

associated with various host physiology states, including disease, diet, and age through the shift of bacterial composition, as well as metabolic and nutritional processes.^{19–23} The ability of probiotics to survive through the intestine and to modulate gut microbiota is a critical factor in determining their potential for health-related outcomes.

There have been a large number of probiotic intervention studies to assess the impact of probiotics on gut microbiota in healthy adults,^{24–34} infants, and children,^{35,36} and in clinical trials on patients with a variety of diseases.^{37,38} Most probiotic intervention studies were carried out by comparison between probiotic-treated groups and placebo controls and examined only one or two samples from periods before and during intervention or post-intervention for each subject. These experimental designs make evaluation of results obscure from a statistical viewpoint due to the high inter-individual variability of gut microbiota.⁴ In addition, most of the analyses focussed on the composition of specific bacterial species or groups by conventional methods such as culturing, quantitative polymerase chain reaction (qPCR), fluorescence *in situ* hybridization, denaturing gradient gel electrophoresis, or terminal-restriction fragment length polymorphism based on the bacterial 16S ribosomal RNA gene (16S). These conventional methodologies may also overlook subtle changes in bacterial community structure and change of species other than targeted species. Thus, the effect of probiotic administration on the overall structure of gut microbiota is largely unknown.

Recently, a high-throughput sequencing-based analysis has been conducted for gut microbiota fed with a probiotic yogurt that provided new insights into probiotics research by utilizing a large-scale dataset.³⁹ However, much more data are required to understand the impact of probiotics on gut microbiota. Recent advances in sequencing technology have enabled us to elucidate complex bacterial communities, including human gut microbiota.^{40,41} Particularly, 454 pyrosequencing of bacterial 16S gene tags coupled with bioinformatics provides a high-throughput and cost-effective approach for the comprehensive analysis of bacterial communities at the species level.^{42–48}

In this study, we developed an analysis pipeline for bacterial communities based on barcoded 454 pyrosequencing of 16S gene tags using modified PCR primers that improved the quantitative accuracy of inferred species composition in human gut microbiota. Using this pipeline, we analysed faecal samples longitudinally collected from individuals with and without probiotic administration to evaluate the effect of probiotics on gut microbiota with respect to species richness and diversity. The results revealed the

robustness and stability of gut microbiota of healthy adults in response to probiotic administration.

2. Materials and methods

2.1. Subjects, faecal sample collection, and probiotic intervention

Eighteen healthy volunteers (age: 22 ± 3.16 yrs, 6 male, 12 female) were recruited through Azabu University, Kanagawa, Japan (Supplementary Table S1). All subjects were informed of the purpose of this study. This study was approved by the ethical committee of Azabu University, and written consent was obtained from all subjects. No subjects were treated with antibiotics during faecal sample collection. The subjects were divided into six groups (three subjects per group), and each group consumed six different commercially available probiotics supplied from Yakult Honsha Co., Ltd, Kagome Co., Ltd, Morinaga Milk Industry Co., Ltd, Takanashi Milk Products Co., Ltd, Meiji Co., Ltd, and Danone Japan Co., Ltd, respectively (Supplementary Table S1). The number of each bacterial strain contained in the probiotic products was estimated as the genome equivalent by qPCR of 16S ribosomal RNA genes using 27Fmod-338R, followed by pyrosequencing of the 16S amplicons (see below). The genome equivalent per gram or millilitre and the total genome equivalent of each bacterial strain in one probiotic product are summarized in Supplementary Table S1. Three subjects in each group consumed the same probiotics daily for 8 weeks according to the schedule of sampling and probiotic intervention (Supplementary Fig. S1). Faecal samples from 4 weeks before (S00) and 8 weeks during probiotic intervention (S01–S04), and 8 weeks after cessation of probiotic intervention (S05–S08), were collected every 2 weeks from each subject. In total, we collected 158 faecal samples from the 18 subjects because we could not collect 1 sample each from 4 of the subjects.

2.2. Recovery of bacteria from faecal samples

Freshly collected faeces (1.0 g) were suspended in 20% glycerol (Wako Pure Chemical Industries, Ltd) and phosphate buffered saline solution (Life Technologies Japan, Ltd, Tokyo, Japan), frozen in liquid nitrogen, and stored at -80°C until ready for use. Bacterial pellets were prepared from frozen faecal samples as described previously.⁴⁹

2.3. DNA isolation from bacteria

Faecal DNA was isolated and purified according to the literature, with minor modifications.⁴⁹ The bacterial pellet was suspended and incubated with 15 mg/ml lysozyme (Sigma-Aldrich Co., LCC) at

37°C for 1 h in TE10. Purified achromopeptidase (Wako Pure Chemical Industries, Ltd) was added at a final concentration of 2000 units/ml and then incubated at 37°C for 30 min. The suspension was treated with 1% (wt/vol) sodium dodecyl sulphate and 1 mg/ml proteinase K (Merck Japan) and incubated at 55°C for 1 h. The lysate was treated with phenol/chloroform/isoamyl alcohol (Life Technologies Japan, Ltd). DNA was precipitated by adding ethanol and pelleted by centrifugation at 3,300 *g* at 4°C for 15 min. The DNA pellet was rinsed with 75% ethanol, dried, and dissolved in 10 mM Tris-HCl/1 mM EDTA (TE). DNA samples were purified by treating with 1 mg/ml RNase A (Wako Pure Chemical Industries, Ltd) at 37°C for 30 min and precipitated by adding equal volumes of 20% polyethylene glycol solution (PEG6000-2.5M NaCl). DNA was pelleted by centrifugation at 8,060 *g* at 4°C, rinsed with 75% ethanol, and dissolved in TE.

2.4. 454 barcoded pyrosequencing of 16S rRNA gene

The V1–V2 region of the 16S rRNA gene was amplified using forward primer (5'-CCATCTCATCCCTGCG TGTCTCCGACTCAGNNNNNNNNNagrgtttgatymtggtcag-3') containing the 454 primer A, a unique 10-bp barcode sequence for each sample (indicated in N), and 27Fmod (5'-agrgtttgatymtggtcag) in which the third base A in the original primer 27F was changed to R, and reverse primer (5'-CCTATCCCCTGTGTGC CTTGGCAGTCTCAGtgctgctcccgtaggagt-3') containing the 454 primer B and reverse primer 338R (5'-tgctgctcccgtaggagt). PCR was performed in 1 × Ex Taq PCR buffer (50 µl), deoxynucleoside triphosphate (2.5 mM), Ex Taq polymerase (Takara Bio, Inc., Shiga), each primer (10 µM), and 40 ng of extracted DNA under conditions of 2 min at 96°C, 20 cycles of 96°C for 30 s, 55°C for 45 s, and 72°C for 1 min, and a final extension of 72°C for 10 min on a 9700 PCR system (Life Technologies Japan, Ltd, Tokyo, Japan). PCR products of approximately 370 bp were confirmed by agarose gel electrophoresis, purified by AMPure XP magnetic purification beads (Beckman Coulter, Inc., Brea, CA, USA), and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies Japan, Ltd, Tokyo, Japan). Mixed samples were prepared by pooling approximately equal amounts of PCR amplicons from each sample and subjected to 454 GS FLX Titanium or 454 GS JUNIOR (Roche Applied Science) sequencing according to the manufacturer's instructions.

2.5. Analysis pipeline for 454 barcoded pyrosequencing of 16S PCR amplicons

We developed an analysis pipeline for 454 barcoded pyrosequencing of PCR amplicons of the V1-2

region amplified by 27Fmod-338R primers. First, 16S reads were assigned to each sample based on the barcode sequence information. Second, 16S reads that did not have PCR primer sequences at both sequence termini and those with an average quality value < 25 were filtered out. Third, 16S reads containing possible chimaeric sequences that had BLAST match lengths of < 90% with reference sequences in the database were removed. Reads removed in these processes accounted for about 35% of all reads, most of which represented reads lacking PCR primer sequences (Supplementary Table S2). Finally, filter-passed reads were obtained for further analysis by trimming off both primer sequences.

All 3000 filter-passed reads of the 16S V1-2 sequences obtained from each subject were deposited in DDBJ/GenBank/EMBL with accession numbers DRA000869–DRA000886.

2.6. Assessment of the quantitative accuracy of 16S data using artificial bacterial communities

Two artificial bacterial communities (designated 'mock01' and 'mock02') were constructed by mixing genomic DNA from 10 and 11 different human gut-associated bacterial strains with an appropriate ratio, respectively (Supplementary Table S3). Genome sequences of these microbes were completely sequenced and are publicly available. From these communities, we amplified the V1-2 region by PCR using 27F-338R and 27Fmod-338R primers, the V5-6 region by 787F-1061R primers, and the V1-9 region by 27F-1492R primers. V1-2 and V5-6 amplicons were subjected to 454 pyrosequencing, and V1-9 amplicons were cloned in *Escherichia coli*, and 3000 clones were sequenced by the Sanger method, and the products were analysed with the ABI3730xl (Life Technologies Japan, Ltd, Tokyo). We also performed duplicate qPCR experiments targeting a specific genomic region of the bacterial strains in the two mock communities. All filter-passed 16S *de novo* sequences and qPCR data were then analysed by principal component analysis (PCA) to compare and assess the quantitative accuracy (Fig. 1).

The error rate of the filter-passed sequences using 27Fmod-338R primers obtained from the two mock communities was estimated by aligning the 16S V1-2 *de novo* sequences with the reference 16S sequences in the two mock communities (Supplementary Table S4).

2.7. Data analysis

2.7.1. Database Two databases were constructed for the analysis of 16S sequences. One is the 16S rRNA gene sequence database constructed

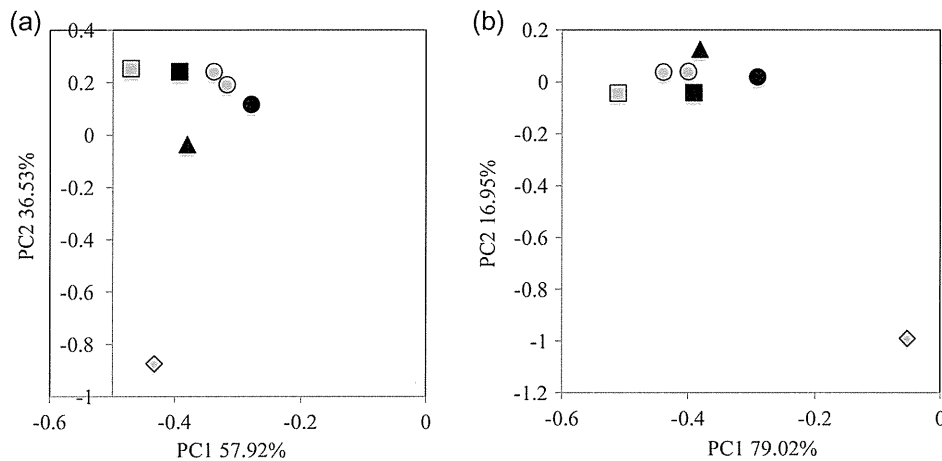


Figure 1. Assessment of the quantitative accuracy of the analysis of the bacterial composition of two mock communities by various methods. PCA analysis of the data was obtained from various methods using mock01 (a) and mock02 (b). Closed circle: expected, open circle: duplicate qPCR, closed square: pyrosequencing of 16S V1-2 region using 27Fmod, open square: pyrosequencing of 16S V1-2 region using 27F, closed triangle: pyrosequencing of 16S V5-6 region, open diamond: Sanger sequencing of nearly full-length 16S clone.

by collecting 16S sequences of ≥ 1200 bp of bacterial isolates from the Ribosomal Database Project v. 10.27. Another database is the reference genome database constructed by collecting genome sequences from the NCBI FTP site (<ftp://ftp.ncbi.nih.gov/genbank/>, Dec 2011) that includes 1482 complete and 605 draft bacterial genomes.

2.7.2. Operational taxonomic unit (OTU) and UniFrac distance analysis We used 3000 filter-passed reads of 16S sequences for operational taxonomic unit (OTU) and UniFrac distance analysis for each sample. For OTU analysis, clustering of 16S reads was done using a 96% pairwise-identity cutoff with the UCLUST program (www.drive5.com). Representative sequences for each OTU were assigned to bacterial species by BLAST search with a 96% pairwise-identity cutoff against the two databases mentioned above. UniFrac distance analysis was used to determine the dissimilarity (distance) between two communities based on the fraction of branch length shared between two communities within a phylogenetic tree constructed from 16S sequence datasets.⁴⁴

2.7.3. Other Estimation of OTU numbers by extrapolation (Chao1 and ACE) was calculated with the vegan package (v2.0-5) for R (v2.15.2).

3. Results and discussion

3.1. Quantitative accuracy of 16S data produced by 454 pyrosequencing

Pyrosequencing of PCR amplicons of bacterial 16S short variable regions is the most popular and a

high-throughput approach to infer and characterize the species composition in bacterial communities.^{42,45,46,48} The 454 pyrosequencing platform, which can produce over 400 bases per read, is also superior to shorter read-length sequencers with respect to sequence accuracy for single-end sequencing.^{50,51} However, this PCR-based method has a problem particularly in quantification of the composition of the genus *Bifidobacterium*, a dominant species in human gut microbiota because the 16S sequence of *Bifidobacterium* has a few base mismatches with the commonly used PCR primer 27F (or 8F), underestimating this genus in the community.⁵²⁻⁵⁵ To improve this, we modified primer 27F to 27Fmod by changing the third base G to R (G or A) in 27F-YM⁵³ that perfectly matched with the annealing site of the *Bifidobacterium* 16S gene (see Materials and methods).

To assess the 16S data using 27Fmod, we compared various 16S sequence and qPCR data obtained from two mock communities (Supplementary Table S3) that are useful to evaluate the quantitative accuracy of 16S-based data and the sequencing error rate.^{56,57} Quantitative accuracy of the overall bacterial composition was evaluated by comparing the similarity of each data to the expected ('Expected') using PCA (Fig. 1). From the PCA data, Euclidean distance was calculated for evaluation of the similarity of each data with the 'Expected'. The results revealed that the order of their similarities with the 'Expected' was the qPCR data \geq the V1-2 data using 27Fmod $>$ the V5-6 data $>$ the V1-2 data using 27F \gg the data of Sanger sequencing-based full-length V1-9, indicating that the use of 27Fmod greatly improved the quantitative accuracy for evaluation of the overall

bacterial composition (Supplementary Table S5). This improvement was largely dependent on the improved estimation of the *Bifidobacterium* content by the use of 27Fmod. The average relative *Bifidobacterium* content in the two mock communities estimated from the data of V1-2 using 27F was only 1.5% of the 'Expected' (100%), whereas the use of 27Fmod increased the relative content to 61% that was also better than that estimated from the data of V5-6 and Sanger full-length analyses (Supplementary Fig. S2). Because qPCR can be used only when genomes of all bacteria in a given community are known, or only for a limited number of specific known species, we concluded that 454 pyrosequencing of the V1-2 region using 27Fmod-338R provided more quantitatively accurate data for bacterial composition in human gut microbiota than that using the conventional 27F primer.

We estimated the average error rate of filter-passed V1-2 data using 27Fmod-338R by aligning the V1-2 and reference 16S sequences of bacterial strains used in the two mock communities. The error rate was estimated to be 0.58 and 0.66% for mock01 and mock02 by local alignment, respectively (Supplementary Table S4). These error rates are similar to the previously published data,^{43,45,50} but lower than in another study.⁵⁸ The latter may be due to differences in the examined alignment length and between local and global alignments. Errors in 454 pyrosequencing data can be the primary cause for overestimation of the OTU number that is an issue which needs to be improved for accurate estimation of species richness in bacterial communities.^{59,60} We compared OTU numbers generated from clustering of various qualities of 16S reads with a 96% and a 97% pair-wise identity cutoff. For this comparison, we made and used three datasets: only primer check-passed reads having the highest error rates, filter-passed reads, and selected filter-passed reads having the lowest error rates. The results indicated that a 96% cutoff clustering of error-rich reads and a 97% cutoff clustering of filter-passed reads gave the worse results than a 96% cutoff clustering of filter-passed and selected filter-passed reads (Supplementary Fig. S3). A 97% cutoff was defined for clustering of highly accurate Sanger full-length 16S sequences.⁶¹ Therefore, in clustering of pyrosequencing data having higher error rate than Sanger data, the use of a cutoff identity lower than 97% and a lower number of reads are reasonable to reduce overestimation of the OTU number. A 96% cutoff clustering of filter-passed reads gave similar OTU numbers up to 30–50 reads to those of filter-passed reads having the lowest error rates. These read numbers are approximately three to five times the number of input strains. After several trials

testing the mock communities, we decided to use 3000–5000 reads per sample for clustering with a 96% cutoff for the analysis of human gut microbiota. Indeed, OTU numbers using a 96% cutoff clustering of 3000 reads decreased about 15% when compared with those using a 97% cutoff clustering.

3.2. Species richness and diversity in human faecal microbiota with probiotic intervention

We randomly selected 3000 reads of 16S V1-2 sequences from all filter-passed reads for each sample (Supplementary Table S2) and used 474 000 reads in total from 158 faecal DNA samples of 18 subjects for the analysis of species richness and composition in human gut microbiota. Clustering of all reads with a 96% pairwise-identity cutoff gave a total of 2758 OTUs. After removing the minority OTUs having <0.1% abundance in any samples, 1175 OTUs having $\geq 0.1\%$ abundance in at least 1 sample, accounting for 99.1% of all 16S reads, were used for further analysis.

3.2.1. Detection of administrated probiotic strains in faecal samples

We investigated whether administrated strains contained in the probiotic products can be detected in faecal DNA. We sequenced the 16S V1-2 region of all bacterial strains contained in probiotic products used in this study. The BLAST search to the databases indicated that except for the *Bifidobacterium longum* strain used in Group III, the 16S sequences of all strains in the probiotic products significantly differed from those of the indigenous species phylogenetically closest to the probiotic strains. The 16S sequence of the *B. longum* strain used in Group III was almost identical to that of an indigenous *Bifidobacterium* species, so that we used a distinguishable additive *Lactococcus lactis* strain in this product for the detection of administrated bacteria in Group III samples. The 16S sequences of these probiotic strains were included in the databases constructed in this study, and the 16S reads assigned to administrated strains had the average similarity between 99.4 and 99.9% identities with the reference sequences (data not shown). The 16S reads assigned to probiotic or additive strains were detected in samples (S01–S04) during probiotic intervention [designated 'Pro(+)] at various frequencies, but almost none were detected in samples (S00 and S05–S08) without probiotic administration [designated 'Pro(-)] (Supplementary Table S6). The administrated probiotic strains were shown to be more frequently detected in samples during the intervention than in the pre- and post-intervention periods using different detection methods such as culturing, targeted PCR, and hybridization.^{24,26–28,30,32,33,62} In

the present study, two probiotic *Lactobacillus* and one additive *Lactococcus* strains were detected in post-intervention samples in three subjects with a minimum count, respectively. The similarity of three 16S sequences was 99.4, 99.7, and 100% identity with those of administrated *Lactobacillus* and *Lactococcus* strains, indicating that these are administrated strains. The survival of some probiotic strains in the post-intervention period was also reported previously.^{28,30} Our data suggested that some probiotic strains seem to be able to persistently colonize the intestine and their survivability may be related to metabolic activity in the intestine.^{63,64} Probiotic *Bifidobacterium* strains were not detected in any Pro(-) samples. However, we found two distinct 16S sequences both assigned to *Bifidobacterium animalis* in two subjects APr37 and APr39. One showed a high similarity of >98% identity with the 16S sequence of the administrated *B. animalis* and was detected with high frequency only in the Pro(+) samples, whereas another showed a low similarity of 96.5–97.4% identity (a mean of 97.2%) with low frequency in both the Pro(-) and Pro(+) samples. These data suggest the presence of unknown indigenous species phylogenetically close to, but distinct from, probiotic *B. animalis* in human gut microbiota. The total number of bacteria contained in each probiotic product was varied between 10^9 and 10^{10} , showing no large difference in quantity among them (Supplementary Table S1). No clear correlation was

also observed between the number of bacteria in the products and the frequency in detection of the administrated strains in the Pro(+) samples. From these observations, the frequency of administrated bacteria detected in faeces may not be largely affected by their amounts in the products. Therefore, detection of *Lactobacillus brevis* and *Lactobacillus delbrueckii* at relatively low level in faeces cannot be simply explained by the difference in a dose, but could be considered the association with several factors such as their survivability in the intestine, diet, or physiological conditions of subjects.

3.2.2. Change of species richness in samples with and without probiotics We analysed species richness (OTU number) in the Pro(+) and Pro(-) samples. Supplementary Figure S4 shows the change in OTU numbers for every sample in each subject, indicating that OTU numbers vary dramatically for every sample. Most of the variation can be attributed to single OTUs representing the minority species. We averaged the OTU numbers of the Pro(-) and Pro(+) samples and compared them for subject, group, type of probiotics (*Lactobacillus* and *Bifidobacterium*), and all combined samples, respectively (Fig. 2). The average OTU numbers in 6 out of 18 subjects were decreased in the range of the ratio of 0.83–0.95 in the Pro(+) samples when compared with the Pro(-) samples, whereas those in other 12 subjects were increased in the range of the ratio of

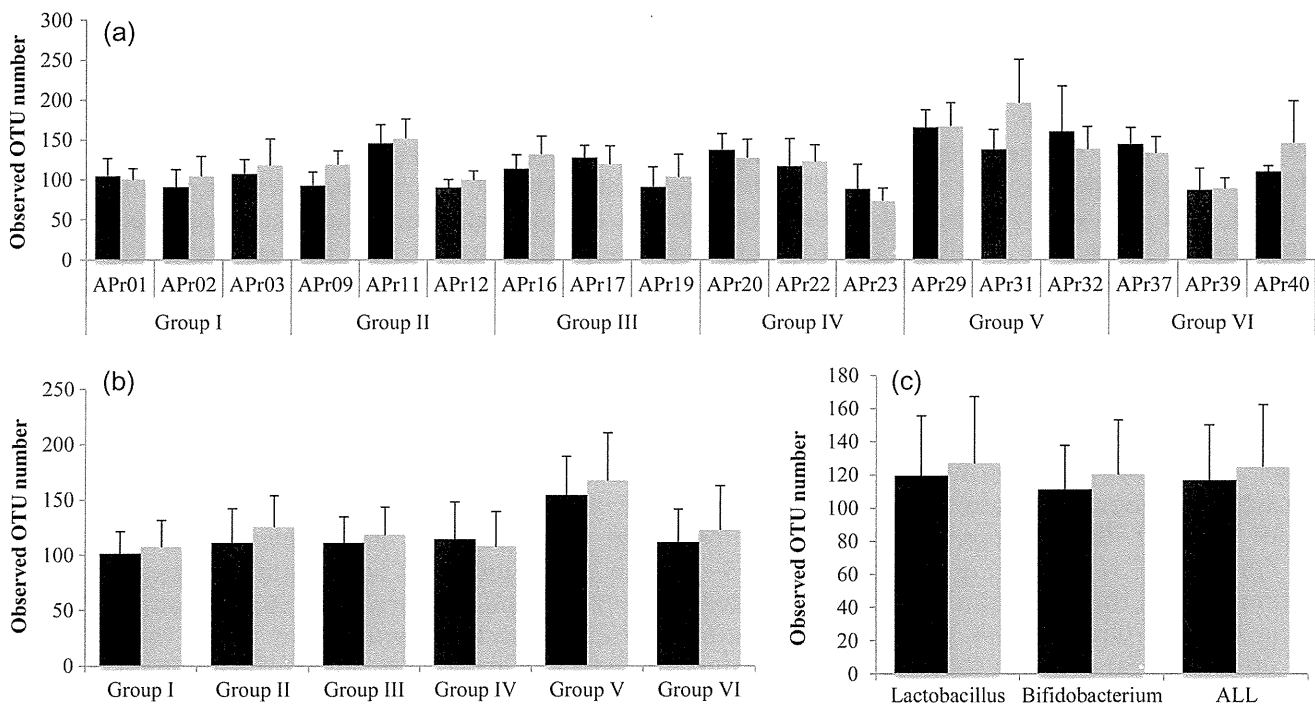


Figure 2. Change in OTU number in faecal microbiota with and without probiotic administration. (a) Individual, (b) group, (c) type of probiotics. Black bar indicates Pro(-) samples. Grey bar indicates Pro(+) samples. The error bars represent standard deviation.

1.01–1.43. For group, only Group IV showed a decrease in the average OTU number in the Pro(+) samples with the ratio of 0.94. For type of probiotics and all samples, the average OTU numbers in the Pro(+) samples were slightly more abundant (approximately 1.07-fold) than those in the Pro(–) samples, but no statistical significance was observed in any dataset. The increase in OTU number in the Pro(+) samples was largely due to the minority species (Supplementary Fig. S4), whereas the abundance of the majority species (OTUs containing ≥ 10 reads) was almost constant over time. We performed the same analysis using different sets of 3000 reads for each subject. The analysis reproducibly showed the similar pattern and the degree of the change in OTU numbers to which the minority species is largely attributed (data not shown). These data indicate that administration of probiotics tends to increase species richness in faecal microbiota that may

be beneficial for the consumer because the species richness in faecal microbiota of subjects afflicted with disease such as inflammatory bowel disease is significantly reduced when compared with that of healthy subjects.⁶⁵

3.2.3. Change of species composition in samples with and without probiotics We obtained the average weighted and unweighted UniFrac distances within Pro(–), within Pro(+), and between Pro(–) and Pro(+) samples for every group, probiotic types, and all subjects, respectively (Fig. 3). High UniFrac distance implies high variability of microbiota structure within and between samples. If the difference between any pair of the three distances is statistically significant, it can be considered that probiotic administration significantly affected the overall microbiota composition. We found the largest difference between weighted UniFrac distances of the Pro(+)

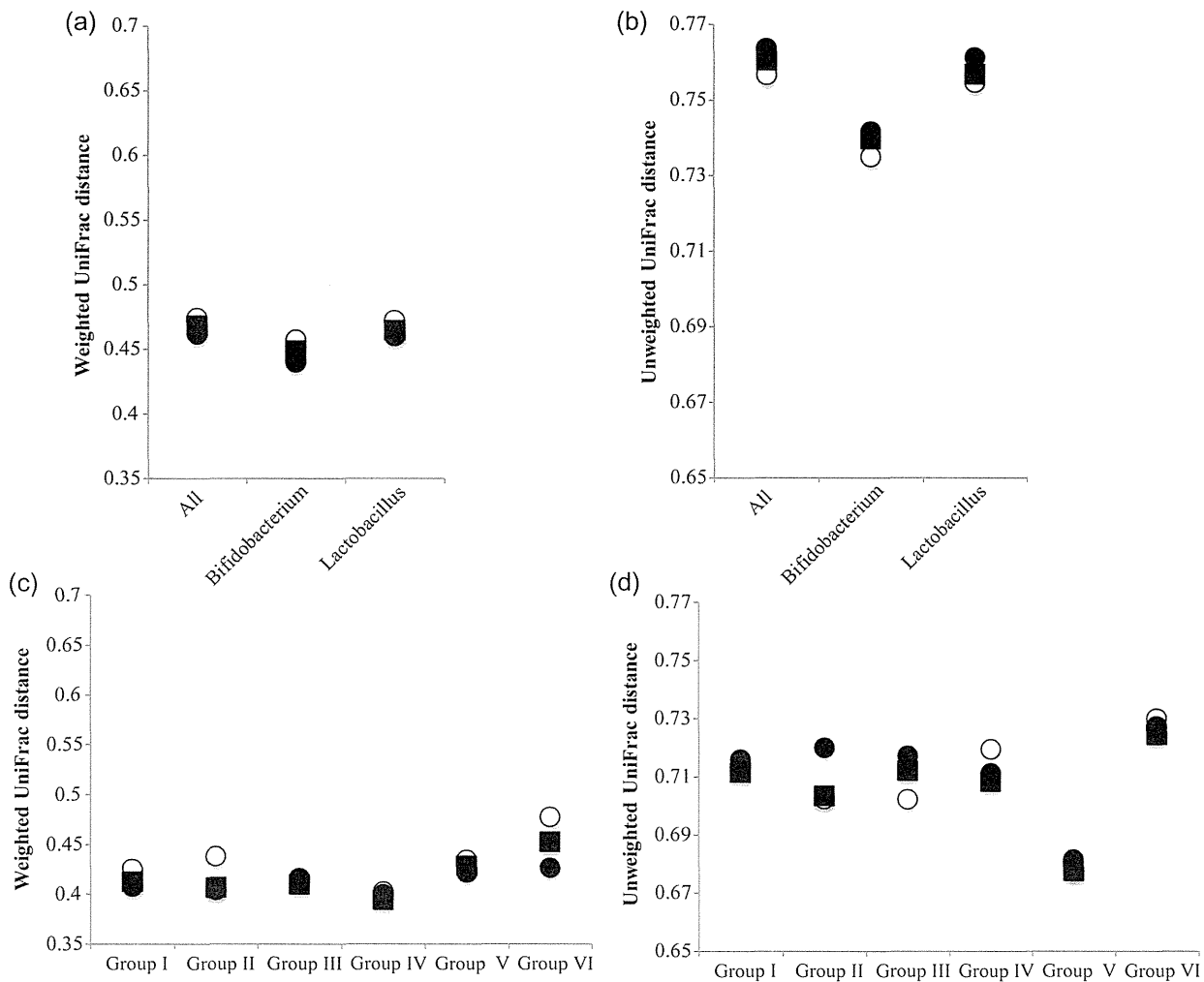


Figure 3. Average UniFrac distance within Pro (–) and Pro(+) and between Pro(–) and Pro(+) for each group, type of probiotics, and all subjects. Average UniFrac distance between any pair of the three distances for type of probiotics and all subjects (a and b), and each group (c and d). Open circle, closed circle, and closed square indicate average UniFrac distance within Pro (–), within Pro (+), and between Pro(–) and Pro(+) samples, respectively.

and Pro(-) samples in Group VI. However, statistical evaluation of this difference by the Student's *t*-test showed no significance (P -value > 0.05) for 781 out of 1000 times (Supplementary Table S7). These data imply high stability of gut microbiota to probiotic administration for all subjects examined. We also analysed UniFrac distances of intra-subject gut microbiota (Fig. 4). Although 5 subjects (APr02, 12, 16, 37, and 39) showed a significant difference in the UniFrac distances between Pro(-) and Pro(+) samples, the results showed that both weighted and unweighted distances between Pro(-) and Pro(+) of all intra-subjects were significantly lower than the average distance of the 18 unrelated subjects. The Welch's *t*-test for these differences showed statistical significance (Supplementary Table S8). We also performed the UniFrac distance analysis using different 16S datasets of 5000 reads for group, type of probiotics, all subjects, and intra-subject. The results similarly showed no statistical significance in differences between any pair of the 3 UniFrac distances and the significantly lower UniFrac distance of each intra-subject than that of the 18 unrelated subjects (data not shown). Thus, these data suggested that the perturbation of microbiota elicited by probiotics in an intra-subject did not overcome the inter-subject variations of gut microbiota, supporting high intra-specificity and stability of gut microbiota.^{66,67} This robustness of gut microbiota of adults is in contrast with the profound effect of antibiotic

administration on adult gut microbiota⁶⁸ and the observed response of gut microbiota of infants fed with probiotics, in which the infant gut microbiota composition was considerably affected by probiotics.³⁶ A short-term dietary intervention study showed that in controlled feeding of the same diet to subjects over 10 days, a marked change was observed within 1 day after the intervention initiation.⁶⁹ In the present study, no significant difference was observed between samples before (S00) and first samples (S01) after the intervention initiation (data not shown). It would be valuable to analyse faecal samples collected within a few days after administration of probiotics for evaluation of the short-term effect of probiotics.

4. Identification of bacterial species showing significant increase or decrease by probiotic administration

Although our results suggested that administration of probiotics had almost no effect on the overall structure of gut microbiota, it is possible to identify bacterial species largely responding to the administered probiotics at the OTU/species level. We surveyed OTUs showing an increase or a decrease between the Pro(+) and Pro(-) samples by comparing the number of 16S reads for each OTU. We first enumerated the OTUs showing ≥ 2 -fold change between the

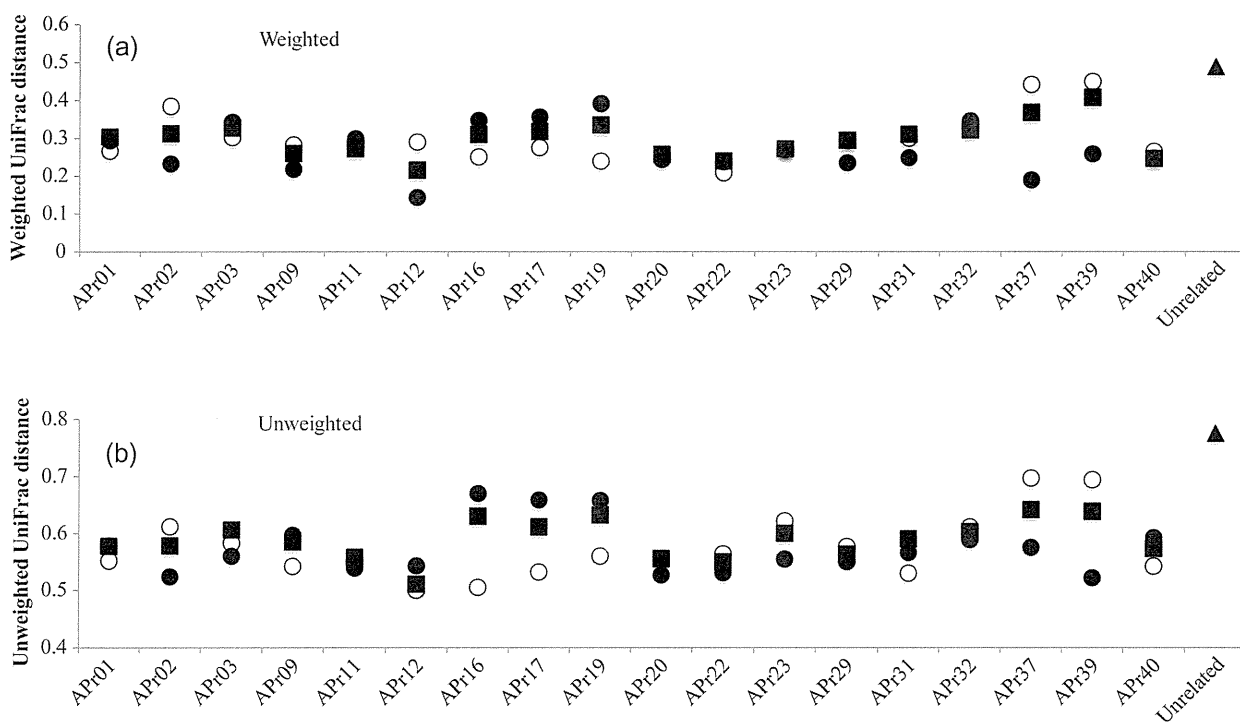


Figure 4. Average UniFrac distance within Pro(-) and Pro(+) and between Pro(-) and Pro(+) for each subject. Open circles, closed circles, and closed squares indicate average UniFrac distance within Pro(-), within Pro(+), and between Pro(-) and Pro(+) samples, respectively. Closed triangles indicate average UniFrac distance between samples (S00) of 18 unrelated individuals.

Pro(-) and Pro(+) samples for each subject, and the quantity difference was also obtained by subtracting the 16S read number of the Pro(+) samples from that of the Pro(-) samples. This is because OTUs showing a high quantity difference, but less fold change may also have substantial influence on gut microbiota composition. We found several OTUs significantly changed by probiotic administration, including OTUs assigned to both the indigenous and administrated strains (Fig. 5). We listed 88 OTUs (7.5% of all analysed 1175 OTUs) showing significant change of ≥ 3 -fold, among which 30 OTUs changed by ≥ 10 -fold (Supplementary Fig. S5). We excluded 6 OTUs assigned to the administrated strains from the 30 OTUs and obtained 24 OTUs assigned to the indigenous species, including OTU00072 assigned to *Streptococcus salivarius* that showed significant change in 2 subjects (Supplementary Table S9). We also found seven OTUs showing significant difference in quantity between both samples (Supplementary Table S10). Of the combined 32 OTUs (2.7%), 18 were increased and 14 were decreased by probiotic administration. Many of the OTUs showing a significant increase were assigned to minority species in the Pro(-) samples, but some increased up to nearly 7% in abundance (e.g. OTU00372 assigned to

Eubacterium rectale). On the other hand, the OTUs showing a significant decrease were almost undetected in the Pro(+) samples. Phylum-level species assignment showed that species belonging to the phylum *Firmicutes* were most largely affected by both probiotics, and all species belonging to the phylum *Bacteroidetes* were affected only by *Lactobacillus* probiotics (Table 1). The 32 OTUs were assigned to 27 indigenous species, among which 4 species (*Clostridium clostridioforme*, *Eubacterium eligens*, *E. rectale*, and *Faecalibacterium prausnitzii*) were assigned by 8 different OTUs and 1 species (*S. salivarius*) was assigned by the 2 same OTUs as described above. All these species except for *S. salivarius* were found to show significant change only in one subject, indicating that response of the indigenous species to probiotics is highly individual specific (Supplementary Fig. S6). Two different OTUs (OTU02677 and OTU02748) assigned to *F. prausnitzii*, of which the reduction is known to be correlated with inflammatory bowel disease,⁷⁰ were found to both decrease and increase in the same subject (APr40) by probiotic administration, suggesting that these two phylogenetically close species may have the diversity of response to probiotic action. We also examined distribution of the 32 OTUs in the subjects. The results revealed that 4 subjects

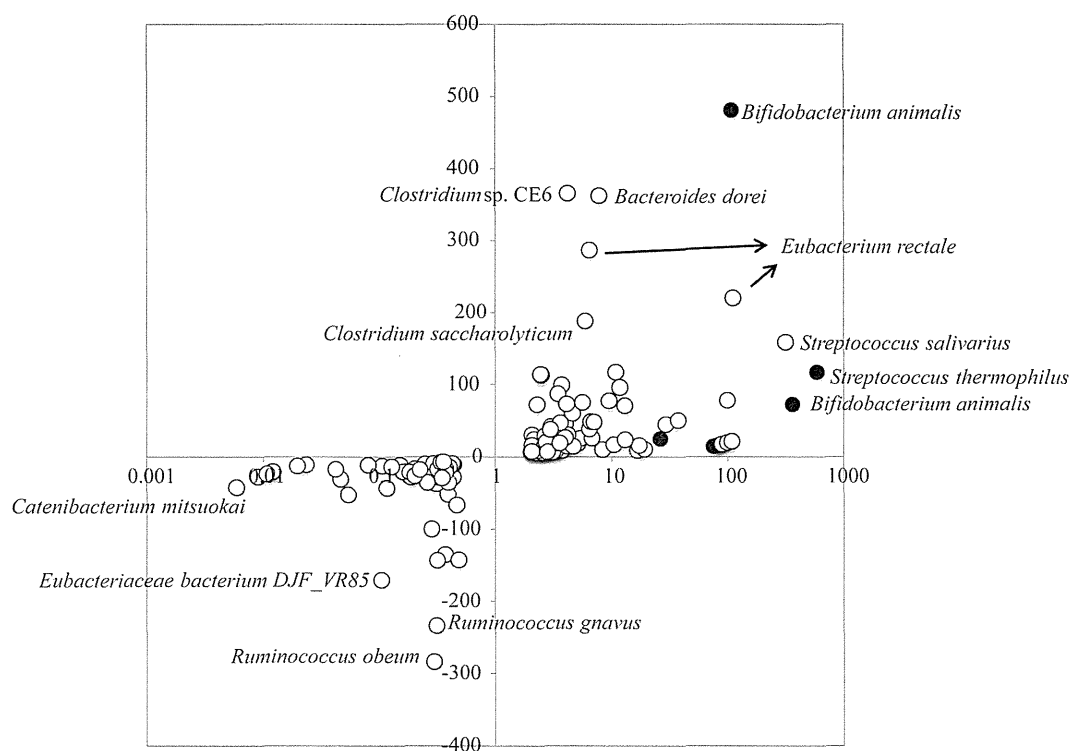


Figure 5. OTUs showing ≥ 2 -fold change and their difference in quantity between the Pro(-) and Pro(+) samples. The x-axis represents the scale of fold change between the Pro(+) and Pro(-) samples. The y-axis represents the difference (number of reads) in quantity between the Pro(+) and Pro(-) samples. Closed and open circles indicate the administrated probiotic and indigenous species, respectively.

Table 1. Phylum-level species assignment of OTUs showing significant fold change or quantity difference by administration of probiotics

Type of probiotics	Change	Number of varied OTUs	Fold change (≥ 10 -fold)				Number of varied OTUs	Difference (≥ 150 reads)	
			Firmicutes	Actinobacteria	Bacteroidetes	Unclassified bacterium		Firmicutes	Bacteroidetes
Lactobacillus	Increase	9	7	0	1	1	3	2	1
	Decrease	7	3	1	3	0	1	1	0
	Total	16	10	1	4	1	4	3	1
Bifidobacterium	Increase	5	5	0	0	0	1	1	0
	Decrease	4	4	0	0	0	2	2	0
	Total	9	9	0	0	0	3	3	0
All	Increase	14	12	0	1	1	4	3	1
	Decrease	11	7	1	3	0	3	3	0
	Total	25	19	1	4	1	7	6	1

^aAdministered probiotic strains were excluded, and only OTUs with a *P*-value < 0.05 are shown.

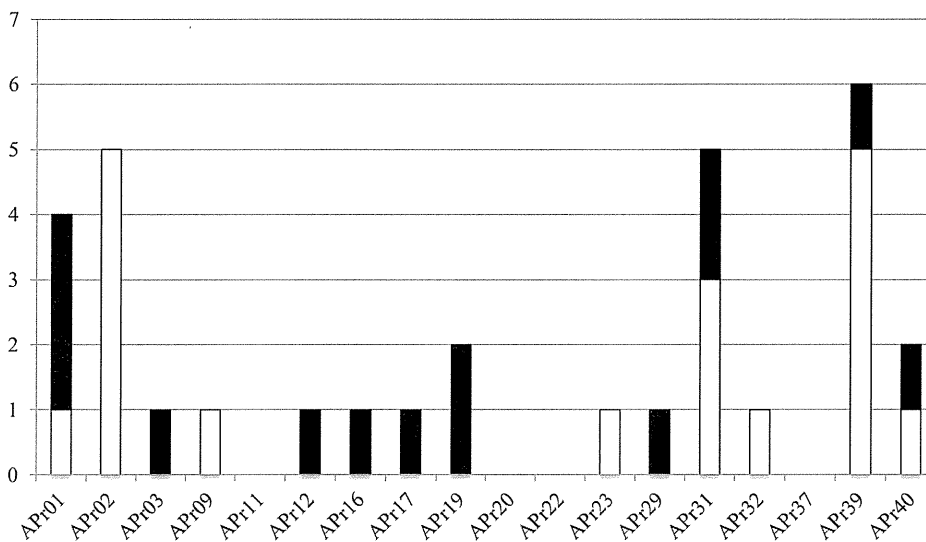


Figure 6. Distribution of 32 OTUs showing a significant change in 18 subjects. The *y*-axis indicates the number of OTUs showing significant change between the Pro(-) and Pro(+) samples in each subject (see Supplementary Tables S9 and S10). Open and closed bars indicate increased and decreased OTUs, respectively.

(APr11, 20, 22, and 37) did not have such OTUs and 8 subjects had only 1 OTU, whereas 4 subjects (APr01, 02, 31, and 39) had more than 4 OTUs showing significant change (Fig. 6), suggesting their uneven distribution in the 18 subjects. These data imply existence of the sensitive and less sensitive responders to probiotic action and if so, it would be interesting to investigate the relation between gut microbiota type and its response to probiotics.

In summary, we analysed changes of the gut microbiota composition of healthy adults fed with probiotics using the 454 pyrosequencing platform with the improved quantitative accuracy for evaluation of the overall bacterial composition. The present study using large datasets enabled us to more comprehensively and precisely evaluate the effect of probiotics on gut microbiota than the previous probiotic intervention researches in which the analysis exclusively

focused on only several limited bacterial species using conventional methods. Our data further support the high inter-subject variability and the high intra-subject stability that is the current common view for the feature of adult gut microbiota. A recent study of gut microbiota in twins demonstrated that probiotics had almost no effect on the community structure, but affected the gene expression of microbiota.³⁹ To more deeply understand the potential function of probiotics, the analysis of bacterial and host cell's transcriptome and intestinal metabolome is required.

Acknowledgements: We thank Dr Todd D. Taylor for critical reading of the manuscript, and K. Furuya, C. Shindo, H. Inaba, E. Iioka, Y. Takayama, E. Ohmori, M. Kiuchi, Y. Hattori (The University of Tokyo), and A. Nakano (Azabu University) for technical support.

Supplementary data: Supplementary Data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported in part by the global COE project of 'Genome Information Big Bang' from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (to M.H. and K.O.), a research project grant from Azabu University to H.M. and by a grant from the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Agency (JST) to K.O.


References

- Preidis, G.A. and Versalovic, J. 2009, Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era, *Gastroenterology*, **136**, 2015–31.
- Patel, R.M. and Lin, P.W. 2010, Developmental biology of gut-probiotic interaction, *Gut Microbes*, **1**, 186–95.
- Gerritsen, J., Smidt, H., Rijkers, G.T. and de Vos, W.M. 2011, Intestinal microbiota in human health and disease: the impact of probiotics, *Genes Nutr.*, **6**, 209–40.
- Sanders, M.E., Heimbach, J.T., Pot, B., et al. 2011, Health claims substantiation for probiotic and prebiotic products, *Gut Microbes*, **2**, 127–33.
- Aureli, P., Capurso, L., Castellazzi, A.M., et al. 2011, Probiotics and health: an evidence-based review, *Pharmacol. Res.*, **63**, 366–76.
- Rauch, M. and Lynch, S.V. 2012, The potential for probiotic manipulation of the gastrointestinal microbiome, *Curr. Opin. Biotechnol.*, **23**, 192–201.
- Fujimura, K.E., Slusher, N.A., Cabana, M.D. and Lynch, S.V. 2010, Role of the gut microbiota in defining human health, *Expert Rev. Anti Infect. Ther.*, **8**, 435–54.
- Deshpande, G.C., Rao, S.C., Keil, A.D. and Patole, S.K. 2011, Evidence-based guidelines for use of probiotics in preterm neonates, *BMC Med.*, **9**, 92.
- Bron, P.A., van Baarlen, P. and Kleerebezem, M. 2011, Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa, *Nat. Rev. Microbiol.*, **10**, 66–78.
- Thomas, D.W., Greer, F.R., American Academy of Pediatrics Committee on Nutrition; American Academy of Pediatrics Section on Gastroenterology, Hepatology, and Nutrition. 2010, Probiotics and prebiotics in pediatrics, *Pediatrics*, **126**, 1217–31.
- Indrio, F. and Neu, J.N. 2011, The intestinal microbiome of infants and the use of probiotics, *Curr. Opin. Pediatr.*, **23**, 145–50.
- Saxelin, M., Tynkkynen, S., Mattila-Sandholm, T. and de Vos, W.M. 2005, Probiotic and other functional microbes: from markets to mechanisms, *Curr. Opin. Biotechnol.*, **16**, 204–11.
- Nagpal, R., Kumar, A., Kumar, M., Behare, P.V., Jain, S. and Yadav, H. 2012, Probiotics, their health benefits and applications for developing healthier foods: a review, *FEMS Microbiol. Lett.*, **334**, 1–15.
- Bisanz, J.E. and Reid, G. 2011, Unraveling how probiotic yogurt works, *Sci. Transl. Med.*, **3**, 106ps41.
- Bron, P.A. and Kleerebezem, M. 2011, Engineering lactic acid bacteria for increased industrial functionality, *Bioeng. Bugs*, **2**, 80–7.
- Kleerebezem, M. and Vaughan, E.E. 2009, Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity, *Annu. Rev. Microbiol.*, **63**, 269–90.
- Ventura, M., O'Flaherty, S., Claesson, M.J., et al. 2009, Genome-scale analyses of health-promoting bacteria: probiogenomics, *Nat. Rev. Microbiol.*, **7**, 61–71.
- Snydman, D.R. 2008, The safety of probiotics, *Clin. Infect. Dis.*, **46**, S104–11.
- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. and Knight, R. 2012, Diversity, stability and resilience of the human gut microbiota, *Nature*, **489**, 220–30.
- Clemente, J.C., Ursell, L.K., Parfrey, L.W. and Knight, R. 2012, The impact of the gut microbiota on human health: an integrative view, *Cell*, **148**, 1258–70.
- Nicholson, J.K., Holmes, E., Kinross, J., et al. 2012, Host-gut microbiota metabolic interactions, *Science*, **336**, 1262–7.
- Walter, J. and Ley, R. 2011, The human gut microbiome: ecology and recent evolutionary changes, *Annu. Rev. Microbiol.*, **65**, 411–29.
- Hooper, L.V., Littman, D.R. and Macpherson, A.J. 2012, Interactions between the microbiota and the immune system, *Science*, **336**, 1268–73.
- Tannock, G.W., Munro, K., Harmsen, H.J., Welling, G.W., Smart, J. and Gopal, P.K. 2000, Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20, *Appl. Environ. Microbiol.*, **66**, 2578–88.
- García-Albiach, R., Pozuelo de Felipe, M.J., Angulo, S., et al. 2008, Molecular analysis of yogurt containing *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus* in human intestinal microbiota, *Am. J. Clin. Nutr.*, **87**, 91–6.
- Alvaro, E., Andrieux, C., Rochet, V., et al. 2007, Composition and metabolism of the intestinal microbiota in consumers and non-consumers of yogurt, *Br. J. Nutr.*, **97**, 126–33.
- Rochet, V., Rigottier-Gois, L., Levenez, F., et al. 2008, Modulation of *Lactobacillus casei* in ileal and fecal samples from healthy volunteers after consumption of a fermented milk containing *Lactobacillus casei* DN-114 001Rif, *Can. J. Microbiol.*, **54**, 660–7.
- Rochet, V., Rigottier-Gois, L., Ledaire, A., et al. 2008, Survival of *Bifidobacterium animalis* DN-173 010 in the faecal microbiota after administration in lyophilized form or in fermented product – a randomised study in healthy adults, *J. Mol. Microbiol. Biotechnol.*, **14**, 128–36.
- Ouweland, A.C., Bergsma, N., Parhiala, R., et al. 2008, Bifidobacterium microbiota and parameters of immune function in elderly subjects, *FEMS Immunol. Med. Microbiol.*, **53**, 18–25.

30. Firmesse, O., Mogenet, A., Bresson, J.L., Corthier, G. and Furet, J.P. 2008, *Lactobacillus rhamnosus* R11 consumed in a food supplement survived human digestive transit without modifying microbiota equilibrium as assessed by real-time polymerase chain reaction, *J. Mol. Microbiol. Biotechnol.*, **14**, 90–9.
31. Lahtinen, S.J., Tammela, L., Korpela, J., et al. 2009, Probiotics modulate the Bifidobacterium microbiota of elderly nursing home residents, *Age (Dordr)*, **31**, 59–66.
32. Savard, P., Lamarche, B., Paradis, M.E., Thiboutot, H., Laurin, É. and Roy, D. 2011, Impact of *Bifidobacterium animalis* subsp. lactis BB-12 and, *Lactobacillus acidophilus* LA-5-containing yoghurt, on fecal bacterial counts of healthy adults, *Int. J. Food Microbiol.*, **149**, 50–7.
33. Yamano, T., Iino, H., Takada, M., Blum, S., Rochat, F. and Fukushima, Y. 2006, Improvement of the human intestinal flora by ingestion of the probiotic strain *Lactobacillus johnsonii* La1, *Br. J. Nutr.*, **95**, 303–12.
34. Engelbrekton, A.L., Korzenik, J.R. and Sanders, M.E. 2006, Analysis of treatment effects on the microbial ecology of the human intestine, *FEMS Microbiol. Ecol.*, **57**, 239–50.
35. Marzotto, M., Maffei, C., Paternoster, T., et al. 2006, *Lactobacillus paracasei* A survives gastrointestinal passage and affects the fecal microbiota of healthy infants, *Res. Microbiol.*, **157**, 857–66.
36. Cox, M.J., Huang, Y.J., Fujimura, K.E., et al. 2010, *Lactobacillus casei* abundance is associated with profound shifts in the infant gut microbiome, *PLoS One*, **5**, e8745.
37. Culligan, E.P., Hill, C. and Sleator, R.D. 2009, Probiotics and gastrointestinal disease: successes, problems and future prospects, *Gut Pathog.*, **1**, 19.
38. Gareau, M.G., Sherman, P.M. and Walker, W.A. 2010, Probiotics and the gut microbiota in intestinal health and disease, *Nat. Rev. Gastroenterol. Hepatol.*, **7**, 503–14.
39. McNulty, N.P., Yatsunenkov, T., Hsiao, A., et al. 2011, The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins, *Sci. Transl. Med.*, **3**, 106ra106.
40. Metzker, M.L. 2010, Sequencing technologies—the next generation, *Nat. Rev. Genet.*, **11**, 31–46.
41. Qin, J., Li, R., Raes, J., et al. 2010, A human gut microbial gene catalogue established by metagenomic sequencing, *Nature*, **464**, 59–65.
42. Huse, S.M., Dethlefsen, L., Huber, J.A., Mark, W.D., Relman, D.A. and Sogin, M.L. 2008, Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing, *PLoS Genet.*, **4**, e1000255.
43. Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. and Welch, D.M. 2007, Accuracy and quality of massively parallel DNA pyrosequencing, *Genome Biol.*, **8**, R143.
44. Hamady, M., Lozupone, C. and Knight, R. 2010, Fast Unifrac: facilitating high-throughput phylogenetic analysis of microbial communities including analysis of pyrosequencing and PhyloChip data, *ISME J.*, **4**, 17–27.
45. Droege, M. and Hill, B. 2008, The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets, *J. Biotechnol.*, **136**, 3–10.
46. Andersson, A.F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P. and Engstrand, L. 2008, Comparative analysis of human gut microbiota by barcoded pyrosequencing, *PLoS One*, **3**, e2836.
47. Kuczynski, J., Lauber, C.L., Walters, W.A., et al. 2011, Experimental and analytical tools for studying the human microbiome, *Nat. Rev. Genet.*, **13**, 47–58.
48. Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J. and Knight, R. 2008, Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex, *Nat. Methods*, **5**, 235–7.
49. Morita, H., Kuwahara, T., Ohshima, K., et al. 2007, An improved isolation method for metagenomic analysis of the microbial flora of the human intestine, *Microbes Environ.*, **22**, 214–22.
50. Claesson, M.J., Wang, Q., O'Sullivan, O., et al. 2010, Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions, *Nucleic Acids Res.*, **38**, e200.
51. Zhou, H.W., Li, D.F., Tam, N.F., et al. 2011, BIPES, a cost-effective high-throughput method for assessing microbial diversity, *ISME J.*, **5**, 741–9.
52. Hattori, M. and Taylor, T.D. 2009, The human intestinal microbiome: a new frontier of human biology, *DNA Res.*, **16**, 1–12.
53. Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A. and Olsen, G.J. 2008, Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes, *Appl. Environ. Microbiol.*, **74**, 2461–70.
54. Hill, J.E., Fernando, W.M., Zello, G.A., Tyler, R.T., Dahl, W.J. and Van Kessel, A.G. 2010, Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the cpn60 universal primer cocktail, *Appl. Environ. Microbiol.*, **76**, 4550–2.
55. Palmer, C. 2007, Development of the human infant intestinal microbiota, *PLoS Biol.*, **5**, 1556–73.
56. Haas, B.J., Gevers, D., Earl, A.M., et al. 2011, Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons, *Genome Res.*, **21**, 494–504.
57. Schloss, P.D., Gevers, D. and Westcott, S.L. 2011, Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies, *PLoS One*, **6**, e27310.
58. Gilles, A., Megléc, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. 2011, Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing, *BMC Genomics*, **12**, 245.
59. Quince, C., Lanzén, A., Curtis, T.P., et al. 2009, Accurate determination of microbial diversity from 454 pyrosequencing data, *Nat. Methods*, **6**, 639–41.
60. Diaz, P.I., Dupuy, A.K., Abusleme, L., et al. 2012, Using high throughput sequencing to explore the biodiversity in oral bacterial communities, *Mol. Oral Microbiol.*, **27**, 182–201.
61. Schloss, P.D. and Handelsman, J. 2005, Introducing DOTUR, a computer program for defining operational

- taxonomic units and estimating species richness, *Appl. Environ. Microbiol.*, **71**, 1501–6.
62. del Campo, R., Bravo, D., Cantón, R., et al. 2005, Scarce evidence of yogurt lactic acid bacteria in human feces after daily yogurt consumption by healthy volunteers, *Appl. Environ. Microbiol.*, **71**, 547–9.
63. Oozeer, R., Leplingard, A., Mater, D.D., et al. 2006, Survival of *Lactobacillus casei* in the human digestive tract after consumption of fermented milk, *Appl. Environ. Microbiol.*, **72**, 5615–7.
64. Marco, M.L., de Vries, M.C., Wels, M., et al. 2010, Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice, *ISME J.*, **4**, 1481–4.
65. Manichanh, C., Rigottier-Gois, L., Bonnaud, E., et al. 2006, Reduced diversity of fecal microbiota in Crohn's disease revealed by a metagenomic approach, *Gut*, **55**, 205–11.
66. Turnbaugh, P.J., Hamady, M.H., Yatsuneko, T., et al. 2009, A core gut microbiome in obese and lean twins, *Nature*, **475**, 480–4.
67. Kurokawa, K., Itoh, T., Kuhahara, T., et al. 2007, Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes, *DNA Res.*, **14**, 169–81.
68. Jernberg, C., Löfmark, S., Edlund, C. and Jansson, J.K. 2007, Long-term ecological impacts of antibiotic administration on the human intestinal microbiota, *ISME J.*, **1**, 56–66.
69. Wu, G.D., Chen, J., Hoffmann, C., et al. 2011, Linking long-term dietary patterns with gut microbial enterotypes, *Science*, **334**, 105–8.
70. Sokol, H., Seksik, P., Furet, J.P., et al. 2009, Low counts of *Faecalibacterium prausnitzii* in colitis microbiota, *Inflamm. Bowel Dis.*, **15**, 1183–9.

Molecular network of chromatin immunoprecipitation followed by deep sequencing-based vitamin D receptor target genes

Multiple Sclerosis Journal
0(0) 1–11
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1352458512471873
msj.sagepub.com


Jun-ichi Satoh and Hiroko Tabunoki

Abstract

Background: Vitamin D is a liposoluble vitamin essential for calcium metabolism. The ligand-bound vitamin D receptor (VDR), heterodimerized with retinoid X receptor, interacts with vitamin D response elements (VDREs) to regulate gene expression. Vitamin D deficiency due to insufficient sunlight exposure confers an increased risk for multiple sclerosis (MS).

Objective: To study a protective role of vitamin D in multiple sclerosis (MS), it is important to characterize the global molecular network of VDR target genes (VDRTGs) in immune cells.

Methods: We identified genome-wide VDRTGs collectively from two distinct chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) datasets of VDR-binding sites derived from calcitriol-treated human cells of B cell and monocyte origins. We mapped short reads of next generation sequencing (NGS) data on hg19 with Bowtie, detected the peaks with Model-based Analysis of ChIP-Seq (MACS), and identified genomic locations by GenomeJack, a novel genome viewer for NGS platforms.

Results: We found 2997 stringent peaks distributed on protein-coding genes, chiefly located in the promoter and the intron on VDRE DR3 sequences. However, the corresponding transcriptome data verified calcitriol-induced upregulation of only a small set of VDRTGs. The molecular network of 1541 calcitriol-responsive VDRTGs showed a significant relationship with leukocyte transendothelial migration, Fcγ receptor-mediated phagocytosis, and transcriptional regulation by VDR, suggesting a pivotal role of genome-wide VDRTGs in immune regulation.

Conclusion: These results suggest the working hypothesis that persistent deficiency of vitamin D might perturb the complex network of VDRTGs in immune cells, being responsible for induction of an autoimmune response causative for MS.

Keywords

ChIP-Seq, GenomeJack, multiple sclerosis, pathway analysis, vitamin D, VDR

Date received: 4th June 2012; revised: 21st November 2012; accepted: 28th November 2012

Introduction

Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system (CNS) presenting with relapsing–remitting and progressive clinical courses. Pathologically, it is characterized by inflammation, demyelination, gliosis, and axonal degeneration, although underlying molecular mechanisms remain largely unknown. Even genetically identical monozygotic twins express a concordance rate of approximately 30% for MS. This suggests a great contribution of environmental factors to MS pathogenesis, such as cigarette smoking, Epstein Barr virus (EBV) infection, and low serum vitamin D levels.¹

Vitamin D is a liposoluble vitamin essential for calcium metabolism, available from dietary sources and also produced endogenously when the sunlight Ultraviolet B

(UVB) triggers its synthesis in the skin, followed by consecutive hydroxylation in the liver and kidney to generate 1 α ,25-dihydroxyvitamin D₃, a biologically active form named calcitriol. The vitamin D receptor (VDR) is expressed on various cells including immune cells, such as lymphocytes, macrophages, monocytes, and dendritic cells.

Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Japan.

Corresponding author:

Jun-ichi Satoh, Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, 2-522-1 Noshio, Kiyose, Tokyo 204-8588, Japan.
Email: satoj@my-pharm.ac.jp

The ligand-bound VDR, heterodimerized with retinoid X receptor (RXR), interacts with vitamin D response elements (VDREs) on regulatory regions of target genes, and recruits protein complexes that alter chromatin structure. They regulate expression of more than 900 vitamin D-responsive genes involved in calcium homeostasis, cell growth, differentiation, apoptosis, and modulation of immune response.² Importantly, the expression of the gene encoding MS-associated MHC class II allele HLA-DRB1*1501 is directly regulated by vitamin D via a VDRE on its promoter.³

A prospective case-control study on millions of US military personnel indicated that MS risk is decreased with increasing serum levels of 25-hydroxyvitamin D among white subjects.⁴ Higher serum levels of 25-hydroxyvitamin D are associated with a greater suppressive capacity of regulatory T (Treg) cells in relapsing-remitting MS patients.⁵ Calcitriol inhibits production of proinflammatory cytokines, such as IFN γ , IL-17, and IL-21, from human CD4⁺ T cells, being responsible for the anti-inflammatory activity of vitamin D.⁶ All of these observations support a protective role of vitamin D in immune cells against development of MS.⁷ To elucidate underlying molecular mechanisms, it is important to characterize the global molecular network of VDR target genes (VDRTGs) presenting with diverse biological roles in immune cells, including those previously underestimated in immunopathogenesis of MS.

Recently, the rapid progress in the next-generation sequencing (NGS) technology has revolutionized the field of genome research. Notably, NGS thoroughly characterized the genetic basis of MS at the level of the whole genome of individual patients.⁸ Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) has been used as a NGS application to achieve genome-wide profiling of DNA-binding proteins, histone modifications, and nucleosomes.⁹ ChIP-Seq with an advantage of higher resolution, less noise, and greater coverage of the genome, compared with the microarray-based ChIP-Chip method, and provides an innovative tool for studying the comprehensive gene regulatory networks. Since NGS analysis produces extremely high throughput experimental data, it is often difficult to extract the biologically meaningful implications. Recent advances in systems biology enable us to illustrate the cell-wide map of the complex molecular interactions by using the literature-based knowledgebase of molecular pathways.¹⁰ The logically-arranged molecular networks construct the whole system, characterized by robustness, which maintains the proper function of the system in the face of genetic and environmental perturbations. Therefore, the integration of high-dimensional NGS data with underlying molecular networks serves as a rational approach to characterize the genome-wide network-based molecular mechanisms of gene regulation.

For an initial step to study molecular mechanisms underlying vitamin D-mediated protection against MS, we

identified thousands of *in vivo* calcitriol-responsive VDRTG candidates collected from two public ChIP-Seq datasets of VDR-binding sites in human cells of B cell and monocyte origins, and investigated their molecular network by using three distinct pathway analysis tools of bioinformatics. We focused on the datasets derived from immune-relevant cell types of B cell and monocyte origins because B cells and monocytes play a key role in MS pathogenesis by serving as a source of proinflammatory and immunoregulatory cytokines and chemokines.¹¹

Datasets and methods

Two distinct ChIP-Seq datasets of VDR-binding sites

We identified a comprehensive set of VDRTGs from two distinct ChIP-Seq datasets of VDR-binding sites numbered SRP002673 and SRP005910, retrieved from DDBJ Sequence Read Archive (DRA) (trace.ddbj.nig.ac.jp/DRASearch). SRP002673 contained the dataset of EBV-transformed human B lymphoblast cell lines (LCLs) isolated from two Centre d'Etude du Polymorphisme Humain (CEPH) individuals GM10855 and GM10861 who participated in the HapMap project.¹² The cells were treated with or without 100 nM calcitriol for 36 h, and then processed for ChIP-Seq using a rabbit anti-VDR antibody (sc-1008; Santa Cruz Biotechnology, USA). They were processed in parallel for transcriptome analysis on a Human Gene 1.0 ST Array (Affymetrix). We retrieved the transcriptome dataset numbered GSE22176 from Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/gds). SRP005910 contained the dataset of THP-1 human monocyte leukemia cell line following exposure to 10 nM calcitriol or vehicle (ethanol) for 40 min for ChIP-Seq with a rabbit anti-VDR antibody (ab3508; Abcam) and for 4 h for transcriptome analysis on a Human-6 v2.0 Expression Beadchip (Illumina, San Diego, California, USA).¹³ We retrieved the transcriptome dataset numbered GSE27270 from GEO. ChIP DNA fragments were processed for deep sequencing at a 35 bp (SRP002673) or a 36 bp (SRP005910) read length on the Genome Analyzer II (Illumina).

We mapped short reads of NGS data on the human genome reference sequence numbered hg19 by using Bowtie 0.12.7 (bowtie-bio.sourceforge.net). Then, we detected the statistically significant peaks of mapped reads by using Model-based Analysis of ChIP-Seq (MACS; liulab.dfci.harvard.edu/MACS) under the stringent condition of fold enrichment (FE) \geq 20 and the false discovery rate (FDR) \leq 1%. Finally, we identified genomic locations of the peaks by importing the processed data into GenomeJack v1.3, a novel genome viewer for NGS platforms (Mitsubishi Space Software; www.mss.co.jp/businessfield/bioinformatics). Based on RefSeq ID, the MACS peaks were categorized into those located on protein-coding genes with mRNA-

coding transcript (NM)-heading numbers, those on non-coding genes with non-coding transcript (NR)-heading numbers, and those located in intergenic regions without relevant neighboring genes. Since proximal promoters are generally located within several hundreds bp upstream of the transcription start site (TSS), the genomic locations of the summits were classified into the promoter region defined by the location within a 5 kb upstream from the 5' end of genes, the 5' untranslated region (5'UTR), the exon, the intron, and the 3'UTR. The locations outside these, which might include the locations of several distal promoters, were tentatively defined as intergenic regions, owing to the inability to accurately assess relevant target genes. The consensus motif sequences were identified by importing a 400 bp-length sequence surrounding the peak summit into MEME-ChIP (meme.sdsc.edu/meme/cgi-bin/meme-chip.cgi).

We also analyzed the transcriptome data of microarray analysis corresponding to the ChIP-Seq datasets. The log₂-converted normalized data were evaluated by Welch's t-test to identify the genes differentially expressed between calcitriol-treated and control samples.

Molecular network analysis

To identify biologically relevant molecular networks and pathways, we imported Entrez Gene IDs of VDRTGs into the Functional Annotation tool of Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (david.abcc.ncifcrf.gov). DAVID identifies the most relevant pathway constructed by Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.kegg.jp), composed of the genes enriched in the given set with an output of statistical significance evaluated by the modified Fisher's exact test. KEGG is a publicly accessible knowledgebase containing manually curated reference pathways that cover a wide range of metabolic, genetic and cellular processes, and human diseases, currently composed of 218,213 pathways generated from 435 reference pathways. We also imported Entrez Gene IDs into Ingenuity Pathways Analysis (IPA) (Ingenuity Systems; www.ingenuity.com) and KeyMolnet (Institute of Medicinal Molecular Design; www.immd.co.jp), both of which are commercial tools for molecular network analysis.

IPA is a knowledgebase that contains approximately 2,500,000 biological and chemical interactions and functional annotations with definite scientific evidence. By uploading the list of Gene IDs and expression values, the network-generation algorithm identifies focused genes integrated in a global molecular network. IPA calculates the score *p*-value, the statistical significance of association between the genes and the networks by the Fisher's exact test.

KeyMolnet contains knowledge-based contents on 150,500 relationships among human genes and proteins, small molecules, diseases, pathways and drugs.¹⁰ They are categorized into the core contents collected from selected

review articles with the highest reliability or the secondary contents extracted from abstracts of PubMed and Human Reference Protein database (HPRD). By importing the list of Gene IDs and expression values, the neighboring network-search algorithm selected one or more molecules as starting points to generate the network of all kinds of molecular interactions around starting molecules, including direct activation/inactivation, transcriptional activation/repression, and the complex formation within the designated number of paths from starting points. The generated network was compared side by side with 484 human canonical pathways of the KeyMolnet library. The algorithm counting the number of overlapping molecular relations between the extracted network and the canonical pathway makes it possible to identify the canonical pathway showing the most significant contribution to the extracted network.

Results

Identification of 2997 VDR target genes from two distinct ChIP-Seq datasets

By analyzing two distinct VDR-binding ChIP-Seq datasets numbered SRP002673 and SRP005910, derived from three samples of calcitriol-treated cells and three samples of non-treated or vehicle-treated (control) cells, we identified totally 4718 stringent ChIP-Seq peaks showing $FE \geq 20$ and $FDR \leq 1\%$ (Table 1). The genomic locations of the peaks were determined by using GenomeJack (Figure 1(a)). VDRE consensus sequences, comprising direct or inverted repeats of hexameric (G/A)G(G/T)TCA motifs with three intervening base pairs (DR3), were detected in the top 50 enriched genes in calcitriol-treated samples (Figure 1(b); Figure 2). After omitting the peaks located on non-coding genes and those located in intergenic regions, we identified 2997 peaks located on protein-coding genes that have their summits on the promoter ($n=1353$), 5'UTR ($n=249$), exon ($n=85$), intron ($n=1253$) or 3'UTR ($n=57$) (Table 1; Supplementary Tables 1–3 online). They are tentatively designated as ChIP-Seq-based VDRTGs.

Calcitriol-induced upregulation of a small set of ChIP-Seq-based VDR target genes

We also analyzed the transcriptome data numbered GSE22176 and GSE27270 corresponding to the ChIP-Seq datasets. Initially, we could not detect any genes differentially expressed between calcitriol-treated and control samples by the microarray-standard statistical evaluation with Welch's t-test followed by Bonferroni correction for multiple comparison, except for CD14 exhibiting a 6.65-fold increase in THP-1 cells by calcitriol treatment ($p=0.0000004$). Therefore, we attempted to extract greater numbers of calcitriol-responsive genes under the statistically less stringent condition of $p \leq 0.05$ by Welch's t-test

Table 1. The overview of vitamin D receptor (VDR) target genes with stringent chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) peaks.

Datasets	Samples and treatment	Peaks on coding genes	Location of summits on coding genes					Peaks on non-coding genes	Peaks on intergenic regions	Total numbers of peaks
			Promoter	5'UTR	Exon	Intron	3'UTR			
SRP002673	GMI0855LCL-control	54	39	8	1	5	1	8	6	68
	GMI0855LCL-calcitriol	200	54	7	4	126	9	23	118	341
	GMI0861LCL-control	402	281	63	7	48	3	41	70	513
	GMI0861LCL-calcitriol	2047	875	161	50	919	42	223	1020	3290
SRP005910	THPI-control	108	55	5	14	34	0	15	63	186
	THPI-calcitriol	186	49	5	9	121	2	17	117	320
Total numbers of peaks		2997	1353	249	85	1253	57	327	1394	4718

By analyzing two distinct ChIP-Seq datasets of VDR-binding sites numbered SRP002673 and SRP005910, composed of three calcitriol-treated and three non-treated or vehicle-treated (control) samples, we identified stringent peaks exhibiting fold enrichment (FE) ≥ 20 and false discovery rate (FDR) $\leq 1\%$, which were categorized into the summits of the peaks located on protein-coding genes, those on non-coding genes and those in intergenic regions, by using Genomejack.

without adjustment by Bonferroni correction and fold change ≥ 1.5 as upregulation or ≤ 0.75 as downregulation. This evaluation results in detection of overall 134 upregulated and 43 downregulated genes (Supplementary Tables 4–6 online). The set of 29 genes (21.6%) among 134 upregulated genes agreed with ChIP-Seq-based VDRTGs, including CLMN, PPP1R2, SPATA13, SULF2, P2RY8, MOBKL2B, CAMP, CD38, FAM116B, SERINC2, TRIM31, SLC25A20, TNFRSF17, ALDH1A1, CTSZ, ASAP2, SLC37A2, LRRRC8A, TMEM37, C20orf197, PDZD7, SFXN3, PRKCH, SEMA4D, SEMA6B, CYTH4, ITGB5, ELF4, and ZMIZ1. In contrast, none of the downregulated genes have met with the list of ChIP-Seq-based VDRTGs.

Identification of 1541 ChIP-Seq-based calcitriol-responsive VDR target genes

Next, we combined all VDRTGs derived from three different calcitriol-treated samples into one, named the treatment set composed of 2432 genes, and those from three non-treated samples into one, named the non-treatment set composed of 564 genes. Thus, we found the constitutive binding of VDR on a genome-wide scale to a substantial number of genes in the absence of calcitriol treatment. Then, we extracted a non-redundant set of 1541 calcitriol-responsive genes that were detected exclusively in the treatment set (Supplementary Table 7 online).

Molecular network of 1541 calcitriol-responsive VDR target genes

By importing Entrez Gene IDs of 1541 calcitriol-responsive VDRTGs into the Functional Annotation

tool of DAVID, we identified the top five most relevant KEGG pathways (Table 2). They include the pathways in cancer (hsa05200; $p=4.92E-05$), leukocyte transendothelial migration (hsa04670; $p=1.17E-04$) (Figure 3), systemic lupus erythematosus (hsa05322; $p=5.57E-04$), focal adhesion (hsa04510; $p=7.30E-04$), and Fc γ receptor-mediated phagocytosis (hsa04666; $p=2.39E-03$).

By using the Core Analysis tool of IPA, we identified the top five most relevant canonical pathways (Supplementary Table 8 online). They include Fc γ receptor-mediated phagocytosis in macrophages and monocytes ($p=6.85E-07$), molecular mechanisms of cancer ($p=1.53E-06$), macropinocytosis signaling ($p=5.61E-06$), glioma signaling ($p=1.44E-05$), and leukocyte extravasation signaling ($p=2.03E-05$). Thus, the results of KEGG and IPA showed a substantial overlap between the pathways constructed by VDRTGs, particularly whose function was involved in extravasation and phagocytosis of immune cells. IPA also extracted the top five most relevant functional networks. They include “Gene Expression, Protein Synthesis, Cellular Assembly and Organization” ($p=1.00E-126$), “Gene Expression, Cancer, Cardiovascular System Development and Function” ($p=1.00E-116$), “Cellular Function and Maintenance, Protein Degradation, Protein Synthesis” ($p=1.00E-92$), “Hematological System Development and Function, Tissue Morphology, Humoral Immune Response” ($p=1.00E-88$), and “Gene Expression, Cellular Development, Tissue Development” ($p=1.00E-74$) (Supplementary Table 9 online). Importantly, VDR serves as a hub molecule, on which the molecular connections are concentrated, in the functional network defined by “Gene Expression, Cellular Development, Tissue Development” (Figure 4).

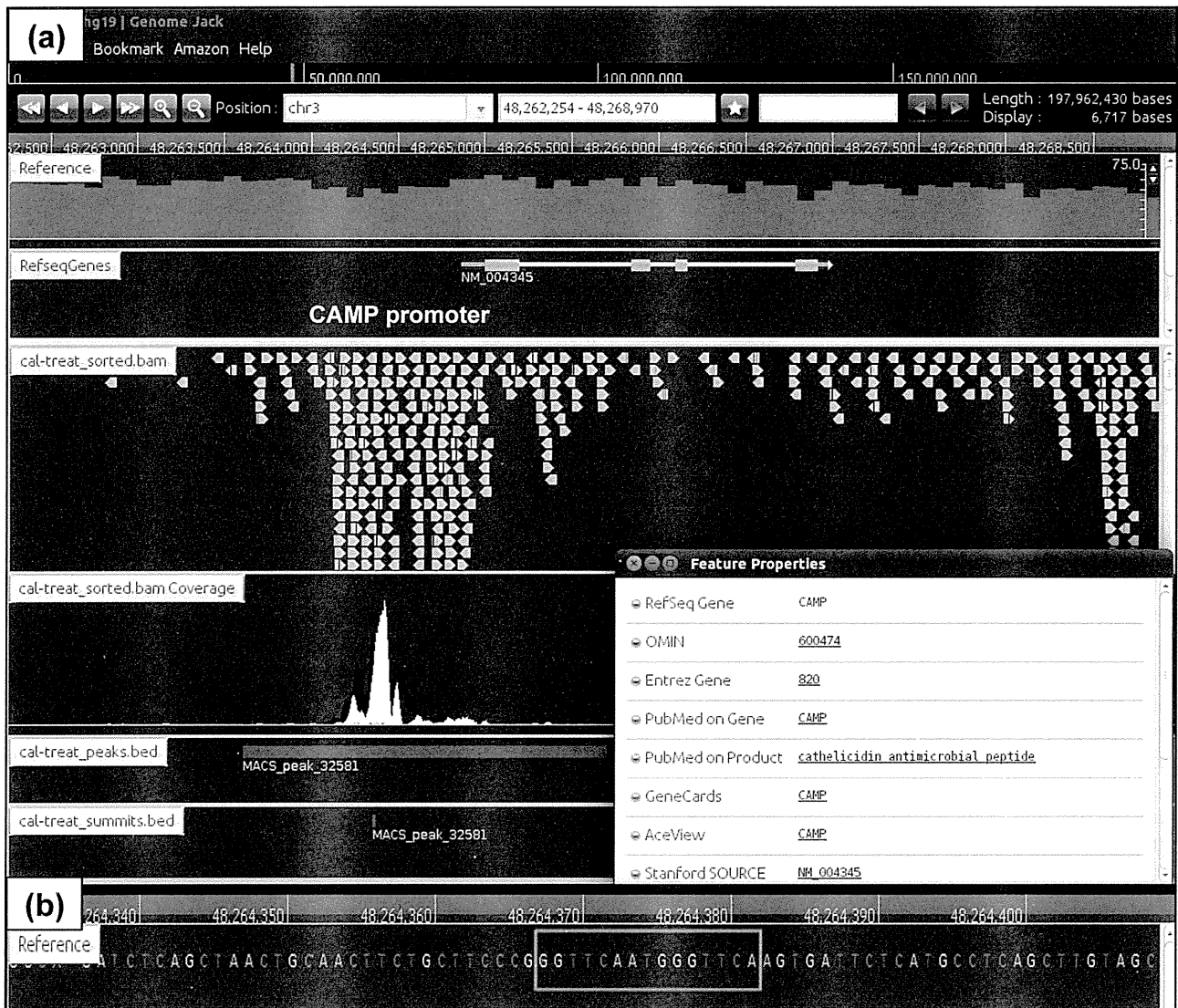


Figure 1. Identification of genomic locations of chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) peaks by GenomeJack.

By analyzing ChIP-Seq datasets of VDR-binding sites numbered SRP002673 and SRP005910, derived from 3 calcitriol-treated and 3 control samples, we identified totally 4718 peaks showing fold enrichment ≥ 20 and $FDR \leq 1\%$ (Table 1). The genomic locations of the peaks were determined by importing the processed data into GenomeJack. An example of a vitamin D receptor (VDR) target gene named cathelicidin antimicrobial peptide (CAMP) (NM_004345) is shown, where a Model-based Analysis of ChIP-Seq (MACS) peak is located on the promoter region (a) with a vitamin D response element (VDRE) consensus sequence highlighted by orange (b). CAMP has been identified as a target gene for $1\alpha,25$ -dihydroxyvitamin D₃.³⁴

By using KeyMolnet, the neighboring network search algorithm extracted a highly complex network, composed of 3300 molecules and 5569 molecular relations. The network exhibited a significant relationship with transcriptional regulation by retinoblastoma protein (Rb)/E2F family proteins ($p=1.65E-182$), transcriptional regulation by VDR ($p=2.74E-171$), transcriptional regulation by interferon regulatory factor (IRF) proteins ($p=7.03E-103$), heat shock protein 90 (HSP90) signaling pathway ($p=9.33E-87$), and histone acetyltransferase (HAT) signaling pathway ($p=3.04E-79$). Again, VDR plays a central role in homeostasis of the complex molecular network

constructed by VDRTGs (Supplementary Figure 1 online).

Discussion

By using different antibodies, two distinct ChIP-Seq studies numbered SRP002673 and SRP005910 in human immune cells exposed to calcitriol identified 2776 and 2340 VDR-binding sites, respectively.^{12,13} They utilized the similar mapping (Bowtie) and peak finding (MACS; $FDR < 1\%$) programs. Regardless of great variation of genetic and epigenetic regulation in distinct cell types,¹⁴

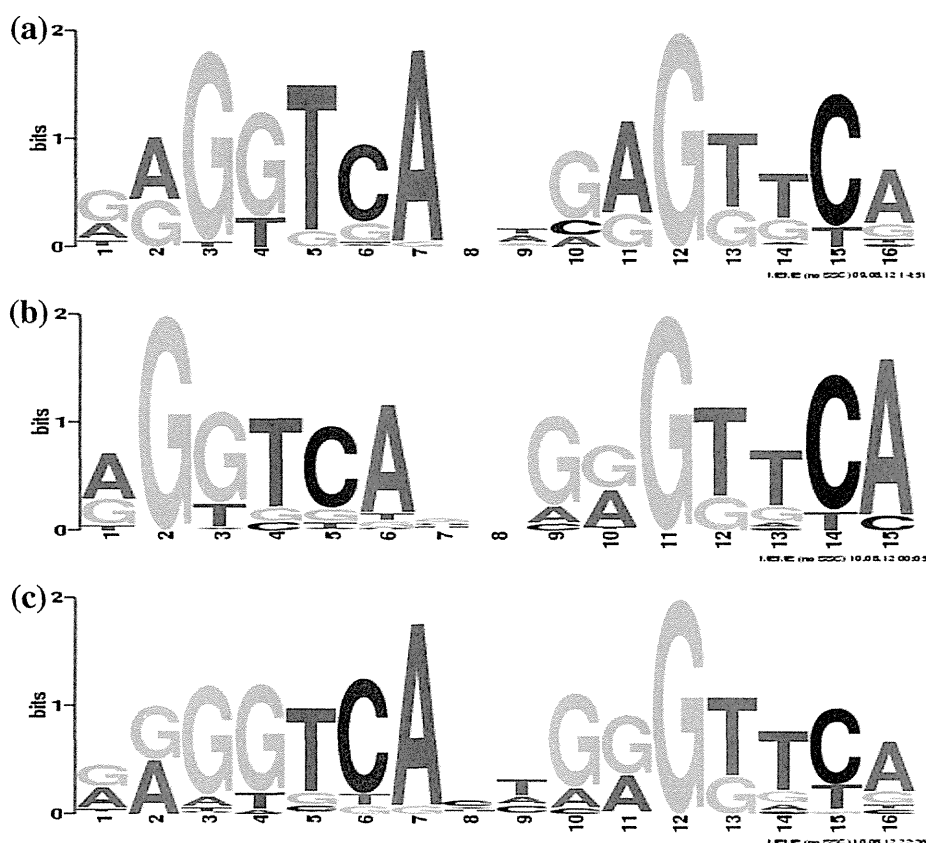


Figure 2. Motif analysis of vitamin D response element (VDRE) consensus sequences on chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) peaks. The consensus motif sequences were identified by importing a 400 bp-length sequence surrounding the summit of Model-based Analysis of ChIP-Seq (MACS) peaks of the top 50 enriched genes into the MEME-ChIP program. The direct repeats of hexameric (G/A)G(G/T)TCA motifs with three intervening base pairs (DR3) are found in calcitriol-treated samples of (a) GMI0855, (b) GMI0861, and (c) THP-1 cells.

Table 2. Top five Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways relevant to 1541 calcitriol-responsive vitamin D receptor (VDR) target genes.

KEGG pathway	Molecules in pathway	p-value
hsa05200: pathways in cancer	ACVRI1B, BCL2, BCL2L1, RUNX1, CDC42, CDK6, COL4A2, CRK, CTBP1, CTNNA1, CTNNB1, E2F2, MECOM, FGF14, GLI3, GRB2, GSK3B, HSP90AB1, ITGA3, LAMA3, LAMA5, LAMB3, LAMC1, MDM2, MSH3, NFKBIA, PDGFA, PLCG1, PLCG2, PRKCA, MAP2K1, PTCHI, PTEN, PTK2, RAC2, RAC3, RXRB, TCF7, TGFB1, HSP90B1, VEGFA, WNT8B, FGF17, TRAF4, PIK3R5, CTNNA3, WNT4, WNT10A, RASSF5, EGLN2	4.92E-05
hsa04670: leukocyte transendothelial migration	RHOH, CDC42, CTNNA1, CTNNB1, CYBA, GRLF1, ICAM1, ITGAL, ITGB2, CD99, NCF4, PLCG1, PLCG2, PRKCA, PTK2, RAC2, RAPIA, VAV1, VAV3, RAPGEF4, PIK3R5, CTNNA3, RASSF5, ESAM	1.17E-04
hsa05322: systemic lupus erythematosus	CD28, CD86, HIST1H2AE, HIST1H2AD, HIST1H2BD, HIST1H2BB, SNRPD3, HIST1H4I, HIST1H2AC, HIST2H2AA3, HIST2H2AC, HIST1H2BG, HIST1H2BL, HIST1H2BM, HIST1H2BF, HIST1H2BI, HIST1H2BO, HIST2H2BE, HIST1H3D, HIST1H4D, HIST1H4K, HIST1H4J, HIST1H4C, HIST1H4H, HIST1H4B, HIST1H4E, HIST2H4A, HIST1H2BJ, HIST1H2AH, HIST1H2BK, HIST2H3C, HIST2H3A, HIST2H2BF, HIST2H2AA4	5.57E-04
hsa04510: focal adhesion	BCL2, CDC42, COL4A2, COL6A3, CRK, CTNNB1, GRB2, GRLF1, GSK3B, TNC, ITGA3, ITGB5, LAMA3, LAMA5, LAMB3, LAMC1, PDGFA, PPPICA, PRKCA, MAP2K1, PTEN, PTK2, RAC2, RAC3, RAPIA, RASGRF1, SRC, VAV1, VEGFA, VAV3, PIK3R5, PDGFD	7.30E-04
hsa04666: Fc gamma R-mediated phagocytosis	CDC42, CRK, INPP5D, PLCG1, PLCG2, PRKCA, PRKCD, PRKCE, MAP2K1, RAC2, VAV1, PIP5K1B, PPAP2C, ASAP2, ARPC2, VAV3, PIK3R5, ARPC5L	2.39E-03

By importing Entrez Gene IDs of 1541 calcitriol-responsive VDR target genes into the Functional Annotation tool of Database for Annotation, Visualization and Integrated Discovery (DAVID), the top five most relevant KEGG pathways were identified. The p-value represents statistical significance evaluated by the modified Fisher's exact test.

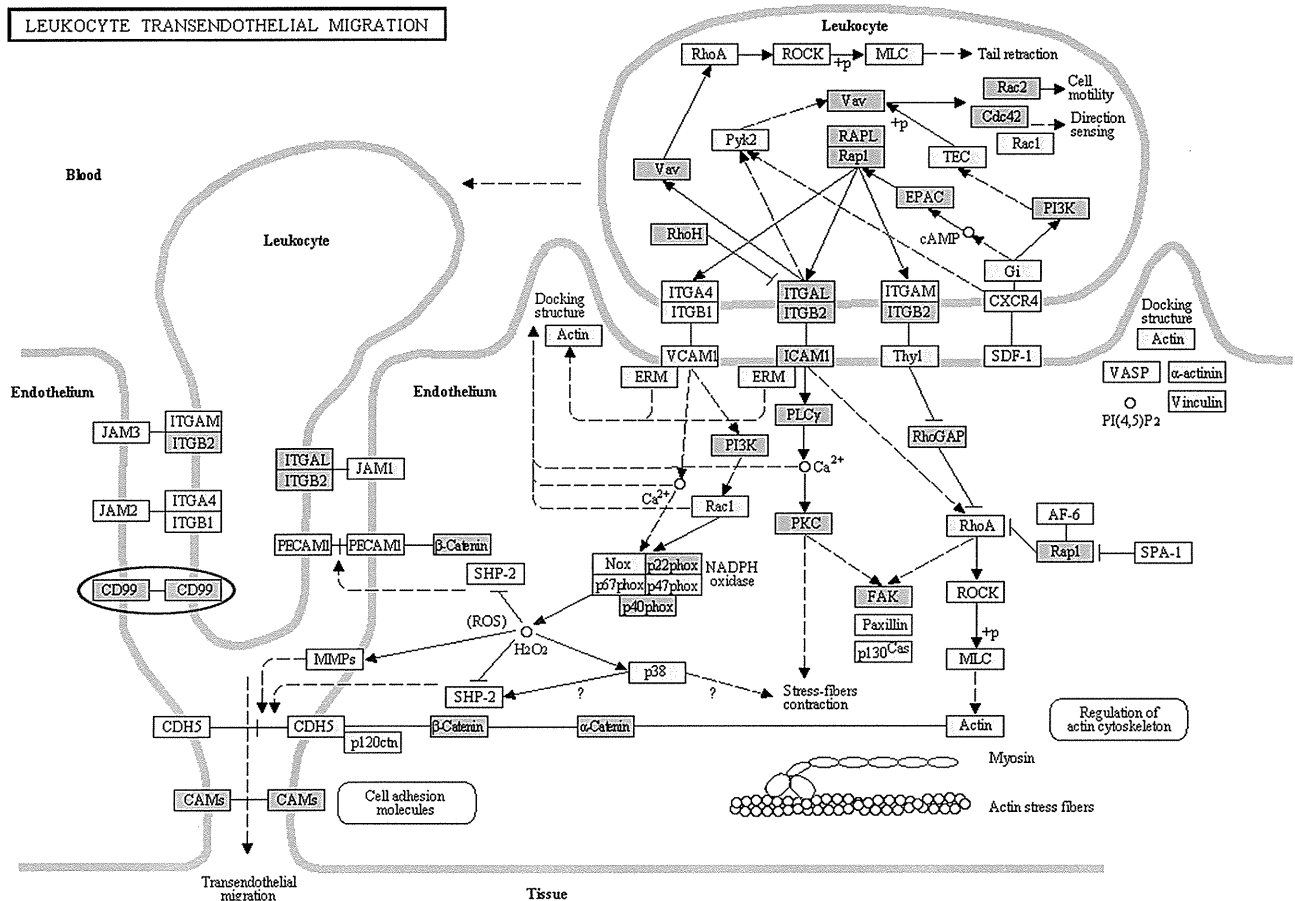


Figure 3. Molecular network of chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)-based vitamin D receptor (VDR) target genes by Kyoto Encyclopedia of Genes and Genomes (KEGG). The second rank relevant pathway defined by “leukocyte transendothelial migration (hsa04670)” is shown, where VDR target genes (VDRTGs) are colored by orange. CD99, a cell-surface glycoprotein involved in T cell migration and apoptosis, is encircled by a blue ellipse.

VDR-binding sites showed a considerable overlap (18.2%) between the cells originated from B cells and monocytes.¹⁵ However, these studies did not clarify the global molecular network of genome-wide VDRTGs, with respect to immunomodulatory function of vitamin D potentially involved in protection against the development of MS. The ChIP-Seq analysis of VDR target genes derived from lymphocytes of MS patients appears to be preferable, if they are available. As an alternative, we reanalyzed the publicly available datasets of two immune-relevant cells, the cells of B cell and monocyte origins, and combined the results of both to retrieve *in vivo* VDRTG candidates as much as possible.

Although ChIP-Seq serves as a highly efficient method for genome-wide profiling of DNA-binding proteins, it requires some technical considerations:¹⁶ the specificity of the antibody, reproducibility of the results, sequencing depth, and the source of controls, along with cell types, developmental stages, and culture conditions capable of affecting epigenetic features, constitute critical factors. In general, DNA-binding of transcription factors is a highly dynamic process. However, the ChIP-Seq data reflect a

snapshot of binding actions. Motif analysis of a defined set of high-quality peaks makes it possible to evaluate the antibody specificity to some extent.¹⁶ We identified VDRE consensus sequences around ChIP-Seq peak summits of all three calcitriol-treated samples, supporting the validity of the experiments. It is not unusual that some biologically important sites show fairly weak ChIP-Seq signals, whereas some sites with very high signals fail to give positive functional output in follow-up experiments.¹⁶ We identified overall 2997 stringent ChIP-Seq peaks located on protein-coding genes. However, by transcriptome analysis, only 20% of upregulated genes and none of downregulated genes in calcitriol-treated cells agreed with ChIP-Seq-based VDRTGs. Most importantly, the molecular network of 1541 calcitriol-responsive VDRTGs is closely related to leukocyte transendothelial migration, Fcγ receptor-mediated phagocytosis, and transcriptional regulation by VDR, suggesting a pivotal role of VDRTGs in immune regulation.

The ligand-bound VDR, heterodimerized with RXR, interacts with VDRE sequences on regulatory regions of target genes.² VDR binding to VDREs is ligand-dependent,

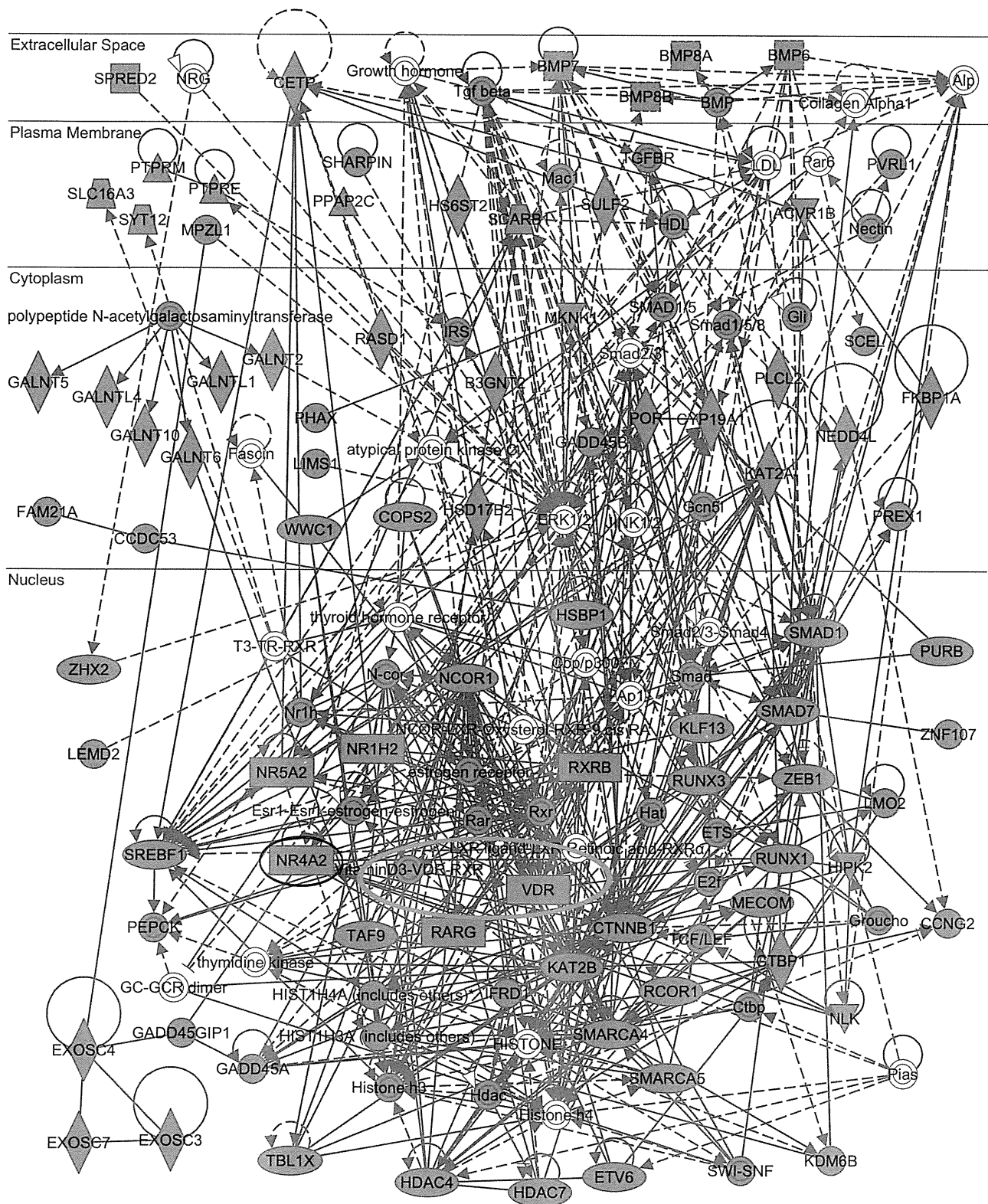


Figure 4. Molecular network of chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)-based vitamin D receptor (VDR) target genes by Ingenuity Pathways Analysis (IPA). By importing Entrez Gene IDs of the set of 1541 calcitriol-responsive VDR target genes (VDRTGs) into the Core Analysis tool of IPA, we identified the top five most relevant functional networks (Supplementary Table 9 online). The fifth rank relevant network defined by “Gene Expression, Cellular Development, Tissue Development” is shown, where VDRTGs are colored by red. VDR and VDR/ retinoid X receptor (RXR) complex are highlighted by an orange ellipse, while NR4A2, a nuclear orphan receptor involved in differentiation of Th1, Th17, and Treg cells, is encircled by a blue ellipse.