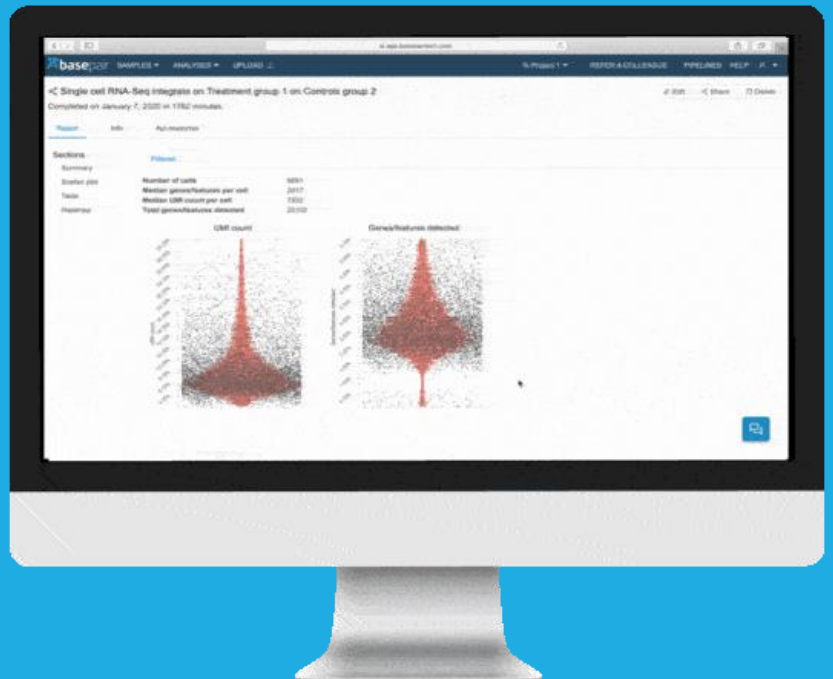
WETもDRYも使える
クラウドバイオインフォマティクス環境



ASTRIDE
薬師寺 秀樹
tech@astride.jp

NGSのデータ解析を行う場合…

NGSのデータ解析を行う場合…

自分でやる

受託会社に依頼する

専用ソフトウェアを使う

NGSのデータ解析を行う場合…

自分でやる

コーディングが…

PCやソフトウェアの設定が…

受託会社に依頼する

前の解析と比較したいのだけど…

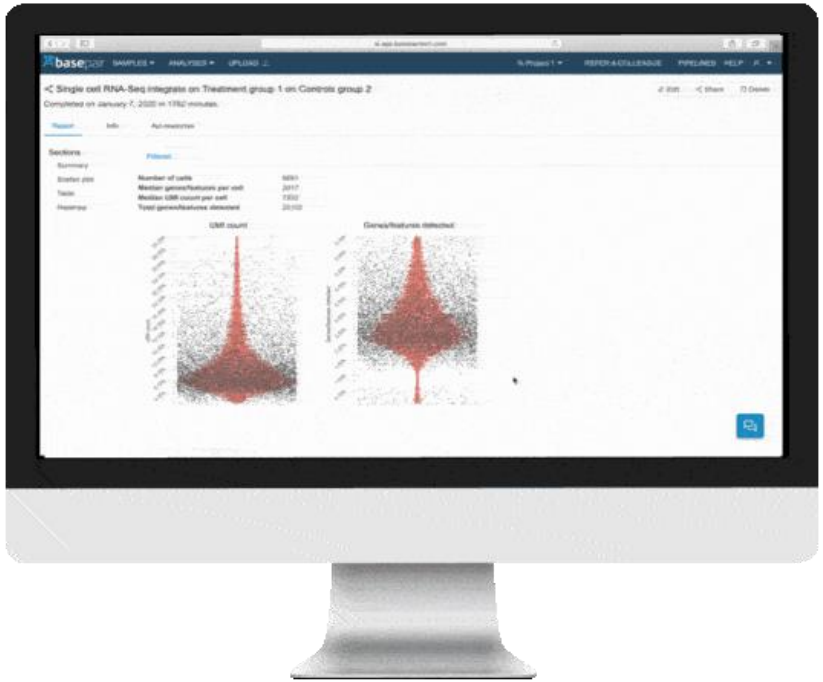
ちょっとした変更なんだけど…

専用ソフトウェアを使う

他の解析ができない…

ライセンスが高い…

basepair


basepair
SAMPLES
ANALYSES
UPLOAD
Single cell RNA-Seq integrate on Treatment group 1 on Controls group 2
Sections
Summary
UMI count
Genes/features detected

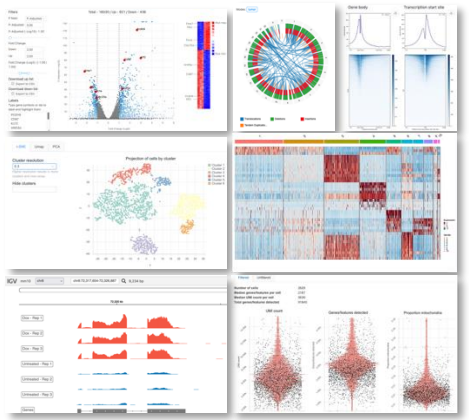
GUIで、QCやアライメント～視覚化まで

Sequencing

RAW Data

FASTA、BAM etc.

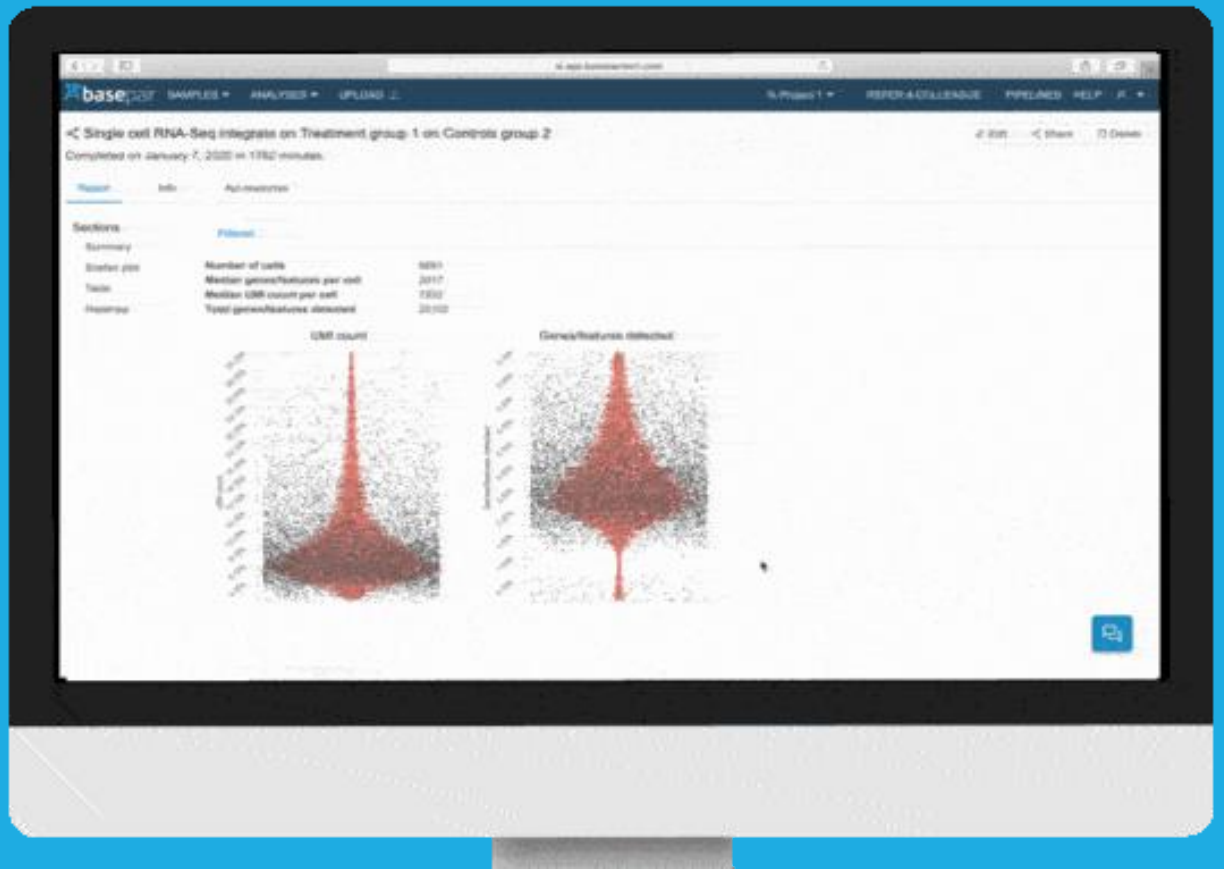
basepair



Analyze

QC, trim, alignment, application related analysis and visualization such as variant call, volcano plot, etc.

マウスで簡単操作（GUI）



対応済みアプリケーション

- ✓ ゲノミクス
  - ✓ WGS, WES, Panels, CRISPR etc.
- ✓ トランスクリプトミクス
  - ✓ bulk and single cell RNA-Seq
- ✓ エピゲノミクス
  - ✓ ATAC-Seq, ChIP-Seq, CUT&RUN, CUT&TAG etc.
- ✓ メタジェノミクス
  - ✓ 16S, Shotgun etc.

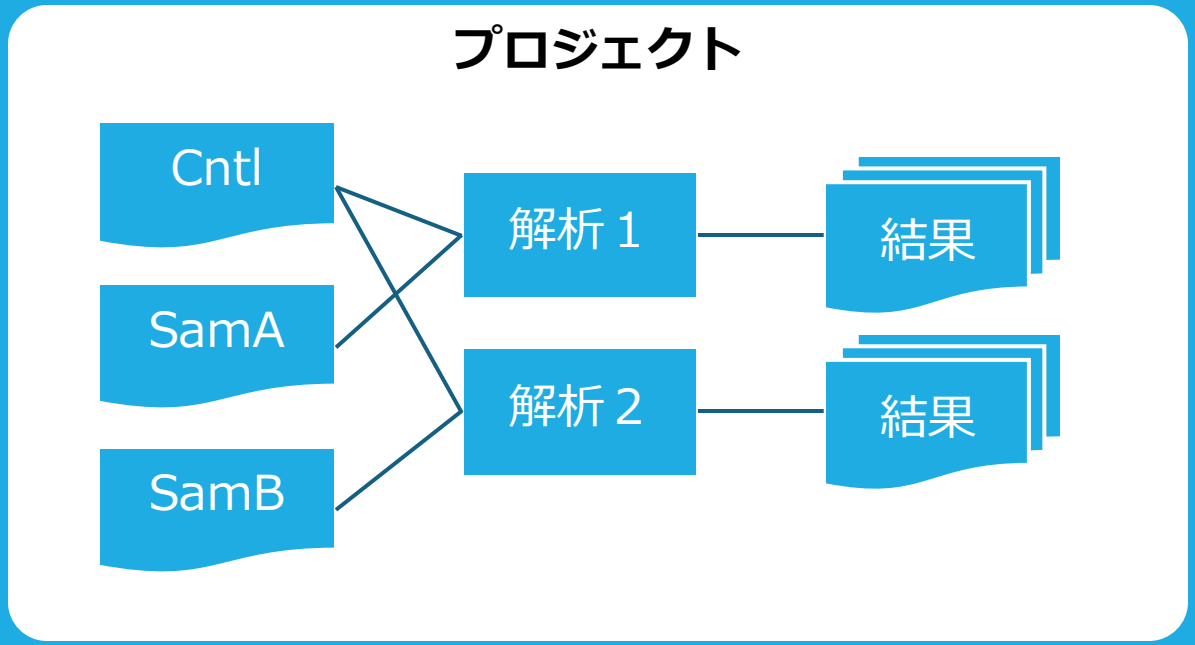
3ステップで解析

1. アップロード

2. 解析

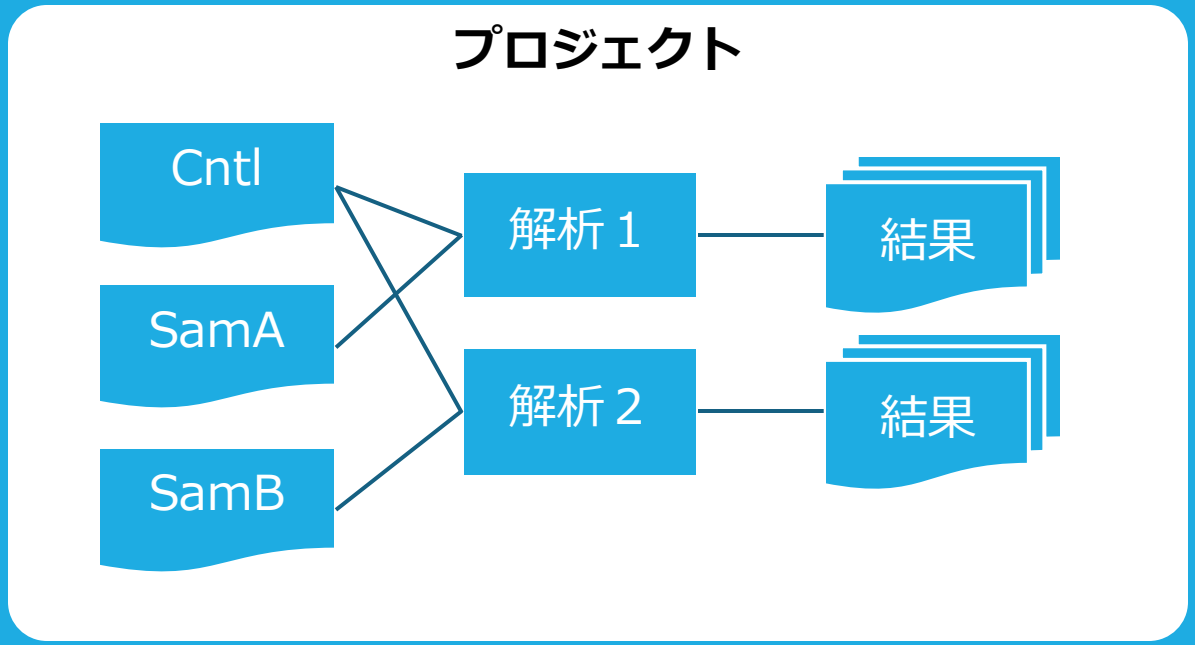
3. 共有

プロジェクト・サンプル・解析

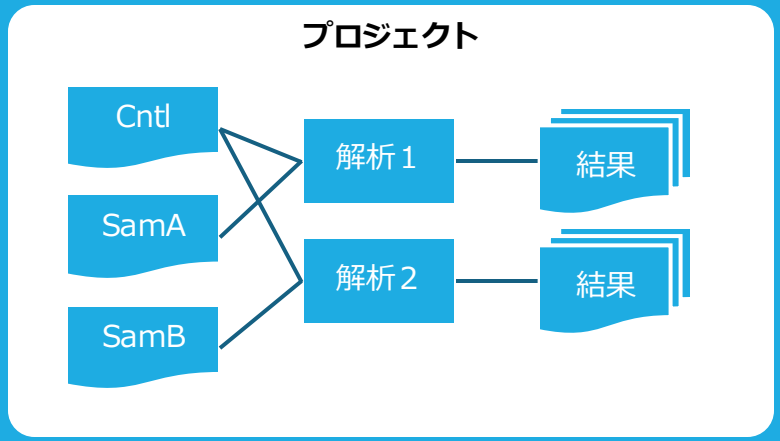


サンプル数により課金・解析数は制限なし

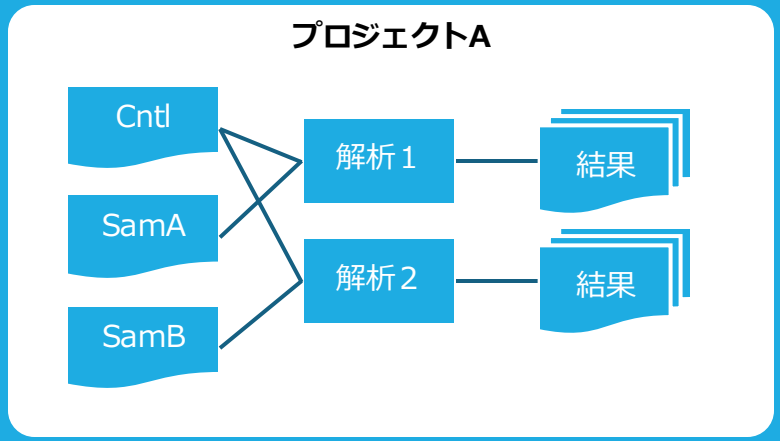
プロジェクト・サンプル・解析

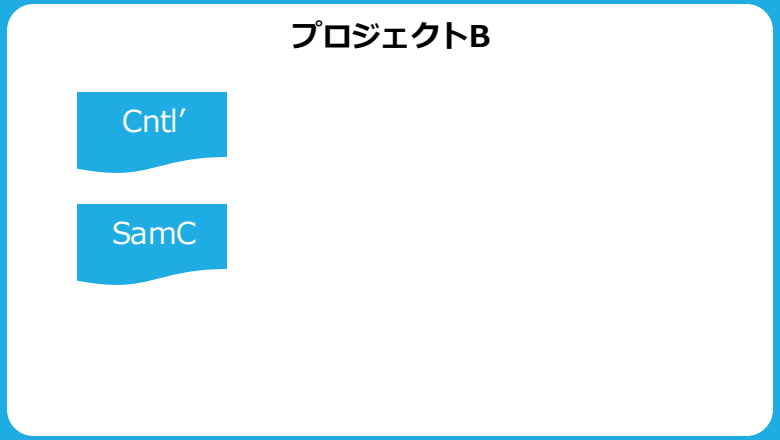


プロジェクト・サンプル・解析



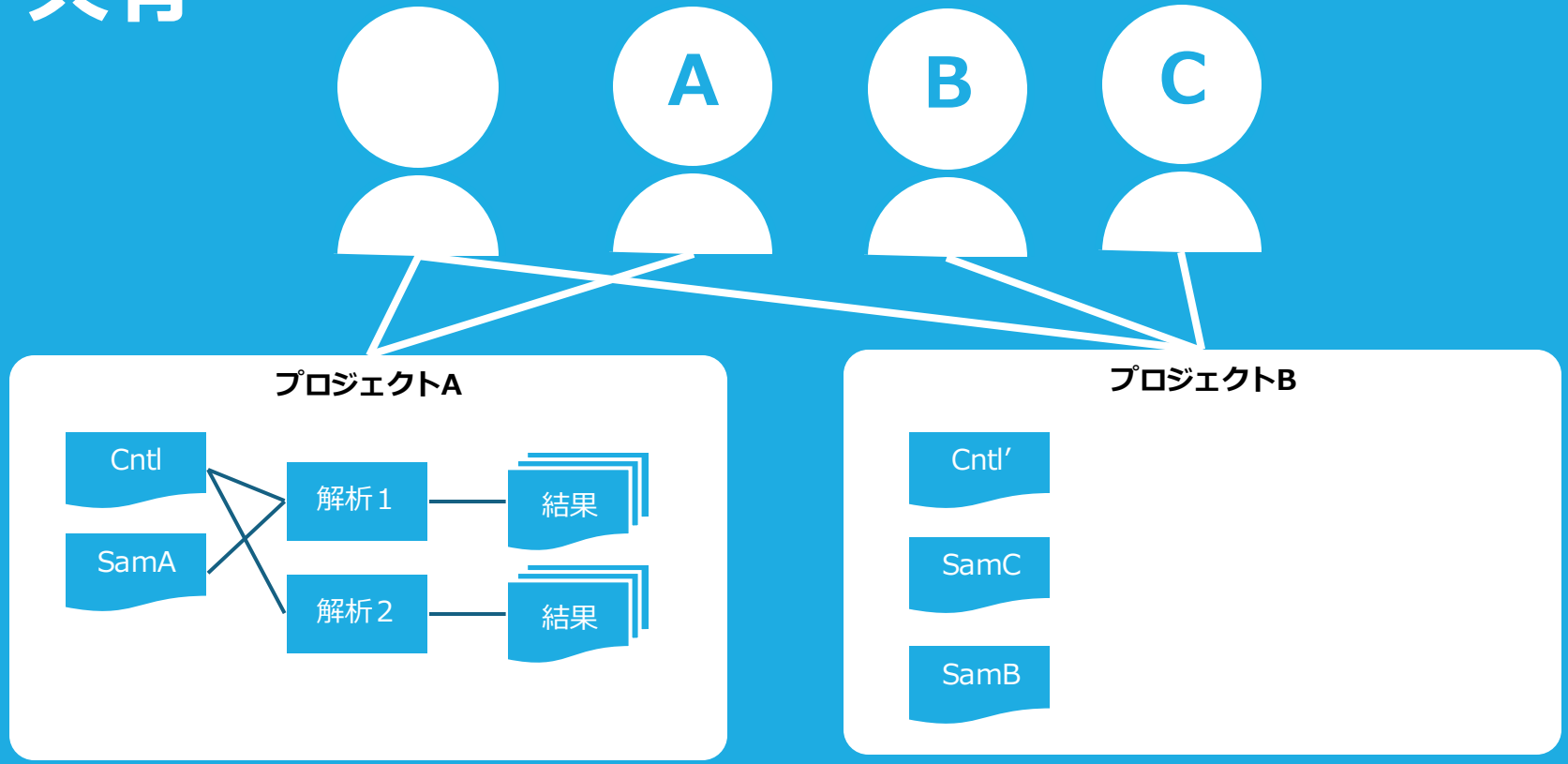
プロジェクト・サンプル・解析





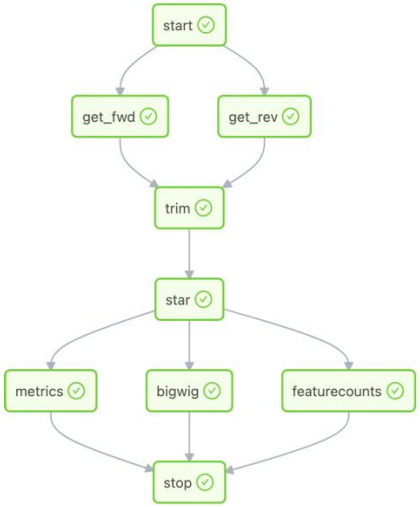
- 複数のプロジェクトを作ることが可能。
- 解析対象は、同じプロジェクト内にある必要がある。
- サンプルのプロジェクト間移動は可能。

共有

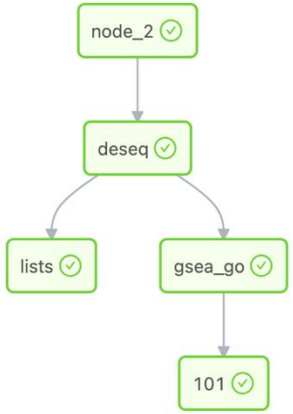


Basepairのパイプライン
（Differential Expressionの場合）

Expressio Count (STAR)

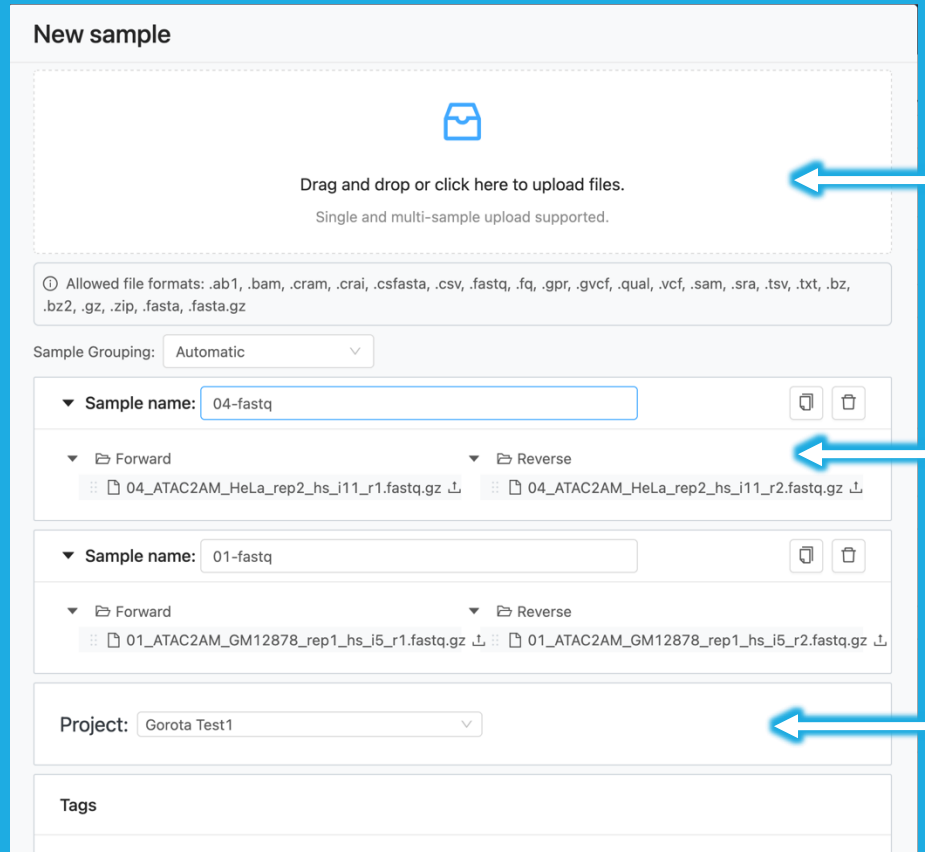


Differential Expression (DESeq2)



FastqからDEGまでに必要なステップを接続

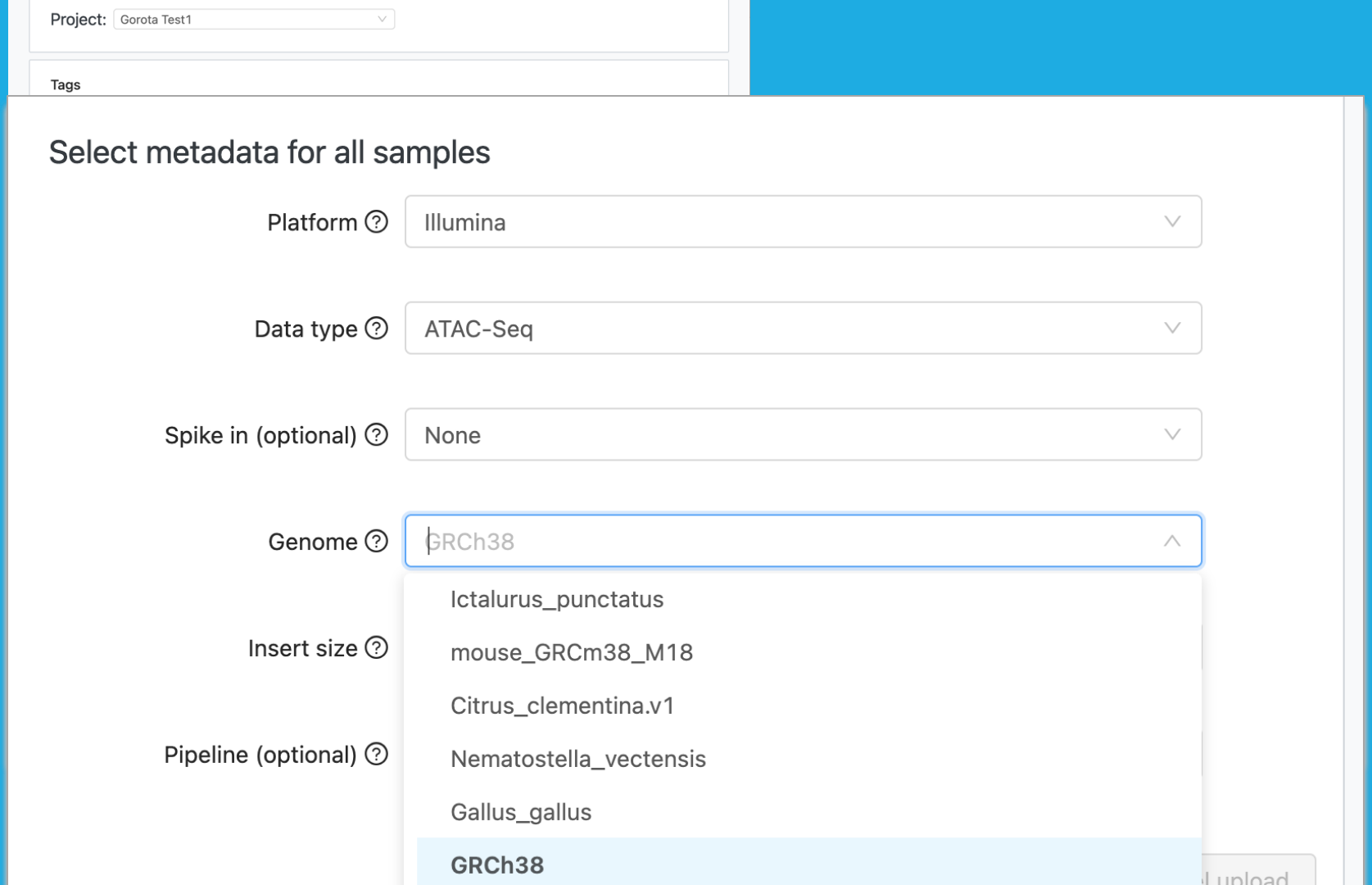
サンプルアップロード



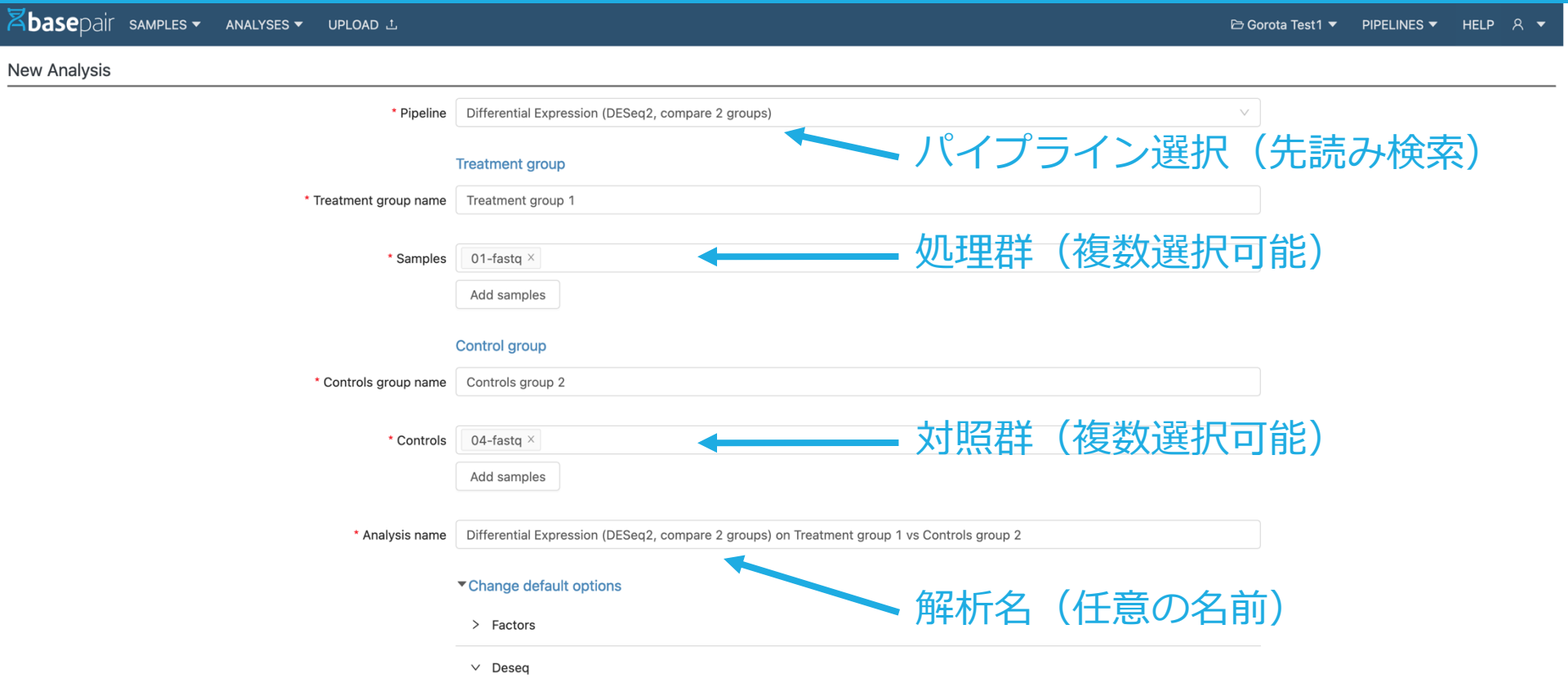
Fastqをドラッグ＆ドロップ
もしくは、ファイルを選択

FW・RV、サンプルグループ
は自動判定

格納する「プロジェクト」を
選択


Project: Gorota Test1
Tags
Select metadata for all samples
Platform
Illumina
Data type
ATAC-Seq
Spike in (optional)
None
Genome
GRCh38
Insert size
Pipeline (optional)
Ictalurus_punctatus
mouse_GRCm38_M18
Citrus_clementina.v1
Nematostella_vectensis
Gallus_gallus
GRCh38

解析（Differential Expressionの場合）



Control group

* Controls group name: Controls group 2

* Controls: 04-fastq ×

Add samples

* Analysis name: Differential Expression (DESeq2, compare 2 groups) on Treatment group 1 vs Controls group 2

Change default options

Factors

Deseq

Alpha: 0.1

Features: genes

Hypothesis: wald

Reducedmodel: ~1

Minimum expression cutoff: 1

pAdjustMethod: BH

Do not use spike-in for normalization: True | False

DESeq2のパラメーター

Run analysis | Cancel

解析開始ボタン

- **抽出条件**
  - cell type : cd4
  - diseasestatus :
    - type1 diabetes
    - healthy control
  - gender : female
- **使用データ**
  - GSM1479441
  - GSM1479457
  - GSM1479463
  - GSM1479502
  - GSM1479509
  - GSM1479523
  - GSM1479543

SAMPLES ANALYSES UPLOAD

Analys

Project:

Import data from GEO

Edit Share Delete Restart

Completed on March 5, 2024 in 120 minutes.

Report Input/Output Execution summary Api response

Sections

Import GEO

Error Output

Import GEO

id	Name	Basepair Id
GSM1479441	lib229	678
GSM1479457	lib245	679
GSM1479463	lib251	680
GSM1479502	lib291	681
GSM1479509	lib298	682
GSM1479523	lib312	685
GSM1479543	lib332	691

Differential Expression (DESeq2, compare 2 groups) on type 1 diabetes vs Healthy Control

Edit Share Delete Restart

Completed on March 5, 2024 in 27 minutes.

Report | Input/Output | Execution summary | Genome Browser | Api response

Sections

- Volcano plot
- Heatmap
- Diff expr genes
- PCA
- Correlation Plot
- GSEA GO
- GSEA Pathway

Volcano Plot

Filters

P field: P-Value

P-Value: 0.0501

P-Value (-Log10): 1.30

Fold Change:

Down 2.00

Up 2.00

Fold Change (Log2): [-1.00 | 1.00]

Download up list

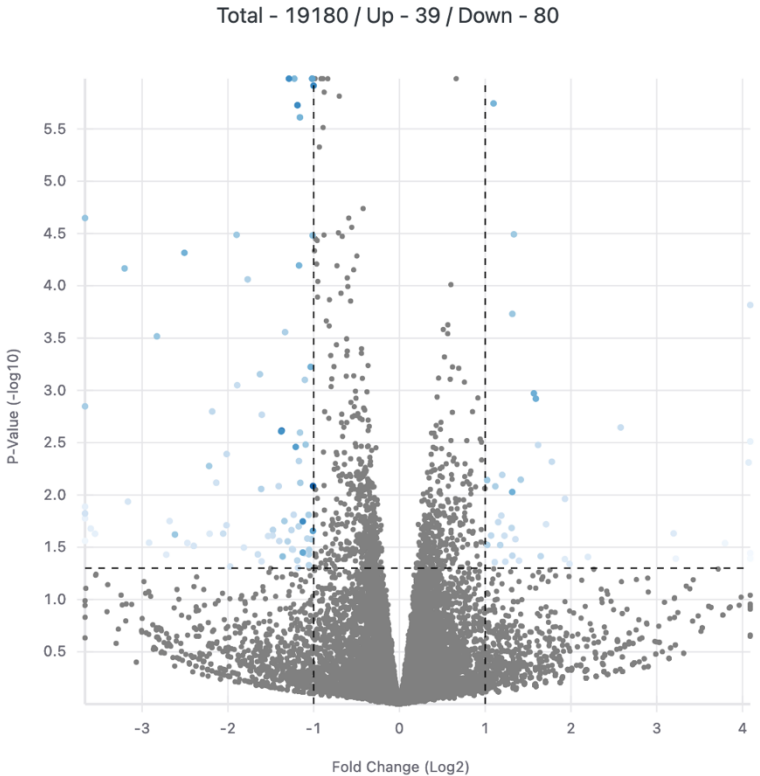
Export to CSV

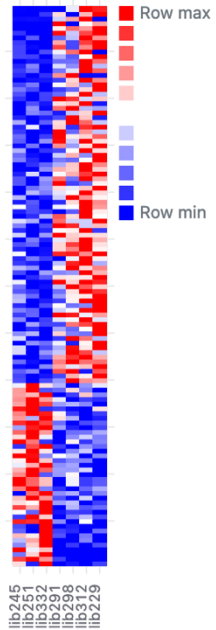
Download down list

Export to CSV

Labels

Type gene symbols or ids to label and highlight them

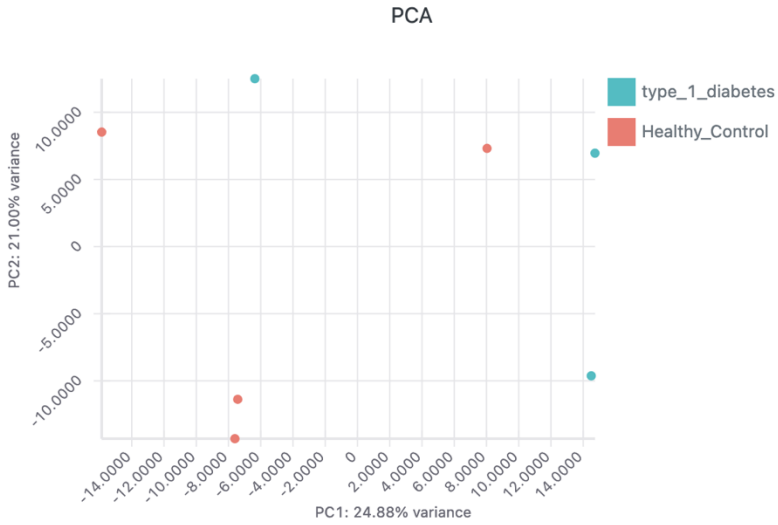




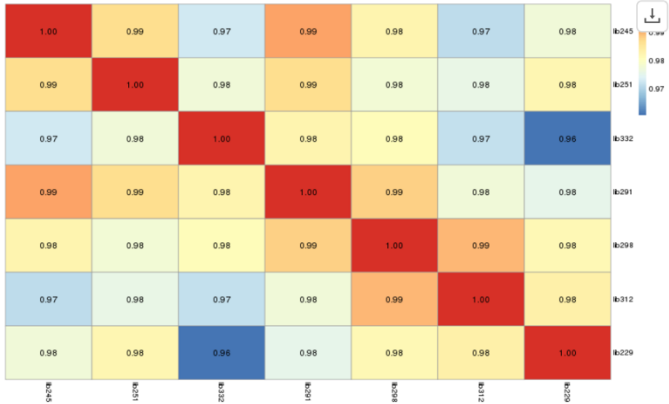
PCA axes

Select the axes you want to plot.

PC1 - PC2



Correlation Plot



GSEA GO

GSEA GO

Download Export to CSV

type_1_diabetes | Healthy_Control

Name	Size	ES	NES	NOM p-val	FDR q-val	FWER p-val	Rank At Max
GOBP_MATURATION_O...	27	0.1268	0.4006	0.9494	0.9996	1.0000	14303
GOBP_MITOCHONDRIA...	157	0.0960	0.3135	0.9494	0.9998	1.0000	6210
HP_FOCAL_T2_HYPER...	32	0.1269	0.3613	0.8562	0.9999	1.0000	4029
GOMF_PROTEIN_DISU...	18	0.1452	0.4054	0.9470	0.9999	1.0000	13802
GOBP_RESPONSE_TO_...	21	0.5945	1.8093	0	1.0000	0.8520	1678
HP_STOMACH_CANCER	20	0.5324	1.7953	0	1.0000	0.8790	1710
GOBP_PREASSEMBLY_...	17	0.6547	1.7618	0	1.0000	0.9360	1437
GOBP_GPI_ANCHOR_M...	31	0.5002	1.7540	0	1.0000	0.9360	2303
GOMF_EXONUCLEASE_...	56	0.3759	1.7183	0	1.0000	1.0000	5746
GOMF_EXORIBONUCLE...	38	0.3536	1.7141	0	1.0000	1.0000	2726
GOBP_RESPONSE_TO_...	15	0.5706	1.6927	0	1.0000	1.0000	3051
GOBP_CYTOSKELETON...	79	0.3263	1.6869	0	1.0000	1.0000	2604
GOBP_MITOTIC_CYTO...	68	0.3407	1.6747	0	1.0000	1.0000	2604
GOCC_CAJAL_BODY	74	0.3755	1.6431	0	1.0000	1.0000	2760
HP_BREAST_CARCINOMA	59	0.3702	1.6337	0	1.0000	1.0000	1678
HP_ABNORMALITY_OF...	15	0.6320	1.6288	0	1.0000	1.0000	3414
HP_RHABDOMYOSARCOMA	21	0.4580	1.6273	0	1.0000	1.0000	2955
GOBP_CELLULAR_RES...	15	0.4852	1.6167	0.0515	1.0000	1.0000	4581
HP_APLASTIC_CLAVICLE	15	0.5400	1.6017	0.1333	1.0000	1.0000	2781
GOBP_BRANCHED_CHA...	22	0.5510	1.5974	0.0568	1.0000	1.0000	6141
GOMF_RNA_POLYMERA...	19	0.4472	1.5923	0	1.0000	1.0000	1752
HP_NEOPLASM_OF_TH...	78	0.3107	1.5799	0	1.0000	1.0000	3029
HP_ABNORMALITY_OF...	32	0.4929	1.5716	0	1.0000	1.0000	249
GOBP_CELLULAR_RES...	36	0.3768	1.5636	0	1.0000	1.0000	2006
GOBP_RNA_PHOSPHOD...	41	0.3527	1.5633	0	1.0000	1.0000	5746

GSEA Pathway

Download Export to CSV

type_1_diabetes | Healthy_Control

Name	Size	ES	NES	NOM p-val	FDR q-val	FWER p-val	Rank At Max
REACTOME_SUMOYLAT...	44	0.4459	1.8516	0.0584	0.4868	0.3220	4813
REACTOME_SYNTHESI...	18	0.6531	1.7351	0	0.7765	0.6050	1437
REACTOME_G_BETA_G...	19	0.1666	0.4836	0.8848	0.9852	1.0000	4665
REACTOME_PROCESSI...	234	0.0976	0.4588	0.8475	0.9862	1.0000	7046
REACTOME_ADP_SIGN...	19	0.1682	0.4931	0.9339	0.9865	1.0000	4665
REACTOME_APC_C_CD...	73	0.1523	0.5282	0.8578	0.9894	1.0000	6266
REACTOME_NRIF_SIG...	15	0.1692	0.4942	0.9460	0.9902	1.0000	6139
REACTOME_HOST_INT...	125	0.0954	0.5017	0.9365	0.9917	1.0000	14556
REACTOME_RNA_POLY...	40	0.1040	0.5315	0.9410	0.9921	1.0000	6070
REACTOME_REGULATI...	67	0.1366	0.5088	0.8191	0.9929	1.0000	662
REACTOME_EPH_EPHR...	47	0.1646	0.5342	0.9244	0.9948	1.0000	5630
REACTOME_PROCESSI...	15	0.2200	0.5388	0.8337	0.9969	1.0000	15000
REACTOME_WNT5A_DE...	15	0.1570	0.5413	0.8668	0.9997	1.0000	6679
REACTOME_G0_AND_E...	26	0.6119	1.6447	0.0643	1.0000	0.8120	4440
REACTOME_HDR_THRO...	36	0.4844	1.6330	0	1.0000	0.8120	5415
REACTOME_TRANSCRI...	19	0.5634	1.5808	0.1255	1.0000	0.8710	4440
REACTOME_INTERACT...	34	0.3207	1.5598	0.1086	1.0000	1.0000	6502
REACTOME_BRANCHED...	21	0.5496	1.5302	0.1201	1.0000	1.0000	5725
REACTOME_G1_S_SPE...	28	0.5967	1.5259	0.0704	1.0000	1.0000	4440
REACTOME_APOPTOTI...	32	0.4033	1.5107	0.0638	1.0000	1.0000	1410
REACTOME_ACTIVATI...	37	0.5205	1.5060	0.0714	1.0000	1.0000	4359
REACTOME_TRANSCRI...	34	0.4299	1.4971	0.1250	1.0000	1.0000	7887
REACTOME_MEIOTIC_...	22	0.4973	1.4681	0	1.0000	1.0000	2977
REACTOME_APOPTOTI...	40	0.3761	1.4655	0.0680	1.0000	1.0000	2558
REACTOME_TRANSCRI...	16	0.5734	1.4369	0.0717	1.0000	1.0000	4440

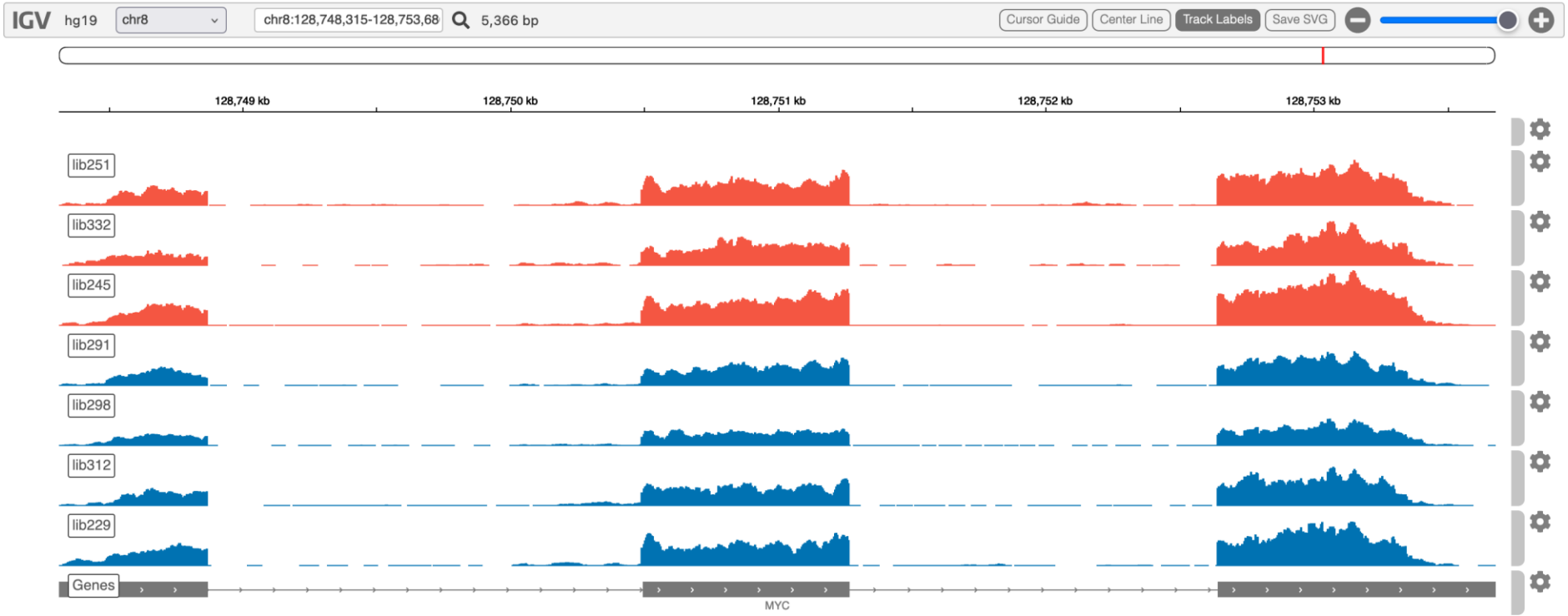
Differential Expression (DESeq2, compare 2 groups) on type 1 diabetes vs Healthy Control

Edit Share Delete Restart

Completed on March 5, 2024 in 27 minutes.

Report | Input/Output | Execution summary | **Genome Browser** | Api response

Equalize Scale | Add track sample



Differential Expression (DESeq2, compare 2 groups) on type 1 diabetes vs Healthy Control

Edit Share Delete Restart

Completed on March 5, 2024 in 27 minutes.

Report | Input/Output | Execution summary | Genome Browser | Api response

Pipeline Differential Expression (DESeq2, compare 2 groups)

Projects DRY解析教本_RNA-Seq

Samples
lib251
lib332
lib245

Controls
lib291
lib298
lib312
lib229

Files

Show hidden

deseq
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.cls
- deseq/corr_plot.png
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.diffexpr.w_symbols.txt
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.norm.gct
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.pca-loadings.txt
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.pca-rotation.txt
- deseq/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.vst.txt

Create Lists
- lists/Diffexpr.min_count_10.pval_0.05.all_ids.txt
- lists/Diffexpr.min_count_10.pval_0.05.down.txt
- lists/Diffexpr.min_count_10.pval_0.05.up.txt

GSEA
- gsea_go/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.GO.Gsea.zip
- gsea_go/Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.pathway.gsea.zip

GSEA Report for Dataset filtered

Enrichment in phenotype: type_1_diabetes (3 samples)

- 1989 / 7514 gene sets are upregulated in phenotype **type_1_diabet**
- 0 gene sets are significant at FDR < 25%
- 81 gene sets are significantly enriched at nominal pvalue < 1%
- 90 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results ← スナップショット
- Detailed enrichment results in html format ← リスト
- Detailed enrichment results in TSV format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: Healthy_Control (4 samples)

- 5525 / 7514 gene sets are upregulated in phenotype **Healthy_Contr**
- 0 gene sets are significantly enriched at FDR < 25%
- 302 gene sets are significantly enriched at nominal pvalue < 1%

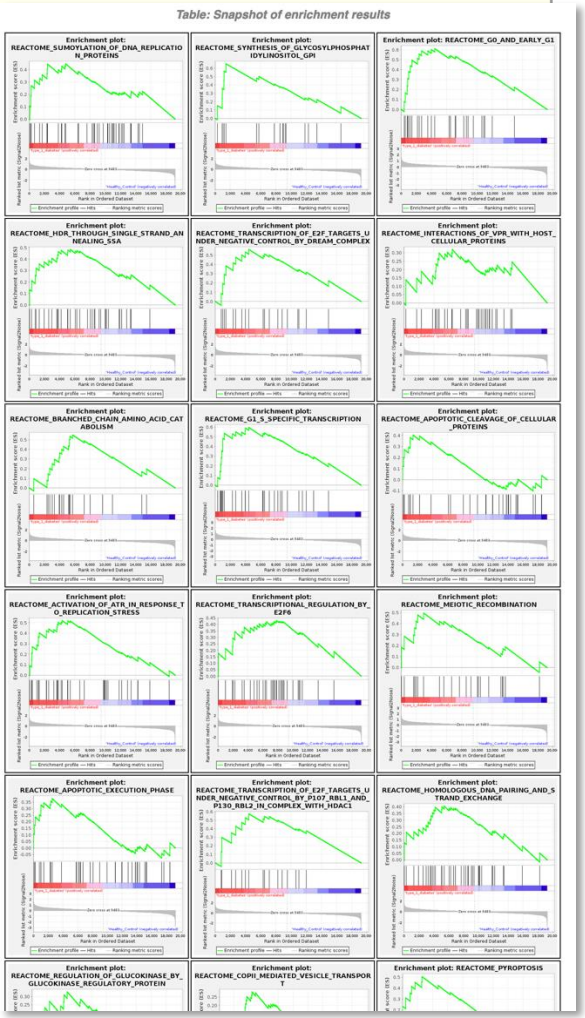


Table: Gene sets enriched in phenotype type_1_diabetes (3 samples) [plain text format]

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
1	REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS	Details ...	44	0.45	1.85	0.058	0.487	0.322	4813
2	REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	Details ...	18	0.65	1.74	0.000	0.777	0.605	1437
3	REACTOME_G0_AND_EARLY_G1	Details ...	26	0.61	1.64	0.064	1.000	0.812	4440
4	REACTOME_HDR_THROUGH_SINGLE_STRAND_ANNEALING_SSA	Details ...	36	0.48	1.63	0.000	1.000	0.812	5415
5	REACTOME_TRANSCRIPTION_OF_E2F_TARGETS_UNDER_NEGATIVE_CONTROL_BY_DREAM_COMPLEX	Details ...	19	0.56	1.58	0.125	1.000	0.871	4440
6	REACTOME_INTERACTIONS_OF_VPR_WITH_HOST_CELLULAR_PROTEINS	Details ...	34	0.32	1.56	0.109	1.000	1.000	6502
7	REACTOME_BRANCHED_CHAIN_AMINO_ACID_CATABOLISM	Details ...	21	0.55	1.53	0.120	1.000	1.000	5725
8	REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	Details ...	28	0.60	1.53	0.070	1.000	1.000	4440
9	REACTOME_APOPTOTIC_CLEAVAGE_OF_CELLULAR_PROTEINS	Details ...	32	0.40	1.51	0.064	1.000	1.000	1410
10	REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	Details ...	37	0.52	1.51	0.071	1.000	1.000	4359
11	REACTOME_TRANSCRIPTIONAL_REGULATION_BY_E2F6	Details ...	34	0.43	1.50	0.125	1.000	1.000	7887
12	REACTOME_MEIOTIC_RECOMBINATION	Details ...	22	0.50	1.47	0.000	1.000	1.000	2977
13	REACTOME_APOPTOTIC_EXECUTION_PHASE	Details ...	40	0.38	1.47	0.068	1.000	1.000	2558
14	REACTOME_TRANSCRIPTION_OF_E2F_TARGETS_UNDER_NEGATIVE_CONTROL_BY_P107_RBL1_AND_P130_RBL2_IN_COMPLEX_WITH_HDAC1	Details ...	16	0.57	1.44	0.072	1.000	1.000	4440
15	REACTOME_HOMOLOGOUS_DNA_PAIRING_AND_STRAND_EXCHANGE	Details ...	41	0.40	1.43	0.178	1.000	1.000	5415
16	REACTOME_REGULATION_OF_GLUCOKINASE_BY_GLUCOKINASE_REGULATORY_PROTEIN	Details ...	30	0.33	1.42	0.000	1.000	1.000	4727
17	REACTOME_COPII_MEDIATED_VESICLE_TRANSPORT	Details ...	65	0.28	1.41	0.064	1.000	1.000	4430
18	REACTOME_PYROPTOSIS	Details ...	23	0.51	1.41	0.107	1.000	1.000	2604
19	REACTOME_SHC_MEDIATED_CASCADE_FGFR4	Details ...	15	0.53	1.39	0.039	1.000	1.000	3086
20	REACTOME_REGULATION_OF_TP53_ACTIVITY_THROUGH_METHYLATION	Details ...	18	0.40	1.39	0.000	1.000	1.000	2955
21	REACTOME_RESOLUTION_OF_D_LOOP_STRUCTURES_THROUGH_SYNTHESIS_DEPENDENT_STRAND_ANNEALING_SDSA	Details ...	25	0.42	1.37	0.128	1.000	1.000	4040
22	REACTOME_INTERACTIONS_OF_REV_WITH_HOST_CELLULAR_PROTEINS	Details ...	35	0.27	1.36	0.212	1.000	1.000	4962
23	REACTOME_RESOLUTION_OF_D_LOOP_STRUCTURES	Details ...	33	0.42	1.36	0.063	1.000	1.000	4040
24	REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	Details ...	33	0.50	1.35	0.129	1.000	1.000	4359
25	REACTOME_GLYOXYLATE_METABOLISM_AND_GLYCINE_DEGRADATION	Details ...	25	0.42	1.34	0.162	1.000	1.000	5725
26	REACTOME_FRS_MEDIATED_FGFR3_SIGNALING	Details ...	16	0.47	1.34	0.167	1.000	1.000	2883
27	REACTOME_CLASS_I_PEROXISOMAL_MEMBRANE_PROTEIN_IMPORT	Details ...	20	0.39	1.34	0.097	1.000	1.000	981
28	REACTOME_NUCLEAR_IMPORT_OF_REV_PROTEIN	Details ...	32	0.29	1.33	0.223	1.000	1.000	4727
29	REACTOME_FRS_MEDIATED_FGFR4_SIGNALING	Details ...	17	0.50	1.33	0.041	1.000	1.000	3086
30	REACTOME_POLO_LIKE_KINASE_MEDIATED_EVENTS	Details ...	16	0.54	1.31	0.203	1.000	1.000	3526
31	REACTOME_REGULATION_OF_TP53_ACTIVITY_THROUGH_PHOSPHORYLATION	Details ...	88	0.28	1.30	0.134	1.000	1.000	3842
32	REACTOME_SUMOYLATION_OF_SUMOYLATION_PROTEINS	Details ...	33	0.27	1.30	0.224	1.000	1.000	4727
33	REACTOME_INTERCONVERSION_OF_NUCLEOTIDE_DI_AND_TRIPHOSPHATES	Details ...	29	0.38	1.29	0.174	1.000	1.000	3888
34	REACTOME_BUTYRATE_RESPONSE_FACTOR_1_BRF1_BINDS_AND_DESTABILIZES_MRNA	Details ...	17	0.36	1.28	0.181	1.000	1.000	6202
35	REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION	Details ...	21	0.39	1.28	0.117	1.000	1.000	881
36	REACTOME_HYALURONAN_METABOLISM	Details ...	16	0.47	1.28	0.185	1.000	1.000	3525
37	REACTOME_TRANSPORT_OF_VITAMINS_NUCLEOSIDES_AND_RELATED_MOLECULES	Details ...	35	0.33	1.27	0.056	1.000	1.000	3163
38	REACTOME_PHASE_0_RAPID_DEPOLARISATION	Details ...	23	0.40	1.27	0.058	1.000	1.000	3556
39	REACTOME_ERCC6_CSB_AND_EHMT2_G9A_POSITIVELY_REGULATE_RRNA_EXPRESSION	Details ...	16	0.43	1.26	0.194	1.000	1.000	6159
40	REACTOME_CELL_CELL_JUNCTION_ORGANIZATION	Details ...	40	0.37	1.25	0.104	1.000	1.000	1842
41	REACTOME_POSTMITOTIC_NUCLEAR_PORE_COMPLEX_NPC_REFORMATION	Details ...	26	0.32	1.24	0.261	1.000	1.000	4727
42	REACTOME_GAMMA_CARBOXYLATION_HYPUSINE_FORMATION_AND_ARYLSULFATASE_ACTIVATION	Details ...	37	0.31	1.24	0.078	1.000	1.000	3213
43	REACTOME_COBALAMIN_CBL_VITAMIN_B12_TRANSPORT_AND_METABOLISM	Details ...	18	0.40	1.23	0.171	1.000	1.000	1197
44	REACTOME_FRS_MEDIATED_FGFR2_SIGNALING	Details ...	18	0.43	1.22	0.199	1.000	1.000	2883
45	REACTOME_BASE_EXCISION_REPAIR_AP_SITE_FORMATION	Details ...	17	0.37	1.22	0.205	1.000	1.000	4649

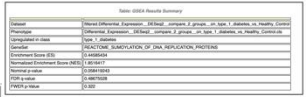


Table: GSEA Results Summary

Dataset	Filtered_Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control
Phenotype	Differential_Expression__DESeq2__compare_2_groups__on_type_1_diabetes_vs_Healthy_Control.cls
Upregulated in class	type_1_diabetes
GeneSet	REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS
Enrichment Score (ES)	0.44585434
Normalized Enrichment Score (NES)	1.8515417
Nominal p-value	0.05841840
FDR q-value	0.48675326
FWER p-value	0.322

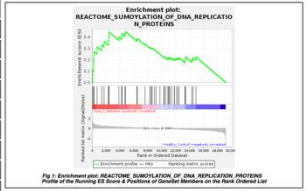


Fig 1: Enrichment plot: REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

Table: GSEA details [plain text format]

1	CDC48	CDCA8	50	1.139	0.1892	Yes
2	BIRC5	BIRC5	223	0.914	0.1884	Yes
3	NUP43	NUP43	276	0.868	0.2704	Yes
4	AURKA	AURKA	714	0.675	0.3134	Yes
5	PAK2	PAK2	1389	0.517	0.3286	Yes
6	NDC1	NDC1	1723	0.481	0.3581	Yes
7	DEHI1	DEHI1	2354	0.421	0.3663	Yes
8	TOP2A	TOP2A	2389	0.418	0.4052	Yes
9	AURKB	AURKB	2452	0.414	0.4436	Yes
10	NUP37	NUP37	3863	0.264	0.3885	Yes
11	NUP107	NUP107	4292	0.269	0.4036	Yes
12	AAK1	AAK1	4308	0.265	0.4050	Yes
13	POM121	POM121	4727	0.237	0.4077	Yes
14	TOP2B	TOP2B	4815	0.232	0.4493	Yes
15	NUP188	NUP188	6409	0.142	0.3793	No
16	NUP50	NUP50	6932	0.140	0.3854	No
17	RAE1	RAE1	7810	0.095	0.3981	No
18	NUP160	NUP160	8206	0.095	0.3937	No
19	NUP133	NUP133	8465	0.048	0.3922	No
20	NUP85	NUP85	8609	0.041	0.2992	No
21	TPR	TPR	8626	0.040	0.3013	No
22	TOP1	TOP1	8981	0.022	0.2849	No
23	PAK4	PAK4	9207	0.011	0.2733	No
24	POM121C	POM121C	9758	-0.012	0.2468	No
25	UBE2I	UBE2I	10020	-0.035	0.2208	No
26	NUP205	NUP205	10370	-0.037	0.2020	No
27	NUP133	NUP133	10394	-0.048	0.2131	No
28	NUP54	NUP54	10917	-0.082	0.2043	No
29	NUP98	NUP98	10930	-0.083	0.2004	No
30	NUP98	NUP98	10947	-0.083	0.2152	No
31	INCENP	INCENP	11068	-0.070	0.2141	No
32	NUP214	NUP214	11181	-0.074	0.2166	No
33	NUP62	NUP62	11441	-0.097	0.2120	No
34	SUMO1	SUMO1	11653	-0.108	0.2000	No
35	NUP133	NUP133	11956	-0.127	0.2113	No
36	HANGAR1	RANGAP1	12136	-0.133	0.2077	No
37	NUP35	NUP35	12339	-0.133	0.2111	No
38	SEC13	SEC13	12507	-0.142	0.2162	No
39	SUMO3	SUMO3	13172	-0.189	0.1896	No
40	SUMO2	SUMO2	13398	-0.191	0.2071	No
41	NUP210	NUP210	13434	-0.193	0.2039	No
42	TNNBP2	TNNBP2	14041	-0.248	0.2009	No
43	NUP93	NUP93	14508	-0.269	0.2174	No
44	PGN	PGN	14605	-0.292	0.2237	No

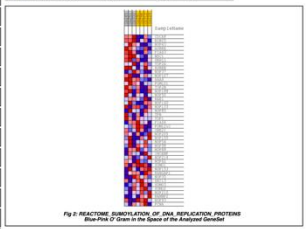


Fig 2: REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS
Blue-Pink O' Gram in the Space of the Analyzed GeneSet

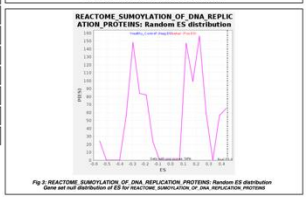


Fig 3: REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS: Random ES distribution
Gene set null distribution of ES for REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS

Expression count (STAR) on lib245

Edit Share Delete Restart

Completed on March 5, 2024 in 28 minutes.

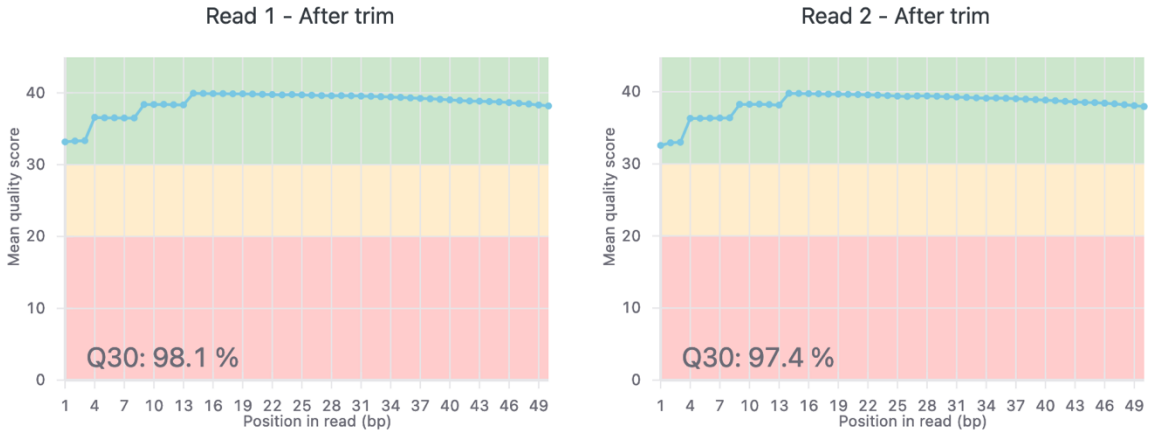
Report | Input/Output | Execution summary | Api response

Sections

- Quality scores
- Number of reads
- Metrics
- Read counts
- Genome browser

Quality scores

Plot: Before trim | After trim

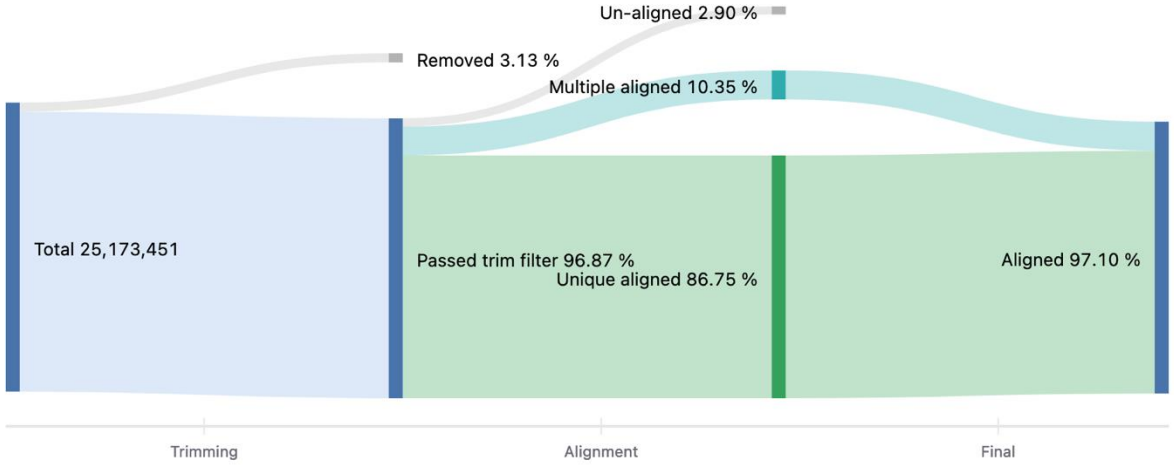


This plot gives you a brief overview of sequence quality along all sequences for your sample. The y-axis in this plots, shows the quality scores.
The higher the score, the better the base call. The background of the graph divides the y-axis into very good quality scores (green), scores of reasonable quality (orange), and reads of poor quality (red).
The numbers in the y-axis represent the **probability of error**. A quality score of 10 means that there's a 10% chance that the base is incorrect, while 40 means that there's only a 0.01% probability of error. Remember, we are working with logs here.

Sections
Quality scores
Number of reads
Metrics
Read counts
Genome browser

incorrect, while 40 means that there's only a 0.01% probability of error. Remember, we are working with logs here.
The sample here shows **96.88%** reads having quality score of 30 for R1 read & **95.48%** for R2 reads before filtering & reads after filtering **98.08%** reads having quality score of 30 for R1 reads & **97.36%** for R2 reads.
Here we observed the number of reads before filtering **50,346,902** & after quality filtering & trimming the number reads is **48,768,672**.

Number of reads



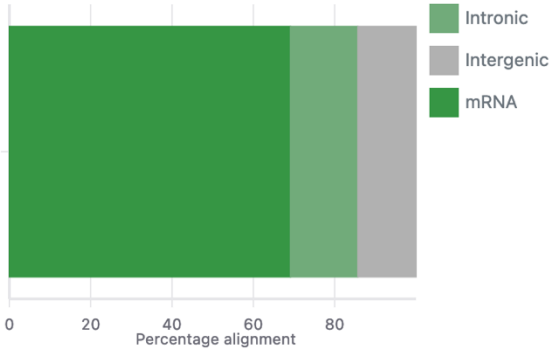
An important mapping quality parameter is the **percentage of mapped reads**, which is a global indicator of the overall sequencing accuracy and of the presence of contaminating DNA. We expect between 70% to 90% of regular RNA-seq reads to map onto reference genome, with a significant fraction of reads mapping to a limited number of identical regions equally well **('multi-mapping reads')**.
A total of **24,384,336** input reads were used for alignment, out of which, **21,154,149 (86%)** reads aligned uniquely.
The alignment seems to be good.

Metrics

Sections
Quality scores
Number of reads
Metrics
Read counts
Genome browser

An important mapping quality parameter is the **percentage of mapped reads**, which is a global indicator of the overall sequencing accuracy and of the presence of contaminating DNA. We expect between 70% to 90% of regular RNA-seq reads to map onto reference genome, with a significant fraction of reads mapping to a limited number of identical regions equally well **('multi-mapping reads')**.
A total of **24,384,336** input reads were used for alignment, out of which, **21,154,149 (86%)** reads aligned uniquely.
The alignment seems to be good.

Metrics



This section reports how many alignments fall into **mRNA**, **intronic** and **intergenic regions**. Even if you have high genomic mapping rate for your sample, check to see where the reads are mapping.
Expect a high proportion of reads mapping to exonic regions **(> 60%)** and lower intronic mapping rates **(20 -30%)**.
Here we observed **69.12%** reads aligned to coding (mRNA) region, **14.29%** to intergenic regions & **16.59%** to intronic regions.

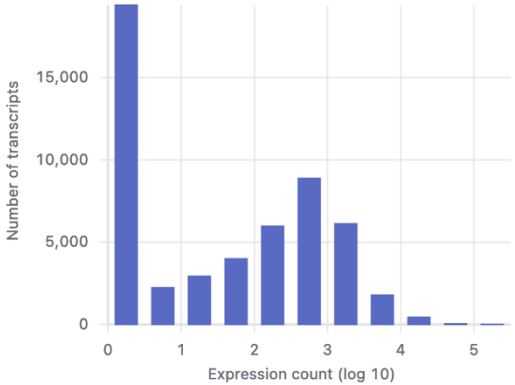
Read counts

Sections
Quality scores
Number of reads
Metrics
Read counts
Genome browser

This section reports how many alignments fall into **mRNA**, **intronic** and **intergenic regions**. Even if you have high genomic mapping rate for your sample, check to see where the reads are mapping.
Expect a high proportion of reads mapping to exonic regions **(> 60%)** and lower intronic mapping rates **(20 -30%)**.
Here we observed **69.12%** reads aligned to coding (mRNA) region, **14.29%** to intergenic regions & **16.59%** to intronic regions.

Read counts

Count distribution: By transcript | By gene



Export to CSV

Id	Symbol	Count	FPKM	TPM	
NM_001402	EEF1A1	178,665	1,410.36	3,149.87	view
NM_001286272	TPT1	105,763	609.08	1,360.30	view
NM_003295	TPT1	105,739	630.55	1,408.26	view
NM_001286273	TPT1	105,654	640.24	1,429.89	view
NM_004048	B2M	100,748	2,829.86	6,320.14	view
NM_000967	RPL3	67,304	1,394.54	3,114.53	view
NM_001033853	RPL3	66,579	1,549.79	3,461.25	view
NM_001961	EEF2	64,597	567.08	1,266.51	view
NR_046235	RNA45S5	63,310	131.40	293.47	view
NM_012423	RPL13A	61,363	1,441.69	3,219.82	view

< 1 2 3 4 5 ··· 5178 >

Sections

- Quality scores
- Number of reads
- Metrics
- Read counts
- Genome browser

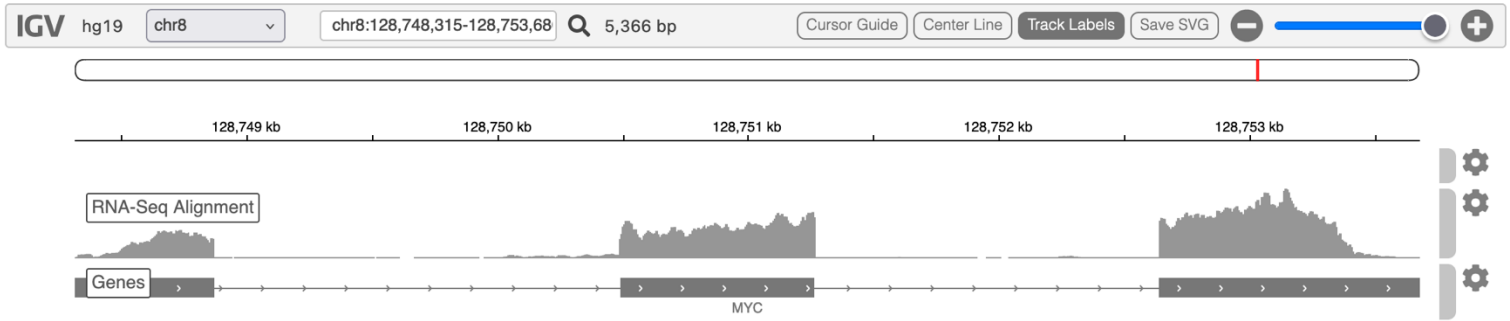
Expression count (log 10)

NM_001961	EEF2	64,597	567.08	1,266.51	view
NR_046235	RNA45S5	63,310	131.40	293.47	view
NM_012423	RPL13A	61,363	1,441.69	3,219.82	view

< 1 2 3 4 5 ... 5178 >

Genome browser

Equalize Scale | Add track sample



所用時間比較

	自分で解析する※	Basepairを使う
所要時間	作業日数：9日 作業時間：13時間40分 サンプルファイル準備：約6日	ダウンロード：120分 STAR：平均24.8分 DESeq2：27分 実作業時間：10分
完了までの日数	30日	1日

※「次世代シークエンサーDRY解析教本」から算出

セットアップ（自分で解析する場合）

Mac Miniの場合	MacBook Pro 14inchの場合
• 10コアCPU、16コアGPU、16コアNeural Engine搭載Apple M2 Proチップ • 32GBユニファイドメモリ • 1TB SSDストレージ	• 12コアCPU、18コアGPU、16コアNeural Engine搭載Apple M3 Proチップ • 36GBユニファイドメモリ • 1TB SSDストレージ
274,800円（税込） キーボード、マウス、ディスプレイは別途	458,800円（税込）

※「次世代シークエンサーDRY解析教本」を参考に独自に選定

必要なパッケージ

- R：統計やグラフを描く
- Homebrew：パッケージ管理プログラム
- wget：ダウンロードするためのプログラム
- Bioconda：Pythonのパッケージ
- SRA Toolkit：SRAが提供するツール

API

他アプリケーションから操作
- サンプル、解析、結果
- アップロード、ダウンロード、実行、削除

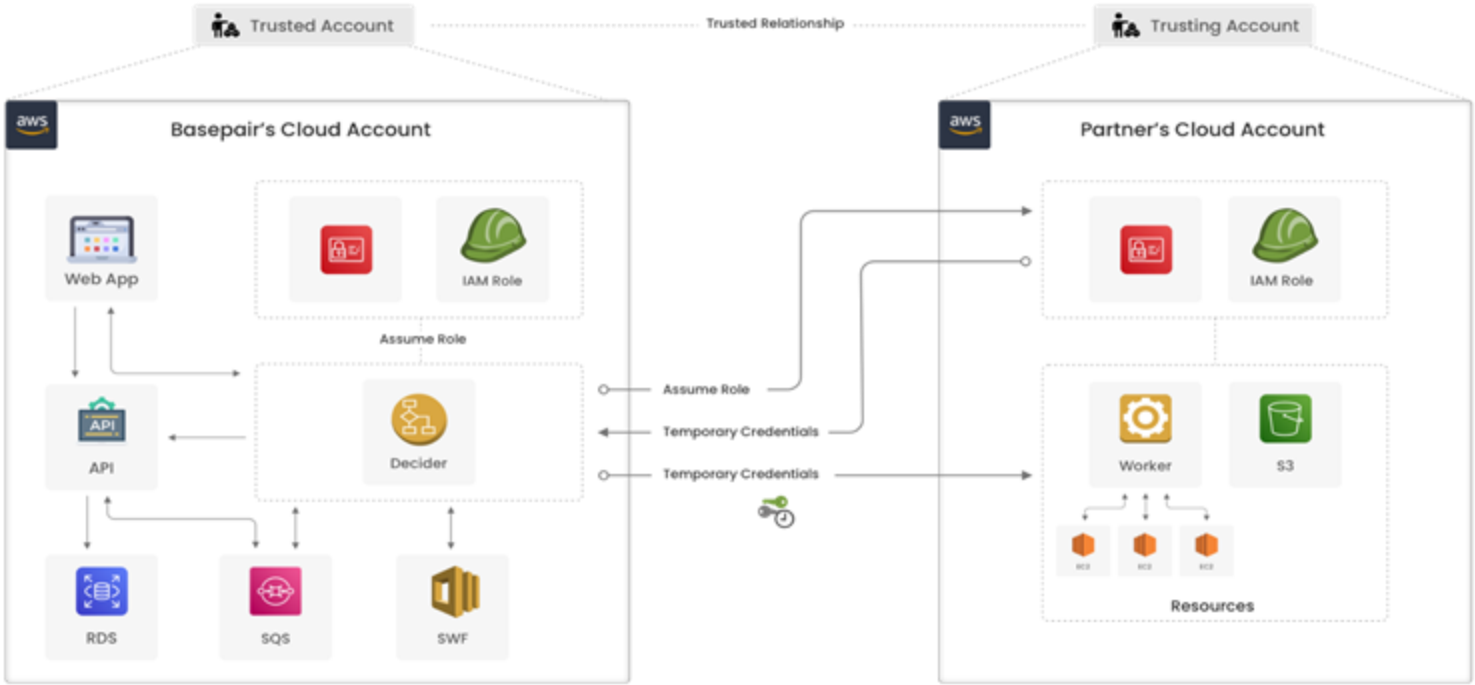
REST API

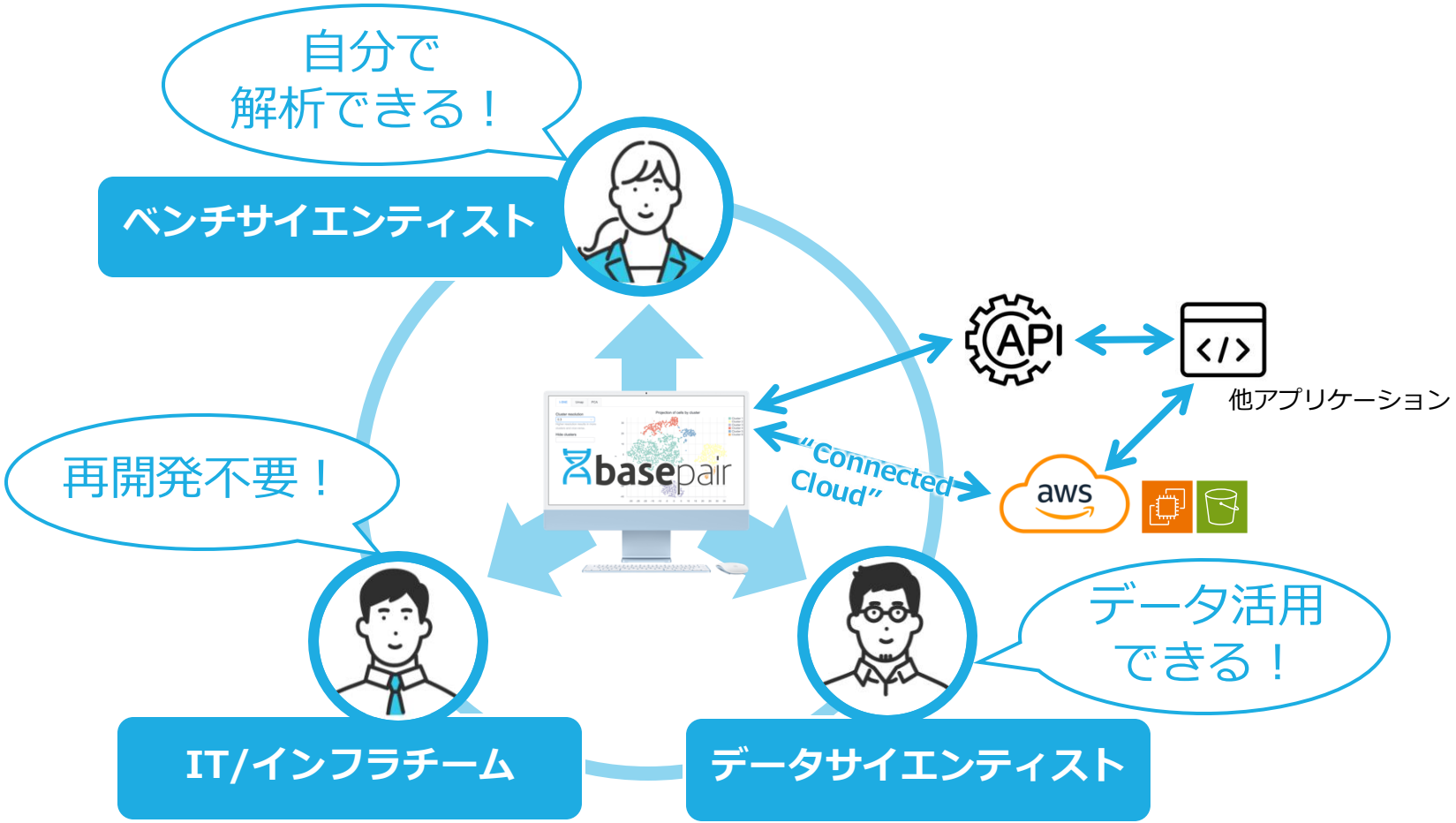
```
import basepair

# CREATE SAMPLE
data = {
    "name": "Sample 10",
    "genome": "hg19",
    "datatype": "dna-seq",
    "file1": "/path/to/file1.fastq.gz",
    "file2": "/path/to/file2.fastq.gz",
}
sample_id = basepair.create_sample(data)


# RUN ANALYSIS
basepair.create_analysis(workflow_id=5, sample_id)
```

```
basepair --action create-sample --name "Sample 5"
\
--datatype chip-seq --genome mm9 --file1
sample_5.1.fastq.gz

basepair --action create-analysis -w 14 -s 4111 -
s 4112 -c 5112 -c 5113 -c 5114
```

Python API

“Connected Cloud”




自分で
解析できる！
ベンチサイエンティスト
再開発不要！
IT/インフラチーム
データサイエンティスト
データ活用
できる！
basepair
“Connected Cloud”
API
他アプリケーション
aws

Basepairの利点

- **セットアップ・インストール不要**
  - ファイルとインターネット環境があればOK
- **手元PCのスペックは不問**
  - サンプル数が多くても大丈夫
  - クラウド上で計算が稼働するので、PCが固まったりしない
  - 並列処理
  - CLIでの操作も可能
- **共有可能**
  - チーム・共同研究相手に共有することができます

basepair

https://basepairtech.jp/