# Supplementary Information

## Activity-Based Metaproteomics Driven Discovery and Enzymological Characterization of Potential α-Galactosidases in the Gut Microbiome

Jianbing Jiang[1,2,7], Diana Czuchry[2,7], Yanxia Ru[5,7], Huipai Peng[2,7], Junfeng Shen[2], Teng Wang[6], Wenjuan Zhao[2], Weihua Chen[6], Sen-Fang Sui[3,4,5], Yaowang Li[5*], Nan Li[2*]

[1] Institute for Inheritance-Based Innovation of Chinese Medicine, School of Pharmacy, Shenzhen University Medical School, Shenzhen University, Shenzhen 518055, China.

[2] Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

[3] State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structure, School of Life Sciences, Tsinghua University, Beijing 100084, China.

[4] Cryo-EM Center, Southern University of Science and Technology, Shenzhen 518055, China.

[5] School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China.

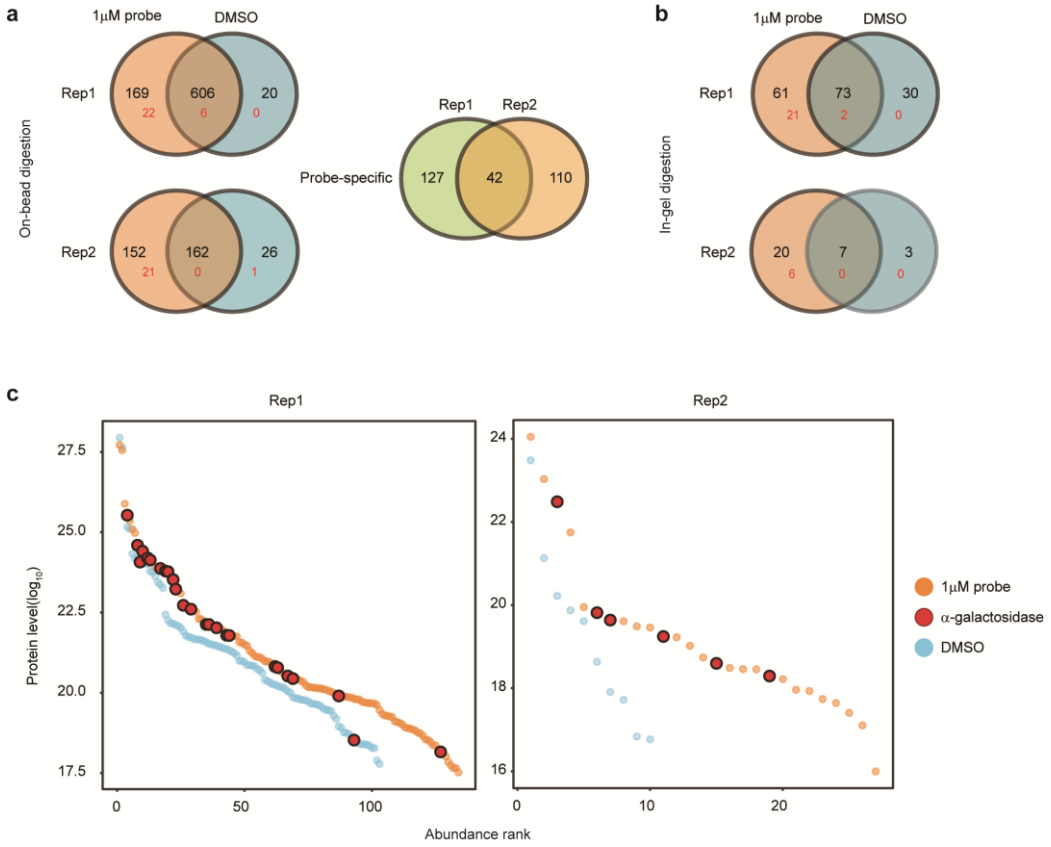
[6] Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular Imaging, Center for Artificial Intelligence Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China.

[7] These authors contributed equally.

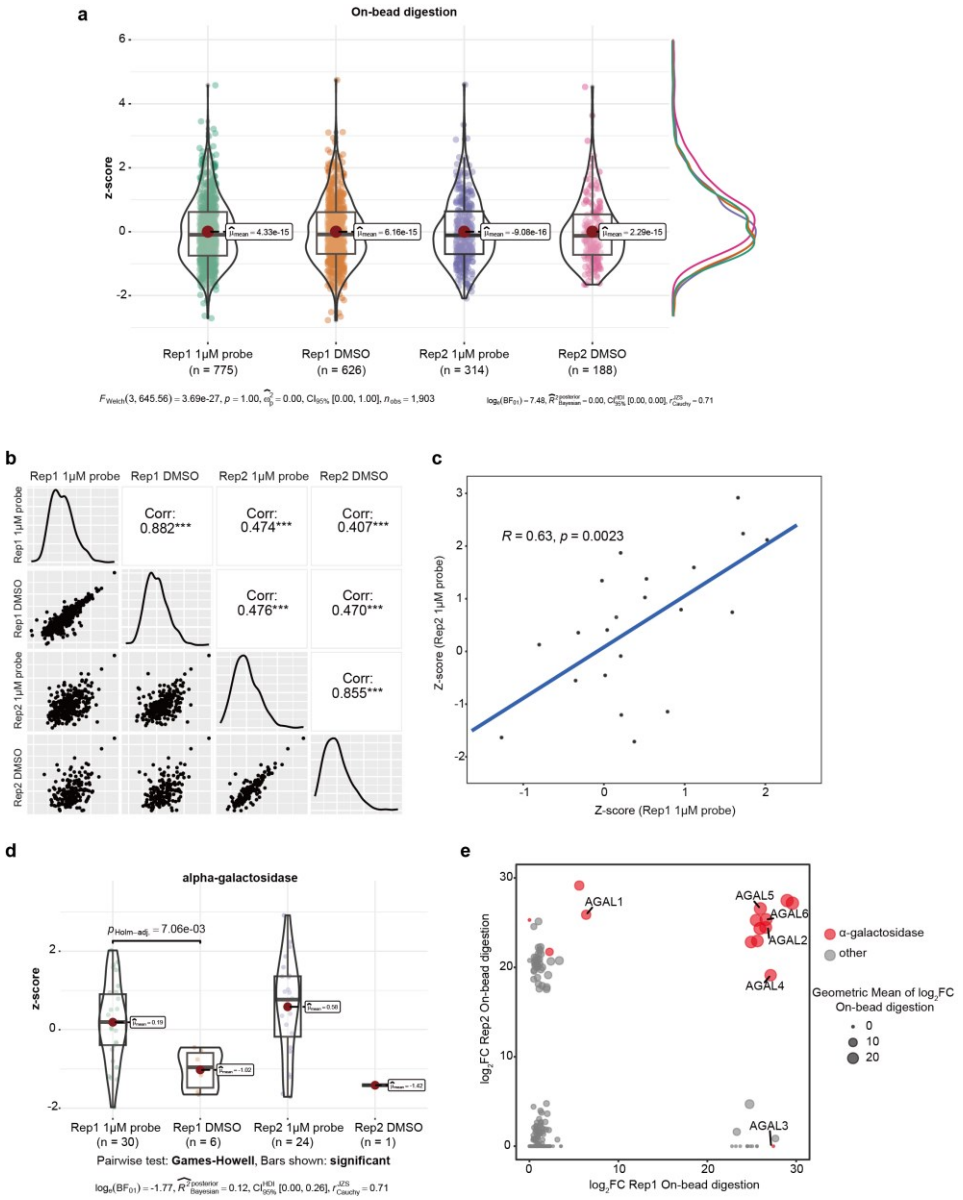
[*] Correspondence: Yaowang Li, liyw@sustech.edu.cn; Nan Li, nan.li@siat.ac.cn

## Table of Contents

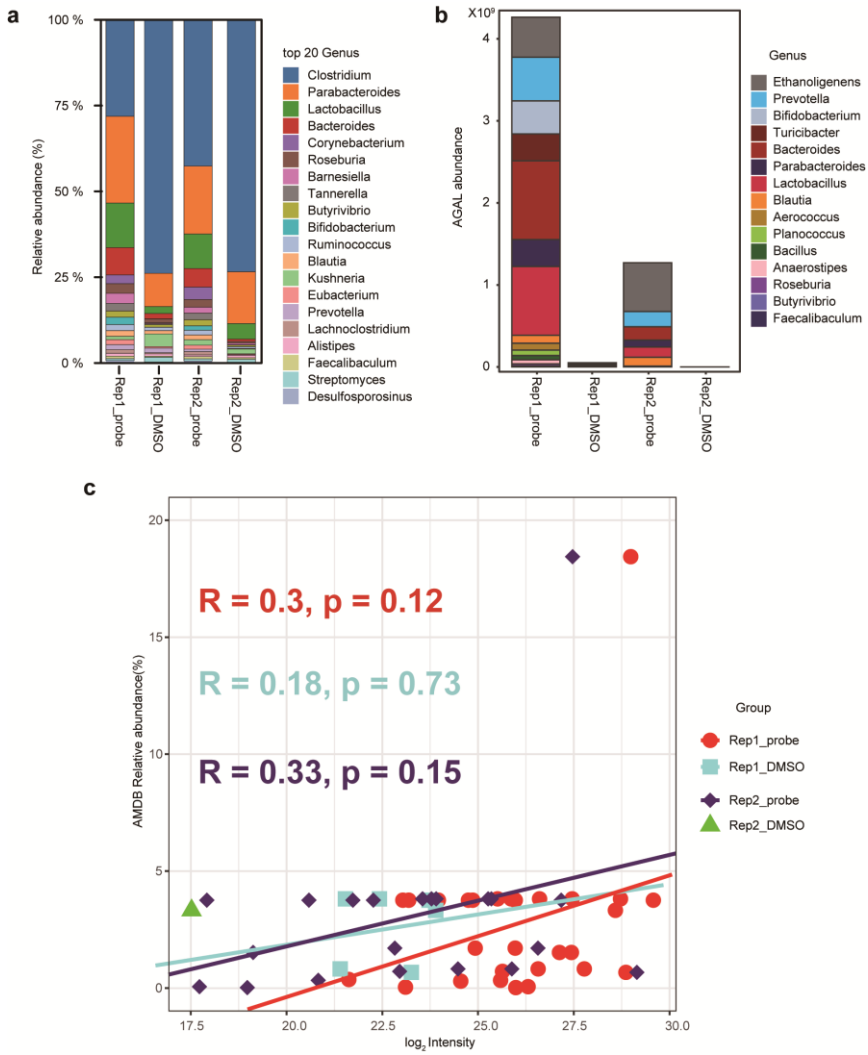
# Supplementary Figures



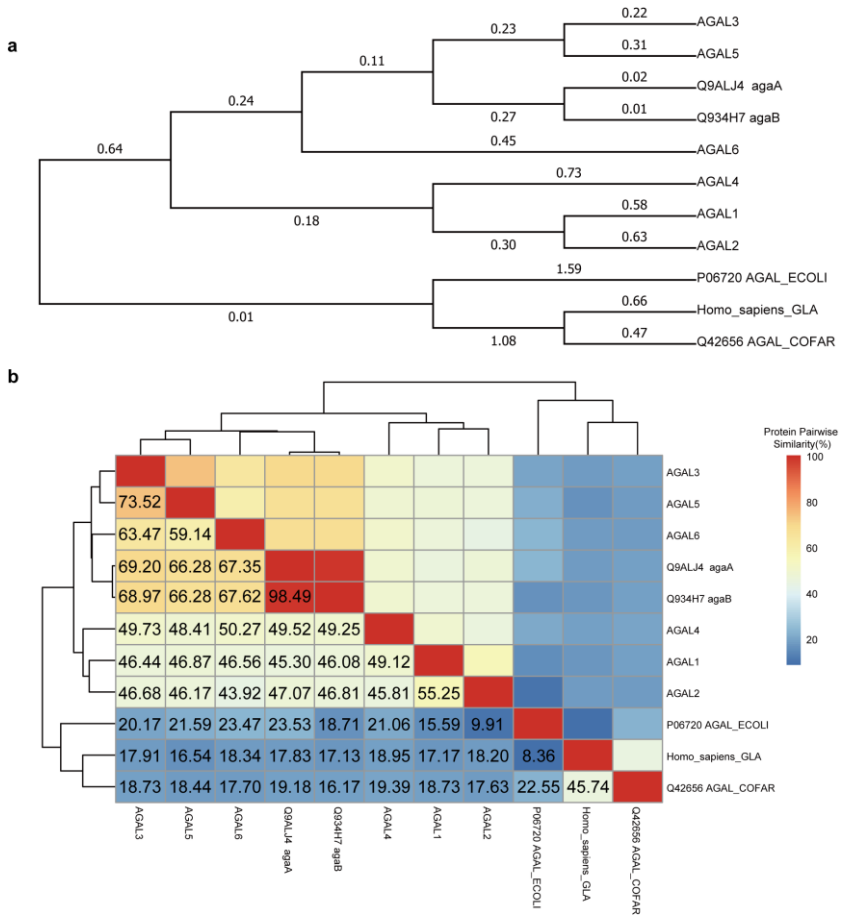

**Supplementary Fig. 1 Distribution of protein groups identified by LC-MS/MS.** (**a**)Venn diagrams illustrate the intersection and difference of protein groups (black number) and AGAL groups (red number) identified by on-bead digestion in two replicates (left), as well as the probe-specific identification in two replicates (right). (**b**) Venn diagram depicts the protein groups and AGAL identified by in-gel digestion. (**c**) Protein abundance distribution is presented as calculated from MS intensities of quantified peptides for each protein.



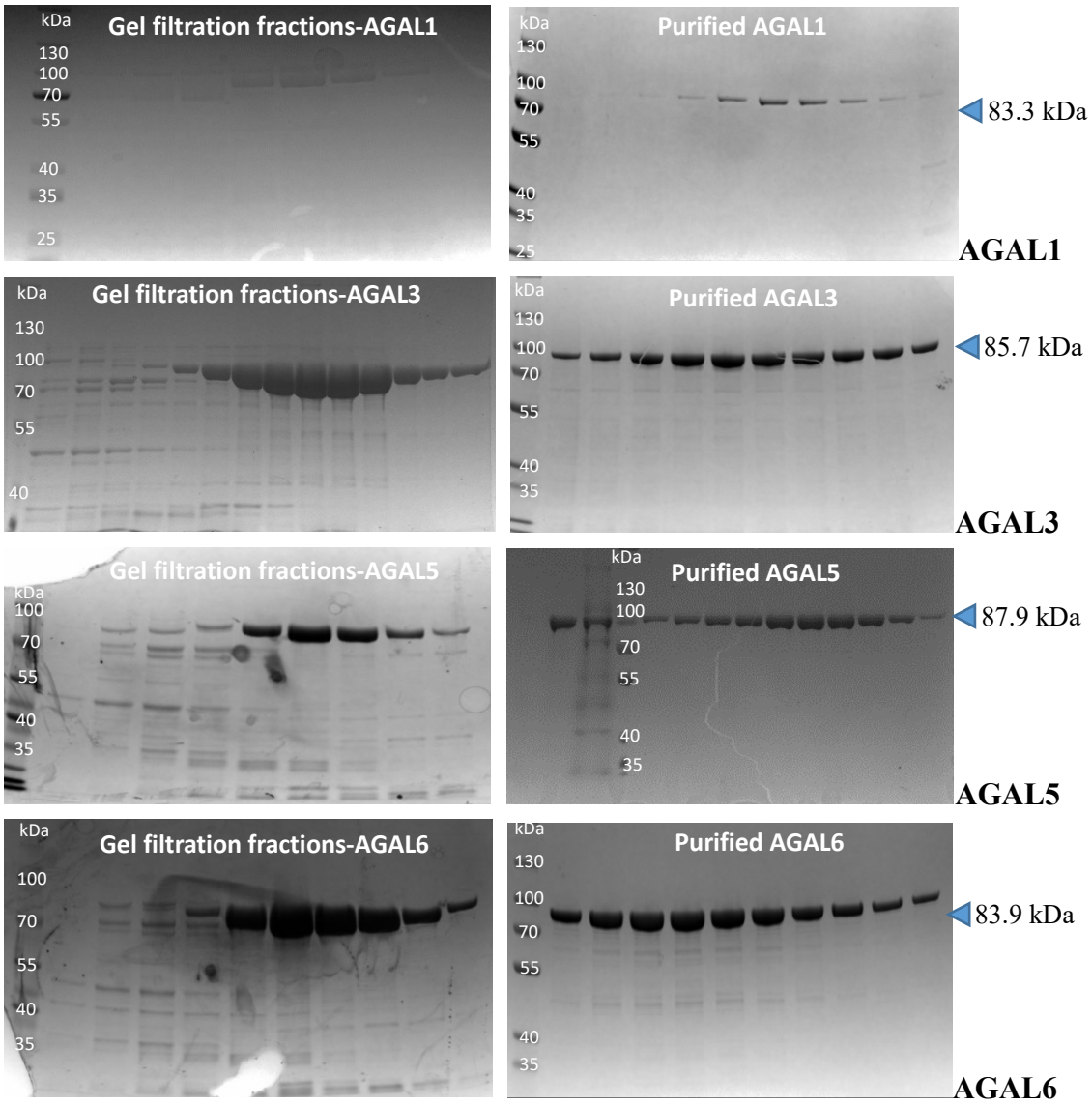

**Supplementary Fig. 2 Normalization, reproducibility, and significant differences analysis in mass spectrometry data of on-bead digestion sample.** (**a**) The Z-score (standard score) is used to normalize the protein intensity in the on each bead-digested sample. By standardizing data to z-scores, we can determine the likelihood of observing a particular value or range of values under the assumption of a normal distribution. This allows us to make statistical inferences and draw conclusions about the data; (**b**) Pearson correlation analysis on on-bead digestion sample to demonstrate data reproducibility. p-values are denoted as follows: ns $p > 0.05$, **$p \leq 0.01$, ****$p \geq 0.0001$; (**c**) Scatter plot shows the distribution of target protein in two on-bead digestion samples. Pearson correlation coefficient: 0.63, p-value < 0.01; (**d**) Violin plots of target protein range of samples between 1μM probe and DMSO controls each sample. The notch in the middle of the box represents the median. The top and bottom of box represent the 75th and 25th quartiles, respectively. The upper and lower whiskers extended 1.5× the interquartile range from the upper edge and lower edge of the box represent maximum and minimum. P values were determined by Welch's ANOVA with Games-Howell's multiple comparisons test. P-values are Holm corrected. P Holm-adj = $5.63 \times 10^{-3} < 0.01$ is significant; (**e**) The scatter plot shows the fold change relationship between two on-bead digestion samples for proteins with unique peptides ≥ 2. red: α-galactosidase, gray: non-target protein.



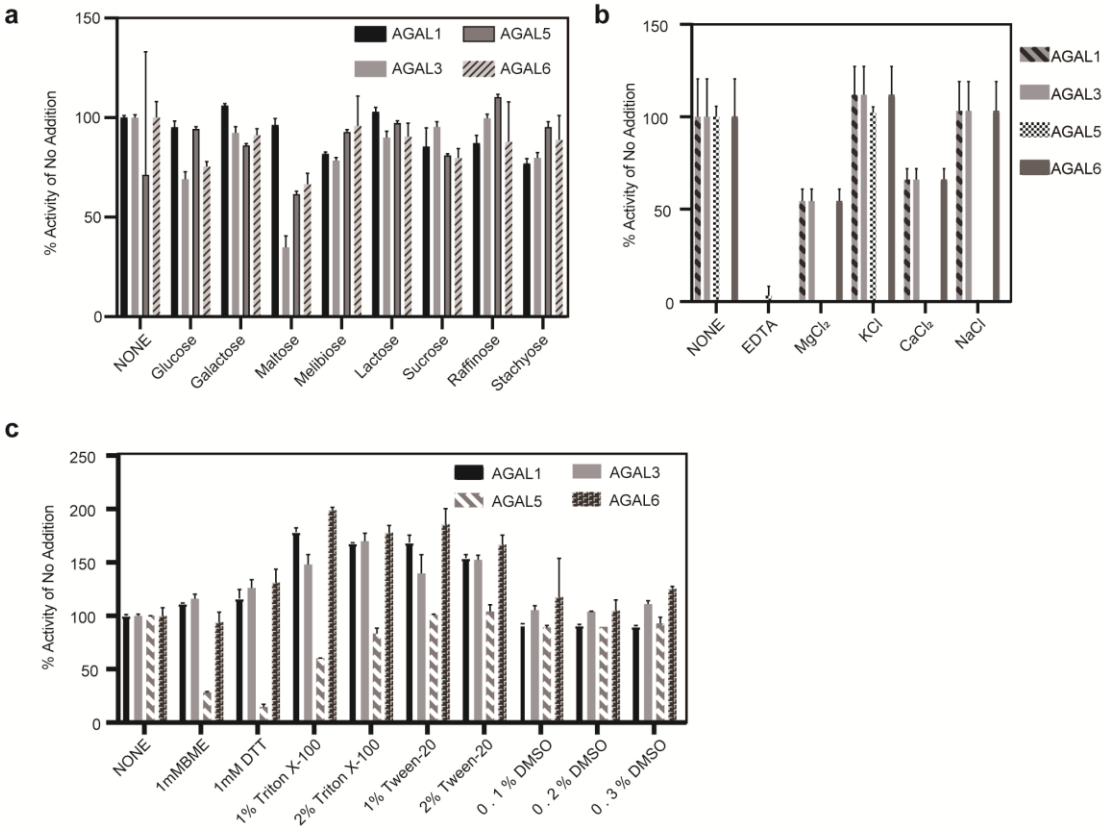

**Supplementary Fig. 3 Taxonomic analysis of on-bead digestion sample.** (**a**) Distribution of the 20 most abundant genera in the on-bead digestion sample; (**b**) Abundance distribution of AGAL at the genus level in the on-bead digestion sample; (**c**) Pearson correlation analysis between the intensity of AGAL proteins and corresponding genus relative abundance. The genus relative abundance data were obtained from the Animal Microbiome Database (AMDB) (Nucleic acids research, 50 (2022) D729-D735), which contains bacterial 16S rRNA gene profiles from various animal species to investigate the relationship between gut microbiota and animal hosts.



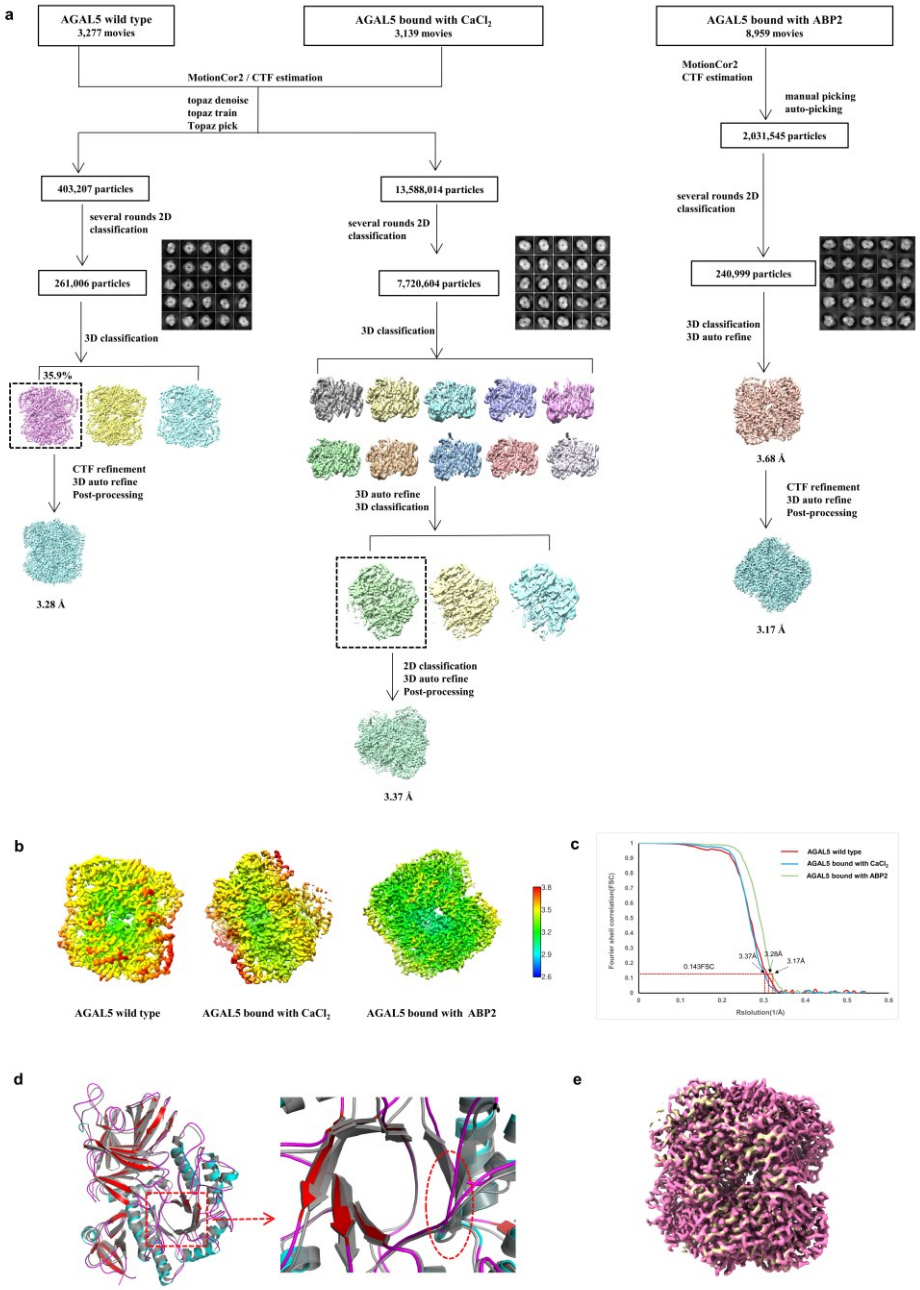

**Supplementary Fig. 4 Similarity and evolutionary analysis of several AGALs. (a)** Phylogenetic clustering was performed using the Maximum Likelihood method and JTT matrix-based model**.** Evolutionary analysis of AGAL1-6, AgaA, AgaB and AGALs from human, coffee bean and *E. coli*. **(b)** Percent amino acid sequence identity value between enzymes using ClustalOmega Multiple Sequence Alignment. This analysis involved 11 amino acid sequences. There are a total of 904 positions in the final dataset. Evolutionary analyses were conducted in MEGA-X. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. This analysis involved 11 amino acid sequences.



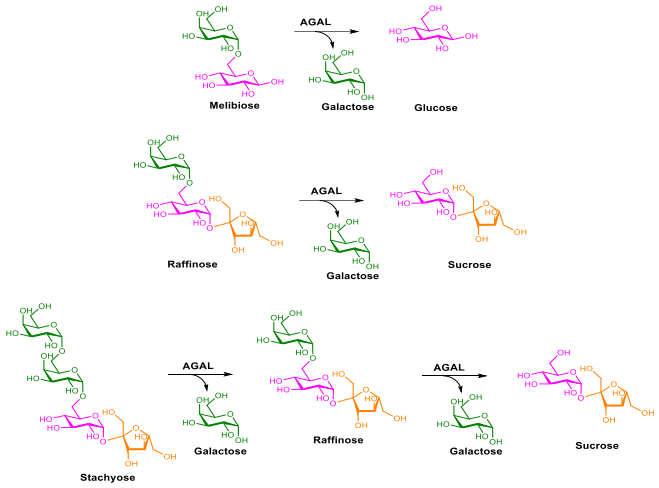

**Supplementary Fig. 5 Expression and purification of recombinant AGAL enzymes.** Recombinant AGAL1, AGAL3, AGAL5, and AGAL6 were expressed and purified using Ni-NTA affinity chromatography followed by gel filtration on fast protein liquid chromatography (FPLC).

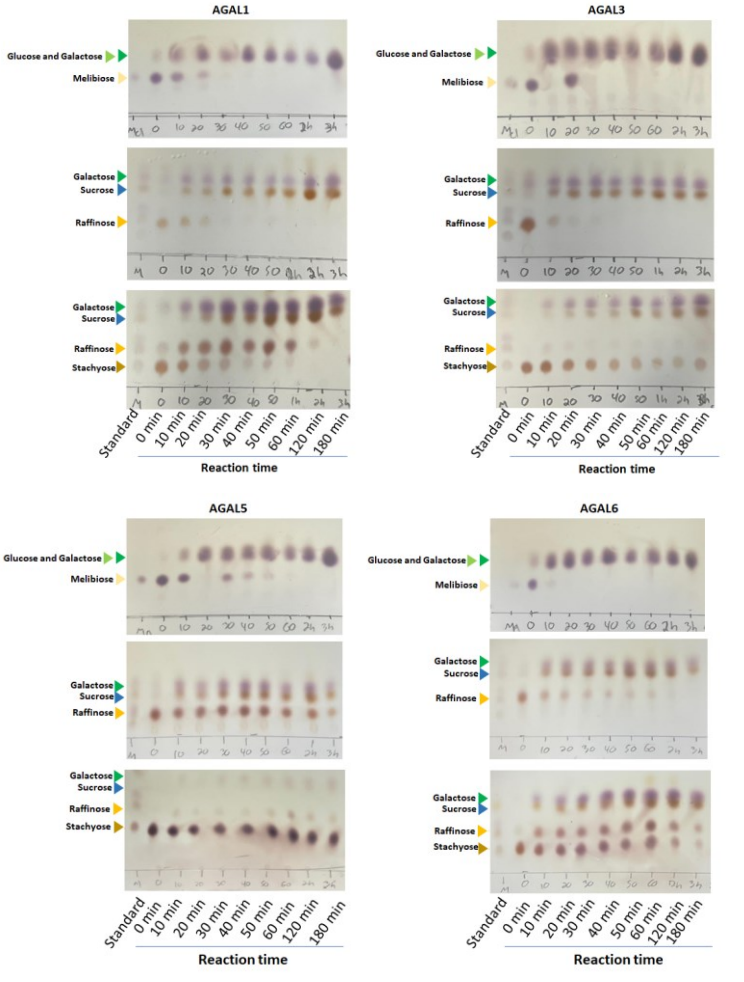



**Supplementary Fig. 6 Characterization of AGALs substrate specificity and enzyme activity modulation by sugars, metal ions, and detergents.** (**a**) Effect of various metal ions and EDTA on galactosidase activity; (**b**) Effect of different sugars on galactosidase activity; (**c**) Substrate-dependent modulation of galactosidase activity.

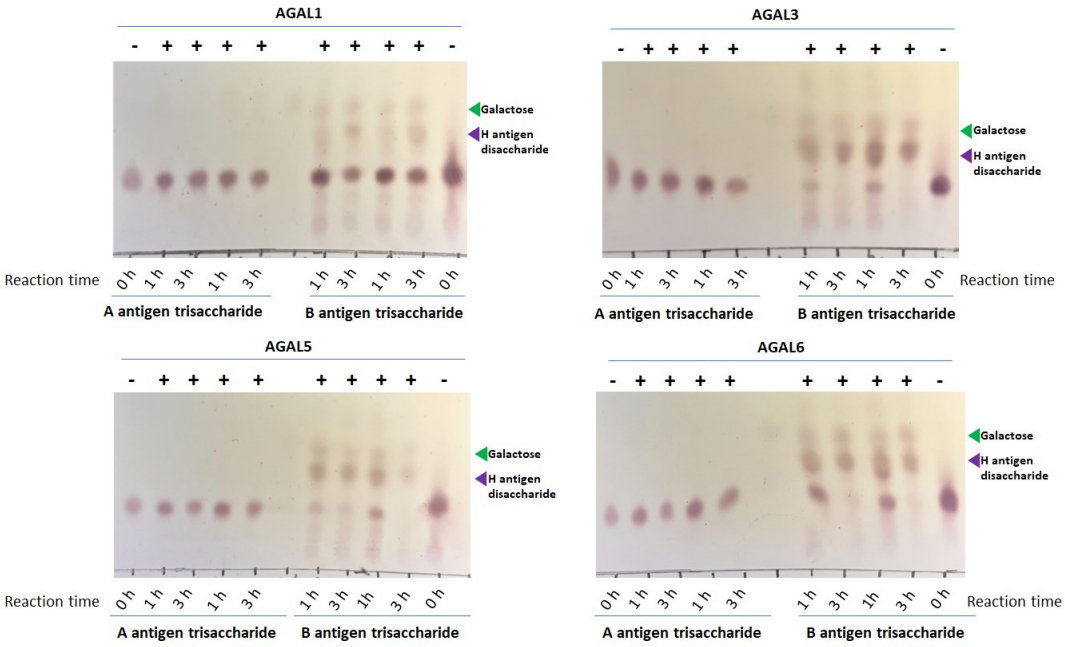



**Supplementary Fig. 7 Cryo-EM single-particle analysis of the AGAL5.** (**a**) Flowchart outlining the electron microscopy image processing used to analyze AGAL5. Micrographs were corrected by motionCor2 and used for defocus and astigmatism estimation using CTF fitting. Protein particles were manually and automatically picked by topaz, followed by 2D and 3D classification. Final maps were reconstructed from selected particles, resulting in three maps with resolutions of 2.74 Å, 3.37 Å, and 3.17 Å, respectively. (**b**) Final map resolutions reported by gold standard FSC plots (FSC = 0.143). (**c**) The gold standard FSC plots were used to report the final resolution for each map (FSC = 0.143). (**d**) Comparison of the monomer structure of native AGAL5 (red) and the crystal structure of AgaB (PDB: 4FNQ, grey) with a zoomed-in view of the catalytic barrel on the right. The AGAL5 catalytic barrel displays fewer secondary structures compared to the crystal structure. (**e**) Superposition of the electron density map of AGAL5 with $CaCl_2$ (yellow) and AGAL5 with ABP**2** (pink) demonstrating that AGAL5 with ABP**2** has a more densely packed structure, indicating tight binding of ABP**2** to AGAL5 and confirming the investigation into the hydrolysis retaining mechanism.

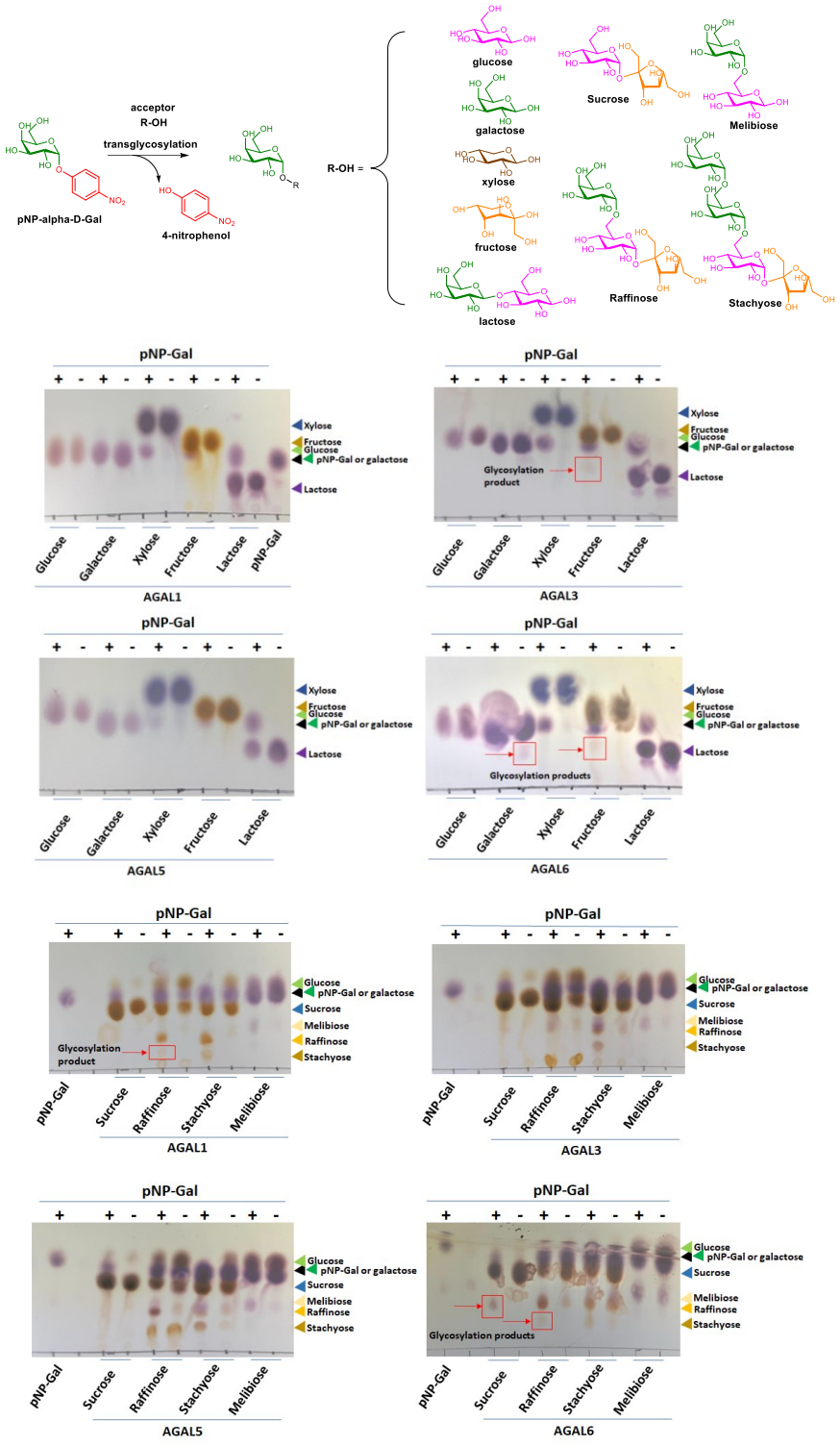



**Supplementary Fig. 8 Hydrolysis of melibiose (top), raffinose (middle), and stachyose (bottom) by AGALs.** TLC analysis was conducted to monitor the reaction process over time. Samples were incubated for 0, 10, 20, 30, 40, 50, 60, 120, and 180 minutes, and analyzed for product formation. The results demonstrate the efficient hydrolysis of melibiose, raffinose, and stachyose by AGALs.




**Supplementary Fig. 9 TLC analysis of AGAL-mediated hydrolysis of A and B antigen trisaccharides.** The results demonstrate successful hydrolysis of both A and B antigen trisaccharides by AGALs, as detected by TLC analysis.




**Supplementary Fig. 10 Transglycosylation reactions of AGALs towards glucose, galactose, xylose, fructose, lactose, sucrose, raffinose, stachyose, and melibiose monitored by TLC analysis.** The reaction process is shown, and the results demonstrate successful transglycosylation by AGALs towards the tested substrates.

# Supplementary Tables

**Supplementary Table 1. Enzymes selected for further characterization from those identified in the gut microbiome using ABMP.**

| Assigned Name | Gene Names | GenBank Accession Number | Organism Species | Number of Amino Acids | Computed Mol. weight [kDa] without polyhis tag | Computed Theoretical pI |
|---|---|---|---|---|---|---|
| AGAL1 | A4V02_01335 | ANU62514.1 | *Muribaculum intestinale YL27* | 731 | 83.32 | 5.57 |
| AGAL2 | A4V02_08955 | WP_068961143.1 | *Muribaculum intestinale YL27* | 727 | 81.90 | 6.36 |
| AGAL3 | AT726_08105 | AMC08872.1 | *Turicibacter sp. H121* | 743 | 85.73 | 5.27 |
| AGAL4 | AT726_08965 | AMC09013.1 | *Turicibacter sp. H121* | 686 | 79.03 | 5.05 |
| AGAL5 | A4V09_19485 | ANU77732.1 | *Blautia pseudococcoides strain YL58* | 763 | 87.92 | 5.16 |
| AGAL6 | N134_04935 | AGR64239.1 | *Limosil ctobacillus reuteri TD1* | 729 | 83.92 | 5.36 |

**Supplementary Table 2. Model refinement and validation statistics**

| Model refinement and validation statistics | | | | |
|---|---|---|---|---|
| | | **AGAL5 wild type** | **AGAL5 mixed with $CaCl_2$** | **AGAL5 bound with inhibitor** |
| Composition | Amino acids | 742 | 673 | 755 |
| | Ligand | 0 | 0 | 1 |
| Mean B-factors | Amino acids | 0 | -100 | -168 |
| | ligand | 0 | 0 | |
| Ramachandran plot | Favored (%) | 93.10% | 93.26% | 94.23% |
| | Allowed (%) | 6.90% | 6.74% | 5.77% |
| | Outliers (%) | 0 | 0 | 0 |
| | Rotamer Outliers (%) | 0 | 0 | 0 |
| | CC (mask) | 0.7842 | 0.8533 | 0.8342 |

**Supplementary Table 3. Root-mean-squared deviation (RMSD) analysis between different structures**

| Model | RMSD value |
|---|---|
| 4FNQ **vs** AGAL5 wild type | 1.334 |
| AGAL5 mixed with $CaCl_2$ **vs** AGAL5 wild type | 29.695 |
| AGAL5 mixed with inhibitor **vs** AGAL5 wild type | 0.406 |
| AGAL5 mixed with $CaCl_2$ **vs** AGAL5 bound with inhibitor | 26.847 |

**Supplementary Table 4. Cryo-EM data collection**

| Cryo-EM data collection | | | |
|---|---|---|---|
| **Protein** | **AGAL5 wild type** | **AGAL5 with $CaCl_2$** | **AGAL5 bound with inhibitor** |
| Microscope | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| Voltage(kV) | 300 | 300 | 300 |
| Detector | Gatan K3 | Gatan K3 | Gatan K3 |
| Magnification | 130K | 130K | 130K |
| Pixcel size(Å/pix) | 0.92 | 0.92 | 0.92 |
| Flux(e-/pix/sce) | 20 | 20 | 20 |
| Frames per exposure | 32 | 32 | 32 |
| Exposure(e-/$Å^2$) | 50 | 50 | 50 |
| Defocus range(μm) | 1.5-2.5 | 1.5-2.5 | 1.0-2.0 |
| Micrographs collected | 3277 | 3139 | 8959 |
| Masked resolution at 0.143 FSC (Å) | 3.28Å | 3.31Å | 3.17Å |