

Research Article

Overcoming off-targets: assessing Western blot signals for Bcnt/Cfdp1, a tentative component of the chromatin remodeling complex

 Shintaro Iwashita¹, Takehiro Suzuki², Yoshimitsu Kiriya¹, Naoshi Dohmae², Yoshiharu Ohoka¹, Si-Young Song¹ and Kentaro Nakashima¹

¹Kagawa School of Pharmaceutical Sciences, Tokushima Bunri University, Shido 1314-1, Sanuki, Kagawa 769-2193, Japan; ²Biomolecular Characterization Unit, RIKEN Center for Sustainable Resource Science, Saitama 351-0198, Japan

Correspondence: Shintaro Iwashita (siwast@kph.bunri-u.ac.jp) or Kentaro Nakashima (nkentaro@kph.bunri-u.ac.jp)



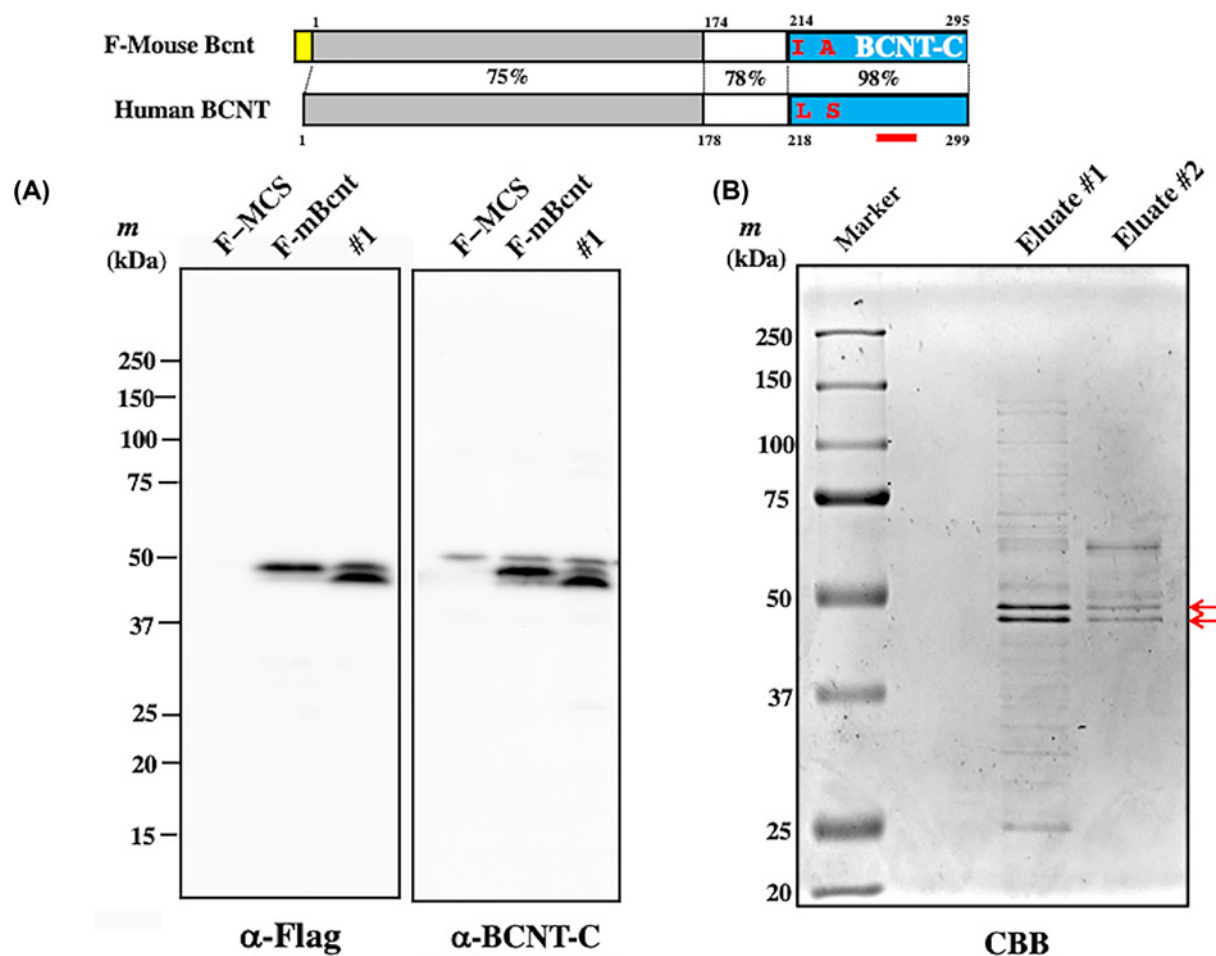
The Bucentaur (BCNT) protein family is characterized by a conserved amino acid sequence at the C-terminus (BCNT-C domain) and plays an essential role in gene expression and chromosomal maintenance in yeast and *Drosophila*. The mammalian Bucentaur/Craniofacial developmental protein 1 (Bcnt/Cfdp1) is also a tentative component of the SNF2-related CBP activator protein (Srcap) chromatin remodeling complex, but little is known about its properties, partly because few antibodies are available to examine the endogenous protein. In this paper, we assigned the Western blot signal against the mouse Bcnt/Cfdp1 as a doublet of approximately 45 kDa using anti-Bcnt/Cfdp1 antibodies, which were generated against either of two unrelated immunogens, BCNT-C domain or mouse N-terminal peptide, and in addition, the Cfdp1 knockdown mouse ES cell line and bovine tissue were used as potential negative controls. Moreover, LC-MS/MS analysis of the corresponding doublet to the Flag-tagged mouse Bcnt/Cfdp1 that was constitutively expressed in a HEK293 cell exhibited that the upper band was much more phosphorylated than the lower band with preferential Ser phosphorylation in the WESF motif of BCNT-C domain. Western blot analysis with these evaluated antibodies indicated a preferential expression of Bcnt/Cfdp1 in the early stages of brain development of mouse and rat, which is consistent with a data file of the expression of Bcnt/Cfdp1 mRNA.

Introduction

The BCNT family members in yeast and *Drosophila* have been shown to play essential roles in gene expression and chromosomal maintenance [1]. Mammalian Bcnt/Cfdp1 is also presumed to be a component of the SNF2-related CBP activator protein (Srcap) chromatin remodeling complex, which is based on the results using fractionation of cultured human cell extracts followed by mass spectrometry [2]. This presumed frame is originally based on the analysis of Swc5, a budding yeast ortholog of Bcnt/Cfdp1, in Swr1 (yeast Srcap) chromatin complex [3–5]. Although Swc5 is not integrated with the Swr1 complex, it participates in the activation of the remodeler ATPase, Srcap, and the ATP-dependent histone exchange reaction. The reaction replaces nucleosomal H2A–H2B with H2A.Z–H2B dimers by recruiting the variant H2A.Z in the transcription and DNA repair [6–8]. Swc5 is not essential for survival, yet its deletion mutant *swc5Δ* cells caused lack in histone replacement activity for Swr1, resulting in genetic instability, hypersensitivity to drugs, and transcriptional misregulation [9], [The *Saccharomyces* Genome Database <https://www.yeastgenome.org/>].

Received: 26 November 2019
Revised: 15 May 2020
Accepted: 20 May 2020

Accepted Manuscript online:
20 May 2020
Version of Record published:
09 June 2020



Members of the Bcnt family generally consist of an acidic N-terminal region, a highly conserved C-terminal region with approximately 80 amino acids (BCNT-C domain), and a hydrophilic region between them (Figure 1 and [1]). Recently, the BCNT-C domain of the budding yeast Swc5 was found to be essential for the histone exchange reaction [10]. On the other hand, a deletion mutant induced by one-base replacement of the *Drosophila* ortholog, *Yeti*, which entirely lacks the BCNT-C domain, shows substantial chromosomal abnormalities, resulting in lethality before pupation [11]. Thus, the metazoan *Bcnt/Cfdp1* is essential for survival in contrast to the yeast *swc5*. Furthermore, the chicken ortholog CENP-29 has been identified to be a kinetochore-associated protein [12]. Given a report that

Table 1 Properties of four anti-Bcnt/Cfdp1 Abs used in the study

Antibody name	Immunized animal	Antigen (Sequences/Location)	Properties of antigens or Abs	Reference
Anti-BCNT-C Ab	Guinea pig	NH ₂ - CEELAIHNRGKEGYIERKA -COOH Human BCNT/CFDP1 259-276 th peptide	Common to almost mammals	[17,18]
Anti-mBcnt-N Ab	Guinea pig	CH ₃ CO- GEEQAEKTKGKRRKAQC -COOH Mouse Bcnt/Cfdp1 41-56 th peptide	Mouse and rat specific	[18]
A305-624A-M (Bethyl)	Rabbit	Human BCNT/CFDP1 249-299 th peptide https://www.bethyl.com/product/a305-624A-M/CFDP1-Antibody	Detection of ~48, ~37, and ~19 kDa bands, immunoprecipitation ability	
26636-1-AP (Proteintech)	Rabbit	Human BCNT/CFDP1 172-299 th Recombinant https://www.ptglab.com/products/CFDP1-Antibody-26636-1-AP.htm	Validated by siRNA knockdown	

Amino acid residues of antigen peptides are indicated in bold letters and their cysteine residues for conjugation with keyhole limpet hemo-cyanin are also shown.

CENP-B protects centromere chromatin integrity by promoting histone deposition [13], these results imply that the Bcnt members may play broader roles in the maintenance of the structure and function of the chromosome.

RNA sequencing analysis, as shown by the dramatic influence caused by genetic mutations of *swc5* in fission yeast [14], is a powerful method to clarify the mechanism of complex and dynamic processes, such as embryonic development and stress adaptation. However, recent studies have shown a discordance between mRNA and protein expression in such dynamic processes and argue that analysis at the transcriptional level is insufficient to predict protein levels [15]. Since members of the Bcnt family may preferentially function to maintain the structure and function of the chromosome in these dynamic processes, analysis of their protein dynamics is essential to understand those processes. However, it is challenging to assign the correct signal by most of the currently available Bcnt/Cfdp1 antibodies (Abs), even in Western blot analysis.

We previously characterized human BCNT/CFDP1 (hBCNT/CFDP1) using a cell line that constitutively expresses its His-tagged molecule (His-hBCNT/CFDP1) [16] as well as Ab generated against an 18-mer peptide (EELAIH-NRGKEGYIERKA) derived from BCNT-C (anti-BCNT-C Ab) (Table 1) [17]. His-hBCNT/CFDP1 has a calculated mass of 34.9 kDa (33.6 kDa plus His-tag), but its immunoreactive signal was detected around 50 kDa as a doublet band on SDS-polyacrylamide gel electrophoresis (SDS/PAGE). We showed that the difference between its calculated and apparent molecular mass is mainly due to the acidic stretch located in the N-terminal region and Ser²⁵⁰ phosphorylation in the BCNT-C domain [16]. However, we could not identify the endogenous hBCNT/CFDP1 due to a high background caused by anti-His Ab reactive proteins, and we also could not accurately assess the specificity of anti-BCNT-C Ab to endogenous hBCNT/CFDP1. Furthermore, we recently found that the anti-BCNT-C Ab cross-reacts with an unrelated target, glutamine synthetase (GS) (mouse GS, NP_032157; EC 6.3.1.2), which is also known as γ -glutamate: ammonia ligase [18]. In this paper, we assigned the Western blot signal against the endogenous mouse Bcnt/Cfdp1 (mBcnt/Cfdp1) using various target-related materials, including flag-tagged mBcnt/Cfdp1 (F-mBcnt/Cfdp1) and *Bcnt/Cfdp1* knockdown ES cells. We also present a scheme to prepare a potential negative control for the western blot analysis to evaluate the Bcnt/Cfdp1 signal. Then, we demonstrate a high expression of Bcnt/Cfdp1 at an early developmental stage of the brains of mice and rats by evaluating the Abs.

Materials and methods

Detailed information on material sources are listed in Supplementary Table S4.

Cell culture

T-REx-293 (a HEK293 cell derivative, T-REx) cells, their subcolonies which express Flag-multi-cloning site (F-MCS) or F-mBcnt, and their clone G11 cells which express His-hBCNT/CFDP1 [16] were routinely maintained in a 5% CO₂ incubator using DMEM-GlutaMAX-1 (DMEM) supplemented with 10% fetal calf serum, 50 μ g/ml gentamicin, and with or without G418 (0.5 mg/ml). These cells were subcultured by incubating cell layers (2 ml medium in a 35-mm dish) with 0.75 ml of Accutase for 5 min at room temperature, dispersed by pipetting with a 1000- μ l pipet tip, and ~1/15 of the suspension was directly plated on a dish preincubated with the medium. G418 was added on the next day when needed. For preparation of cellular protein extract or transfection, after washing the cell layers (5 ml of medium in a 60-mm dish or 10 ml of medium in a 100-mm dish) with prewarmed Hepes buffered saline

(HBS [10 mM Hepes-NaOH, pH 7.5, 150 mM NaCl]), their cell suspension were prepared as described above with the medium. TrypLE Express was also used in the earlier stage of the study (see below).

ES cell culture

35-mm dishes were precoated by incubating with recombinant human Laminin (iMatrix-511) at a final concentration of 5 µg/ml in phosphate-buffered saline (PBS) either at room temperature or at 37°C. Cfdp1-K1 (cell No: AyuK7D01) and vdr2-4 (cell No: JCRB1658) cells were obtained from Japanese Collection of Research Bioresources (JCRB), and were grown on precoated 35-mm dishes in ESGRO Complete Clonal Grade Medium plus GSK3β Inhibitor (50 µl/100 ml), further supplemented with gentamicin (50 µg/ml). After washing with prewarmed HBS, cell layers (2 ml medium in a 35-mm dish) were incubated with 0.75 ml of Accutase at room temperature and dispersed by pipetting with a 1000-µl pipet tip after 5 min, and 1 ml of DMEM containing 0.1% polyvinyl alcohol (DMEM-PVA) was added. The cells were collected into a 15-ml tube after washing the dish with DMEM-PVA. While counting its cell number, the suspension was centrifuged (100 × *g*, 3 min, room temperature), resuspended in a new medium and seeded at the density of $\sim 2.5 \times 10^5$ cells per 35-mm dish. In order to prepare protein extracts or total RNA, cell suspensions were washed twice with chilled HBS, dispensed at $\sim 1 \times 10^6$ cells per 1.5-ml tube, and centrifuged (100 × *g*, 5 min, 4°C). After removing the buffer, cell pellets were softly vortexed, snap-frozen in liquid nitrogen, and stocked at −80°C until use.

Generation of T-REx colonies expressing Flag-mBcnt

T-REx cells ($\sim 2 \times 10^6$ per 100-mm dish) were culture for one day, transfected with each 5 µg of Flag-MCS-pcDNA3.1 (Accession No. LC311018) or Flag-MCS-pcDNA3.1 plasmid carrying *mBcnt* using Lipofectamine 3000 according to the manufacturer's protocol. Just before transfection, each half medium was once removed, saved, and returned to the culture ~ 5 h after transfection. After culturing for a total of 44–48 h, each cell layer was washed with prewarmed HBS, dispersed with TrypLE Express for 10 min at 37°C, harvested with the medium into a 15-ml tube, centrifuged, and resuspended in the medium. The number of resuspended cells was counted, and ~ 60 cells were seeded in a 48-well plate preincubated with the medium. After 4 h, G418 was added to a final concentration of 0.75 mg/ml. Colonies growing in 48-well plates were sequentially expanded to 24-well plates and 35-mm dishes. Each colony was then cultured in the presence of 0.5 mg/ml G418.

Preparation of protein extracts of T-REx cells and their subcolonies

After washing with chilled HBS, cells on a 100-mm dish were lysed by adding 0.5 ml of lysis buffer (20 mM Hepes-NaOH, pH 7.5, 150 mM NaCl, designated L-buffer) supplemented with both inhibitors of proteinases and phosphatases, then cells were collected with a cell scraper (17-mm width). The suspension was transferred into a 1.5-ml tube, sonicated by a Bioruptor (BM Equipment) in an ice-water bath (15 × 10-s pulses at 10-s intervals), and centrifuged (25000 × *g*, 30 min, 4°C, Kubota 3780, rotor AF-2536A). The supernatants were aliquoted, snap-frozen in liquid nitrogen, and stored at −80°C until use. The obtained pellets were dissolved in 50 µl of lysis buffer containing SDS (1% SDS, 1 mM EDTA in 10-mM Hepes-NaOH, pH 7.5, designated LS-buffer), sonicated, and centrifuged (10000 × *g*, 1 min, 4°C. After measurement of the protein concentration, the sample was boiled in SDS/PAGE sample buffer.

Isolation of Flag-mBcnt using anti-Flag antibody-conjugated agarose beads

The frozen supernatant of the subcolony #1 cells was thawed, and Nonidet P-40 (NP-40) was added to a final concentration of 0.05%. After sonication for 30 s in an ice-water bath followed by centrifuging at 10000 × *g* for 1 min at 4°C, the supernatant (1.2 mg in 1 ml) was mixed with anti-Flag-conjugated agarose beads (20 µl settled volume) in a 1.5-ml siliconized tube and incubated in a rotary shaker for 2 h, at 4°C. The mixture was centrifuged for 30 s, and the supernatant was saved as the unbound fraction. The pellet was resuspended in 100 µl of L-buffer containing 0.05% NP-40 plus inhibitors and transferred to a spin column using a 200-µl wide-bore tip. The tube was once more washed with L-buffer plus NP-40, and the suspension was recovered to the spin column. The flow-through fraction obtained by centrifugation was saved as the first wash fraction. The protein-bound agarose was washed twice with the same buffer, followed by washing once with HBS, and then the bound proteins were eluted by incubating with 50 µl of Flag peptide solution (150 µg in HBS) at 4°C for 30 min (Eluate #1) and another 5 min (Eluate #2), sequentially. The agarose was further treated with 50 µl of glycine-HCl (50 mM, pH 2.5), and its eluate and the agarose in the column were immediately neutralized with 2 M Tris. Finally, 50 µl of 1 × SDS/PAGE sample buffer was added to the

column, vortexed, and boiled for 5 min. All of the fractionated samples except the fraction eluted with SDS/PAGE sample buffer, were boiled in 1 × SDS/PAGE sample buffer after being adjusted with 4 × SDS/PAGE buffer. Each sample was separated on 12.5 % SDS/PAGE and followed by western blot analysis using anti-Flag Ab. For LC-MS/MS analysis, Eluate #1 and Eluate #2 described above (each ~40 µl) were concentrated with acetone according to a protocol (<http://tools.thermofisher.com/content/sfs/brochures/TR0049-Acetone-precipitation.pdf>). The dry pellet was once dissolved in 6 µl of LS-buffer, and 2 µl of 4 × SDS/PAGE sample buffer was added, followed by boiling for 3 min. After separating the sample on 12.5% gel SDS/PAGE 10 min longer than usual to ensure the separation of the upper and lower bands, the gel was fixed with 50% MeOH-10% acetic acid solution, stained with 0.25% CBB, and then de-stained in 10% MeOH-7% acetic acid solution.

Preparation of protein extracts from mouse and rat brain

The whole brain and cerebellum were dissected from C57BL/6J mice (P0, male) and Wistar rats (P0-P56, male), respectively, after euthanasia by diethyl ether anesthesia. These samples were snap-frozen in liquid nitrogen and stored at −80°C until use. Frozen samples were crushed with a hammer on dry ice and immediately transferred to a glass-Teflon homogenizer containing LS-buffer (1 ml per 100 mg of samples). Then, tissues were homogenized at 600 rpm using a digital homogenizer and boiled for 5 min. The homogenates were sonicated (12 × 10-s pulses at 20-s interval) and centrifuged (28000 × g, 30 min, 20°C). In a case of a mouse brain of the embryonic day around 17, a frozen piece (~40 mg) was wrapped with aluminum foil, crushed with a plier cooled with liquid nitrogen, transferred to a 1.5-ml Bio Masher, and soaked in 200 µl of chilled LS-buffer. Then, the tissues were homogenized cooling in ice, boiled for 5 min, sonicated, and centrifuged (15000 × g, 10 min). The supernatants were used for Western blot analysis. The protein concentrations of the supernatants were estimated using a Bicinchoninic Acid protein assay kit (BCA kit) with bovine serum albumin as a standard.

Enrichment of bovine Bcnt/Cfdp1 using Phos-tag agarose

Bovine placenta (Holstein on 116th day of the pregnancy) as small frozen pieces, which were a gift from Dr Kazuyuki Hashizume (Iwate University, Morioka), were stored in liquid nitrogen until use. Tiny frozen tissue was obtained by cutting the pieces with a knife (~30 mg); then, they were further minced with a razor blade, and proteins were extracted in the following two ways. The minced pieces were placed in a 1.5-ml BioMasher, soaked in 200 µl of cold L-buffer plus inhibitors, homogenized, and then mixed with 10 µl of 20% SDS, followed by boiling for 5 min. The extract was sonicated for 2.5 min, centrifuged (15000 × g, 10 min, 22°C), and the supernatant was used as a whole extract. For enrichment of Bcnt/Cfdp1 content, the minced tissue pieces in the 1.5-ml BioMasher were homogenized in 100 µl of chilled RIPA buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% sodium deoxycholate, 1% NP-40, 1 mM EDTA) plus both inhibitors of proteinases/phosphatases. After washing with another 100 µl of RIPA buffer, the suspension was centrifuged (15000 × g, 10 min, 4°C) and the supernatant was aliquoted, snap-frozen in liquid nitrogen and stored in −80°C until use. After dilution of the supernatant with RIPA buffer to 200 µg/100 µl, the suspension was mixed with Phos-tag agarose (100 µl settled volume) in a spin column, and the protein-bound beads were washed three times with each 200 µl of washing buffer and eluted three times with each 100 µl of elution buffer per tube according to the manufacturer's protocol. Each eluate was precipitated with TCA and washed twice with acetone according to a protocol. After heating at 95°C, 50 µl of 1 × SDS/PAGE sample buffer was added and solubilized by a mixer (Tomy MT-360) for the Western blot analysis. Bcnt/Cfdp1 content in the above pellets (PPT) was estimated by adding 6.7 µl of 4 × SDS/PAGE sample buffer to PPT (~20 µl), boiling for 5 min, and sonicating for 2.5 min. For larger scale preparation, the above 700 µg of the extract was used.

Immunoblotting

Procedures of SDS/PAGE using 12.5% or 15% gel and blotting onto membranes were virtually the same as previously described [16,18]. Abs against Bcnt/Cfdp1 are listed in Table 1. All immunoreactivity was visualized by a chemiluminescence method using horseradish peroxidase (HRP)-conjugated secondary Abs as described previously [18], except for using HRP-conjugated goat anti-mouse IgG light chain instead of anti-mouse IgG (H+L).

Mass spectroscopy analysis

The upper and lower bands of a doublet of F-mBcnt, which were detected by CBB staining, were separately cut out from the gel and digested with API, AspN, and chymotrypsin. Each digest was analyzed by nano-LC-MS/MS using a Q Exactive mass spectrometer, virtually the same as described previously [16]. Peptides derived from the upper and lower bands were quantified by a label-free quantification method using Proteome Discoverer Ver 2.2.0.388 (Thermo

Fisher Scientific). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [19] partner repository with the dataset identifier PXD019016 and 10.6019/PXD019016.

Isolation of total RNA

Frozen ES cells were homogenized with TRI reagent, and total RNAs were purified using PureLink RNA Mini kit according to the manufacturer's protocol. Purified total RNAs were treated with TURBO DNase to eliminate contaminating genomic DNA, extracted with phenol/chloroform/isoamyl alcohol (pH 5.2), and re-purified using RNA Clean & Concentrator -25 kit. The concentration of the total RNA was determined by the absorbance at 260 and 280 nm using NanoDrop One (Thermo Fisher Scientific), and the quality was estimated using a 2100 Bioanalyzer (Agilent). RNA sequencing was outsourced using each total RNA from Cfdp1-K1 and vdR2-4, respectively (A269/280: 2.07 and 2.03; RIN: 9.9 and 9.6~9.9, respectively) (Macrogen Japan Corp., Kyoto).

Reverse transcription-PCR, plasmid construction, and PCR product analysis

cDNAs were synthesized from purified total RNAs of Cfdp1-K1, vdR2-4 cells, and whole mouse brain of a P56 C57BL/6J male by Superscript III First-Strand Synthesis SuperMix according to the manufacturer's protocol using oligo-(dT) 20. The full-length ORF of *mBcnt* or the fragment of *mBcnt* exons 1-5 fused with *Hph* (hygromycin phosphotransferase) was amplified from each cDNA (1 ng of total RNA) by PCR using KAPA HiFi HotStart DNA polymerase, and PCR products were analyzed as described previously [18]. The primer sequences for detection of the gene trapped *mBcnt* are 5'-gagctcggatccgccgcatggaggaattcgactccgaag-3' (*SacI*-*Bam*HI-*mBcnt*, forward) and 5'-gagctctcgcgagccgatgcaagtgccgataaac-3' (*hph*-*Xho*I-*Sac*I, reverse), respectively.

Transcriptome analysis

The following is a summary of the report from Macrogen Corp Japan (Supplementary Appendix S1). The two cDNA libraries from the RNA of Cfdp1-K1 or vdR2-4 cells were prepared, and their sequences were obtained. The trimmed paired-end reads (read length 101) of 45,262,214 from Cfdp1-K1 or 52,245,252 from vdR2-4 were mapped to a mouse reference genome (UCSC GRCm38.p4/mm10, annotation RefSeq_2017_06_12).

Results

Detection of Flag-tagged *mBcnt* as a doublet band with differential phosphorylation

To characterize mammalian endogenous *Bcnt*/Cfdp1, we previously expressed exogenous His-hBCNT/CFDP1 as a reference. However, proteins that crossed the anti-His tag Ab provided a relatively high background, making it difficult to assign the signals of the endogenous target in the Western blot [16]. This time, as another reference, we expressed F-*mBcnt* in T-REx cells.

mBcnt/Cfdp1 is composed of 295 amino acids (aa), which is 4 aa less than the human counterpart (Figure 1, top panel, [1]). While 82 aa sequences of both BCNT-C domains are identical except for two aa residues, the N-terminal region has low homology between mouse and human (75%) and can be used as species-specific immunogens. T-REx cell colonies constitutively expressing F-*mBcnt* were isolated using G418 selection. Both the number and size of the antibiotic-resistant colonies expressing F-*mBcnt* were significantly lower and smaller than those expressing F-MCS as a control. The extracts from several colonies of each transfectant were prepared and evaluated with either anti-Flag Ab or anti-BCNT-C Ab (Table 1) using Western blot analysis. The signal with anti-Flag Ab showed no or two bands, whereas anti-BCNT-C Ab detected single or three bands (Figure 1A; Supplementary Figure S1A and B).

When comparing the doublet pattern between transiently expressing cells and constitutively expressing cells, the upper band of the transient expression was significantly stronger than that of the constitutive expression (Figure 1A). These features are similar to those of His-hBCNT/CFDP1, as previously reported [16].

To reveal the molecular difference between the upper and lower bands, they both were isolated from lysates of T-REx-derived colony #1, which constitutively express F-*mBcnt*, using anti-Flag Ab-conjugated agarose beads (Supplementary Figure S2 and Figure 1B). Each band excised from the gel was digested with three different proteases, and each digest was subjected to LC-MS/MS analysis (PXD019016 and 10.6019/PXD019016). The ratios of the upper to lower bands for each set of fragments and their phosphorylation sites are systematically presented in Supplementary Table S1 and Figure S3. The upper band is much more phosphorylated than the lower band. In particular, serine 246th *mBcnt*/Cfdp1, which corresponds to S250 of hBCNT/CFDP1 [16], is a preferential site of phosphorylation.

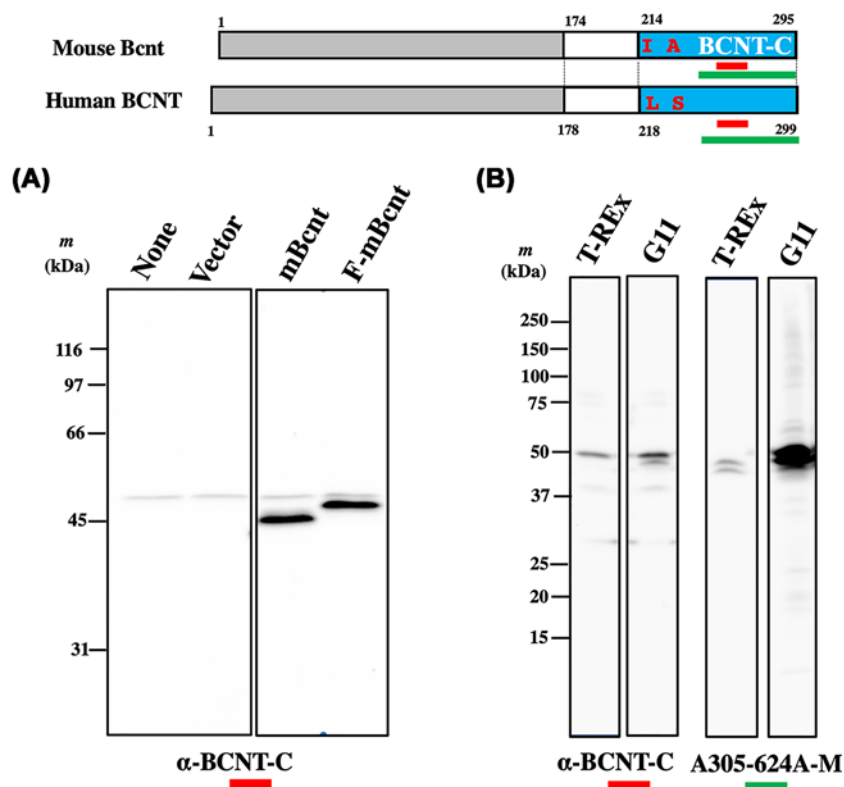


Figure 2. Comparative assessment of Western blot signals between two antibodies: anti-BCNT-C and A305-624A-M

Top panel: Structures of human BCNT/CFDP1 (Human BCNT) and mouse Bcnt/Cfdp1 (Mouse Bcnt) are shown as in the top panel of Figure 1, and the location of each immunogen of anti-BCNT-C antibody (Ab) (red bar) or A305-624A-M (green bar) is presented. **(A)** Comparison of mobilities between Flag-tagged (F-mBcnt) and tag-free mouse Bcnt/Cfdp1 (mBcnt) on SDS/PAGE. HEK293T cells were transfected with Flag-MCS-pcDNA3.1 plasmid carrying mBcnt or mBcnt in BsrGI-MCS-pcDNA3.1 (Accession No. LC311017), and their extracts were subjected to Western blot analysis with anti-BCNT-C Ab. Left panel: None and Vector indicate untreated and a blank vector control, respectively. The results published in [18] were rearranged and shown. **(B)** The supernatants were prepared from each cell lysate of T-REx or G11 cells (His-hBCNT/CFDP1 expressing clone) by centrifugation. Equal amounts of protein (20 µg) were subjected to Western blot analysis with either anti-BCNT-C Ab (left two lanes) or A305-624A-M (right two lanes).

Evaluation of anti-BCNT-C antibody

Based on the results of the Western blot analyses with anti-BCNT-C Ab [16], we previously presumed that the ~50-kDa signal corresponded to endogenous hBCNT/CFDP1 in both T-REx cells and its G11 clone, the latter of which constitutively expressed exogenous His-hBCNT/CFDP1. F-mBcnt is 306 aa in length, which is 7 aa longer than hBCNT/CFDP1 and is similarly phosphorylated (Supplementary Table S1 and Figure S3). Furthermore, the effect of acid Flag tag (DYKDDDDK) on mobility in SDS/PAGE is more significant than that of the 11 aa difference (1.2 kDa) between F-mBcnt and mBcnt (Figure 2A and [18]). Thus, it should be larger than His-hBCNT/CFDP1. However, F-mBcnt appeared below the ~50-kDa band (Figure 1). From these results, the ~50-kDa is a band that is not likely derived from endogenous hBCNT/CFDP1.

Thus, we reexamined the previous Western blot data of His-hBCNT/CFDP1 [16] by introducing two other commercially available anti-Bcnt/Cfdp1 Abs: 26636-1-AP and A305-624A-M (Table 1). We chose these two Abs from the following criteria. The main antigen regions are from BCNT-C domain, and the candidate signal (s) is reported in a region that is significantly smaller than 50 kDa from the Western blot analysis. However, as we detected an extra ~50 kDa signal in 26636-1-AP (shown below), we compared the western signal patterns of cell lysates between anti-BCNT-C Ab and A305-624A-M. Both Abs recognized exogenously expressed His-hBCNT/CFDP1 in the G11 extract but showed distinctly different patterns in the parental T-REx extract (Figure 2B). It is noteworthy that anti-BCNT-C Ab reacted to a band above the doublet detected with A305-624A-M. The difference between the two patterns was confirmed by reprobing with each of the replaced Abs (Supplementary Figure S4). Whereas anti-BCNT-C Ab revealed a few other bands, A305-624A-M detected mainly the doublet band; however, so far, no

direct evidence suggests that it corresponds to the doublet of F-mBcnt shown in Figure 1. Since we confirmed the ability of A305-624A-M to immunoprecipitate His-hBCNT/CFDP1 using G11 cell extract and anti-His Ab, we inferred that the distinct ~50-kDa band detected with the anti-BCNT-C Ab shown in Figure 1 and Supplementary Figure S1 is an off-target signal.

Assignment of the candidate signal of mBcnt/Cfdp1 using ES mutant cells

To assign the endogenous Bcnt/Cfdp1 signal using Western blot analysis, we utilized a mouse embryonic stem (ES) cell line (Cfdp1-K1) listed as homozygous mBcnt/Cfdp1 mutant cells (i.e. double-knockout cell line) [20]. The gene trap vector is reported to be inserted in intron 5 of mBcnt/Cfdp1 (Figure 3A) (GenBank: accession number AG999723.1). As BCNT-C is encoded by exons 6 and 7 (Figure 3A), the Cfdp1-K1 cell lysate can be used as a potential negative control (knockout cell extracts) to evaluate the Abs generated against antigens derived from BCNT-C if the cells are homozygous mBcnt/Cfdp1 mutants.

First, we examined by reverse transcription-polymerase chain reaction (RT-PCR) whether Bcnt/Cfdp1 is knocked out in Cfdp1-K1. After subculturing in the presence or absence of G418 and puromycin, both of which can delete the feeder layer cells, we prepared cDNAs from Cfdp1-K1 (mutant) and its parental cells, vdr2-4 (wild-type). We then examined the expression of mBcnt/Cfdp1 mRNA by RT-PCR and DNA sequencing of the products. In the cDNA from Cfdp1-K1, we detected PCR products corresponding to the full-length mBcnt/Cfdp1 ORF as well as the fusion gene coding Bcnt/Cfdp1 exons 1-5 and Hph derived from the gene trap vector (Figure 3B). This result indicates that Bcnt/Cfdp1 was not double knockout in Cfdp1-K1 cells, even though the gene trap vector was inserted adequately into intron 5. Further comparative transcriptome RNA sequencing between Cfdp1-K1 and vdr2-4 revealed that 694 genes were expressed with more than a 2-fold difference, while 188 genes were up-regulated, and 506 genes were down-regulated (Supplementary Appendix S1 and Table S2). Among them, Bcnt/Cfdp1 mRNA in the Cfdp1-K1 cells was reduced to 74.4% to that of the parent cells (101.75 vs. 136.79 FPKM, the top row of Supplementary Table S3). Besides, the mRNA of the flanking genes Bcar1/Cas and Tmem170 (<https://www.ncbi.nlm.nih.gov/gene/23837>) and several housekeeping genes frequently used as internal controls were also significantly altered (Supplementary Table S3).

Next, we examined whether anti-BCNT-C Ab and A305-624A-M can detect the differential expression of Bcnt/Cfdp1 between mutant and parental cells using western blot analysis. Of the several bands detected by anti-BCNT-C Ab, one band of ~45-kDa (indicated by a red arrow in Figure 3C) was distinctively reduced in the Cfdp1-K1 cells compared with vdr2-4 cells, which was also observed with A305-624A-M. These results suggest that the ~45-kDa band is a candidate for the endogenous mBcnt/Cfdp1 in mouse ES cells.

Evaluation of the anti-mBcnt-N antibody and assignment of the mBcnt/Cfdp1 signal

As mentioned above, the C-terminal region of BCNT members is highly conserved, but the N-terminal region is variable among species. To further investigate whether the ~45-kDa band is a valid signal, an anti-mBcnt/Cfdp1 Ab was generated using a mouse N-terminal peptide as an immunogen (anti-mBcnt-N Ab, Table 1 and top panel of Figure 4). Rat Bcnt/Cfdp1 has the same sequence as the immunogenic peptide derived from mBcnt/Cfdp1, but the bovine counterpart has a distinctively different amino acid sequence (Figure 4, middle panel), suggesting that the cross-reactivity with the anti-mBcnt-N Ab was expected to be low. Thus, we can use tissue extracts of bovine and rat as a potential negative control and positive control, respectively, to evaluate the specificity of anti-mBcnt-N Ab against the endogenous mBcnt/Cfdp1 in the Western blot analysis. After concentrating the Bcnt/Cfdp1 content in the bovine extract with Phos-tag agarose [21] to enrich the phosphorylated proteins (Supplementary Figure S5), we showed that anti-mBcnt-N Ab detected the ~45-kDa signal in mice and rats but not in cattle (Figure 4A, left filter), while A305-624A-M detected the ~45-kDa signal in all tissue extracts (Figure 4A, right filter). This result strongly suggests that anti-mBcnt-N Ab specifically recognizes endogenous mBcnt/Cfdp1 despite nonspecific cross-reactions with multiple unknown proteins. We further confirmed that the ~45-kDa signal was a valid target signal using another anti-hBCNT/CFDP1 Ab, 26636-1-AP (Figure 4B, right filter).

Using anti-mBcnt-N Ab or A305-624A-M, we performed Western blot analysis with two kinds of Cfdp1-K1 cell lysates prepared after passages in the presence or absence of G418 / puromycin (Figure 5A). The results showed that both Abs exhibited a ~45-kDa band with an intensity that is distinctively reduced in both Cfdp1-K1 lysates as compared with that in the parental cell lysate. Finally, we confirmed that all four anti-Bcnt/Cfdp1 Abs used in the

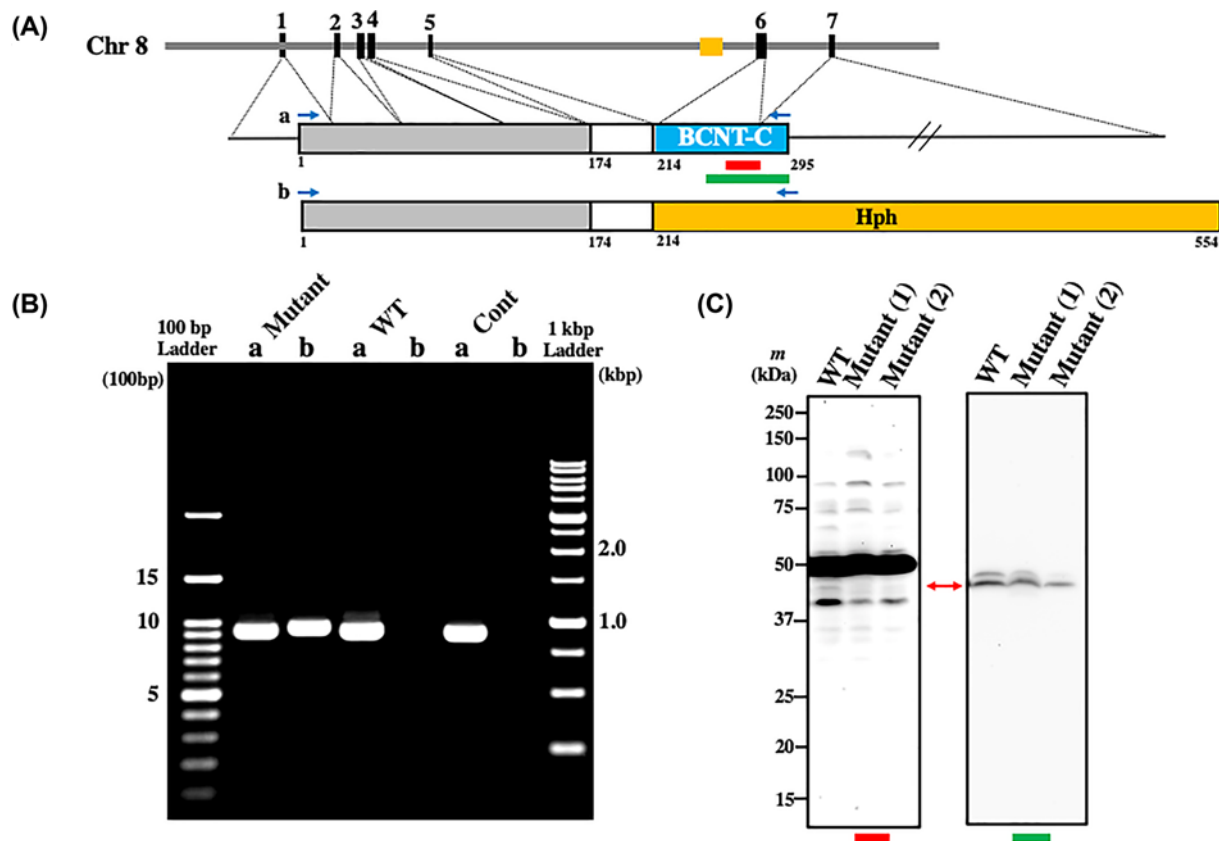


Figure 3. Assignment of the endogenous Bcnt/Cfdp1 signal using mutant ES cells by Western blot analysis

(A) Location of the inserted gene trap vector in mouse *Bcnt/Cfdp1* and the expected fusion protein. The mouse *Bcnt/Cfdp1* gene is located in the chromosome 8 (Chr 8) and consists of 7 exons. Dashed lines indicate corresponding regions of each exon to *Bcnt/Cfdp1*. A brown box on Chr 8 indicates the gene trap vector. Red and green bars under *Bcnt/Cfdp1* indicate each location of immunogens to raise anti-BCNT-C antibody (Ab) and A305-624A-M, respectively. At the top, the structure of expected fusion protein is shown, which is encoded by *Bcnt/Cfdp1* exons 1-5 and *Hph* (hygromycin phosphotransferase) derived from the gene trap vector. Four blue arrows at the top of each structure of the fusion protein and *Bcnt/Cfdp1* indicate the location of two sets of PCR primers for RT-PCR analysis. (B) RT-PCR analysis of *Bcnt/Cfdp1* mRNA. Using each cDNA from *Cfdp1*-K1 (Mutant), its parental cells *vdR2-4* (wild-type, WT), or mouse brain (Cont), RT-PCR was carried out. Lanes a and b indicate RT-PCR products corresponding to the full-length ORF of *mBcnt/Cfdp1* (928 bp) (a) and the fused gene of *mBcnt* exons 1-5 and a part of *Hph* in the gene trap vector (948 bp) (b), respectively. The PCR products and DNA size ladder markers were separated in agarose gel, followed by staining with ethidium bromide. (C) Assessment of Western blot signal of endogenous *Bcnt/Cfdp1* using extracts of *Cfdp1*-K1 and its parental cells. Cell extracts from an equal number ($\sim 2 \times 10^5$ cells) of *vdR2-4* (WT) and *Cfdp1*-K1 (Mutant) cells, latter which had been serially passaged in the presence [Mutant (2)] or absence [Mutant (1)] of G418/puromycin were subjected to Western blot analysis with either anti-BCNT-C Ab (left filter) or 305-624A-M (right filter). A red two-headed arrow indicates a candidate signal of the endogenous mouse *Bcnt/Cfdp1*.

present study exhibited reasonable signals of ~ 45 -kDa (Figure 5B). From these results, we conclude that the ~ 45 -kDa signal corresponds to the endogenous *mBcnt/Cfdp1*.

Expression of *mBcnt/Cfdp1* in the early stage of brain development

RNA profiling data of mouse and rat (<https://www.ebi.ac.uk/gxa/home>) show that *Bcnt/Cfdp1* mRNA expresses ubiquitously and preferentially in the early stage of development. However, recent studies have revealed pervasive discordance between mRNA levels and protein levels, especially in embryonic development [15]. Therefore, we examined *Bcnt/Cfdp1* expression during the developmental stages of mouse brain and rat cerebrum using the evaluated anti-*Bcnt/Cfdp1* Abs described above. The results showed that *Bcnt/Cfdp1* preferentially expresses in the early stages and significantly decreased according to the postnatal stages in the rat cerebrum (Figure 6).

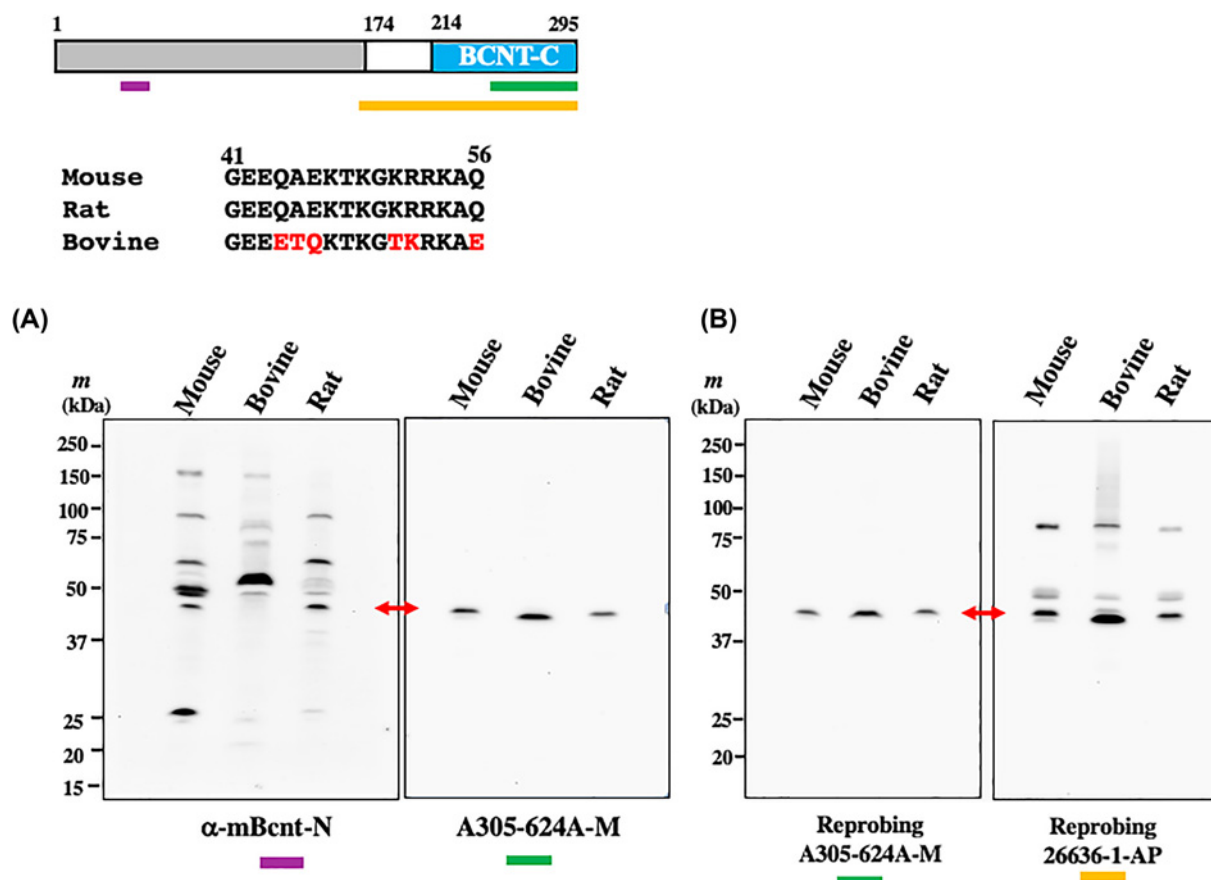


Figure 4. Validation of the anti-mBcnt-N antibody

Top panel: Structure of mouse Bcnt/Cfdp1 and amino acid sequence alignment of the immunogen to raise the anti-mBcnt-N antibody. The structure of mBcnt/Cfdp1 is schematically shown. Three colored bars present the locations of each immunogens to generate the anti-mBcnt-N antibody (Ab) (purple), A305-624A-M (green), and 26636-1-AP (ocher), respectively. Middle panel: Alignment of the immunogen for the anti-mBcnt-N Ab generation and its counterparts of rat and bovine. Red letters indicate different amino acids from the immunogen peptide. (A and B) The specificity of anti-mBcnt-N Ab concerning a ~45 kDa band. (A) Bovine placenta extract enriched for Bcnt/Cfdp1 content (20 or 30 μ g), and brain extracts of mouse and rat (15 or 20 μ g) were separated on SDS/PAGE, followed by Western blot analyses with either anti-mBcnt-N Ab (left filter, the larger amounts of protein, i.e. 20 or 30 μ g) or A305-624A-M (right filter, the smaller amounts of protein, i.e. 15 or 20 μ g). (B) Each filter was reprobed with A305-624A-M (left filter) and 26636-1-AP (right filter), respectively. Two red two-headed arrows indicate the candidate signals of the endogenous Bcnt/Cfdp1.

Discussion

In the present study, we assigned the signal of the endogenous mBcnt/Cfdp1 as a ~45-kDa protein using Western blot analysis by utilizing various target-related materials, and we showed that mouse and rat Bcnt/Cfdp1 are expressed preferentially at early stages of brain development. Moreover, based on the problems encountered during the present study, we discuss immune-cross reactions with off-target proteins.

We generated F-mBcnt-expressing cells instead of those of His-hBCNT/CFDP1 because the latter turned out to be unsuited to assign the endogenous hBCNT/CFDP1 band due to the difficulty to distinguish anti-BCNT-C Ab positive signals from the high background of the anti-His tag Ab positive proteins [16] such as Nono (p54rnb), which contains the HHQHHH sequence in the N-terminus region (NP_001138880) (Supplementary Figure S6). F-mBcnt appeared as a doublet using Western blot analysis, and the upper band was much more phosphorylated than the lower band. In particular, the phosphorylation of the Ser²⁴⁶ residue in the WESF motif of BCNT-C domain was characteristic. The doublet of mBcnt/Cfdp1 was probably caused by the presence or absence of Ser²⁴⁶ phosphorylation as shown in

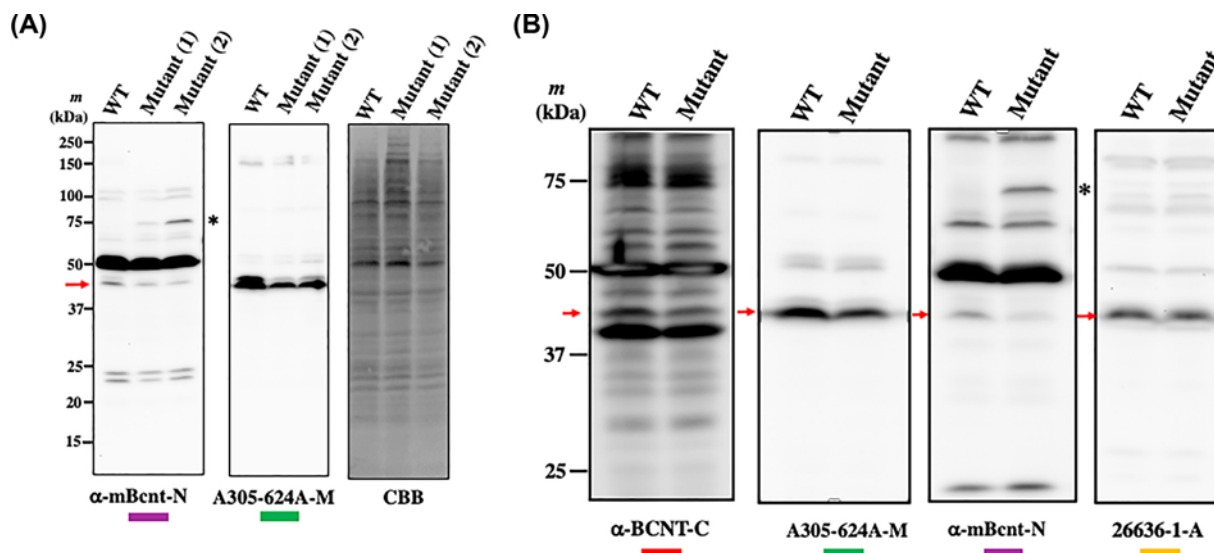


Figure 5. Assignment of the mouse Bcnt/Cfdp1 signals by Western blot analysis

(A) Equal amounts of cell extract (20 μ g protein) of vdR2-4 (WT) and Cfdp1-K1 (Mutant) cells, latter which were serially passaged in the presence [Mutant (2)] or absence [Mutant (1)] of G418/puromycin, were subjected to a Western blot analysis with anti-mBcnt-N antibody (Ab) (left filter) or A305-624A-M (middle filter). A red arrow indicates a signal of mouse Bcnt/Cfdp1. A filter was stained with Coomassie Brilliant Blue (CBB) to check the amounts of loading proteins (right filter). (B) Four filters were prepared as the same as (A) but without [Mutant (1)], and Western blotting was performed with four anti-Bcnt/Cfdp1 Abs used in the present study shown below each filter. Red arrows indicate signals of mouse Bcnt/Cfdp1. A band detected with the anti-mBcnt-N Ab at \sim 75 kDa (shown as *) is probably the fusion protein of a part of mBcnt/Cfdp1 (exon 1-5) and hygromycin phosphotransferase, as schematically shown at the top of Figure 3A, which has a calculated molecular mass of 63.9 kDa, but may run slowly on SDS/PAGE due to the acid stretch located in the N-terminal region of the mouse Bcnt/Cfdp1 [16].

hBCNT/CFDP1 [16] because the characteristic phosphorylation is similar [16]. However, the ratio of the upper and lower bands was variable, and the reason for this is currently unknown.

We first evaluated a custom-made anti-BCNT-C Ab that detected an extra band in F-mBcnt-expressing cell lysates (Figure 1 and Supplementary Figure S1). We initially presumed that the band was hBCNT/CFDP1. We used A305-624A-M, which was confirmed for immunoprecipitation ability using the His-hBCNT/CFDP1-expressing G11 cell extract with anti-His tag Ab (data not shown). We compared the immune-positive signals detected with anti-BCNT-C Ab and A305-624-M and found that their patterns were distinctively different in the parental cell extract, whereas both Abs recognized the exogenous hBCNT/CFDP1 in the extract of the G11 clone (Figure 2B). Therefore, we utilized a Cfdp1-K1 cell line that was listed as a mouse *Bcnt/Cfdp1* homozygous mutant. The cell is a product of a library of random mutations introduced by gene trap vector insertion in Bloom-deficient ES cells and was selected for populations of homozygous mutant cells following mitotic recombination [20]. However, RT-PCR analysis of the cDNA from Cfdp1-K1 revealed the presence of mRNA corresponding to the full-length ORF of the mBcnt/Cfdp1. Indeed, a comparative analysis of the transcriptome RNA sequencing results showed that *Bcnt/Cfdp1* mRNA levels of Cfdp1-K1 were 74.4% of its parental cells. Splicing may efficiently occur by skipping the acceptor site in the gene trap vector, which is located in the intron 5 that spans over 50 kb (Figure 3A). Although Cfdp1-K1 was not a double-knockout cell line of mBcnt/Cfdp1, it was useful as a mBcnt/Cfdp1 knockdown cell to select a promising candidate signal of mBcnt/Cfdp1 at \sim 45 kDa.

On the other hand, an attempt was made to generate *Bcnt/Cfdp1* knockout MDCK (Madin-Darby Canine Kidney) cells by targeting its exon 1 by CRISPR-Cas9 technology as in production of β - and γ -catenin double knockout cells [22]; the expected cells were not obtained so far (W. Kobayashi, personal communication). By using Cfdp1-K1 and its parental cells, we were finally able to assign the signal of endogenous mBcnt/Cfdp1 detected by two Abs raised against unrelated immunogens: a mouse N-terminal peptide and a peptide derived from BCNT-C domain, the latter of which is highly conserved in mammalian Bcnt/Cfdp1.

The following pieces of evidence show that the \sim 45-kDa band is the signal of endogenous Bcnt/Cfdp1 using Western blot analysis. First, among several signals detected by anti-BCNT-C Ab, the \sim 45-kDa signal was distinctively

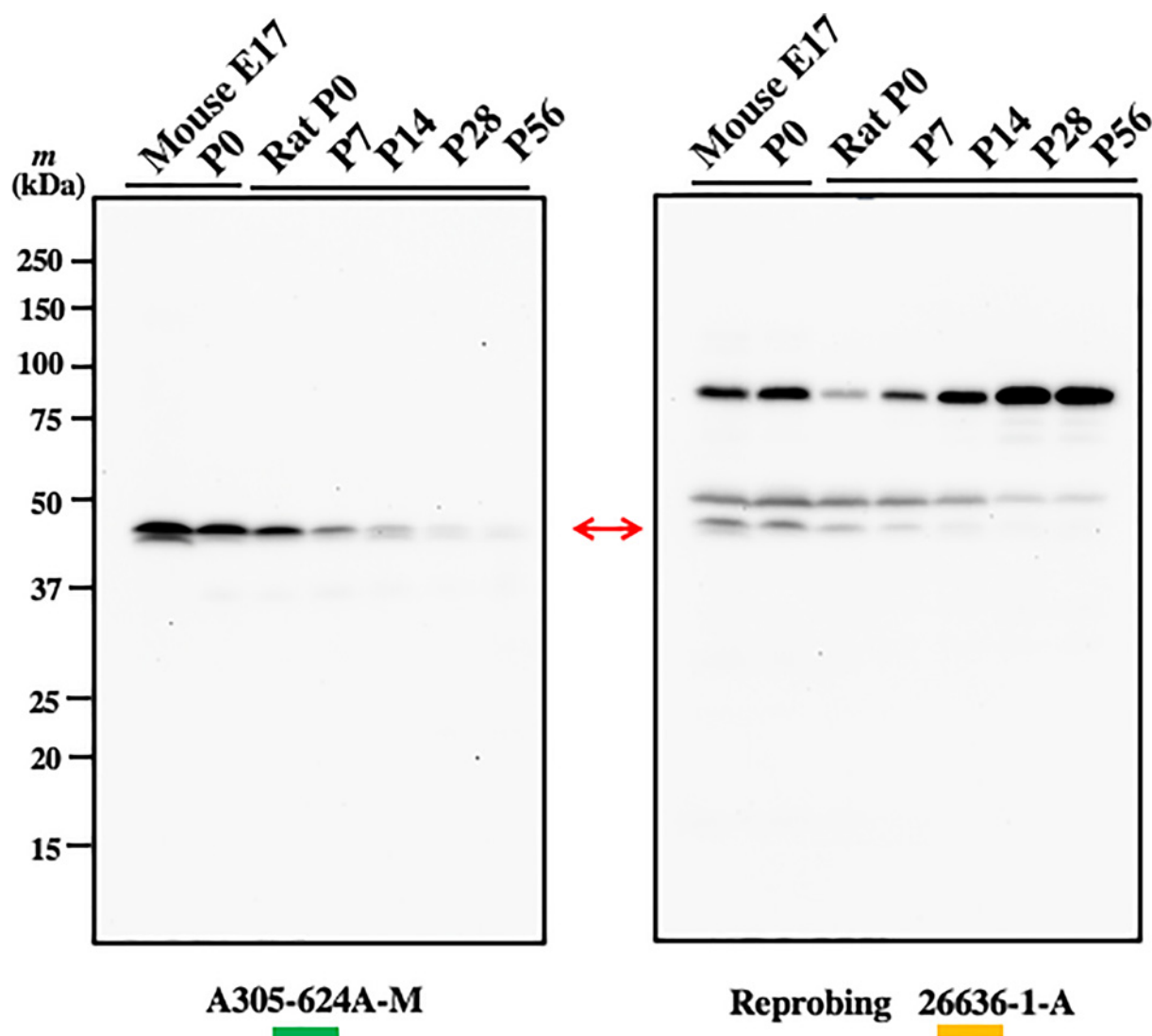


Figure 6. Preferential expression of Bcnt/Cfdp1 in the early developmental stages of mouse and rat brains

Mouse brain extracts of approximately embryonic day17 (E17) and postnatal day 0 (P0), and of rat cerebral extracts from postnatal day samples (denoted by P# on the top of each lane) were loaded (each 20 μ g amounts of protein) and subjected to a Western blotting analysis with A305-624A-M (left panel). The filter was stripped and reprobed with 26636-1-AP (right panel). A red two-headed arrow indicates the signals of the endogenous Bcnt/Cfdp1.

reduced in the Cfdp1-K1 cell extract compared with the signal of the parental cell extract (Figures 3C and 5). Second, the differential \sim 45-kDa signal was also detected by three other antibodies for Bcnt/Cfdp1 (i.e. anti-mBcnt-N Ab, A305-624A-M, and 26636-1-AP), two of which were generated using mutually unrelated immunogens (Figures 4 and 5). We confirmed the immunoprecipitation ability of A305-624A-M among these Abs. More importantly, we validated the specificity of the anti-mBcnt-N Ab concerning the \sim 45-kDa signal using a potential negative control by preparation of enriched Bcnt/Cfdp1 content in the bovine tissue with Phos-tag agarose (Figure 4 and Supplementary Figure S5).

The apparent molecular size of the target band (\sim 45-kDa) on SDS/PAGE appears significantly smaller than those of signals reported by many available anti-Bcnt/Cfdp1 Abs, including anti-BCNT-C Ab. The molecular behavior of the \sim 50-kDa band(s) was significantly different from that of endogenous Bcnt/Cfdp1; therefore, the 50-kDa protein(s) is probably an off-target(s). Our initial conjectures based on the results using anti-BCNT-C Ab in our previous articles are all misdirected, including conjectures on the 50-kDa signal (refer to Abstract in [16]), the 43-kDa signal

(glutamine synthetase, [18] refer to Figure 3C in [17] and Figure 1B in [23]), and the intracellular localization of Bcnt/Cfdp1 by immunostaining (refer to Figure 6B in [24]).

We recently introduced two-dimensional gel electrophoresis for the separation of Cfdp-K1 cell extracts and detected a broad smear and doublet bands around 45 kDa with an expected isoelectric point using A305-624-M (Calculated IP value: 4.80, Iwashita and Nakashima, unpublished data). On the other hand, we did not detect two bands at ~37 and ~19 kDa with A305-624A-M, which is shown in its catalog.

It was evident that anti-BCNT-C Ab detected a weak signal to endogenous Bcnt/Cfdp1 with many proteins that are distinct for the target (Figures 3C and 5). A similarly severe situation that Abs have no ability to detect the endogenous target protein due to many nonspecific bands have been reported (e.g. [25,26]). Concerning the troublesome evaluation of anti-Bcnt/Cfdp1 Abs, which has caused inappropriate assignments using Western blot analysis, several factors could be avoided, while others would be difficult to avoid. First, the expression information of Bcnt/Cfdp1 mRNA should be used more efficiently. It is highly expressed in mouse embryo brains, which is much more suitable than adult samples to identify the target and to evaluate anti-Bcnt/Cfdp1 Abs. Second, although other independent anti-Bcnt/Cfdp1 Abs were required and had been prepared using such as the N-terminal region as immunogens, no Abs was obtained, which satisfied the criteria of a single band at the appropriate molecular range. However, it is not necessary to follow single-band criteria, if we have appropriate information concerning the interest target(s) in the Western blotting analysis. In spite of many nonspecific proteins (Figure 4), we were able to show the specificity of anti-mBcnt-N Ab concerning the target ~45 kDa band. Many similar cases such as anti-Rho Ab [27] have been reported. Third, although nonspecific proteins are generally various, which are recognized by multiple Abs raised against different immunogens of the same target, a relatively strong signal(s) near 50-kDa seemed to be commonly detected by many available anti-Bcnt/Cfdp1 Abs including anti-BCNT-C Ab and 26636-1-AP, resulting in an incorrect assignment. The second and third points may be related to the properties of Bcnt/Cfdp1, which mainly consists of structurally disordered regions. It is noteworthy that the epitopes from the antigen of disordered regions are smaller than those of the ordered regions and that the former epitopes interact more efficiently with their Abs [28]. Thus, the commonly available anti-Bcnt/Cfdp1 Abs that have been consequently generated against immunogens of its structurally disordered regions may cause cross-reactions with various proteins with high affinity, resulting in the poor quality of Ab. Fourth, complex migration of Bcnt/Cfdp1 on SDS/PAGE made it challenging to assign a valid signal. hBCNT/CFDP1 as well as mBcnt/Cfdp1 is expressed as a doublet band and migrates more slowly on SDS/PAGE than expected from the calculated molecular mass (33.6 kDa of hBCNT/CFDP1 and 32.7 kDa of mBcnt/Cfdp1). This feature may be mainly due to the acid stretch of the N-terminal region and the Ser phosphorylation of BCNT-C domain [16].

Off-target problems have been widely discussed in many Ab validation studies (e.g. [25,26]), including specific in-depth efforts for Ab evaluation (e.g. [29]). Moreover, a strategy for Ab validation has been proposed [30]. However, the focus of this discussion appears to be blurred, at least regarding Western blot analysis. Abs do not act under the confined all-or-nothing modes via specific reactions. It is noteworthy that Abs recognize target molecules based on their physicochemical environment, which is the chemical entity of the epitope and does not necessarily correspond to the linear amino acid sequence [18,31,32]. For example, the following two cases reflect topological or stereochemical similarity of spatially limited environments that determine the common epitopes between completely different proteins: Bcnt/Cfdp1 and glutamine synthetase [18], and phosphoducin and β -actin [33], respectively.

Although Western blot analysis is much more powerful to examine the target proteins of extracts than mass spectroscopy analysis, these tools are fundamentally different—that is, individual Ab is not a tool to identify a molecule but rather a tool for checking inconsistencies such as with differential tissue distribution of mRNA expression. The specificity of an Ab depends largely on the relative expression level of the target molecule. Therefore, it is critical to understand the efficacy and limitations of the Ab used in any experiments.

Recently, mass analysis tools have become remarkably advanced and widely available. Thus, it is now much easier to identify molecules that have been considered to be false targets due to nonspecific cross-reaction with Abs. These trials to identify the off-targets may provide byproducts of the excellent Abs, such as conformation-specific antibodies against the proteins.

As a more reliable anti-Bcnt/Cfdp1 Ab (e.g. A305-624A-M) becomes available, it is then possible to more accurately characterize endogenous Bcnt/Cfdp1, including subcellular localization and tissue distribution. On the other hand, the comparative transcriptome analysis of Cfdp1-K1 cells revealed that only a 25% decrease in Bcnt/Cfdp1 mRNA resulted in a marked up-regulation or down-regulation of many genes (Supplementary Tables S2 and S3). The finding is consistent with a report by Messina that suggests that extensive effects on vital cellular functions were caused by RNAi-depletion of *Bcnt/Cfdp1* in HeLa cells [34].

Recently, the budding yeast Swc5 has been clearly shown to preferentially bind to Histone H2A via the tandem DEF/Y motifs in the N-terminus [35], while an additional regulatory factor(s) might be involved in vertebrate Bcnt/Cfdp1. This result implicates that the broad effect of Bcnt/Cfdp1 deficit caused by RNAi-depletion [34] or shown as in genome-mediated knockdown of Cfdp1-K1 cells may be directly caused by impairing the histone exchange reaction of chromatin remodeling complex. However, the natural genetic variation also indicates that extensive abnormal expression of mRNA, antisense and noncoding RNAs may be caused by a mutation in the fission yeast *Swc5* that does not encode the DEF/Y motif in the N-terminus [14]. To further elucidate the functional role(s) of Bcnt/Cfdp1, comprehensive analyses are required not only—at the protein levels with more appropriate Abs but also at RNA and DNA levels—using different materials, such as knockdown and knockout cells or mice and further disease model animals.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This research was partially supported by a general maintenance grant from Tokushima Bunri University to Drs Kentaro Nakashima and Si-Young Song.

Author Contribution

S.I. and N.K. designed, performed the experiments, and wrote the manuscript. T.S. and N.D. performed mass analysis and interpretation. Y.K. performed the experiments and discussion. Y.O. contributed reagents and discussion. S.-Y.S. performed the experiments and wrote the manuscript.

Ethics Approval

This study was approved by the Genetic Recombination Experiment Safety Committee and the Animal Care and Use Committee of Tokushima Bunri University (Numbers, 28P-10, 30P-1, KP10-82-003, and KP18-81-007). Experiments with animals were conducted at Kagawa School of Pharmaceutical Sciences, Tokushima Bunri University.

Acknowledgements

We wish to thank Dr Kyoji Horie and Dr Eiji Kinoshita for technical suggestions on ES cell culture and Phos-tag agarose handling, respectively, Dr Kazuyoshi Hashizume for a kind gift of bovine placenta, Dr Wakako Kobayashi for sharing her interest in generating Bcnt/Cfdp1 knockout cells, and Dr Sang Wan Kim for discussion on the transcriptome analysis. We are grateful to Dr Salvador Eugenio Caoili, Dr Christopher A. MacRaid, and Dr Brian McWilliams for their particularly useful discussions and opinions. We also thank Dr Ed Luk and Dr Takashi Yasuda for discussions on the chromatin remodeling complex and on DNA damage repair, respectively. We were indebted to Professor Hiromi Nochi for her encouragement and to Dr Bradley L. Barnhart for his patient editing.

Abbreviations

Ab, antibody; anti-BCNT-C Ab, Ab generated using a peptide of 18 amino acid residues located in BCNT-C; BCNT-C, The C-terminal region of Bcnt/Cfdp1; Bcnt/Cfdp1, Bucentaur/Craniofacial developmental protein 1; F-mBcnt, Flag-tagged mouse Bcnt/Cfdp1; hBCNT/CFDP1, human BCNT/CFDP1; His-hBCNT/CFDP1, His-tagged hBCNT/CFDP1; mBcnt/Cfdp1, mouse Bcnt/Cfdp1; MCS, multi cloning site in a vector; SDS/PAGE, SDS-polyacrylamide gel electrophoresis; T-REx cell, a HEK293 cell derivative; WT, wild-type.

References

- Messina, G., Celauro, E., Atterato, M.T., Giordano, E., Iwashita, S. and Dimitri, P. (2015) The Bucentaur (BCNT) protein family: a long-neglected class of essential proteins required for chromatin/chromosome organization and function. *Chromosoma* **124**, 153–162, <https://doi.org/10.1007/s00412-014-0503-8>
- Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z. et al. (2012) A census of human soluble protein complexes. *Cell* **150**, 1068–1081, <https://doi.org/10.1016/j.cell.2012.08.011>
- Krogan, N.J., Keogh, M.C., Datta, N., Sawa, C., Ryan, O.W., Ding, H. et al. (2003) A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol. Cell* **12**, 1565–1576, [https://doi.org/10.1016/S1097-2765\(03\)00497-0](https://doi.org/10.1016/S1097-2765(03)00497-0)
- Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S. and Wu, C. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348, <https://doi.org/10.1126/science.1090701>

- 5 Nguyen, V.Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C. et al. (2013) Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. *Cell* **154**, 1220–1231, <https://doi.org/10.1016/j.cell.2013.08.018>
- 6 Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G. et al. (2005) Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nat. Struct. Mol. Biol.* **12**, 1064–1071, <https://doi.org/10.1038/nsmb1023>
- 7 Tramantano, M., Sun, L., Au, C., Labuz, D., Liu, Z., Chou, M. et al. (2016) Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *Elife* **5**, pii: e14243, <https://doi.org/10.7554/eLife.14243>
- 8 Lin, C.L., Chaban, Y., Rees, D.M., McCormack, E.A., Ocloo, L. and Wigley, D.B. (2017) Functional characterization and architecture of recombinant yeast SWR1 histone exchange complex. *Nucleic Acids Res.* **45**, 7249–7260, <https://doi.org/10.1093/nar/gkx414>
- 9 Morillo-Huesca, M., Clemente-Ruiz, M., Andújar, E. and Prado, F. (2010) The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. *PLoS ONE* **5**, e12143, <https://doi.org/10.1371/journal.pone.0012143>
- 10 Sun, L. and Luk, E. (2017) Dual function of Swc5 in SWR remodeling ATPase activation and histone H2A eviction. *Nucleic Acids Res.* **45**, 9931–9946, <https://doi.org/10.1093/nar/gkx589>
- 11 Messina, G., Damia, E., Fantì, L., Atterrato, M.T., Celauro, E., Mariotti, F.R. et al. (2014) Yeti, an essential *Drosophila melanogaster* gene, encodes a protein required for chromatin organization. *J. Cell Sci.* **127**, 2577–2588, <https://doi.org/10.1242/jcs.150243>
- 12 Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves, F.L., Wood, L., Chen, Z.A. et al. (2010) The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**, 810–821, <https://doi.org/10.1016/j.cell.2010.07.047>
- 13 Morozov, V.M., Giovannazzi, S. and Ishov, A.M. (2017) CENP-B protects centromere chromatin integrity by facilitating histone deposition via the H3.3-specific chaperone Daxx. *Epigenetics Chromatin* **10**, 63, <https://doi.org/10.1186/s13072-017-0164-y>
- 14 Clément-Ziza, M., Marsellach, F.X., Codlin, S., Papadakis, M.A., Reinhardt, S., Rodríguez-López, M. et al. (2014) Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Mol. Syst. Biol.* **10**, 764, <https://doi.org/10.15252/msb.20145123>
- 15 Liu, Y., Beyer, A. and Aebersold, R. (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550, <https://doi.org/10.1016/j.cell.2016.03.014>
- 16 Iwashita, S., Suzuki, T., Yasuda, T., Nakashima, K., Sakamoto, T., Kohno, T. et al. (2015) Mammalian Bcnt/Cfdp1, a potential epigenetic factor characterized by an acidic stretch in the disordered N-terminal and Ser250 phosphorylation in the conserved C-terminal regions. *Biosci. Rep.* **35**, pii: e00228, <https://doi.org/10.1042/BSR20150111>
- 17 Iwashita, S., Osada, N., Itoh, T., Sezaki, M., Oshima, K., Hashimoto, E. et al. (2003) A transposable element-mediated gene divergence that directly produces a novel type bovine Bcnt protein including the endonuclease domain of RTE-1. *Mol. Biol. Evol.* **20**, 1556–1563, <https://doi.org/10.1093/molbev/msg168>
- 18 Nakashima, K., Iwashita, S., Suzuki, T., Kato, C., Kohno, T., Kamei, Y. et al. (2019) A spatial similarity of stereochemical environments formed by amino acid residues defines a common epitope of two non-homologous proteins. *Sci. Rep.* **9**, 14818, <https://doi.org/10.1038/s41598-019-51350-2>
- 19 Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J. et al. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450, <https://doi.org/10.1093/nar/gky1106>
- 20 Horie, K., Kokubu, C., Yoshida, J., Akagi, K., Isotani, A., Oshitani, A. et al. (2011) A homozygous mutant embryonic stem cell bank applicable for phenotype-driven genetic screening. *Nat. Methods* **8**, 1071–1077, <https://doi.org/10.1038/nmeth.1739>
- 21 Kinoshita-Kikuta, E., Yamada, A., Inoue, C., Kinoshita, E. and Koike, E. (2011) A novel phosphate-affinity bead with immobilized Phos-tag for separation and enrichment of phosphopeptides and phosphoproteins. *JIOMICS* **1**, 157–169
- 22 Kobayashi, W. and Ozawa, M. (2018) The epithelial-mesenchymal transition induced by transcription factor LEF-1 is independent of β -catenin. *Biochem. Biophys. Rep.* **15**, 13–18
- 23 Iwashita, S., Ueno, S., Nakashima, K., Song, S.Y., Ohshima, K., Tanaka, K. et al. (2006) A tandem gene duplication followed by recruitment of a retrotransposon created the paralogous bucentaur gene (bcntp97) in the ancestral ruminant. *Mol. Biol. Evol.* **23**, 798–806, <https://doi.org/10.1093/molbev/msj088>
- 24 Iwashita, S., Nakashima, K., Sasaki, M., Osada, N. and Song, S.Y. (2009) Multiple duplication of the bucentaur gene family, which recruits the APE-like domain of retrotransposon: Identification of a novel homolog and distinct cellular expression. *Gene* **435**, 88–95, <https://doi.org/10.1016/j.gene.2009.01.012>
- 25 Vanli, G., Cuesta-Marban, A. and Widmann, C. (2017) Evaluation and validation of commercial antibodies for the detection of Shb. *PLoS ONE* **12**, e0188311, <https://doi.org/10.1371/journal.pone.0188311>
- 26 Qi, W., Davidson, B.A., Nguyen, M., Lindstrom, T., Grey, R.J., Burnett, R. et al. (2019) Validation of anti-glucocerebrosidase antibodies for western blot analysis on protein lysates of murine and human cells. *Biochem. J.* **476**, 261–274, <https://doi.org/10.1042/BCJ20180708>
- 27 Jank, T., Bogdanović, X., Wirth, C., Haaf, E., Spoerner, M., Böhmer, K.E. et al. (2013) A bacterial toxin catalyzing tyrosine glycosylation of Rho and deamidation of Gq and Gi proteins. *Nat. Struct. Mol. Biol.* **20**, 1273–1280, <https://doi.org/10.1038/nsmb.2688>
- 28 MacRaid, C.A., Richards, J.S., Anders, R.F. and Norton, R.S. (2016) Antibody Recognition of Disordered Antigens. *Structure* **24**, 148–157, <https://doi.org/10.1016/j.str.2015.10.028>
- 29 McIntush, E.W. (2013) Response: 'Antibody crossreactivity between the tumour suppressor PHLPP1 and the proto-oncogene β -catenin'. *EMBO Rep.* **14**, 494–496, <https://doi.org/10.1038/embor.2013.68>
- 30 Uhlen, M., Bandrowski, A., Carr, S., Edwards, J., Lundberg, E. et al. (2016) A proposal for validation of antibodies. *Nat. Methods* **13**, 823–827, <https://doi.org/10.1038/nmeth.3995>
- 31 Caoili, S.E. (2016) Expressing Redundancy among Linear-Epitope Sequence Data Based on Residue-Level Physicochemical Similarity in the Context of Antigenic Cross-Reaction. *Adv. Bioinformatics* 1276594
- 32 Dalkas, G.A. and Rooman, M. (2017) SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. *BMC Bioinformatics* **18**, 95, <https://doi.org/10.1186/s12859-017-1528-9>

- 33 Piotrowska, U. and Adler, G. (2010) Phosducin and monomeric β -actin have common epitope recognized by anti-phosducin antibodies. *Immunol. Lett.* **134**, 62–68, <https://doi.org/10.1016/j.imlet.2010.08.010>
- 34 Messina, G., Atterato, M.T., Prozzillo, Y., Piacentini, L., Losada, A. and Dimitri, P. (2017) The human Cranio Facial Development Protein 1 (Cfdp1) gene encodes a protein required for the maintenance of higher-order chromatin organization. *Sci. Rep.* **7**, 45022, <https://doi.org/10.1038/srep45022>
- 35 Huang, Y., Sun, L., Pierrakeas, L., Dai, L., Pan, L., Luk, E. et al. (2020) Role of a DEF/Y motif in histone H2A-H2B recognition and nucleosome editing. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 3543–3550, <https://doi.org/10.1073/pnas.1914313117>

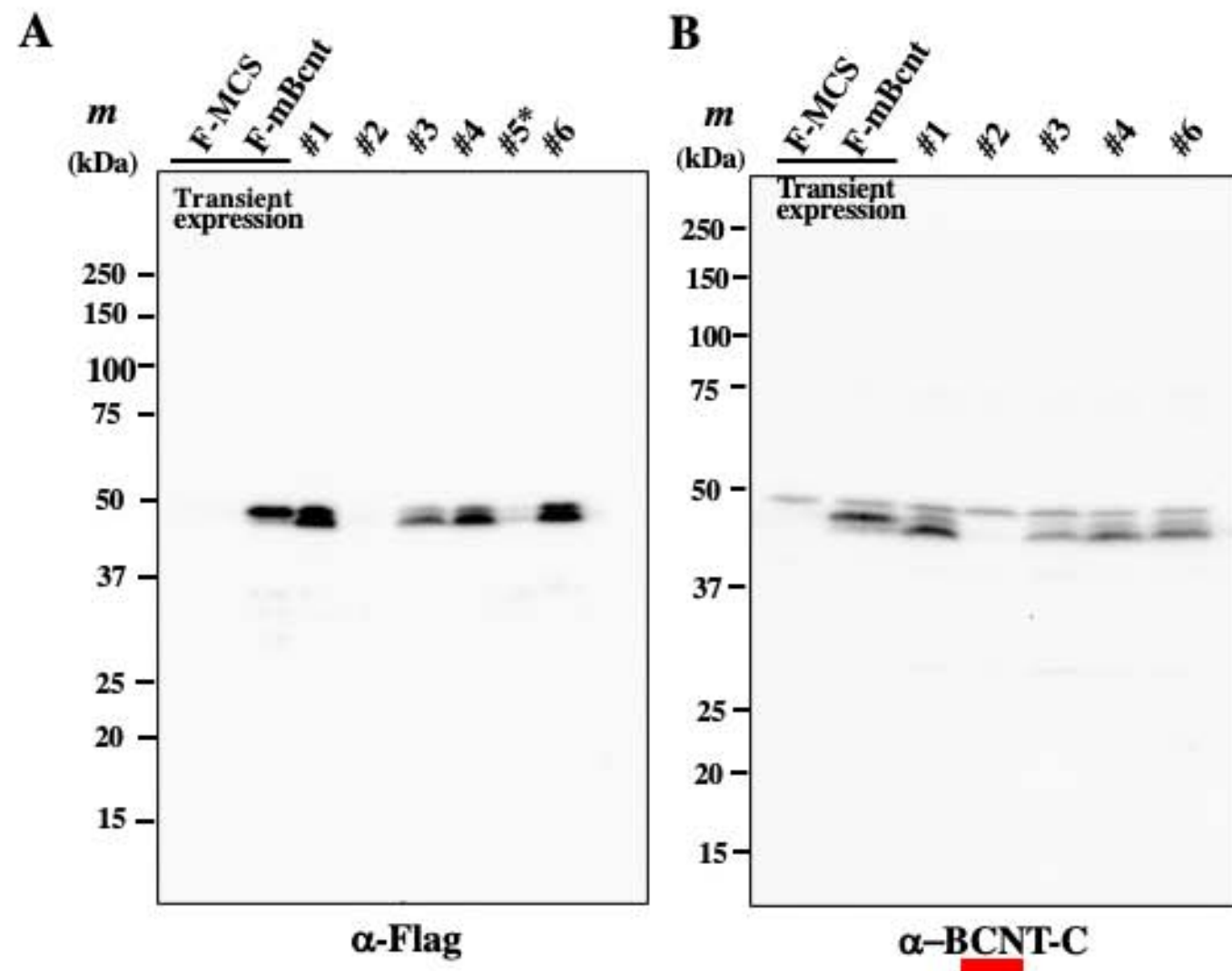
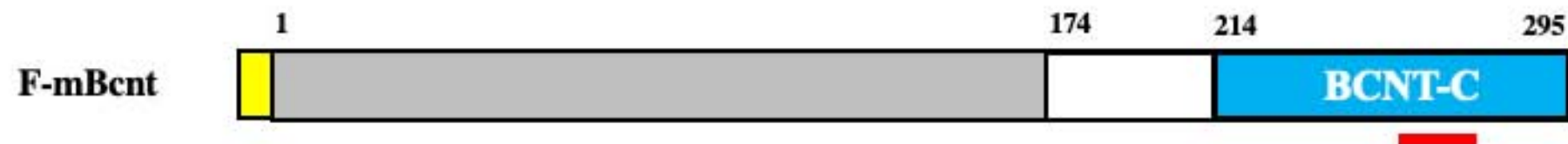


Figure S1. Screening of F-mBcnt expressing T-REx cell colonies by Western blot.

Top panel: Molecular architecture of Flag-tagged mouse Bcnt/Cfdp1. The protein shown in a large outline comprises the acidic N-terminal region, Lys/Glu/Pro-rich 40 amino acids(white box), a highly conserved C-terminal region (BCNT-C domain, blue box), and Flag-tag at the N-terminus (yellow box). The numbers above the outline show the amino acid residues of mBcnt/Cfdp1. The underlined red bar presents the location of the immunogen for generation of anti-BCNT-C Ab. Cell extracts of several G418 resistant colonies from F-mBcnttransfectants(#1~6), and of MCS- or F-mBcnt transiently expressed cultures (all equivalent to $\sim 5 \times 10^4$ cells per lane, shown at the top of each lane), were subjected to Western blot analysis using anti-Flag Tag Ab (A) or anti-BCNT-C Ab (B). #5* indicates a loss of the sample during application to the gel (A).

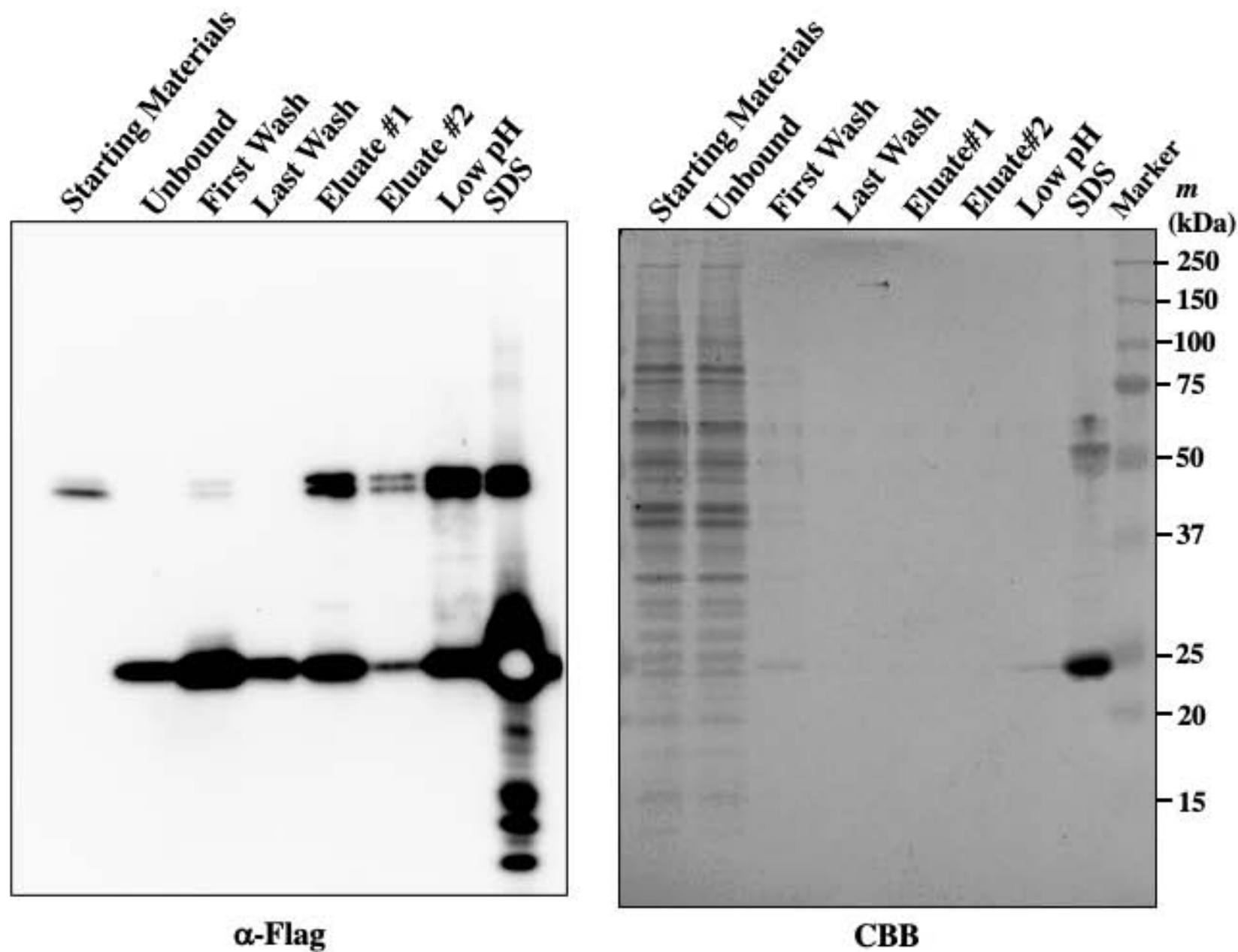


Figure S2. Isolation of F-mBent from T-RExcolony #1 using anti-Flag Ab-conjugated agarose beads.

The supernatant of the #1 colony extract was mixed with anti-Flag-tag antibody-conjugated agarose beads and incubated. The bound fraction was washed, eluted with Flag (DYKDDDDK) peptide sequentially (Eluate #1 and #2), and followed by glycine-HCl, pH 2.5 buffer (Low pH). Finally, the agarose was boiled in SDS/PAGE sample buffer and its supernatant was obtained (SDS). All of the samples were subjected to Western blot analysis with anti-Flag Tag Ab followed by HRP-conjugated anti-mouse IgG light chain specific Ab. Finally, the filter was stained with Coomassie Brilliant Blue (CBB, right panel).

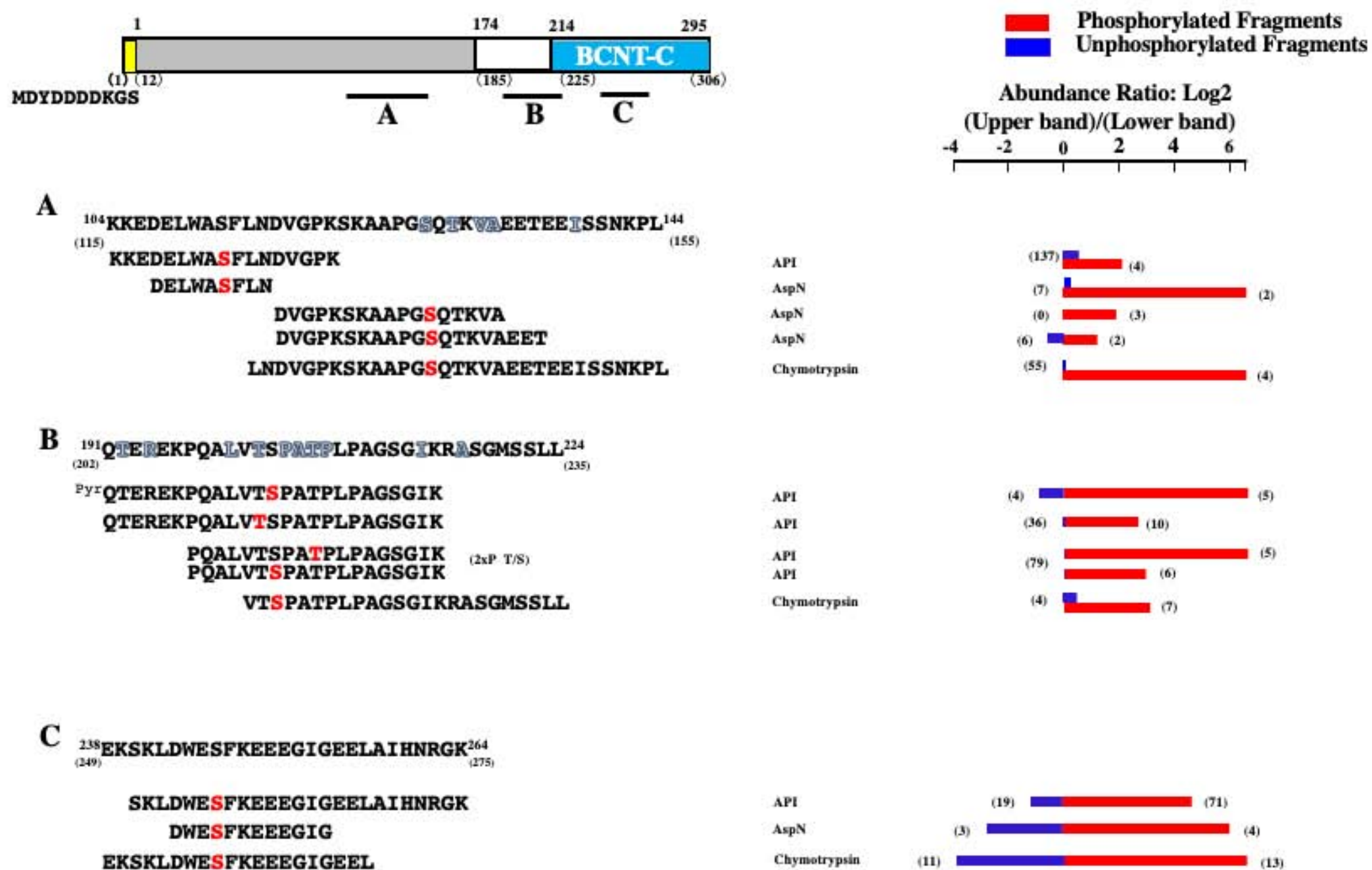


Figure S3. Differential phosphorylation between the upper and lower bands of F-mBcnt.

Top panel: Molecular architecture of Flag-tagged mouse Bcnt/Cfdp1 as shown in Supplementary figure S1. The numbers above and below the F-mBcnt structure show its amino acid residues without and with Flag tag, respectively. Three black bars A, B, and C indicate the grouping regions of focused phosphorylation sites. Each upper and lower band of F-mBcnt from Fig. 1B was digested with three proteases (API, AspN, and Chymotrypsin) and subjected to LC-MS/MS analysis. All of the identified peptides are listed in Supplementary Table S1, and their typical phosphorylated fragments and their unphosphorylated counterparts are represented in this figure. Amino acid sequence shown in each top line of A, B, and C, represents each focused region, and their numbers at the N-terminus and the C-terminus correspond to the amino acid residue of mBcnt, respectively. **Xs** are different amino acid residues from human BCNT/CFDP1 and red letters show the identified phosphorylated sites. In the right side of the panel, red bars indicate the ratios of the amounts of identified phosphorylated fragments in the upper band compared to those in the lower band, and blue bars show the corresponding ratios of their unphosphorylated fragments, respectively. The numbers of peptide-spectrum match (PSM) values are presented in parentheses.

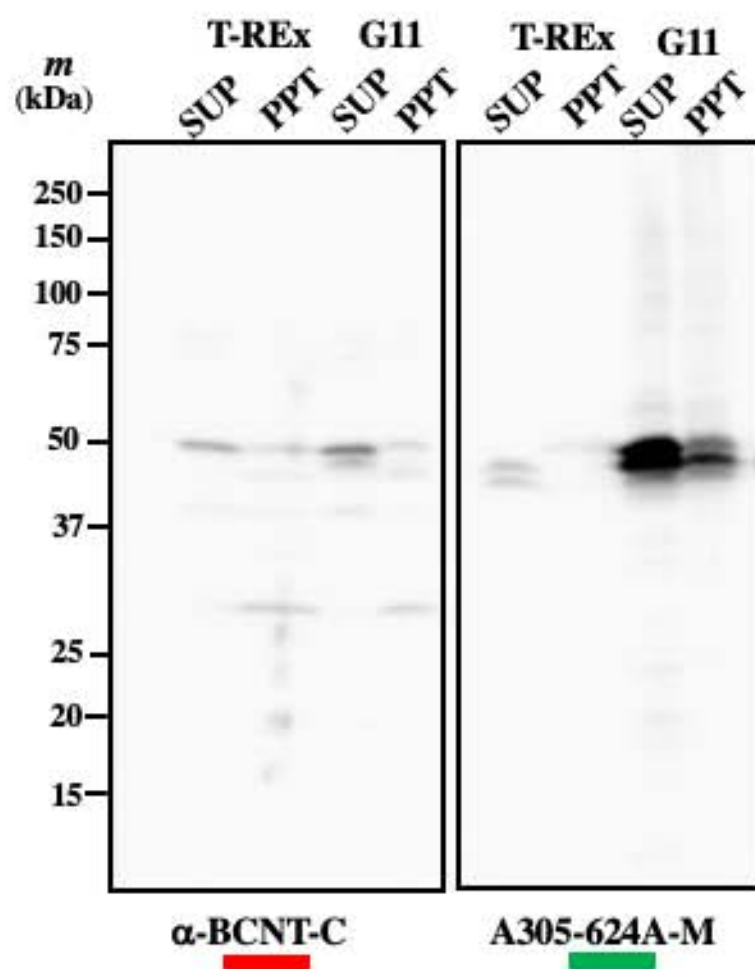
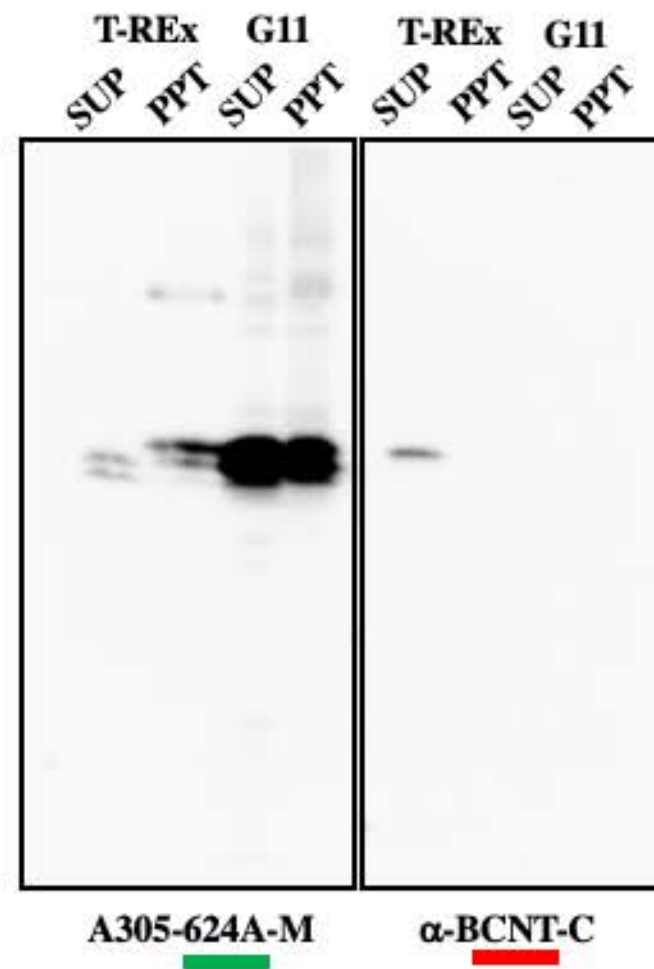
A**B**

Figure S4. Comparative assessment of western blot signals between the anti-BCNT-C antibody and A305-624A-M.

The supernatants (SUP) and their pellets (PPT) of T-REx or G11 cells (His-tagged human BCNT/CFDP1 constitutively expressing clone) were prepared from each cell lysate by centrifugation at 25000 x g. Equal amounts of protein (20 µg) were subjected to a western blot analysis with either anti-BCNT-C Ab or A305-624A-M (**A**). After obtaining their images, each filter was stripped and reprobed with the exchanged Abs (**B**).

Sup Fig S5

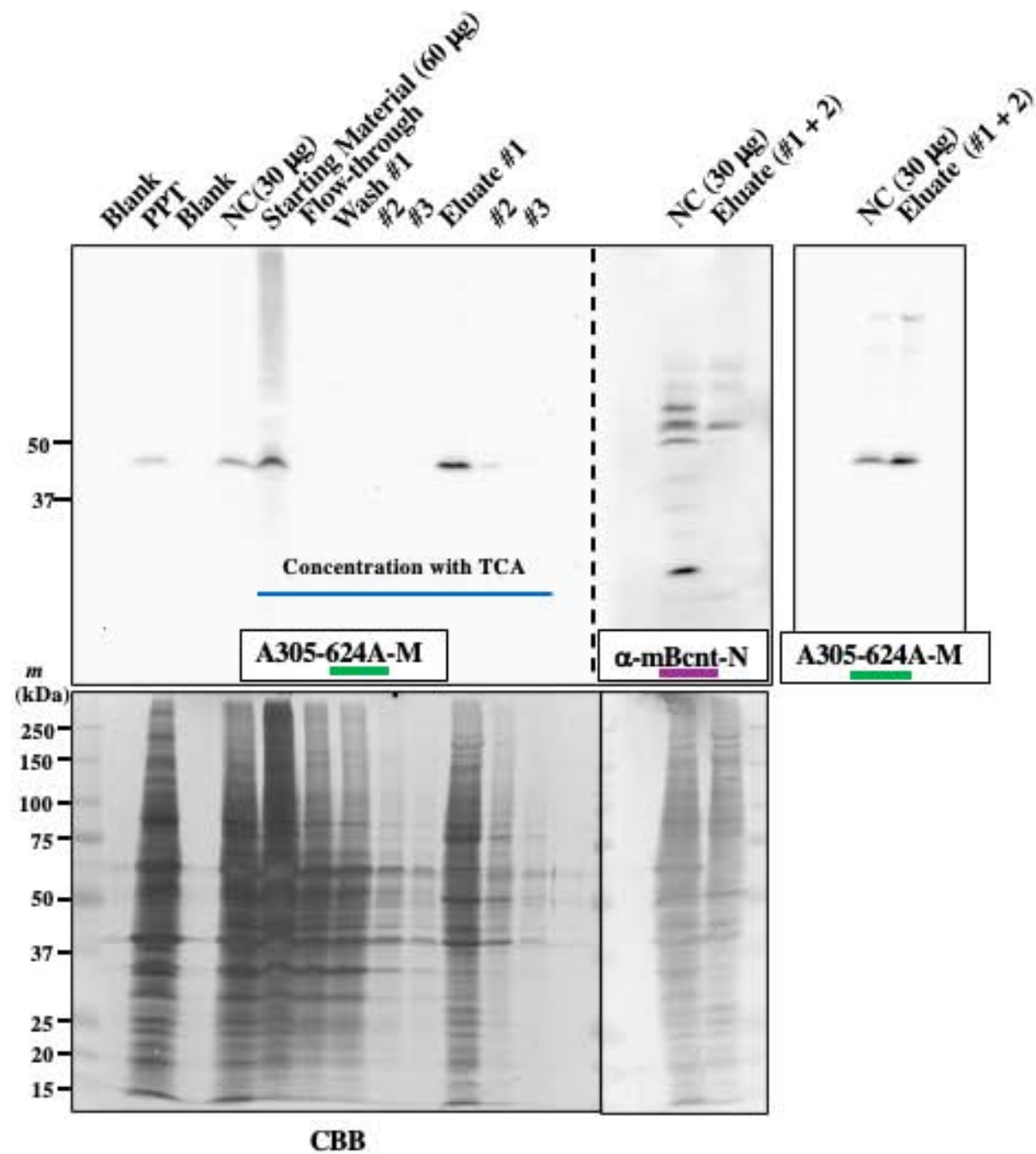


Figure S5. Enrichment of bovine Bcnt/Cfdp1 with Phos-tag agarose.

The supernatant of the bovine placenta extract in RIPA buffer (Starting material, SM) was mixed with Phos-tag agarose and incubated. After washing three times (Wash #1, 2, 3), the bound portion was sequentially eluted (Eluate #1, 2, 3), and each sample was concentrated with TCA. On the other hand, the pellet fraction, which contained ~9-fold concentrated protein compared to the supernatant, was prepared (PPT). All of the samples were boiled in SDS/PAGE sample buffer and then subjected to a western blot analysis with A305-624-M (top, left filter). The samples before and after Phos-tag treatment, corresponding to Non-concentrated (NC) and the pooled Eluates (#1 / 2), respectively, were also analyzed with anti-mBcnt-N Ab (top, middle filter) followed by reprobing with A305-624-M (top, right filter). Finally, the filters were stained with Coomassie Brilliant Blue (CBB, bottom panel).

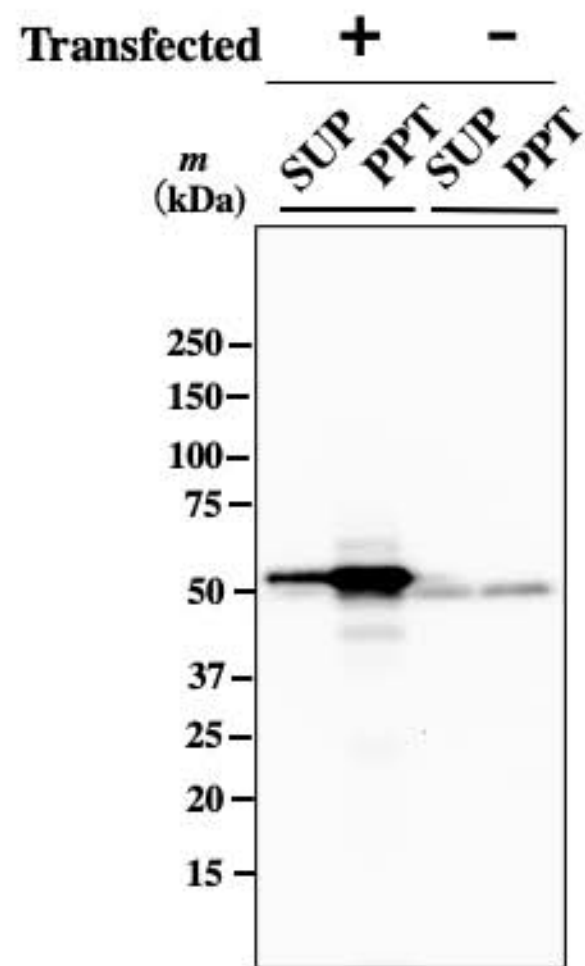
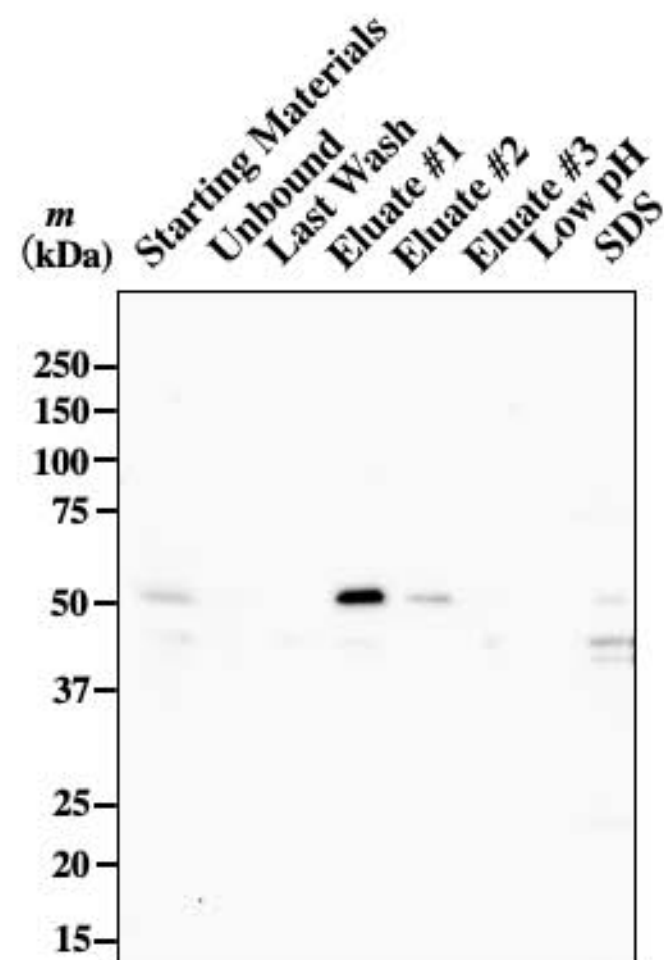
A**B**

Figure S6. Isolation of Nono (p54nrb) using anti-HisAb-conjugated agarose.

(A) Evaluation of anti-p54nrb Ab. T-REx 293 Clone G11 cells were transfected with Hemagglutinin-tagged P54nrb pcDNA3.1 (a gift from Dr. Atsushi Yokoyama, Tohoku University, Sendai) and the cell lysates were prepared after culturing for total 70 h, and centrifuged at 25000 x g, 30 min. The protein ratio between the supernatant (SUP) and the particulate (PPT) was 8:1. Their constant amounts of protein (18 µg/lane) were subjected to a Western blot analysis with anti-p54nrb antibody (250 ng/mL, GeneTex, GTX101419). **(B)** Isolation of Nonowith anti-His tag antibody-conjugated agarose beads. The supernatant of T-REx cell lysate was applied to anti-His-tag antibody-conjugated agarose beads, and after washing, the bound proteins to agarose were eluted with His-tag peptide as previously described [16], followed by with glycine-HCl buffer (pH 2.5), and finally boiled in SDS/PAGE sample buffer as shown in Supplementary figure S2. All of the samples were treated in SDS/PAGE sample buffer and subjected to a western blot analysis as described in (A).

Mus musculus Transcriptome Sequencing Report

March 2019



Project Information

Client Name	MacroGen Japan
Company/Institution	MacroGen Corp. Japan
Order Number	HN00101712
Species	<i>Mus musculus</i>
Reference	UCSC mm10
Annotation	RefSeq_2017_06_12
Read Length	101
Number of Samples	2
Library Kit	TruSeq Stranded mRNA LT Sample Prep Kit
Library Protocol	TruSeq Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E
Reagent	NovaSeq 6000 S4 Reagent Kit
Sequencing Protocol	NovaSeq 6000 System User Guide Document # 1000000019358 v02
Type of Sequencer	NovaSeq
Sequencing Control Software	1000000019358 v02

Project Results Summary

In this study, *Mus musculus* whole transcriptome sequencing was performed in order to examine the different gene expression profiles, and to perform gene annotation on set of useful genes based on gene ontology pathway information.

Analyses were successfully performed on all 2 paired-ends samples. Figure 1 shows the throughput of raw data and trimmed data. Figure 2 shows the Q30 percentage (% of bases with quality over phred score 30) of each sample's raw and trimmed data.

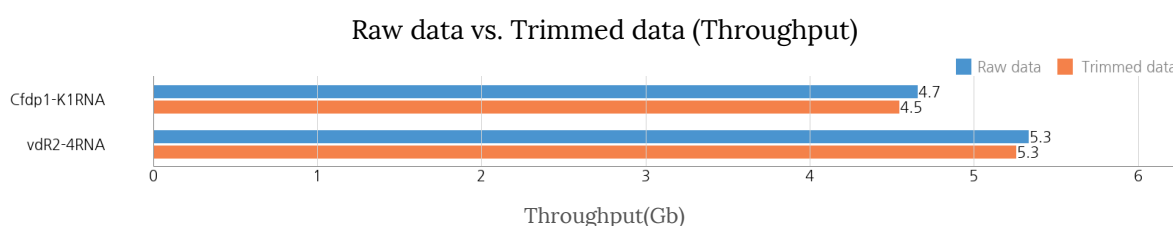


Figure 1. Throughput output of Raw and Trimmed data

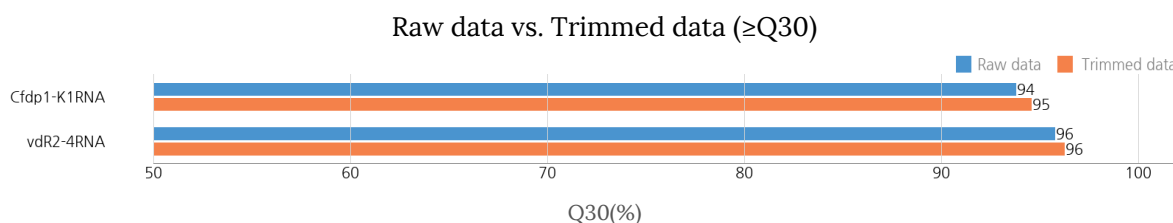


Figure 2. Q30 score of Raw and Trimmed data

Trimmed reads are mapped to reference genome with HISAT2. Figure 3 shows the overall read mapping ratio, the ratio of mapped reads to trimmed reads.

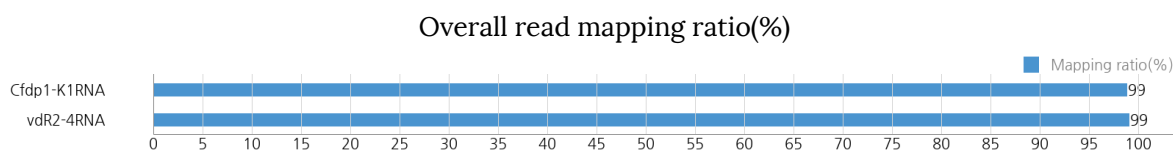


Figure 3. Overall read mapping ratio(%)

After the read mapping, Stringtie was used for transcript assembly. Expression profile was calculated for each sample and transcript/gene as read count and FPKM (Fragment per Kilobase of transcript per Million mapped reads).

DEG (Differentially Expressed Genes) analysis was performed on a comparison pair (Cfdp1-K1RNA_vs_vdR2-4RNA) as requested using FPKM. The results showed 694 genes which satisfied $|fc| \geq 2$ conditions in comparison pair.

Figure 4 shows the result of hierarchical clustering (distance metric= Euclidean distance, linkage method= complete) analysis. It graphically represents the similarity of expression patterns between samples and genes.

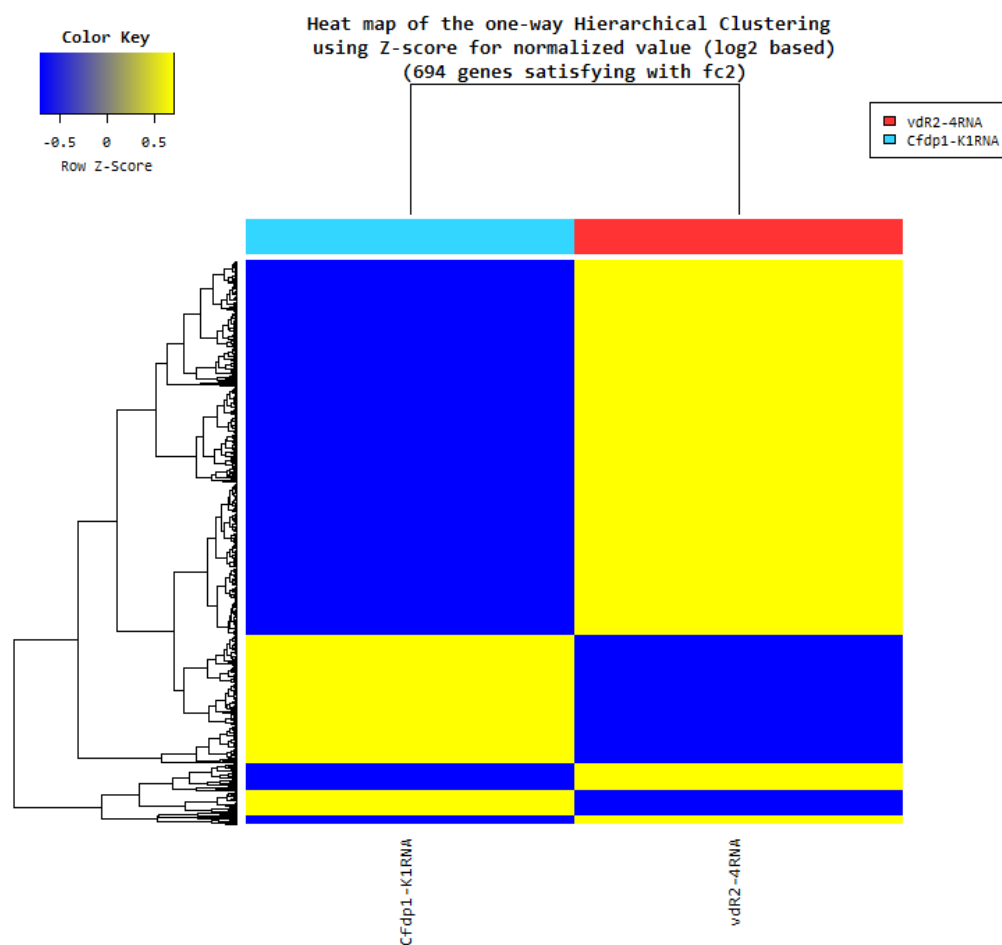


Figure 4. Heatmap for DEG list

DEG list was further analyzed with Gene Ontology (<http://geneontology.org/>) for gene set enrichment analysis per biological process (BP), cellular component (CC) and molecular function (MF). The Figure 5, 6 and 7 show the significant gene set by each category.

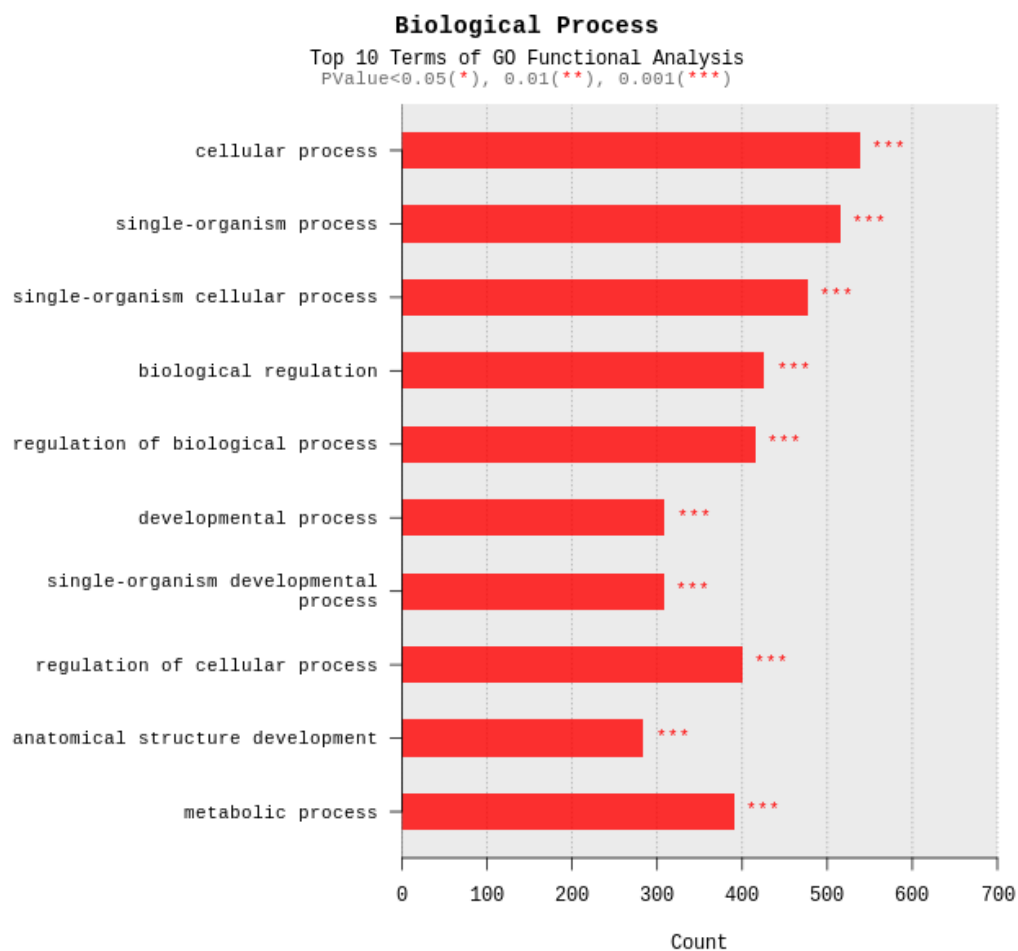


Figure 5. Gene Ontology terms related to Biological Process

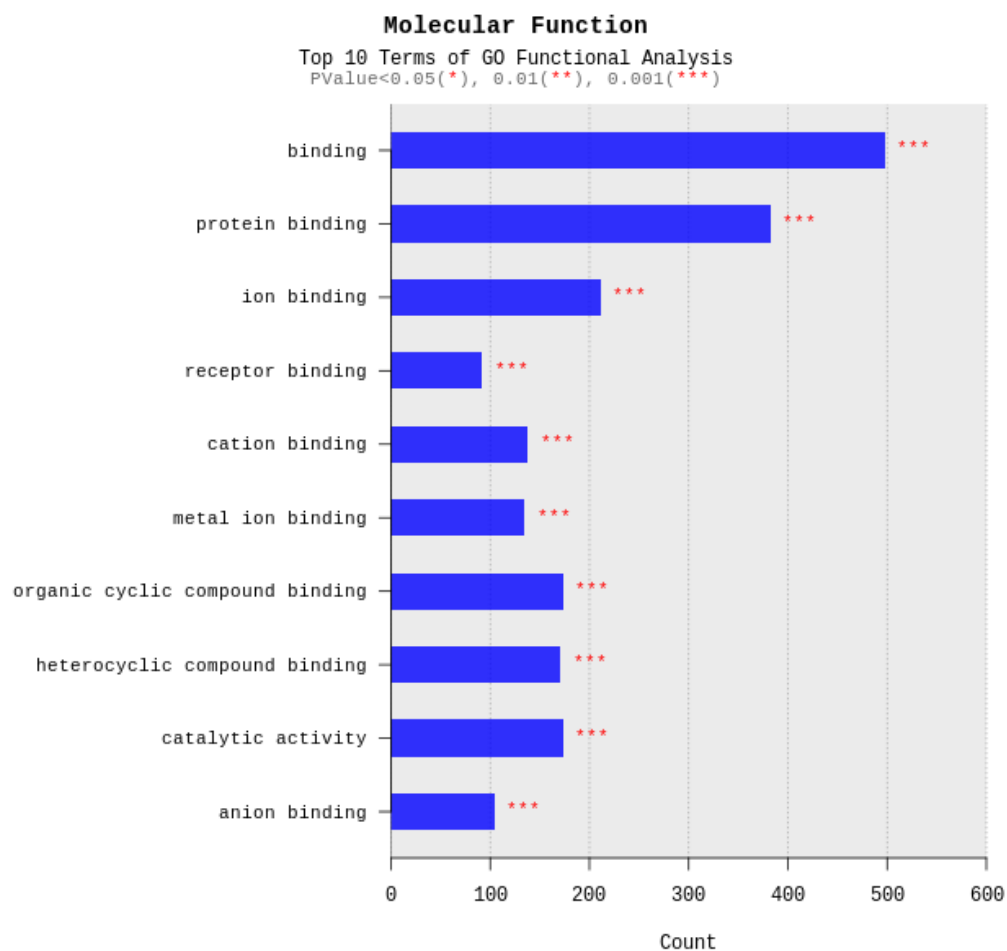


Figure 6. Gene Ontology Terms related to Molecular Function

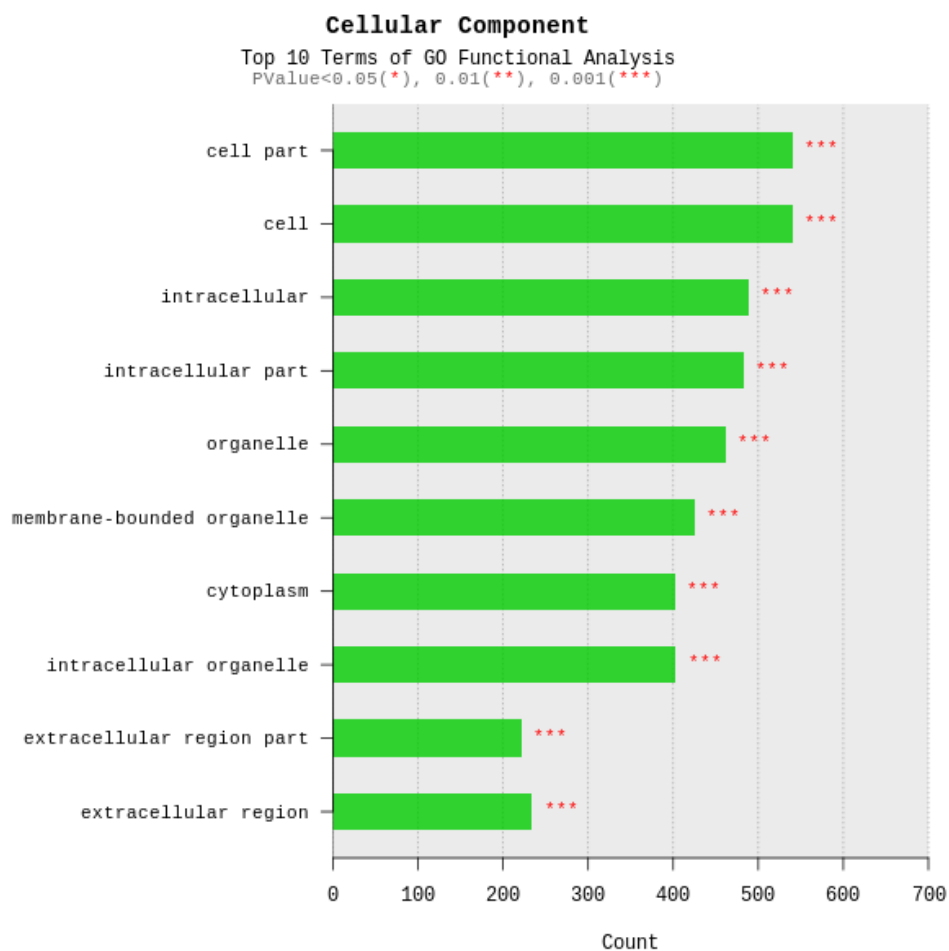


Figure 7. Gene Ontology Terms related to Cellular Component

Table of Contents

Project Information	02
Project Results Summary	03
1. Experimental Methods and Workflow	10
2. Analysis Methods and Workflow	11
3. Summary of Data Production	12
3. 1. Raw Data Statistics	12
3. 2. Average Base Quality at Each Cycle	13
3. 3. Trimming Data Statistics	14
3. 4. Average Base Quality at Each Cycle after Trimming	15
4. Reference Mapping and Assembly Results	16
4. 1. Mapping Data Statistics	16
4. 2. Expression Profiling	17
5. Differentially Expressed Gene Analysis Results	19
5. 1. Data Analysis Quality Check and Preprocessing	19
5. 2. Differentially Expressed Gene Analysis Workflow	24
5. 3. Significant Gene Results	25
5. 4. GO Enrichment Analysis	30
6. Data Download Information	36
6. 1. Raw Data	36
6. 2. Analysis Results	36
7. Appendix	39
7. 1. Phred Quality Score Chart	39
7. 2. Programs used in Analysis	40
7. 3. References	41

1. Experimental Methods and Workflow

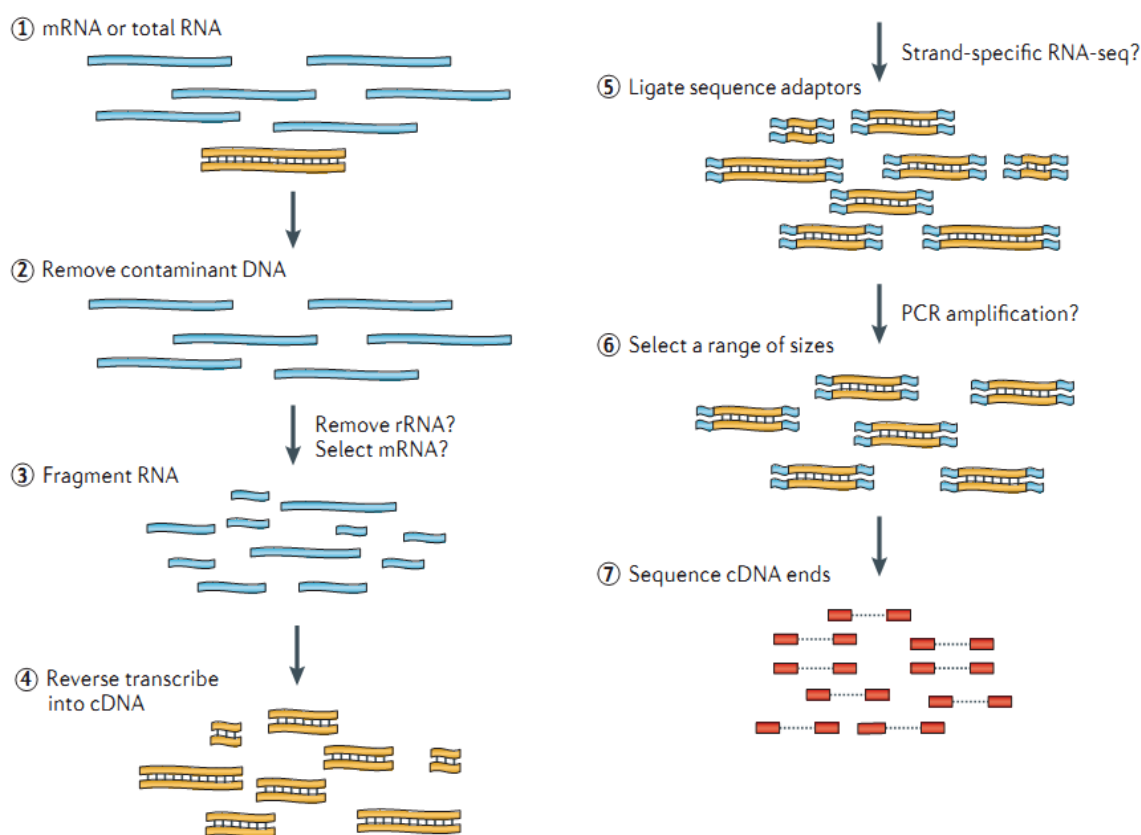


Figure 8. RNA Sequencing Experiment Workflow

REFERENCE • Nat Rev Genet. 2011 Sep 7;12(10):671-82

- 1) Isolate the Total RNA from Sample of interest (Cell or Tissue).
- 2) Eliminate DNA contamination using DNase.
- 3) Choose an appropriate kit for library prep process depending on the types of RNA. For mRNA with poly-A tail, use mRNA purification kit; for non-coding RNAs, such as lincRNA, use ribo-zero RNA removal Kit to purify RNA of interest.
- 4) Randomly fragment purified RNA for short read sequencing.
- 5) Reverse transcribe fragmented RNA into cDNA.
- 6) Ligate adapters onto both ends of the cDNA fragments.
- 7) After amplifying fragments using PCR, select fragments with insert sizes between 200-400 bp. For paired-end sequencing, both ends of the cDNA is sequenced by the read length.

2. Analysis Methods and Workflow



Figure 9. Analysis Workflow

- 1) Analyze the quality control of the sequenced raw reads. Overall reads' quality, total bases, total reads, GC (%) and basic statistics are calculated.
- 2) In order to reduce biases in analysis, artifacts such as low quality reads, adaptor sequence, contaminant DNA, or PCR duplicates are removed.
- 3) Trimmed reads are mapped to reference genome with HISAT2, splice-aware aligner.
- 4) Transcript is assembled by StringTie with aligned reads.
- 5) Expression profiles are represented as read count and normalization value which is based on transcript length and depth of coverage. The FPKM (Fragments Per Kilobase of transcript per Million Mapped reads) value or the RPKM (Reads Per Kilobase of transcript per Million mapped reads) is used as a normalization value.
- 6) In groups with different conditions, genes or transcripts that express differentially are filtered out through statistical hypothesis testing.
- 7) In case of known gene annotation, functional annotation and gene-set enrichment analysis are performed using GO and KEGG database on differentially expressed genes.

3. Summary of Data Production

3.1. Raw Data Statistics

(Refer to Path: result_RNAseq/Analysis_statistics/rawData/raw_throughput.stats)

The total number of bases, reads, GC (%), Q20 (%), Q30 (%) are calculated for 2 samples. For example, in Cfdp1-K1RNA, 46,100,372 reads are produced, and total read bases are 4.7Gbp. The GC content (%) is 49.57% and Q30 is 93.77%.

Table 1. Raw data stats

Index	Sample id	Total read bases*	Total reads	GC (%)	Q20 (%)	Q30 (%)
1	Cfdp1-K1RNA	4,656,137,572	46,100,372	49.57	97.87	93.77
2	vdR2-4RNA	5,331,793,636	52,790,036	50.52	98.66	95.75

(* 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

3. 2. Average Base Quality at Each Cycle

(Refer to Path: Analysis_statistics/rawData/A_fastqc/)

The quality of produced data is determined by the phred quality score at each cycle. Box plot containing the average quality at each cycle is created with FastQC.

The x-axis shows number of cycles and y-axis shows phred quality score. Phred quality score 20 means 99% accuracy and reads over score of 20 are accepted as good quality.

LINK <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

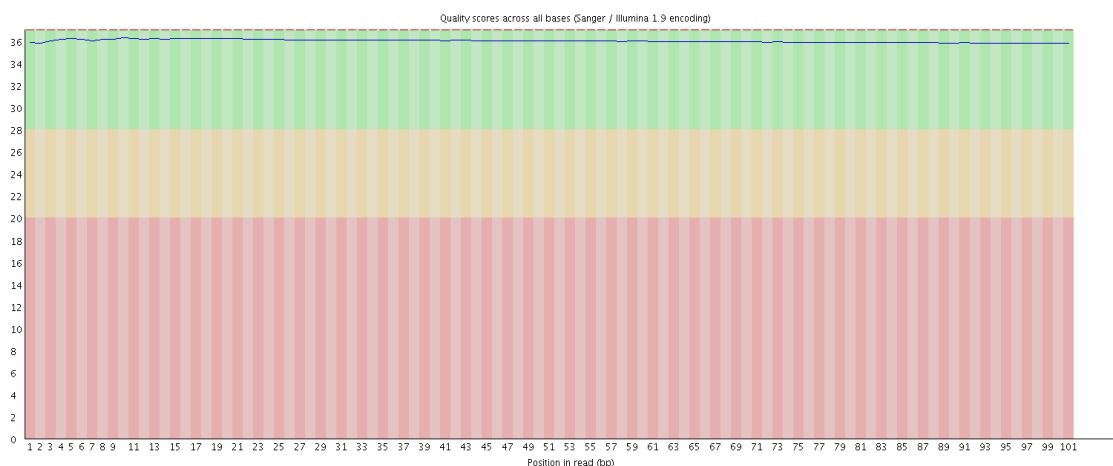


Figure 10. Read quality at each cycle of Cfdp1-K1RNA (read1)

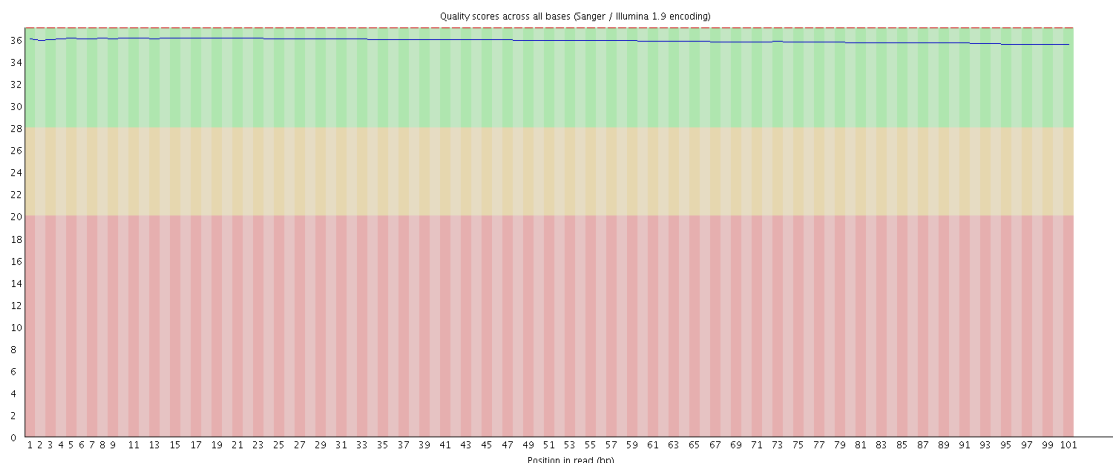


Figure 11. Read quality at each cycle of Cfdp1-K1RNA (read2)

- 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

3. 3. Trimming Data Statistics

(Refer to Path: result_RNAseq/Analysis_statistics/trimmedData/trim_throughput.stats)

Trimmomatic program is used to remove adapter sequences and bases with base quality lower than three from the ends. Also using sliding window method, bases of reads that does not qualify for window size 4, and mean quality 15 are trimmed. Afterwards, reads with length shorter than 36bp are dropped to produce trimmed data.

Table 2. Trimming Data Stats

Index	Sample id	Total read bases	Total reads	GC(%)	Q20(%)	Q30(%)
1	Cfdp1-K1RNA	4,543,927,517	45,262,214	49.59	98.39	94.57
2	vdR2-4RNA	5,253,493,756	52,245,252	50.53	98.99	96.25

- Total read bases: Total number of read bases after trimming
- Total reads: Total number of reads after trimming
- 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

3. 4. Average Base Quality at Each Cycle after Trimming

(Refer to Path: result_RNAseq/Analysis_statistics/trimmedData/A_fastqc/)

Figure 12 and 13 show average base quality at each cycle after trimming.

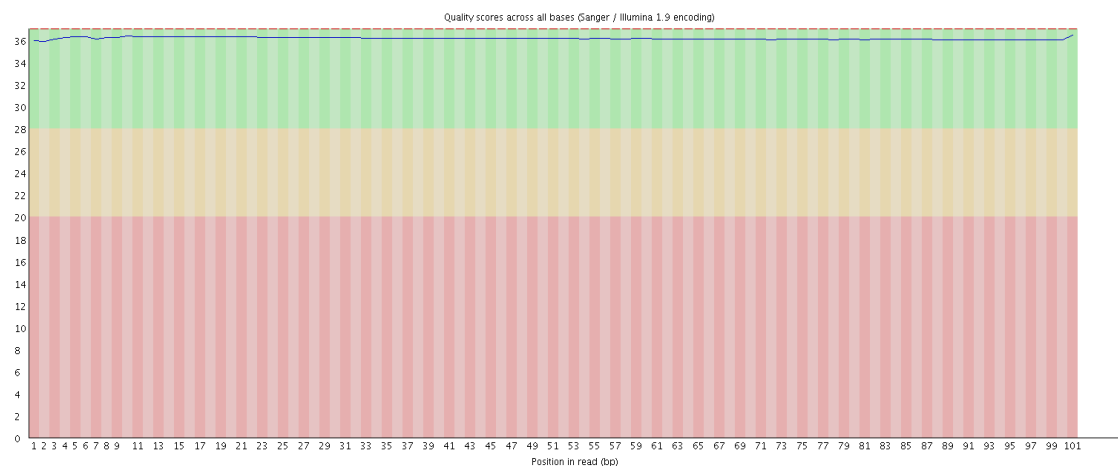


Figure 12. Average base quality of Cfdp1-K1RNA (read1) at each cycle after trimming

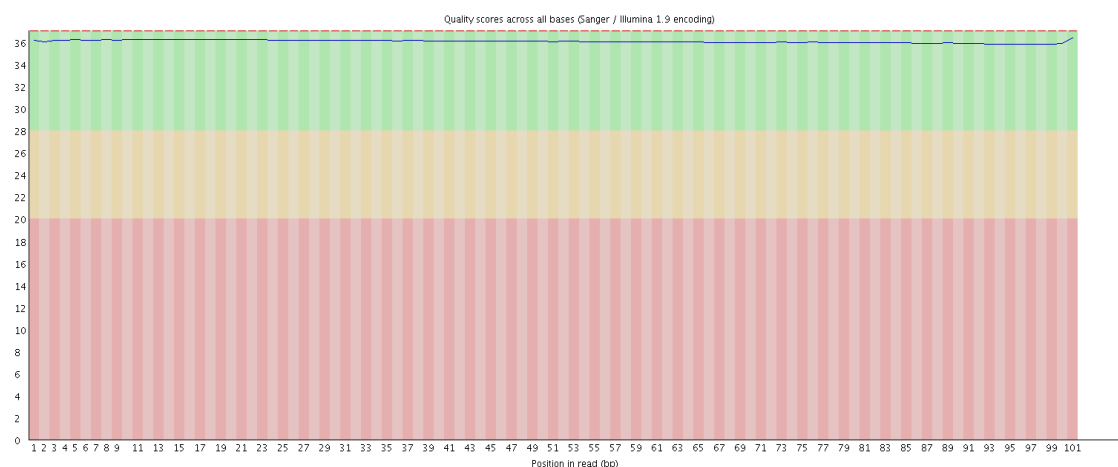


Figure 13. Average base quality of Cfdp1-K1RNA (read2) at each cycle after trimming

- 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

4. 1. Mapping Data Statistics

(Refer to Path: result_RNAseq/Analysis_statistics/mapping.hisat.stats)

In order to map cDNA fragments obtained from RNA sequencing, UCSC mm10 was used as a reference genome. Table 3 shows the statistic obtained from HISAT2, which is known to handle spliced read mapping through Bowtie2 aligner. You can check number of processed reads, mapped reads.

Table 3. Mapped Data Stats

Sample ID	# of processed reads	# of mapped reads (%)	# of unmapped reads (%)
Cfdp1-K1RNA	45,262,214	44,725,812 (98.81%)	536,402 (1.19%)
vdR2-4RNA	52,245,252	51,767,696 (99.09%)	477,556 (0.91%)

- Processed reads: Number of cleaned reads after trimming
- Mapped reads: Number of reads mapped to reference
- Unmapped reads: Number of reads that failed to align

4. 2. Expression Profiling

Known genes and transcripts are assembled with StringTie based on reference genome model.

After assembly, the abundance of gene/transcript is calculated in the read count and normalized value as FPKM (Fragments Per Kilobase of transcript per Million mapped reads) for a sample.

4. 2. 1. Known Transcripts Expression Level

(Refer to Path: result_RNAseq_excel/Expression_profile/StringTie/Expression_Profile.mm10.transcript.xlsx)

Table 4 is an example of known transcript expression level per sample in expression value. This result is obtained by -e option of StringTie does not consider novel transcript assembly.

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
NM_001302545	14	AAMP	angio associated migratory cell protei	chr2:219128852-219134	1835	898	987	12.220251	12.415353
NM_001087	14	AAMP	angio associated migratory cell protei	chr2:219128852-219134	1832	4678	6437	63.774269	81.140015
NM_001166579	15	AANAT	aralkylamine N-acetyltransferase, tra	chr17:74449433-744661	1913	46	30	0.599741	0.352587
NR_110548	15	AANAT	aralkylamine N-acetyltransferase, tra	chr17:74463630-744661	1082	9	9	0.192813	0.186779
NM_001101	60	ACTB	actin beta	chr7:5566779-5570232	1812	93591	129901	1290.007935	1655.640503
NM_001161572	23764	MAFF	MAF bZIP transcription factor F, trans	chr22:38597939-386125	2465	1	150	0.002107	1.397431
NM_012323	23764	MAFF	MAF bZIP transcription factor F, trans	chr22:38597939-386125	2439	1682	2109	17.222849	19.96483
NM_001161574	23764	MAFF	MAF bZIP transcription factor F, trans	chr22:38597939-386125	2372	0	0	0	0
NM_001161573	23764	MAFF	MAF bZIP transcription factor F, trans	chr22:38599027-386125	2223	44	25	0.485203	0.252227
NM_001289905	23765	IL17RA	interleukin 17 receptor A, transcript v	chr22:17565849-175965	8506	1303	975	3.825815	2.644646
NM_014339	23765	IL17RA	interleukin 17 receptor A, transcript v	chr22:17565849-175965	8608	3241	1998	9.402107	5.359576
NR_028287	23766	GABARAPL3	GABA type A receptor associated pr	chr15:90889763-908926	1885	3	6	0.036076	0.073511
NM_001017526	23779	ARHGAP8	Rho GTPase activating protein 8, tra	chr22:45148438-452586	1725	460	641	6.645803	8.576918
NM_181335	23779	ARHGAP8	Rho GTPase activating protein 8, tra	chr22:45148438-452586	1632	1979	2405	30.27355	34.027134
NM_001198726	23779	ARHGAP8	Rho GTPase activating protein 8, tra	chr22:45148438-452586	1528	84	59	1.366953	0.889118
NM_030882	23780	APOL2	apolipoprotein L2, transcript variant a	chr22:36622255-366356	2545	559	1155	5.482551	10.474212
NM_145637	23780	APOL2	apolipoprotein L2, transcript variant b	chr22:36622255-366360	2686	1212	0	11.260728	0

- 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

4. 2. 2. Known Genes Expression Level

(Refer to Path: result_RNAseq_excel/Expression_profile/StringTie/
Expression_Profile.mm10.gene.xlsx)

Table 5 is an example of known gene expression level per sample in expression value. This result is obtained by -e option of StringTie does not consider novel transcript assembly.

Table 5. Known genes Expression Level (example)

Gene_ID	Transcript_ID	Gene_Symbol	Description	AM Read_Count	BM Read_Count	AM_FPKM	BM_FPKM
60	NM_001101	ACTB	actin beta	93591	129901	1290.007935	1655.640503
70	NM_005159	ACTC1	actin, alpha, cardiac muscle 1	20	6	0.1339	0.031949
175	NM_000027, NM_001171988, NR_011013	AGA	aspartylglucosaminidase	252	279	2.995219	3.071083
176	NM_001135, NM_013227	ACAN	aggrecan	8	0	0.022519	0
177	NM_001136, NM_001206929, NM_001206930	AGER	advanced glycosylation end-product specific	3332	3124	51.224842	44.355004
178	NM_000028, NM_000642, NM_000643	AGL	amylase, alpha-1, 6-glucosidase, 4-alpha-gluc	4919	3679	16.662192	11.52329
191	NM_000687, NM_001161766, NM_001161767	AHCY	adenosylhomocysteinase	12053	13891	129.59984	138.005572
245	NR_002710, NR_120453	ALOX12P2	arachidonate 12-lipoxygenase pseudogene	8	5	0.070872	0.041258
246	NM_001140	ALOX15	arachidonate 15-lipoxygenase	785	710	7.302354	6.108678
247	NM_001039130, NM_001039131, NM_001039132	ALOX15B	arachidonate 15-lipoxygenase, type B	6	0	0.049592	0
248	NM_001831	ALPI	alkaline phosphatase, intestinal	13	3	0.098671	0.021092
249	NM_000478, NM_001127501, NM_001127502	ALPL	alkaline phosphatase, liver/bone/kidney	9	19	0.085416	0.164094
250	NM_001632	ALPP	alkaline phosphatase, placental	464	142	3.894943	1.098701
251	NM_031313	ALPPL2	alkaline phosphatase, placental like 2	88	12	0.876858	0.106491
257	NM_006492	ALX3	ALX homeobox 3	310	319	5.229297	4.975804
258	NM_016519	AMBN	ameloblastin	0	0	0	0
259	NM_001633	AMBP	alpha-1-microglobulin/bikunin precursor	0	0	0	0

- 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

5.1. Data Analysis Quality Check and Preprocessing

There is a process that sorts differentially expressed gene among samples by FPKM value of known genes. In preprocessing, there are data quality and similarity checks among samples in case of biological replicates exist.

(Refer to Path: result_RNAseq_excel/DEG_result/Analysis_Result.html)

5.1.1. Sample Information and Analysis Design

Total of 2 samples was used for analysis. For more information of samples and comparison pair, please refer to Sample.Info.txt file.

Index	Sample.ID	Sample.Group
1	vdR2-4RNA	vdR2-4RNA
2	Cfdp1-K1RNA	Cfdp1-K1RNA

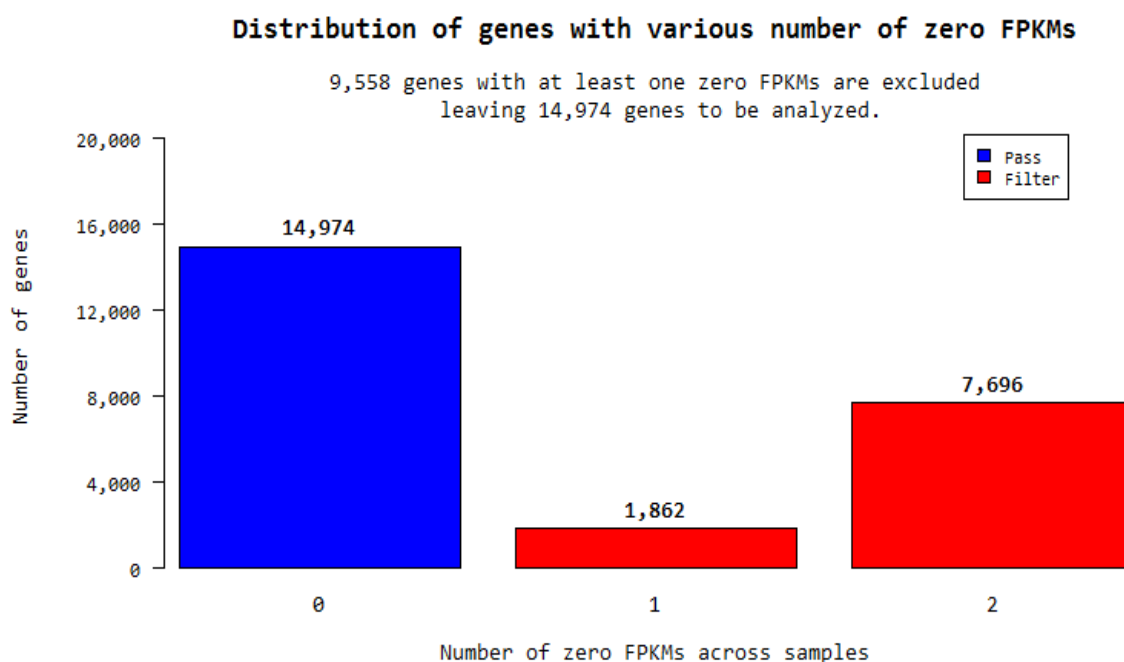
Comparison pair and statistical method for each pair are shown below.

Index	Test vs. Control	Statistical Method
1	Cfdp1-K1RNA vs. vdR2-4RNA	Fold Change, Hierarchical Clustering

5. 1. 2. DATA Quality Check

(Refer to Path: result_RNAseq_excel/DEG_result/Data Quality Check/)

For 2 samples, if more than one FPKM value was 0, it was not included in the analysis. Therefore, from total of 24,532 genes, 9,558 were excluded and only 14,974 genes were used for statistic analysis.

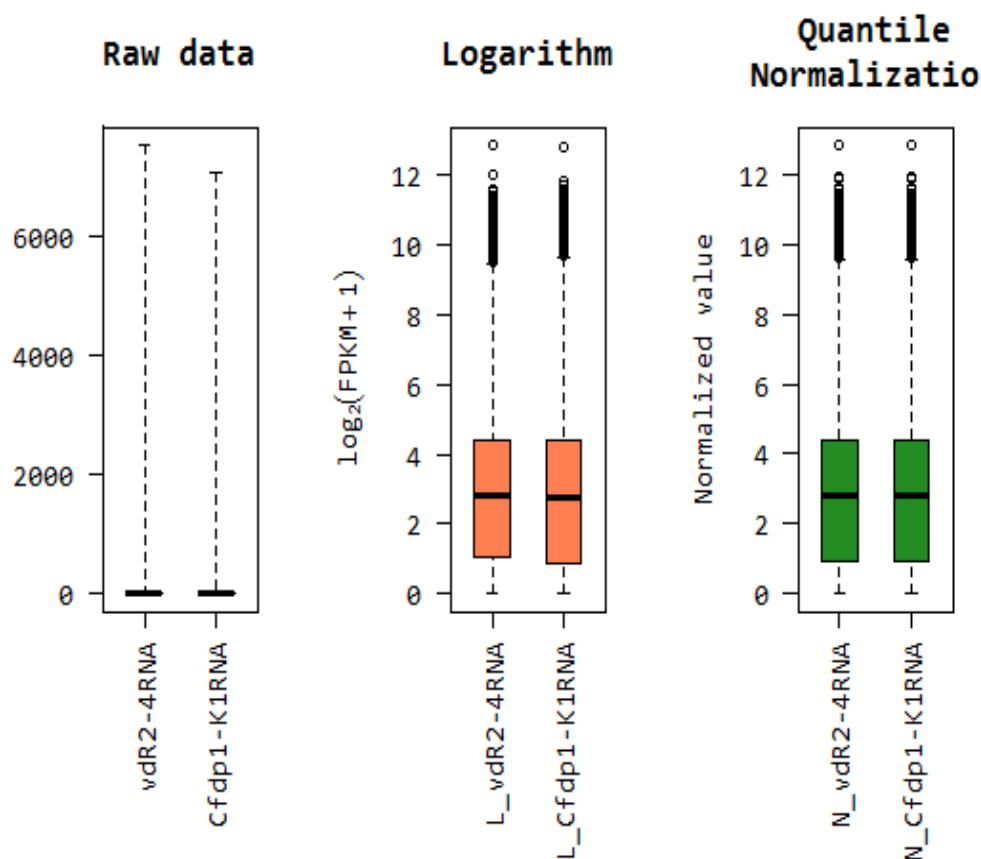


5. 1. 3. Data Transformation and Normalization

To facilitate log2 transformation, 1 was added to the raw signal (FPKM). This process is performed because raw signals are scattered along wide range and most signals are concentrated on the low signal value, so log transformation reduces the range of the signals and produces more even data distribution. After log transformation, in order to reduce systematic bias, quantile normalization is used with preprocessCore' R library.

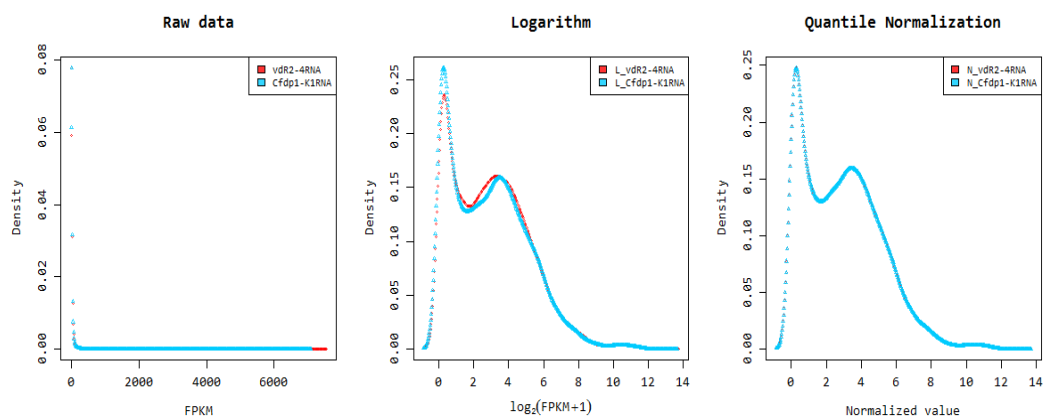
5. 1. 3. 1. Boxplot of Expression Difference between samples.

Below boxplots show the corresponding sample's expression distribution based on percentile (median, 50 percentile, 75 percentile, maximum and minimum) based on raw signal (FPKM), Log2 transformation of FPKM+1 and Quantile Normalization.



5.1.3.2. Expression Density Plot per sample

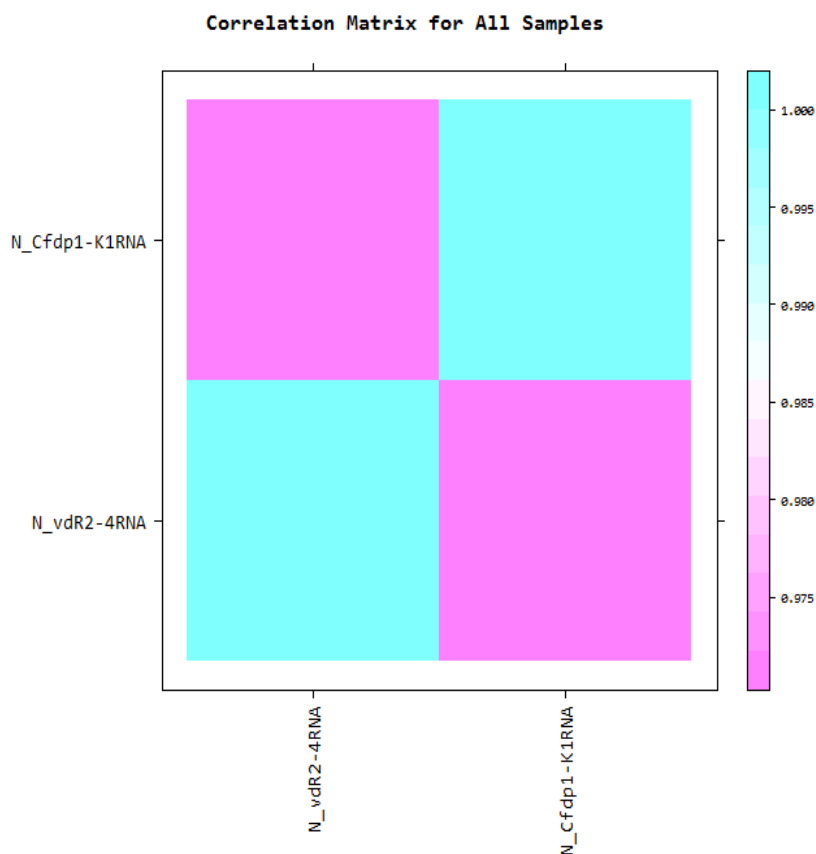
Below density plots show the corresponding samples expression distribution before and after of raw signal (FPKM), Log2 transformation of FPKM+1 and Quantile Normalization.



5. 1. 4. Correlation Analysis between samples

The similarity between samples are obtained through Pearson's coefficient of the normalized value. For range: $-1 \leq r \leq 1$, the closer the value is to 1, the more similar the samples are.

Correlation matrix of all samples is as follows.



5. 2. Differentially Expressed Gene Analysis Workflow

Below shows the orders of DEG (Differentially Expressed Genes) analysis.

- 1) the FPKM value of known genes obtained through -e option of the StringTie were used as the original raw data.

- Raw data

(Refer to Path: result_RNAseq_excel/Expression_profile/StringTie/
Expression_Profile.mm10.gene.xlsx)

: 24,532 genes, 2 samples

- 2) During data preprocessing, low quality transcripts are filtered. Afterwards, log2 transformation of FPKM+1 and quantile normalization are performed.

- Processed data

(Refer to Path: result_RNAseq_excel/DEG_result/data2.xlsx)

: 14,974 genes, 2 samples

- 3) Statistical analysis is performed using Fold Change per comparison pair.
The significant results are selected on conditions of $|fc| \geq 2$.

- Significant data

(Refer to Path: result_RNAseq_excel/DEG_result/data3_fc2.xlsx)

: 694 genes

- 4) For significant lists, hierarchical clustering analysis is performed to group the similar samples and genes. These results are graphically depicted using heatmap and dendrogram.

- Hierarchical Clustering (Euclidean Distance, Complete Linkage)

(Refer to Path: result_RNAseq_excel/DEG_result/Cluster image/)

- 5) For significant lists, gene-set enrichment analysis was performed based on gene ontology(
<http://geneontology.org/>).

Please refer to the gomap_stat sheet and the gomap_genes sheet of data3 file.

Following result are provided.

- gomap_stat
- gomap_genes

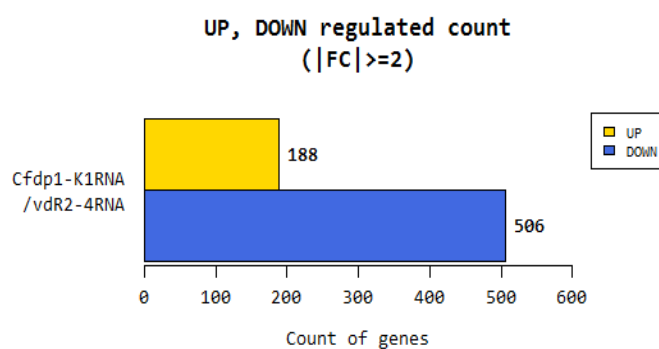
5. 3. Significant Gene Results

(Refer to Path: result_RNAseq_excel/DEG_result/Plots/)

These are DEG result of Cfdp1-K1RNA_vs_vdR2-4RNA meeting fc2 by example.

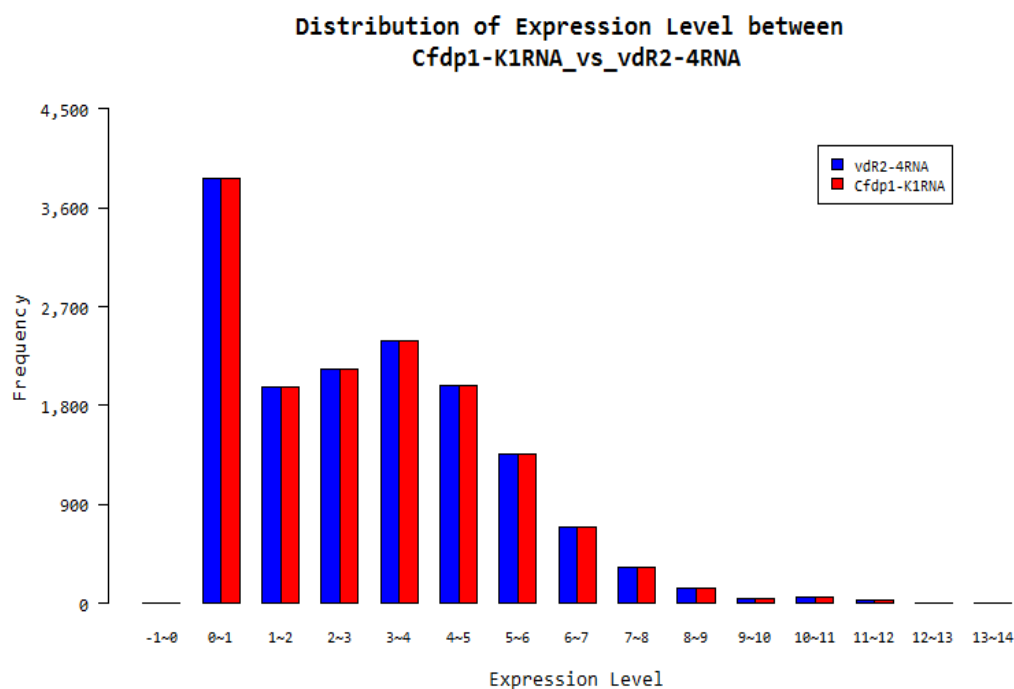
5. 3. 1. Up, Down Regulated Count by Fold Change

Shows number of up and down regulated genes based on fold change of comparison pair.



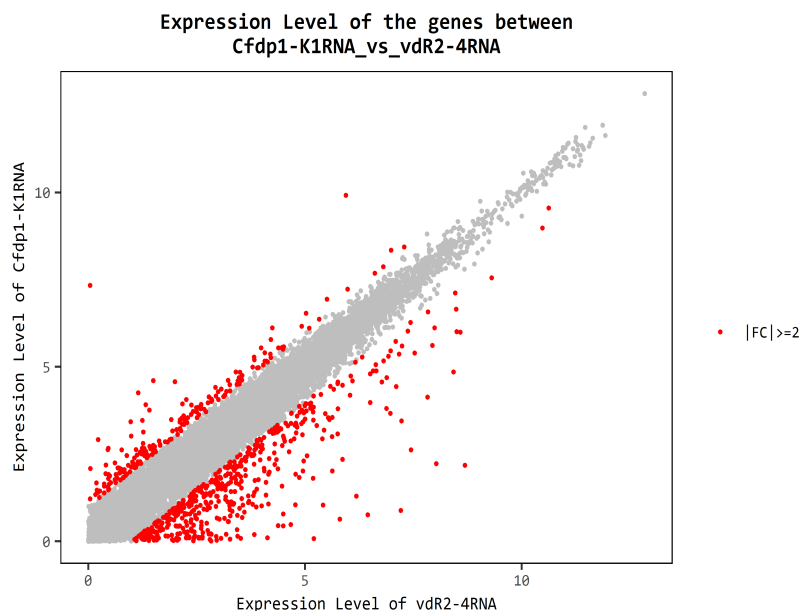
5. 3. 2. Distribution of Expression Level between two groups

Shows distribution of normalized value of each group for comparison pair.



5. 3. 3. Scatter Plot of Expression Level between two groups

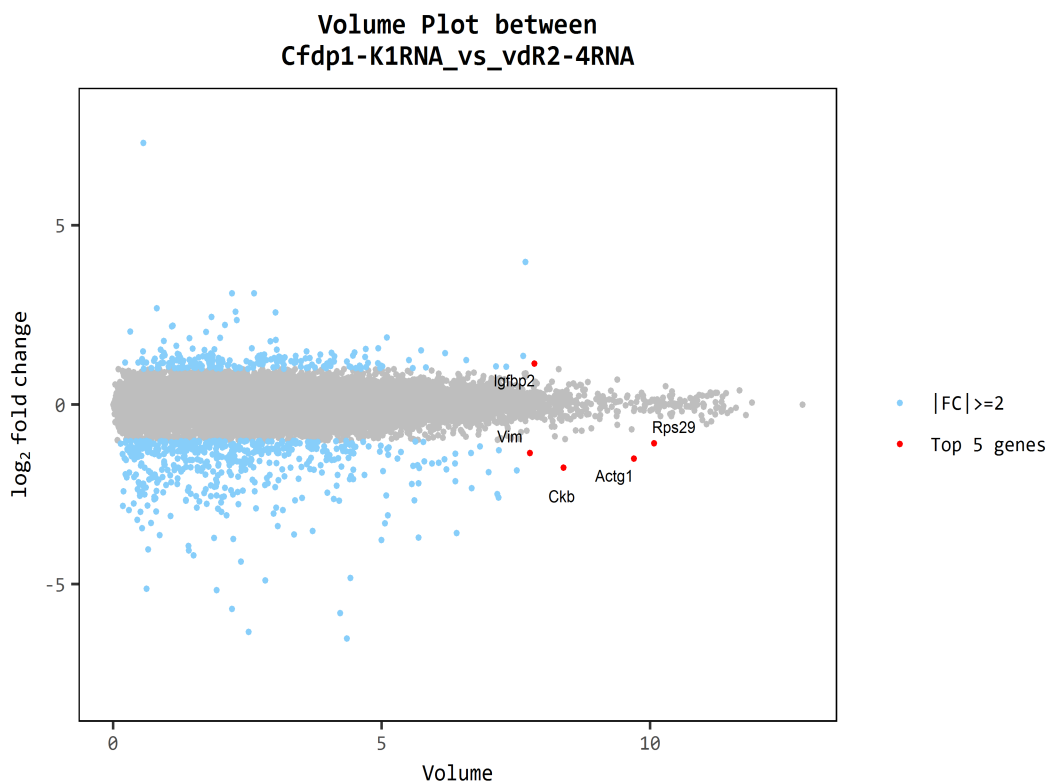
Shows expression levels between comparison pair as a scatter plot. X-axis is control and Y-axis is average normalized value of the group.



5. 3. 4. Volume Plot

Expression volume is defined as the geometric mean of two group's expression level. In order to confirm the genes that show higher expression difference compared to the control according to expression volume, volume plot is drawn. (X-axis: Volume, Y-axis: log₂ Fold Change).

For example, even though fold change might be different by two-fold, the genes with higher volume may be more credible.

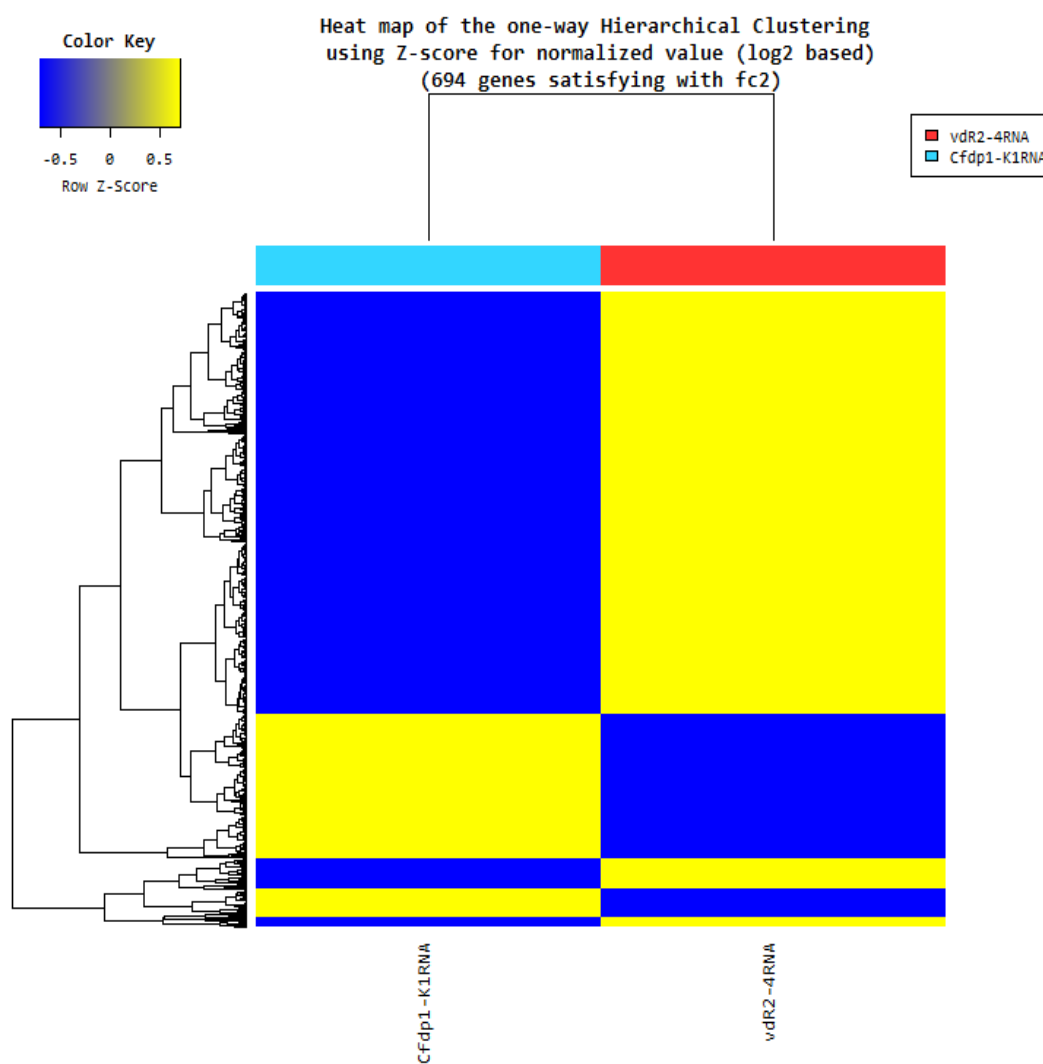


- Red dot: Top five genes by volume which satisfies, $|fc| \geq 2$

5. 3. 5. Hierarchical Clustering Analysis

(Refer to Path: result_RNAseq_excel/DEG_result/Cluster image/)

Heatmap shows result of hierarchical clustering analysis (Euclidean Method, Complete Linkage) which clusters the similarity of genes and samples by expression level (normalized value) from significant list.



5. 4. GO Enrichment Analysis

(Refer to Path: result_RNAseq_excel/DEG_result/GO)

For Gene-Enrichment test which based on Gene Ontology (<http://geneontology.org/>) DB was conducted with significant gene list.

Progressing about 3 categories of GO. The gene or gene product, molecule associated with GO ID was summarized by parsing the ontology file and the annotation file (multispecies annotation provided by Uniprot, or the annotation provided by each type reference DB for the GO consortium) for the GO graph structure.

- Link for the ontology documentation: <http://geneontology.org/page/ontology-documentation>
- Link for the ontology files: <http://geneontology.org/page/download-ontology>
- Link for the annotation files: <http://geneontology.org/page/download-annotations>

The two results are provided for enrichment analysis.

- gomap_stat
- gomap_genes

5. 4. 1. gomap_stat Sheet

The result of associated gene and enrichment test was summarized by GO ID. The significance of specific GO ID in enrichment test with DEG set was calculated by modified fisher's exact test.

Namespace	GOID	Term	count	Genes	Sig.NotIn.GO	Genome.In.GO	Genome.NotIn.GO	PValue	Bonferroni	FDR
cellular_component	GO:0005623	cell	736	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	205	115	59368	0	0	0
cellular_component	GO:0044464	cell part	735	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	206	1286	58197	0	0	0
biological_process	GO:0009987	cellular process	719	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	222	7204	52279	0	0	0
biological_process	GO:0046699	single-organism process	684	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	257	5148	54335	0	0	0
cellular_component	GO:0005622	intracellular	678	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	263	1147	58336	0	0	0
cellular_component	GO:0043226	organelle	638	AARS, ABAT, ABCA3, ABCA7, ABCB1, ABC	303	1343	58140	0	0	0
biological_process	GO:0044763	single-organism cellular process	631	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	310	1973	57510	0	0	0
cellular_component	GO:0043227	membrane-bounded organelle	612	AARS, ABAT, ABCA3, ABCA7, ABCB1, ABC	329	4914	54569	0	0	0
cellular_component	GO:0043229	intracellular organelle	576	AARS, ABAT, ABCA3, ABCA7, ABCB9, ABC	365	21	59462	0	0	0
cellular_component	GO:0005737	cytoplasm	567	AAMDC, AARS, ABAT, ABCA3, ABCA7, AB	374	4904	54579	0	0	0
biological_process	GO:0065007	biological regulation	555	AAMDC, AARS, ABAT, ABCA7, ABCG1, AB	386	94	59389	0	0	0
cellular_component	GO:0043231	intracellular membrane-bounded organelle	541	AARS, ABAT, ABCA3, ABCA7, ABCB9, ABC	400	1895	57588	0	0	0
biological_process	GO:0071704	organic substance metabolic process	524	AAMDC, AARS, ABAT, ABCA7, ABCG1, AB	417	2	59481	0	0	0
biological_process	GO:0044237	cellular metabolic process	510	AAMDC, AARS, ABAT, ABCA7, ABCG1, AB	431	1195	58288	0	0	0
biological_process	GO:0044238	primary metabolic process	507	AAMDC, AARS, ABAT, ABCA7, ABCG1, AB	434	18	59465	0	0	0
biological_process	GO:0050794	regulation of cellular process	498	AAMDC, AARS, ABAT, ABCA7, ABCG1, AB	443	2	59481	0	0	0
biological_process	GO:0050896	response to stimulus	447	AARS, ABAT, ABCA3, ABCA7, ABCB1, ABC	494	64	59419	0	0	0
cellular_component	GO:0044444	cytoplasmic part	443	AARS, ABAT, ABCA3, ABCA7, ABCB9, ABC	498	7	59476	0	0	0
cellular_component	GO:0016020	membrane	415	AARS, ABCA3, ABCA7, ABCB1, ABCB9, AB	526	601	58882	0	0	0
cellular_component	GO:0044422	organelle part	413	ABAT, ABCA3, ABCA7, ABCB9, ABCG1, AB	528	87	59396	0	0	0
cellular_component	GO:0044446	intracellular organelle part	407	ABAT, ABCA3, ABCA7, ABCB9, ABCG1, AB	534	16	59467	0	0	0
biological_process	GO:0044260	cellular macromolecule metabolic process	379	AAMDC, AARS, ABCA7, ABCG1, ABT81, A4	562	14	59469	0	0	0
biological_process	GO:0032501	multicellular organismal process	369	AARS, ABAT, ABCA7, ABCG1, ACE, ACTA2	572	338	59145	0	0	0
biological_process	GO:0044707	single-multicellular organism process	367	AARS, ABAT, ABCA7, ABCG1, ACE, ACTA2	574	8	59475	0	0	0

- Namespace: 3 categories of Gene ontology (Cellular Component, Molecular Function, Biological Process)
- GOID: Gene ontology ID
- Term: Gene ontology term
- count: The number of unique genes associated with GO ID
- Genes: Associated genes with GO ID (connect by comma)
- Sig.NotIn.GO: The number of genes which are not associated GO ID
- Genome.In.GO: The number of genes associated with GO ID of total sample of gene species
- Genome.NotIn.GO: The number of genes which are not associated with GO ID of total sample of gene species
- PValue: Raw p-value was calculated by modified fisher's exact test
- Bonferroni: Adjusted p-value by Bonferroni
- FDR: Adjusted p-value by FDR

5. 4. 2. gomap_genes Sheet

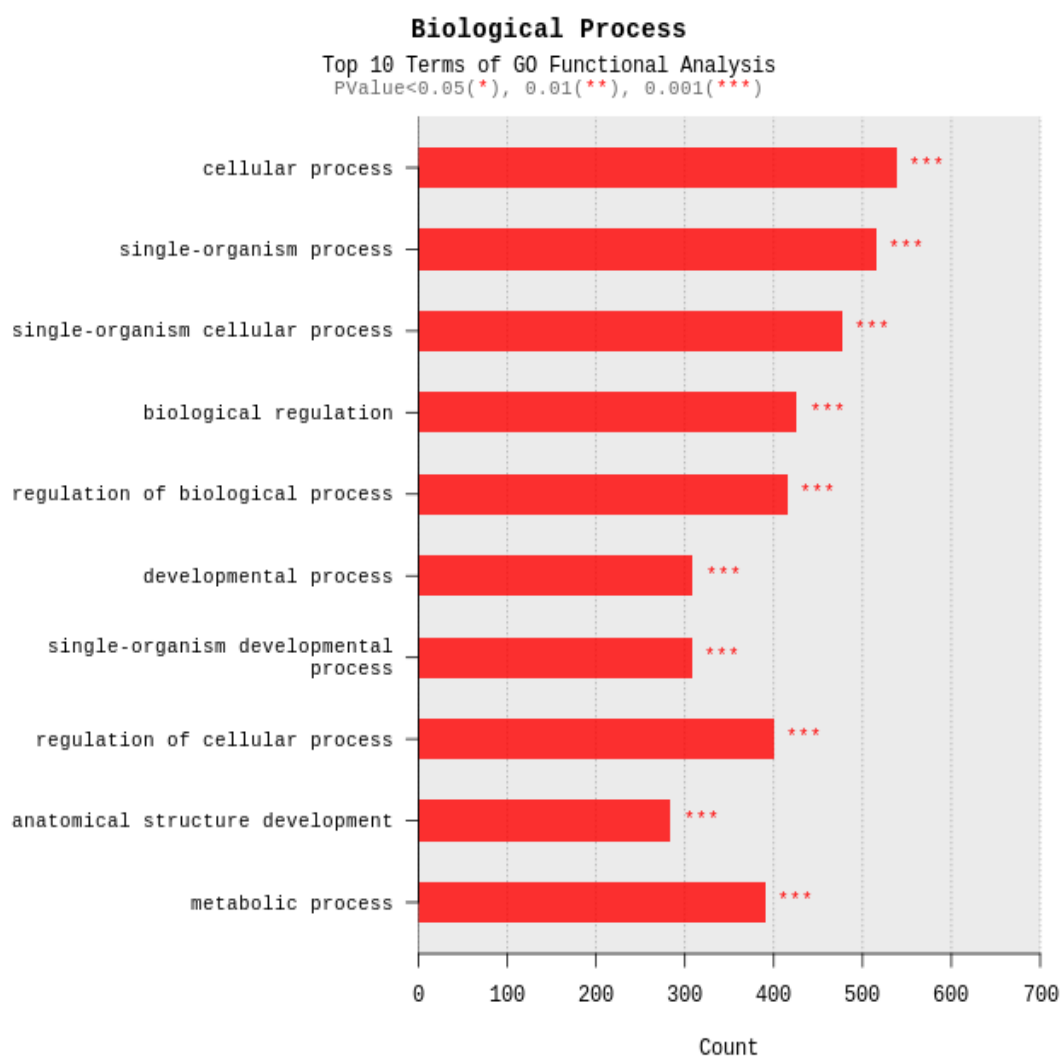
The result of associated GO ID and DEG analysis was summarized based on gene. The GO ID which associated with specific gene was summarized with statistic such as fold change, p-value, volume, and normalized value.

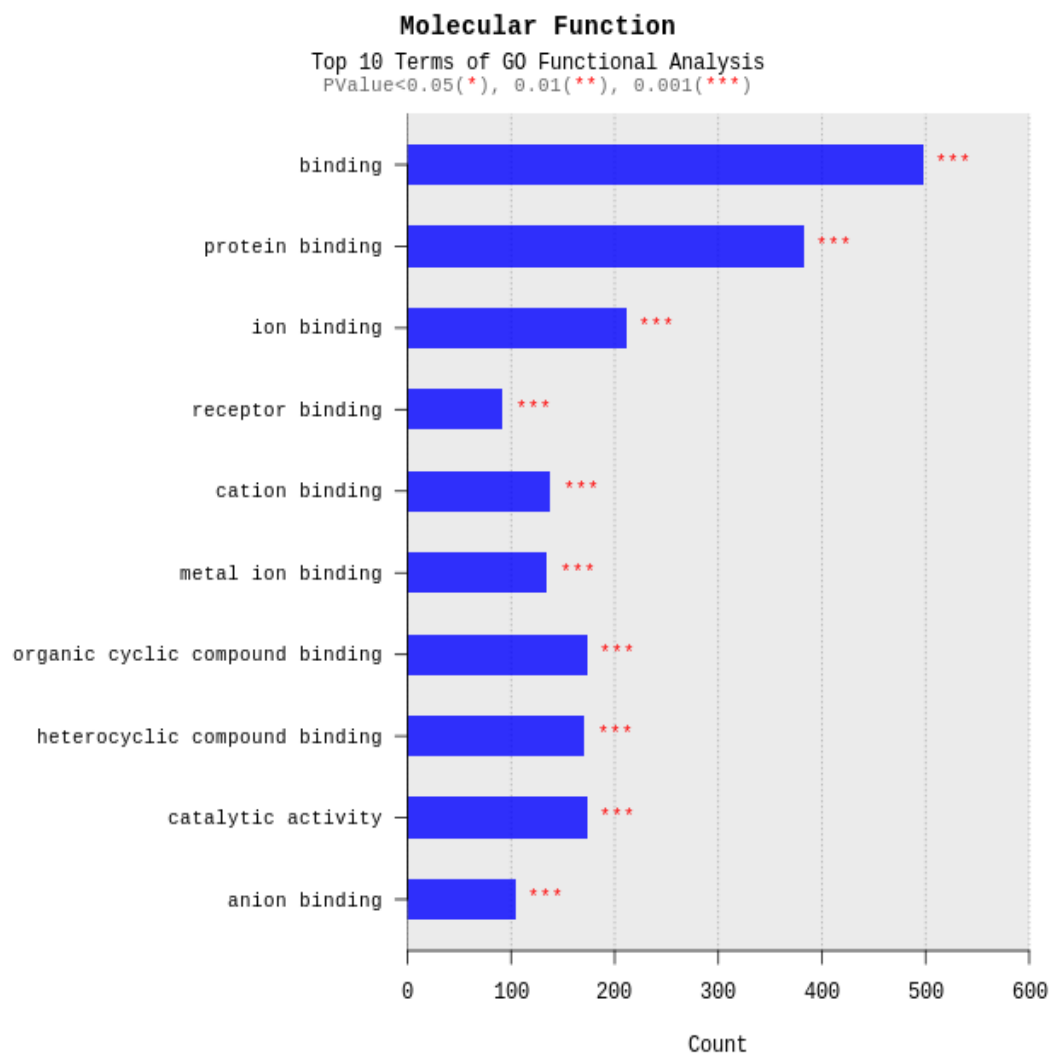
InID	OutID	GOID	Term	Namespace	PValue	Bonferroni	FDR	Gene_ID	Gene_Symbol	test/control.fc	test/control.volume	N_test	N_control
AAAMDC	AAAMDC	GO:0005488	binding	molecular_function	4.4834E-208	3.9544E-204	2.8864E-206	28971	AAAMDC	1.899756	4.918387	5.885185	5.970135
AAAMDC	AAAMDC	GO:0005515	protein binding	molecular_function	3.8301E-167	3.3782E-163	2.0108E-165	28971	AAAMDC	1.899756	4.918387	5.885185	5.970135
AARS	AARS	GO:0000049	tRNA binding	molecular_function	1	1	1	16	AARS	1.080088	6.531080	7.518749	7.629339
AARS	AARS	GO:0001666	nucleotide binding	molecular_function	0.006121778	1	0.017222993	16	AARS	1.080088	6.531080	7.518749	7.629339
AARS	AARS	GO:0001101	response to acid chemical	biological_process	1.61547E-30	1.42485E-26	1.98447E-29	16	AARS	1.080088	6.531080	7.518749	7.629339
AARS	AARS	GO:0001882	nucleoside binding	molecular_function	4.5182E-146	4.0026E-142	2.0632E-144	16	AARS	1.080088	6.531080	7.518749	7.629339
HNF1A	HNF1A	GO:0007267	cell-cell signaling	biological_process	1.20893E-96	1.06627E-94	3.63916E-97	6927	HNF1A	-1.423774	3.386822	2.322587	2.153258
HNF1A	HNF1A	GO:0007275	multicellular organismal development	biological_process	3.9889E-257	3.5182E-253	3.3829E-255	6927	HNF1A	-1.423774	3.386822	2.322587	2.153258
RARRES2	RARRES2	GO:0036211	protein modification process	biological_process	6.1358E-279	5.4118E-275	6.0131E-277	5919	RARRES2	-1.070216	6.260522	5.196256	4.595104
VDR	VDR	GO:0003707	steroid hormone receptor activity	molecular_function	1.96245E-06	0.01730879	8.29418E-06	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0004871	signal transducer activity	molecular_function	9.73774E-51	8.58869E-47	2.27214E-49	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0004872	receptor activity	molecular_function	2.4367E-51	2.14917E-47	4.47744E-50	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0004879	RNA polymerase II transcription factor activity	molecular_function	2.87475E-07	0.002535094	1.31352E-06	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0005102	receptor binding	molecular_function	2.1529E-175	1.8988E-171	1.1868E-173	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0005488	binding	molecular_function	4.4834E-208	3.9544E-204	2.8864E-206	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0005496	steroid binding	molecular_function	1.3282E-09	1.17147E-05	7.29435E-09	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0005499	vitamin D binding	molecular_function	0.045949424	1	0.106960653	7421	VDR	1.664240	2.647718	3.457308	3.607864
VDR	VDR	GO:0005515	protein binding	molecular_function	3.8301E-167	3.3782E-163	2.0108E-165	7421	VDR	1.664240	2.647718	3.457308	3.607864
ZFP36	ZFP36	GO:0006952	defense response	biological_process	6.5961E-158	5.8178E-154	3.2321E-156	7538	ZFP36	1.566431	2.191879	2.986505	3.233095
ZFP36	ZFP36	GO:0006954	inflammatory response	biological_process	1.20554E-36	1.06328E-32	1.69313E-35	7538	ZFP36	1.566431	2.191879	2.986505	3.233095
ZFP36	ZFP36	GO:0007154	cell communication	biological_process	0	0	0	7538	ZFP36	1.566431	2.191879	2.986505	3.233095
ZFP36	ZFP36	GO:0007165	signal transduction	biological_process	0	0	0	7538	ZFP36	1.566431	2.191879	2.986505	3.233095
ZFP36	ZFP36	GO:0007275	multicellular organismal development	biological_process	3.9889E-257	3.5182E-253	3.3829E-255	7538	ZFP36	1.566431	2.191879	2.986505	3.233095
ZFP36	ZFP36	GO:0008152	metabolic process	biological_process	2.2907E-215	2.0204E-211	1.5423E-213	7538	ZFP36	1.566431	2.191879	2.986505	3.233095

- InID: Input ID for GO enrichment analysis
- OutID: Mapping ID as gene symbol from input ID through GO enrichment analysis
- GOID: Gene ontology ID
- Term: Gene ontology term
- Namespace: 3 categories of gene ontology (Cellular Component, Molecular Function, Biological Process)
- PValue: Raw p-value was calculated by modified fisher's exact test
- Bonferroni: Adjusted p-value by Bonferroni
- FDR: Adjusted p-value by FDR

The bar plot below shows the results of the enrichment analysis based on Gene Ontology DB for significant genes.

(These plots were made based on gomap_stat result.)

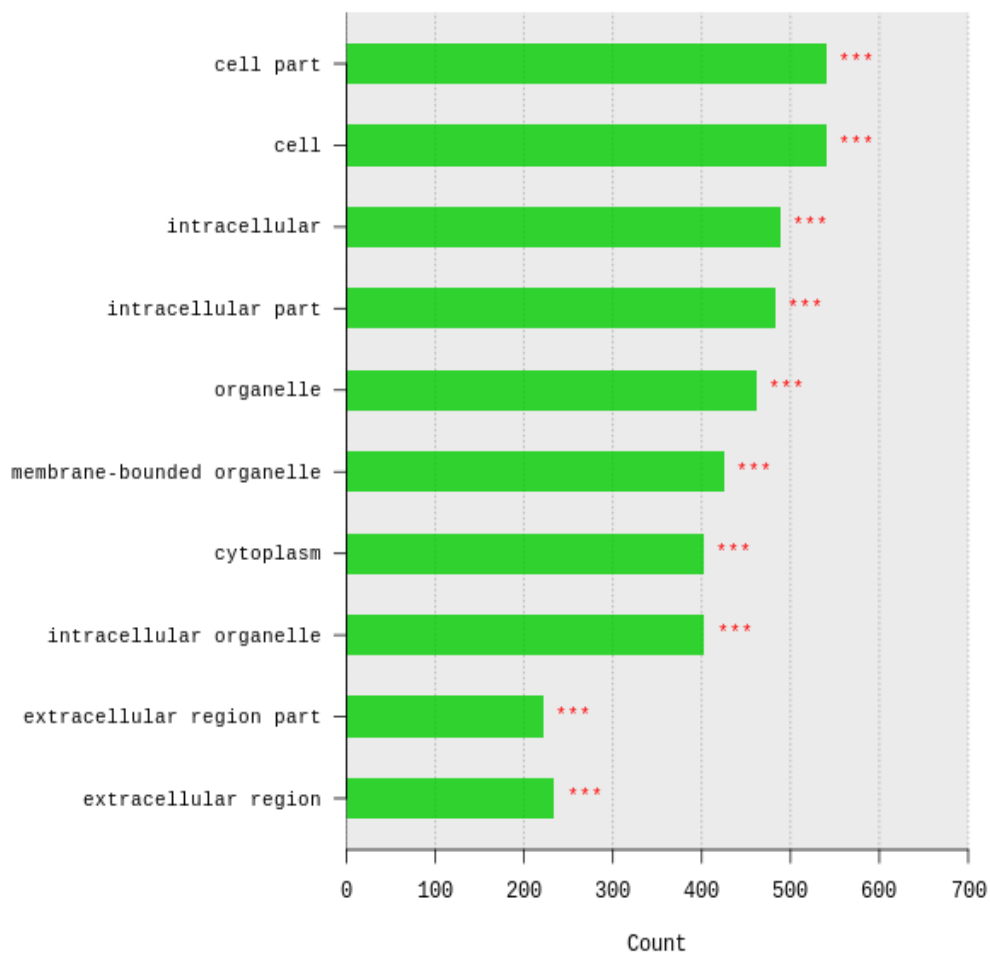




Cellular Component

Top 10 Terms of GO Functional Analysis

PValue<0.05(*), 0.01(**), 0.001(***)



6. Data Download Information

6.1. Raw Data

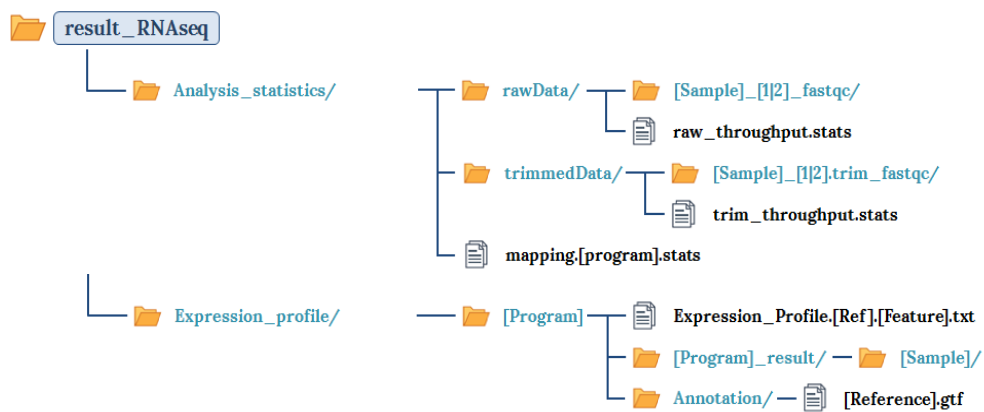
Raw data is the FASTQ file that isn't trimmed adapter sequence.

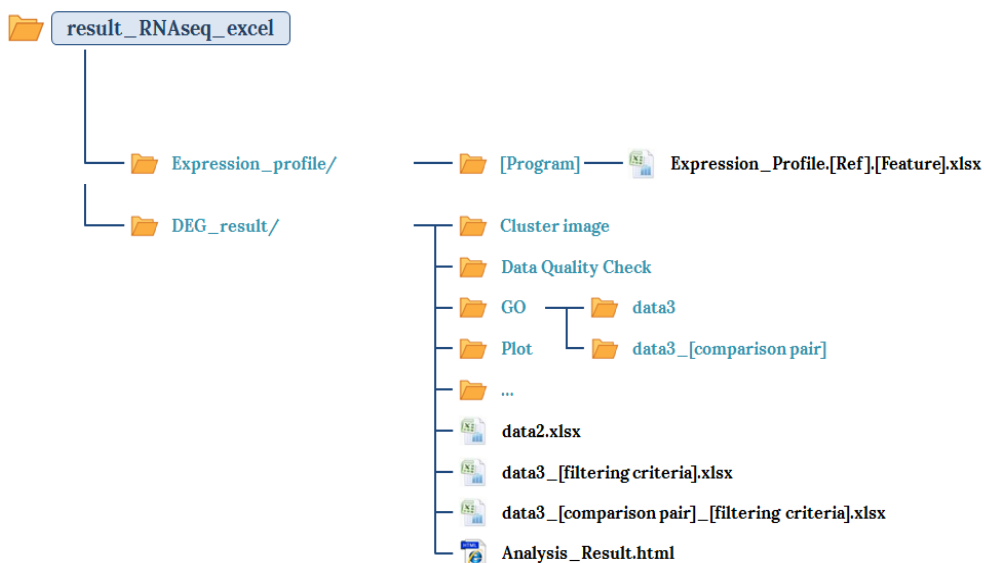
Download link	File size	md5sum
Cfdp1-K1RNA_1.fastq.gz	1.2G	34d313aab0592c10d9a8c1181e14dfb5
Cfdp1-K1RNA_2.fastq.gz	1.23G	68288067dba6036df06abc1a1fef372e
vdR2-4RNA_1.fastq.gz	1.31G	a793383ab0387867c2d5edaae7627c38
vdR2-4RNA_2.fastq.gz	1.35G	ed193bf6585ada34d4a1d450d52971f1

- fastq.gz : This is a zip file of raw data used in analysis.
- md5sum : In order to verify the integrity of files, md5sum is used. If the values of md5sum are the same, there is no forgery, modification or omission.


6.2. Analysis Results

Download link	File size
HN00101712_result_RNAseq.zip (md5sum: 7f33ecd78a315ed837bf635cddbec896)	30.45M
HN00101712_result_RNAseq_excel.zip (md5sum: 621bddc7040d0df14b635ddd4db9ce28)	14.35M





Folder
File

 The data retention period is three months,
please send an e-mail (ngs@macrogen.com)
or contact representative if you want longer retention period.

7. Appendix

7.1. Phred Quality Score Chart

Phred quality score numerically express the accuracy of each nucleotide. Higher Q number signifies higher accuracy. For example, if Phred assigns a quality score of 30 to a base, the chances of having base call error are 1 in 1000.

Quality of phred score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%

Phred Quality Score Q is calculated with $-10\log_{10}P$, where P is probability of erroneous base call.

Q-Score Binning

Illumina NovaSeq sequencer groups quality scores into specific ranges, or bins, and assigns a value to each range. Q-scores is typically updated when significant characteristics of the sequencing platform changes, such as new hardware, software, or chemistry versions.

7. 2. Programs used in Analysis

7. 2. 1. FastQC v0.11.7

LINK <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

FastQC is a program that performs quality check on the raw sequences before analysis to make sure data integrity. The main function is importing BAM, SAM, FastQ files and providing quick overview on which section has problems. It provides such results as graphs and tables in html files.

7. 2. 2. Trimmomatic 0.38

LINK <http://www.usadellab.org/cms/?page=trimmomatic>

Trimmomatic is a program that performs trimming depending on various parameters on illumina paired-end or single-end.

- ILLUMINACLIP: Cut adapter and other illumina-specific sequences from the read.
- SLIDINGWINDOW: Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold.
- LEADING: Cut bases off the start of a read, if below a threshold quality.
- TRAILING: Cut bases off the end of a read, if below a threshold quality.
- CROP: Cut the read to a specified length.
- HEADCROP: Cut the specified number of bases from the start of the read.
- MINLEN: Drop the read if it is below a specified length.
- TOPHRED33: Change quality score to phred33.
- TOPHRED64: Change quality score to phred64.

7. 2. 3. HISAT2 version 2.1.0, Bowtie2 2.3.4.1

LINK <https://ccb.jhu.edu/software/hisat2/index.shtml>

HISAT2 is a fast and sensitive alignment program for mapping next-generation sequencing reads to genomes. Its first implementation based on an extension of BWT for graphs, designed a graph FM index (GFM). In addition to using one global GFM index, HISAT2 uses a large set of small GFM indexes that collectively cover the whole genome (each index representing a genomic region of 56 Kbp, with 55,000 indexes needed to cover the human population). These small indexes (called local indexes), combined with several alignment strategies, enable rapid and accurate alignment of sequencing reads. This new indexing scheme is called a Hierarchical Graph FM index (HGFM).

7. 2. 4. StringTie version 1.3.4d

LINK <https://ccb.jhu.edu/software/stringtie/>

StringTie is a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts. It uses a novel network flow algorithm as well as an optional de novo assembly step to assemble and quantitate full-length transcripts representing multiple splice variants for each gene locus.

7. 3. References

1. BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014, btu170.
2. KIM, Daehwan; LANGMEAD, Ben; SALZBERG, Steven L. HISAT: a fast spliced aligner with low memory requirements. Nature methods, 2015, 12.4: 357-360.
3. LI, Heng, et al. The sequence alignment/map format and SAMtools. Bioinformatics, 2009, 25.16: 2078-2079.
4. PERTEA, Mihaela, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology, 2015, 33.3: 290-295.
5. PERTEA, Mihaela, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols, 2016, 11.9: 1650-1667.

