolic part	2.71165E-08	252	1921	40	18797(6134_6206_6155_6204_6168_19921)	
GO:BP	GO:0012921	multicellular organismal process	1.22013E-08	7718	1824	922	17816(6134_6206_6405_5670_5054_7079_81)	

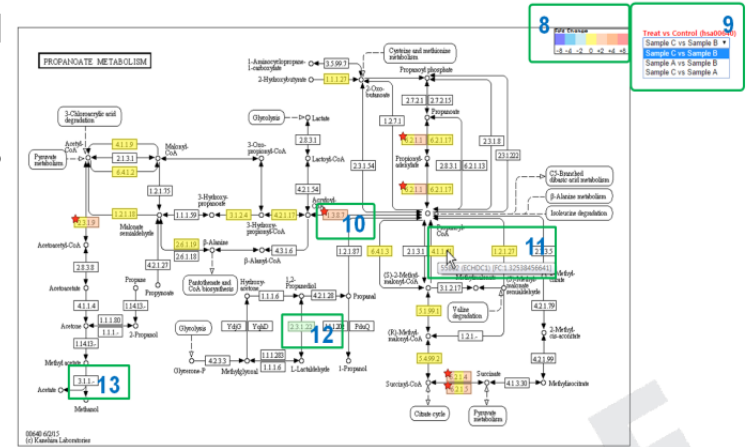
- source: Code for the data source. Ex> GO:BP | GO:CC | GO:MF ...
- term_id: ID for the enriched term/functional category
- term_name: Readable name for the enriched term
- adjusted_p_value: Adjusted p-value by FDR
- query_size: The number of unique DEG that are annotated to the data source (the functional category).
- intersection_size: The number of unique DEG that are annotated to the term_id
- term_size: The number of genes of species that are annotated to the term_id.
- effective_domain_size: The number of genes of species that are annotated to the data source (the functional category).
- intersections: list of unique DEG that are annotated to the term_id

⑥KEGG Enrichment Analysis

※有償オプション。
生物種によっては対応できないため事前にご相談ください。

KEGG(https://www.genome.jp/kegg/kegg_ja.html) という database の情報に、各比較パターンでの発現量変動を当てはめることで、各Pathway上でどのような遺伝子に有意な変化があったかを図示しています。

例1



例1 Description of KEGG pathway map frame

各Pathwayで有意差のあった遺伝子の発現量比を8のような色分けでPathwayマップ図内に表しています。また、各pathwayで重要なモジュールには★マークを付けています。

Figure 16. Description of KEGG pathway map frame

※ PathwayマップはKEGGデータベースに登録の情報を使用しているため、お客様の希望に合わせての作図・改定の対応は致しかねます。

- Block 8: Fold change values of DEG are shown in colors.
- Block 9: You can change to different combination within the current KEGG pathway page. The combination shown in the box is currently shown combination.
- Block 10: Significant pathway module is marked with red star (based on data3 file of significance).
- Block 11: The name and fold change value of the gene are shown when mouse is over. (genes are separated with comma). If the gene id exists but there is no FC value on the title of module, then the gene does not exist in data2 file that is processed QC filtering step.
- Block 12: Green color box of pathway map is modules that are not mapped. Gene is in the pathway map but the expression is not shown.
- Block 13: White box of pathway map is module that is not relevant to the species.

データ解析で使用しております数値データのRaw dataは以下の内容で” data3_fc2_&_raw.p.xlsx”に記載しています。

- KEGG_stats
- KEGG_genes

例2

例2 KEGG_stat Sheet

Example of KEGG pathway enrichment analysis result

MapID	MapName	Number of SigGenes	Genes	Sig.NotIn.KEGG	Genome.In.KEGG	Genome.NotIn.KEGG	PValue	Bonferroni	FDR
01100	Metabolic pathways	86	10229,10622,10797,10998,11105	281	1220	58263	8.6376E-01	2.29709E-50	2.29709E-50
01110	Biosynthesis of antibiotics	25	113679,1491,2026,2027,22994	342	214	59209	5.67107E-22	1.5061E-19	7.54252E-20
05203	Viral carcinogenesis	22	1021,1026,1030,3017,3106,3113	345	206	59277	1.32649E-18	3.52434E-16	1.74749E-16
04151	PKA-Akt signaling pathway	25	10110,1021,1026,1280,2057,22	342	347	59136	1.79170E-17	4.76608E-15	1.91131E-15
04142	Lysosome	18	10577,138050,1514,175,1777,21	349	123	59860	2.54025E-17	6.75707E-15	1.35141E-15
05200	Pathways in cancer	26	1021,1026,1030,1121,1,2094,22	341	398	59085	1.66913E-17	8.42988E-15	1.40498E-15
05205	Proteoglycans in cancer	20	1026,1121,1514,1839,3678,40	347	204	59279	2.77936E-16	7.28235E-14	1.04031E-14
01230	Biosynthesis of amino acids	14	113679,1491,2026,2027,22994	353	74	59409	9.20418E-15	2.44835E-12	1.90404E-13
05166	HTLV-I infection	20	1026,1030,1121,1,1958,214,23	347	261	59222	1.77887E-14	4.7318E-12	5.25796E-13
01200	Carbon metabolism	15	113679,2026,2027,22994,230,21	352	113	59370	6.6255E-14	1.76238E-11	1.76238E-12
04210	MAPK signaling pathway	19	1848,1847,2348,2261,2264,235	348	257	59226	1.62278E-13	4.31166E-11	3.92418E-12
04390	Hippo signaling pathway	16	1121,126374,1490,16824,271	351	154	59229	2.11892E-13	5.51631E-11	4.69694E-12
04115	p53 signaling pathway	12	1021,1026,2713,5094,51246,5	355	68	59425	2.40937E-12	6.38498E-10	4.91131E-11
04145	Phagosome	14	10381,11151,1514,155066,3106	353	155	59328	4.8663E-11	1.29976E-08	9.28397E-10
05206	MicroRNAs in cancer	17	1021,1026,2261,3162,3371,3671	350	297	59186	1.46683E-10	3.90177E-08	2.60118E-09
04550	Signaling pathways regulating ovariectomy	13	1121,1261,1294,3625,5600,56	354	142	59341	3.51263E-10	6.6836E-08	4.17725E-09
04048	Th1 signaling pathway	12	1051,1090,2325,3726,323,468	355	110	59373	2.8844E-10	1.27774E-08	4.2222E-09
05168	Herpes simplex infection	14	2353,3106,3133,3665,406,4938	353	186	59297	4.01970E-10	1.00926E-07	5.94034E-09
00280	Glycine, serine and threonine metabolism	9	113679,1491,211,23464,2593,21	358	40	59443	5.52329E-10	1.40973E-07	7.73541E-09
04110	Cell cycle	12	1021,1026,10274,1028,1030,53	355	124	59359	8.7649E-10	2.31146E-07	1.16573E-08
04015	Ras1 signaling pathway	14	2248,2261,2264,2770,3000,360	353	211	59272	1.79860E-09	4.54501E-07	2.16431E-08
04048	FoxO signaling pathway	12	10110,1026,1030,10305,2370,1	355	134	59349	1.81036E-09	4.6937E-07	2.24095E-08
04060	Cytokine-cytokine receptor interaction	15	2057,3578,3590,3625,5130,51	352	265	59218	2.64579E-09	7.01811E-07	3.89920E-08
05169	Epstein-Barr virus infection	13	1026,10622,3106,3113,3115,37	354	201	59282	1.01835E-08	2.68751E-06	1.1198E-07

- MapID: KEGG map ID
- MapName: KEGG map name
- Number_of_SigGenes: Number of (uniquely) differentially expressed genes that are included in the pathway
- Genes: List of gene that are included in the pathway (comma delimited)
- Sig.NotIn.KEGG: Number of (uniquely) differentially expressed genes that are not included in the pathway
- Genome.NotIn.KEGG: Number of genes that are not associated to this pathway among the genes in given species
- PValue: Raw p-value from the modified fisher's exact test
- Bonferroni: Corrected p-value by bonferroni method
- FDR: Corrected p-value by FDR method

6. SNP and Indel Discovery

※有償オプション。
生物種によっては対応できないため事前にご相談ください。

シーケンス結果をリファレンス情報にマッピング、SNV calling後、Annotationを付与した結果を納品します。

解析結果は”SNV_Call_サンプル名.xlsx”のExcel形式にて格納されています。

例1 SNV納品結果例

各項目の詳細につきまして、PDFファイルにまとめがあります。
例2で示しておりますリンクが納品物PDFにありますので、こちらよりご取得・ご確認ください。

例1

Table 12. An example of annotation of individually discovered SNV

CHROM	chr1	chr1	chr1	chr1	chr1	chr1
POS	981131	982573	982994	1650787	2335969	19413261
REF	A	C	T	T	C	T
[Sample1]_ALT	G		C	C	G	A
[Sample1]_Zygoty	HOM		HOM	HOM	HOM	HET
[Sample1]_QUAL	41.74		45.74	62.74	21.77	126.77
[Sample1]_DP	2		2	2	2	9
[Sample1]_AD	2		2	2	2	5
[Sample1]_MQ	60		60	60	60	60
[Sample1]_FILTER	PASS		PASS	SnpCluster	PASS	PASS
[Sample2]_ALT		T	C		G	A
[Sample2]_Zygoty		HOM	HOM		HOM	HET
[Sample2]_QUAL		96.03	125.9		45.74	35.77
[Sample2]_DP		4	5		2	3
[Sample2]_AD		4	5		2	2
[Sample2]_MQ		60	60		60	60
[Sample2]_FILTER		PASS	PASS		PASS	PASS
[SampleN]...
Effect	missense_variant	sequence_feature	synonymous_variant	missense_variant	3_prime_UTR_variant	missense_variant
Putative_Impact	MODERATE	LOW	LOW	MODERATE	MODIFIER	MODERATE
Gene_Name	AGRN	AGRN	AGRN	CDK11B	RER1	UBR4
Feature_Type	transcript	domain:SEA	transcript	transcript	transcript	transcript
Feature_ID	NM_001305275.1	NM_198576.3	NM_001305275.1	NM_001787.2	NM_007033.4	NM_020765.2
Transcript_BioType	protein_coding	protein_coding	protein_coding	protein_coding	protein_coding	protein_coding
Rank/Total	15/38	19/35	21/38	4/20	7/7	100/106
HGVS.c	c.2555A>G	c.3389-134C>T	c.3558T>C	c.335A>G	c.*1406C>G	c.14599A>T
HGVS.p	p.Gln852Arg	.	p.Phe1186Phe	p.His112Arg	.	p.Met4867Leu
REF_AA	Q	-	F	H	-	M
ALT_AA	R	-	F	R	-	L
...
dbSNP151_ID	rs9697293	rs3813192	rs10267	rs1137003	rs12085089	rs12584
p3_1000G_AF	0.0345447	0.028155	0.835863	.	0.321286	0.601438
...
ESP6500_MAF_EA	G:0.002326	.	T:0.081279	.	.	T:0.434186
...
CLINVAR_CLNSIG	Benign	.	Benign	.	.	.
...

例2

PDF file linked below.

[LINK AnnotDescription.pdf](#)

7. Fusion Gene Prediction Results

※有償オプション。
生物種によっては対応できないため事前にご相談ください。

以下3種類のProgramを使用した予測結果を納品しています。

• Defuse program

Defuse programを使用し、マッピング状況から2遺伝子間を跨ぐリードを抽出し、融合遺伝子を予測します。
※Excel形式での納品例(例1)。

• Fusion Catcher

Fusion Catcherを使用し、マッピング状況から既知融合遺伝子であるか否か、また、転座、キメラかを予測します。
※Excel形式での納品例(例2)。

例1 Table 13. Example of Fusion Gene Prediction Results

Sample	AM	AM	BM	BM
Splitr_Sequence	ATAAATCTGACACTATG GACTTCAGACATGCAG GGTGACIGGTGGTGA GCTGGTAAAGGTTACG AAGATTAATGTGAGTG	TCGAGGATACTCACCA GAAACCGAAAATGCC GAAACCAICATTACTC ACGGTGAACCTCAGCC ATGAGAACCAGAAA	ATAAATCTGACACTATG GACTTCAGACATGCAG GGTGACIGGTGGTGA GCTGGTAAAGGTTACG AAGATTAATGTGAGTG	TCGAGGATACTCACCA GAAACCGAAAATGCC GAAACCAICATTACTC ACGGTGAACCTCAGCC ATGAGAACCAGAAA
Splitr_Count	39	31	15	138
Span_Count	17	12	6	15
Adjacent	Y	N	Y	N
Gene1	ENSG00000108953	ENSG00000092820	ENSG00000108953	ENSG00000092820
Gene2	ENSG00000167193	ENSG00000058335	ENSG00000167193	ENSG00000058335
Gene1_Description	tyrosine 3-monooxygenase	tyrosine 3-monooxygenase	tyrosine 3-monooxygenase	tyrosine 3-monooxygenase
Gene2_Description	v-cr1 avian sarcoma virus	Ras protein-specific guan	v-cr1 avian sarcoma virus	Ras protein-specific guan
Gene1_Name	YWHAE	EZR	YWHAE	EZR
Gene2_Name	CRK	RASGRF1	CRK	RASGRF1
Gene1_Strand	-	-	-	-
Gene2_Strand	-	-	-	-
Gene1_Chrom	17	6	17	6
Gene2_Chrom	17	15	17	15
Gene1_Start	1247566	159186773	1247566	159186773
Gene2_Start	1323983	79252289	1323983	79252289
Gene1_End	1303672	159240444	1303672	159240444
Gene2_End	1366456	79383115	1366456	79383115
Genomic_Strand1	-	-	-	-
Genomic_Strand2	+	+	+	+
Genomic_Break_Position1	1257505	159239114	1257505	159239114
Genomic_Break_Position2	1326944	79356868	1326944	79356868
Probability	0.883417506	0.985006948	0.84040979	0.986824427

例2 Table 14. Example of Fusion Gene Prediction Results

Sample	AM	AM	BM	BM
Gene_1_symbol (Send_fusion_partner)	RPS13	EZR	RPS13	EZR
Gene_2_symbol (Send_fusion_partner)	PLEKHA7	RASGRF1	PLEKHA7	RASGRF1
Fusion_description	adjacent.ribosomal_prot		adjacent.ribosomal_prot ein.10K<gap>100K.read through	
Counts_of_common_mapping_reads	0	0	0	0
Spanning_pairs	18	15	33	104
Spanning_unique_reads	15	9	20	34
Longest_anchor_found	30	30	30	48
Fusion_finding_method	BOWTIE.BOWTIE+BLAT	BOWTIE.BOWTIE+BLAT	BOWTIE.BOWTIE+BLAT	BOWTIE.BOWTIE+BLAT
Fusion_point_for_gene_1 (Send_fusion_partner)	11.17098715-	6.159239114-	11.17098715-	6.159239114-
Fusion_point_for_gene_2 (Send_fusion_partner)	11.16892729-	15.79356868-	11.16892729-	15.79356868-
Gene_1_id (Send_fusion_partner)	ENSG00000110700	ENSG00000092820	ENSG00000110700	ENSG00000092820
Gene_2_id (Send_fusion_partner)	ENSG00000166689	ENSG00000058335	ENSG00000166689	ENSG00000058335
Gene_1_Description	ribosomal protein S13	S13	ribosomal protein S13	S13
Gene_2_Description	pleckstrin homology dom	Ras protein-specific guan	pleckstrin homology dom	Ras protein-specific guan
Exon_1_id (Send_fusion_partner)	ENSE00000351366	ENSE00001212701	ENSE00000351366	ENSE00001212701
Exon_2_id (Send_fusion_partner)	ENSE000003671290	ENSE00001665313	ENSE000003671290	ENSE00001665313
Fusion_sequence	ATTACAAACTGGCCCA ACAGAGSCCTTACTCT TTCCAGAGATCGCGATA ACCAGCAGACCACAG CATTCCAGCATCCTGT GADGGGA	GGGGATCGAGGATAC TCACAGCAACCCAAA ATGCCGAAACCA*CAT TACTTCAGGTGAACCT CAGCCATGAGAACC GAAAGCCT	ATTACAAACTGGCCCA ACAGAGSCCTTACTCT TTCCAGAGATCGCGATA ACCAGCAGACCACAG CATTCCAGCATCCTGT GACGGGA	TGTTTTGGGGATCGCA GGACTACTCACCAGAAA CCGAAAATGCCGAAA CCA*CACTTCTCACGG TGAACCTCAGCCATGA GAACCCAGAAAGCGTTG GAGCT
Predicted_effect	out-of-frame	in-frame	out-of-frame	in-frame
Predicted_fused_transcripts	ENST00000228140.176/ ENST00000531066.264. ENST00000228140.176/ ENST00000355661.233. ENST00000533969.157/ ENST00000531066.264. ENST00000533969.157/ ENST00000355661.233. ENST00000525634.297/ ENST00000531066.264. ENST00000525634.297/ ENST00000355661.233	ENST00000367075.181/ ENST00000558480.543. ENST00000367075.181/ ENST00000419573.552. ENST00000337147.146/ ENST00000533969.157/ ENST00000355661.233. ENST00000337147.146/ ENST00000525634.297/ ENST00000419573.552. ENST00000525634.297/ ENST00000355661.233	ENST00000228140.176/ ENST00000531066.264. ENST00000228140.176/ ENST00000355661.233. ENST00000533969.157/ ENST00000531066.264. ENST00000533969.157/ ENST00000355661.233. ENST00000525634.297/ ENST00000531066.264. ENST00000525634.297/ ENST00000355661.233	ENST00000367075.181/ ENST00000558480.543. ENST00000367075.181/ ENST00000419573.552. ENST00000337147.146/ ENST00000533969.157/ ENST00000355661.233. ENST00000337147.146/ ENST00000525634.297/ ENST00000419573.552. ENST00000525634.297/ ENST00000355661.233
Predicted_fused_proteins	MGRMHAPKGLSQSA LPRYSRVPYTWLKLSD DVKEQIYLAKKGLTPS QAITSRRPQHSGLL	MPKPHYTFVNFSHENQ KALELRTEADKDCDEW VAJAHASRYRLA_DG SFVMDDESLYESSLRE PKLPT.	MGRMHAPKGLSQSA LPRYSRVPYTWLKLSD DVKEQIYLAKKGLTPS QAITSRRPQHSGLL	MPKPHYTFVNFSHENQ KALELRTEADKDCDEW VAJAHASRYRLA_DG SFVMDDESLYESSLRE PKLPT.

7. Fusion Gene Prediction Results

• Arriba

Arribaを使用し、STAR alignerでのマッピング状況から融合遺伝子を予測、および融合までの動態を図示しています。

※Arribaが使用できるのはhumanのみとなります。

※PDFファイル、Excel 形式での納品。

例3 Arriba による解析結果

1. 融合遺伝子形成までの模式図

※染色体情報、Coverage、転写様式、配列、BreakPointの内容が確認できます。

2. Circus Plot

※各染色体ごとに全ゲノム情報を環状に表記し、融合している部位を線つなぎで表示しています。

3. 予測融合遺伝子のドメイン構造の模式図。

※Excel 形式での納品例(例4)。

例3

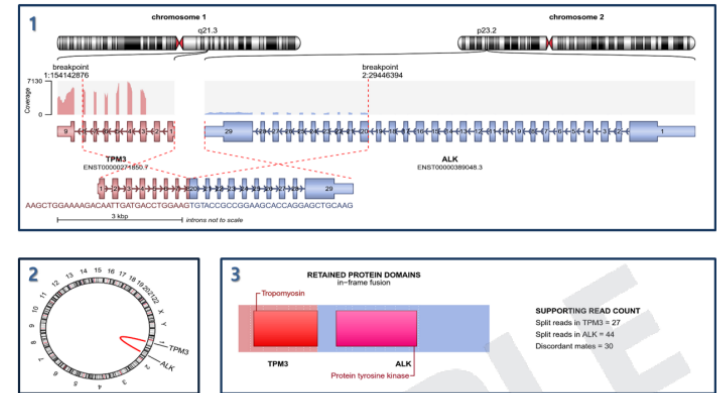


Figure 17. An example of a predicted fusion

- The structure of fusion gene and its basic information (chromosome, transcript, coverage, sequence and break point).
- CircosPlot of fusion gene containing its location on chromosome.
- Retained protein domains and supporting read information associated with fusion gene.
 - If there is no associated protein domain, it is marked as blank.
 - Split reads in [geneA], Split reads in [geneB]: The number of split reads in each gene.
 - Discordant mates: The number of discordant mate reads (spanning reads or bridge reads)

例4

Table 15. Example of Fusion Gene Prediction Results

Sample	MG_CTRL_1	MG_CTRL_1	MG_CTRL_2	MG_CTRL_3	
gene1	AC_TN3	HRX1	ENH1	PKNOX	
gene2	EVE1	RTT1B	L3RC2B14	LOC105170614(ENH1),PKNOX(13417)	
strand1(gene/fusion)	+/-	+/-	-/-	-/-	
strand2(gene/fusion)	+/-	+/-	-/-	-/-	
breakpoint1	1938647907	2248955446	1225507951	1835478282	
breakpoint2	1938540943	1741439952	1229102875	1835456916	
site1	splice-site	intron	splice-site	splice-site	
site2	splice-site	intron	splice-site	intergenic	
type	duplication	translocation(1:5)	duplication	deletion/read-through	
direction1	downstream	downstream	upstream	upstream	
direction2	upstream	upstream	downstream	downstream	
split_read1	15	0	0	0	
split_read2	39	2	1	0	
discordant_mates	7	2	4	4	
coverage1	374	32	469	165	
coverage2	37	4	2	1	
confidence	high	medium	high	low	
filters	duplicate(s),low_entropy(3)	duplicate(s)	duplicate(s)	misassembles(1)	
fusion_transcript	GCGGGAGCTGAGCGCGGAGCGGACAG CTGGTGGCGGAGCGGAGCGGCGGGA TGGTGGACTACACCGCGGAGCGGCG STACCAGTACCGCCCGGAGCGGCGG CAATGGCGCTGGCGGCGGCGGCGAT GGGGCGACTACATGCGCCAGGAGGAG CTGGAGCGGAGCTGCTGGGAGCGG GCGCTGAGGAGGAGCGGAGCGGAGTGT TACCTTTTCATGATGAGCGGCGGCG GCGTTCGGGAGGCGGCGGCGGAGGAT GAGAGCGGCGGCGGCGGCGGAGGAG GCGCTACAGGCGGCGGCGGAGGAGG TCTCTCTTCGATGCTGCTGAGCGGCG CCATCATCAGGCGGCGGCGGAGGAGG CTTTGGTGGAGCGGCGGAGGAGGAG GCGAGTGGAGGAGGAGGAGGAGGAG AGTGTCTTCTCTGTGGAGGAGGAGG TACTTTGATACGAGCGGCGGAGGAG GACTACAGCGGCGGAGGAGGAGGAG GCC	CTCCCACTTGGCGCTCCAAATGTTGG ATTATAGCGCTGAGCGGCGGAGGAG CTGAAATTTTACAGCGGCGGAGGAG GTAAAGAAATAAAGAGGAGGAGGAG TGAAGGAGCAAGATATGCTGCTGAT TAATGAGGAGTTTCTCTCATCTTCCAT CAGCTCTGATGTTGATGAAATTTACTGT AGAGAAATCAGGAGGAGGAGGAGGAG TGAATCCCTCACTCAAGAAATGTTGG	CAACATAGAAACAGCAAAAAGAGGA CAAGGTT_GAAGATTCAGAGCGTGAAC TTCTAAGCGCTCTTCAACAGTACACCT AAGCAGAGAGAAACCTTGGGAGGAG CAATACATTTGAGGAGGAGGAGGAG GTTATCTCAGTCATTGCTCGGAGGAT CTCTGAAAGTCAAGGAGCGGCGGCGG TGGGCGGCGGCGGCGGAGGAGGAGG GATGGCGGCGGCGGCGGAGGAGGAGG AACTGACATCACAGGAGGAGGAGGAG	ACCTGGAAGGAGGAGGAGGAGGAGGAG CTCTGGAAGGAGGAGGAGGAGGAGGAG AAGCGTCCAACTTACTCAGTATTGATG AAGCAGAGAGAAACCTTGGGAGGAG CTCTCCTCCTTAAAGGAGGAGGAGGAG CAATACATTTGAGGAGGAGGAGGAGG GTTATCTCAGTCATTGCTCGGAGGAT AGTGCAGGAGGAGGAGGAGGAGGAGG AGCTCCCTCTCAGGCGTCAAGGAGGAG TGGGCGGCGGCGGAGGAGGAGGAGG TCACTCCTCAGGCGGCGGAGGAGGAG T_GAAGCAGGAGGAGGAGGAGGAGGAG GCTGCATCAATGAGGAGGAGGAGGAG	
reading_frame	in-frame		out-of-frame	out-of-frame	
peptide_sequence	MYTHIANQDQYQYPSAGAGAGGAGGGS MCDYHMQEEDWWRKLLDPAWKKQQRK CYLPHMVGVRAGGCGDEEDPAGDEYEL RYVDFDTFFVVLVLAQGLDAGELRDQQ EQWEDMEKCKRCGSSDYDTPHGHETH ELEEHL			TETQKEDKEDSEPVTSKASSTSPPEPRK PNRKHTMNGKSPVSLQDQWQSPGAPG WQVQVH*	SGLLgtpplpplcsgllahcnlpjggsnphsl*

8. Data Download Information

Raw dataおよび解析結果のダウンロードリンクを記載しています。

Raw data(fastqファイル)は論文投稿時に行うデータベース登録時に、また、弊社でのデータ保管期間経過後に、追加データ解析をご希望の場合に必要なファイルです。

ダウンロードできる期間は約2週間となっておりますので、**必ず**期間中に全てのファイルをダウンロードして下さい。

また、ダウンロードしたfastq.gzファイルはファイルに破損がないかの確認のため、**必ず**”md5sum値”の照合を行ってください。
ご確認をお願いしております。

確認、照合方法につきましては冒頭”CAUTION”をご確認ください。

9. Appendix

参考情報、解析に使用したプログラム、参考文献をまとめています。

Thank You

For your attention



NGS



CES



Clinical
Diagnostics



DTC/PGS



Microarray



Oligo



GEM



CRISPR