

Table of Contents

<i>Supplementary Methods</i>	2
Whole Genome Sequencing	2
Genome Assembly and Quality Control	2
Core-genome Phylogeny Inference	2
Clonality Determination	2
Cluster analysis and visualizations	3
Molecular Clock Analysis	3
Supplementary references	4
<i>Extended Discussion</i>	7
Pathogen Trends in BSIs	7
Infection Prevention Measures	7
Limitations	7
<i>Supplementary figures and tables</i>	8
Table S1: Phylogenetically corrected SNP Thresholds for Hierarchical Groups	8
Table S2: Characteristics of 69 Identified Transmission Clusters	10
Figure S1: Multi-SNP Threshold Plots for Hierarchical Groups containing at Least Two Genomes	11
Figure S2: Distribution of Transmission-acquired Colonization and Infection Events	12
Figure S3: NICU Floorplans Demonstrating Patient Movements in the Identified Clusters	13
Figure S4: Temporal analysis for each identified transmission cluster by location	81
Figure S5: Temporal analysis for each identified transmission cluster by treatment team assignments	149
Figure S6: NICU Floorplan Indicating Environmental Sampling Locations	217

Supplementary Methods

Whole Genome Sequencing

Single colony isolates from the CHOP IDDL were stored at -80°C at the CHOP microbial ARchive and Cryo-collection (microbialARC) strain bank and were sequenced at the CHOP Microbiome Center. DNA from all isolates were extracted using the Illumina Direct from Colony protocol. Extracted DNA was quantified with the Quant-iT PicoGreen® dsDNA assay kit (ThermoFisher) before library generation. Libraries were generated from 7.5 ng DNA using the Illumina DNA Prep kit and IDT for Illumina unique dual indexes at 1:4 scale reaction volumes. Library success was assessed by Quant-iT PicoGreen® dsDNA assay kit and samples with library yields < 1 ng/ul were re-prepped as needed. After all samples for a given pool were prepped, an equal volume of library was pooled from every sample and then the pool was sequenced using a 300 cycle Nano kit on the Illumina MiSeq. Libraries were then repooled based on the demultiplexing statistics of the MiSeq Nano run. Final libraries were sequenced on an Illumina NovaSeq 6000 v1.5 flow cells, producing 2x150 bp paired end reads. Extraction blanks and nucleic acid-free water were processed along with experimental samples to empirically assess environmental and reagent contamination. A laboratory-generated mock community consisting of DNA from *Vibrio campbellii* and Lambda phage were included as a positive sequencing control.

Genome Assembly and Quality Control

Raw reads underwent quality control using Sunbeam 4.3.7, including adapter trimming and host decontamination¹. Isolate genomes were assembled using SPAdes v3.15.5², while 102 metagenomic samples were processed through the Anvi'o v8 pipeline using MEGAHIT v1.2.9^{3,4}. Assembly quality was assessed using CheckM with the following criteria: $\geq 95\%$ completeness, $\leq 5\%$ contamination, and “Staphylococcus (UID301)” lineage classification. Species-level contamination was evaluated using Mash⁵. Assemblies were filtered to retain genomes between 2.55 Mb and 3.15 Mb. Of 1,670 assembled genomes, 1,446 passed quality criteria and were used for downstream analyses (https://github.com/qianxuans/CHOP-NICU-Scripts/blob/main/Genome_assembly_and_quality_control/failed_genomes.xlsx). Genomes rejected from quality control appeared to be a random subset of the data; we found no temporal, patient-based, location, sequence type, or anatomical collection site associations (https://github.com/qianxuans/CHOP-NICU-Scripts/blob/main/Genome_assembly_and_quality_control/failed_genomes.xlsx). All passed assemblies are available on NCBI (Accession numbers available in https://github.com/qianxuans/CHOP-NICU-Scripts/blob/main/Genome_assembly_and_quality_control/failed_genomes.xlsx).

The genome assemblies demonstrated robust sequencing depth with mean coverage of 90·22X (median: 76·79X; IQR: 42·78-115·92X) and assembly contiguity as reflected by contig length averaging 18,541.07 base pairs (median: 18·361 bp; IQR: 13,357·5-22,566·5 bp). This comprehensive genomic dataset allowed high-resolution investigation of clonality and transmission dynamics. The genome assemblies deposited in the NCBI GenBank database may exhibit minor variations from those employed in our analyses, due to GenBank's standard processing pipeline which includes sequence trimming and the exclusion of contigs below 200 base pairs.

Core-genome Phylogeny Inference

The 1,446 high-quality genome assemblies were annotated using Bakta v1.9.2 with Database v5.1⁶. The resulting gff3 files were utilized for core-genome alignment, which was carried out using panaroo 1.5.0⁷. This alignment was then used to infer the maximum likelihood phylogeny in IQ-TREE version 2.3.0⁸, employing the general time-reversible (GTR) substitution model⁹ and accounting for among-site rate heterogeneity using the Gamma distribution and four rate categories¹⁰. The phylogeny inference was performed with 100 initial parsimony trees. Unbiased branch support values were provided by UFBoot using 10,000 bootstrap replicates¹¹. SH-like approximate likelihood ratio test was implemented with 10,000 replicates to assess branch support¹². The resulting phylogeny was midpoint-rooted, and the Interactive Tree of Life (iTOL) was used for visualization and annotation¹³.

Clonality Determination

Two analyses were performed to produce the input for the analysis to determine strain compositions.

On one hand, pairwise SNP distances were calculated using the wrapper dnadiff in MUMmer 4.0.0rc¹⁴, and the gSNP in the dnadiff output was used as the number for SNP distance. We constructed pairwise SNP distance matrices comparing NICU genomes against: (1) themselves (NICU matrix), (2) a local database of 349 pediatric *S. aureus* genomes collected in Philadelphia (Local matrix)¹⁵, and (3) a curated global database of 68,298 high-quality publicly available *S. aureus* genomes¹⁶. A close clonal relationship can be observed in the SNP distance matrices, where only NICU matrix presents a unique cluster below 100 SNP distances.

On the other hand, we used a hierarchical clustering approach to partition all genomes into 28 groups using the core-genome phylogenetic tree described previously in **Core-genome phylogeny inference**. During the hierarchical clustering process, we started by computing the pairwise distances between any pair of genomes using branch lengths in the phylogenetic tree using the `cophenetic.phylo` function in R package `ape`¹⁷. The generating distance matrix was used to test every possible grouping pattern using function `hclust` from `stats`, a built-in R package. To evaluate the quality of the grouping pattern, we calculated the average silhouette scores using function `silhouette` from R package `cluster`¹⁸, and grouping pattern with the highest average silhouette scores was used to

partition the genomes. Every group with no less than two genomes was supported with 100 ultrafast bootstrap support¹¹ and 100 SH-like approximate likelihood ratio test support (SH-aLRT)¹⁹, except for Group2 with 97 SH-aLRT support (appendix p6). In total 28 groups were determined with hierarchical clustering. A quality check step was conducted before the downstream analysis. If there is only one genome in a group, that genome is considered a singleton. For the groups with no less than 2 genomes, an arbitrary threshold at 1000 SNP, informed by the natural gaps in the patterns of SNP distance matrices, was used to detect the potential singleton members in the group. In a group, if any member has no less than 1000 SNP distances to any of the rest members, the member is considered a singleton. If every member has no less than 1000 SNP distance to any of the rest members, every member is considered singleton. Removing those singletons minimize the possibility of the presence of extra-long branches in the phylogenetic trees due to the outlier members. Those long branches could compromise the accuracy of the analysis. Twenty-four out of 28 groups consisted of no less than two genomes, averaging 60 genomes (median 26, IQR 4-73). After excluding three singletons from three groups, subsequent fine-grained phylogenetic analysis was implemented. For each of the 24 groups, we used the curated database of 68,298 high-quality *S. aureus* genomes mentioned previously to identify five most closely related publicly available genomes for every study genome in the group. Those publicly available genomes restricted the thresholds to focus on potential transmission clusters within the NICU. Excessively high thresholds including public genomes in the strains were considered unsuitable, as they would likely extend beyond the scope of intra-NICU transmission. The multi-sequence alignments containing study genomes and closest publicly available genomes were subsequently produced using Snippy 4.6.0²⁰, and the study genome with highest N50 in each group was used as the reference genome. The resulting alignments were used as the input for inferring the reference-based maximum-likelihood phylogenetic trees with IQ-TREE version 2.3.0⁸, employing the general time-reversible (GTR) substitution model⁹ and accounting for among-site rate heterogeneity using the Gamma distribution and four rate categories¹⁰, and supported with standard bootstrap supports²¹. We implemented a single linkage clustering (SLC) algorithm to determine closely related strains at every possible potential SNP threshold from the minimal distance in the group to 500. At each SNP threshold tested, the strain compositions determined by pairwise SNP distances were mapped to the corresponding reference-based phylogenetic tree of the group. Interestingly, even at conservative SNP thresholds the strains determined by SNP thresholds were sometimes not monophyletic on the phylogenetic tree (Multi-SNP-Threshold Plot). To correct for this problem, we expanded our strains to include the smallest clade that included all SLC members as well as any non-SLC members in that clade when the bootstrap support of this clade is no less than 70. We then selected the SNP threshold at which the number and composition of strains plateaued. The plateaus of 24 groups averaged 21.58 (median 18.5, IQR 10.75-27.25), and 486 strains were determined (223 singletons and 263 clones). The analytical codes are publicly available in the project's GitHub repository (<https://github.com/qianxuans/CHOP-NICU-Scripts>). The ready-to-use pipeline performing the analysis is publicly available for download at <https://github.com/microbialARC/THRESHER>

Cluster analysis and visualizations

If any bacteremia isolate was present in a transmission cluster, it was defined as an invasive cluster; otherwise, it was referred to as a colonizing cluster.

Methicillin resistance in transmission clusters was confirmed through clinical laboratory testing and detection of the *mecA* gene using BLASTx 2.14.0+ against a customized protein database containing 13 PBP2a family β -lactam-resistant peptidoglycan transpeptidase amino acid sequences from the NCBI protein database^{22,23}. The accession numbers of the sequences were WP_057521704.1, WP_063852670.1, WP_063852677.1, WP_063852683.1, WP_000721309.1, WP_000721306.1, WP_063852626.1, WP_063852617.1, WP_000721310.1, WP_063851348.1, WP_012655867.1, WP_063852710.1, and WP_000725529.1.

We used customized R scripts to create visualizations. These included plots depicting patient metadata, invasive status, methicillin resistance, and cluster persistence. Additionally, we generated swimmer plots and annotated NICU floor plans using separate customized R scripts. All scripts are available from <https://github.com/qianxuans/CHOP-NICU-Scripts>.

Molecular Clock Analysis

For genomes identified as belonging to the same strain, we generated whole-genome alignments and masked putative recombination regions using ClonalFrameML²⁴. Time to most recent common ancestor (tMRCA) was estimated using BEAST v2.7.6²⁵, employing an HKY substitution model and strict molecular clock. The analysis incorporated a coalescent constant population model with clock rates following log-normal distributions (1.0E-6 to 1.0E-7) as Bayesian priors. Markov Chain Monte Carlo (MCMC) chains were run up to 500 million steps to ensure effective sample sizes exceeded 200 for all parameters. The resulting phylogenetic trees were generated in NEXUS format and can be accessed at https://github.com/qianxuans/CHOP-NICU-Scripts/tree/main/Molecular_Clock_Analysis/BEAST2_Tree_Visualization/nexus. All analysis scripts and visualization tools are publicly available in the project's GitHub repository (<https://github.com/qianxuans/CHOP-NICU-Scripts>).

Supplementary references

1. Erik L. Clarke, Louis J. Taylor, Chunyu Zhao, et al. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome*. 2019;7(1):46. doi:10.1186/s40168-019-0658-x
2. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*. 2020;70(1):e102–n/a. doi:10.1002/cpbi.102
3. Li D, Liu C, Luo R, Sadakane K, Lam T. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–1676. doi:10.1093/bioinformatics/btv033
4. A. Murat Eren, Evan Kiefl, Alon Shaiber, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology*. 2020;6(1):3–6. doi:10.1038/s41564-020-00834-3
5. Brian D. Ondov, Todd J. Treangen, Páll Melsted, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*. 2016;17(1):132. doi:10.1186/s13059-016-0997-x
6. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial genomics*. 2021;7(11). doi:10.1099/mgen.0.000685
7. Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol*. 2020;21:1–21
8. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*. 2020;37(5):1530–1534. doi:10.1093/molbev/msaa015
9. Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. *J Mol Evol*. 1984;20:86–93
10. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol*. 1994;39:306–314
11. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*. 2013;30(5):1188–1195. doi:10.1093/molbev/mst024

12. Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–321
13. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic acids research.* 2024. doi:10.1093/nar/gkae268
14. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS computational biology.* 2018;14(1):e1005944
15. She Q, Zackular J, Moustafa A, Planet P. Local database of 349 pediatric *S. aureus* genomes. doi:10.5281/zenodo.14795788
16. Moustafa AM, Planet PJ. WhatsGNU: a tool for identifying proteomic novelty. *Genome Biology.* 2020;21(1):58. doi:10.1186/s13059-020-01965-w
17. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35(3):526–528
18. Struyf A, Hubert M, Rousseeuw P. Clustering in an object-oriented environment. *Journal of Statistical Software.* 1997;1:1–30
19. Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–321
20. Seemann T (2015)
snippy: fast bacterial variant calling from NGS reads
<https://github.com/tseemann/snippy>.
21. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985;39(4):783–791
22. Sayers EW, Beck J, Bolton EE, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research.* 2021;49(D1):10. doi:10.1093/nar/gkaa892
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–410
24. Xavier Didelot, Daniel J. Wilson. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology.* 2015;11(2):e1. doi:10.1371/journal.pcbi.1004041

25. Bouckaert R, Vaughan TG, Barido-Sottani J, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*. 2019;15(4):e1–e1006650. doi:10.1371/journal.pcbi.1006650

Extended Discussion

Pathogen Trends in BSIs

While bloodstream infections due to coagulase-negative staphylococci have declined in our NICU, *S. aureus* (mostly MSSA) remains the predominant pathogen. Point prevalence testing in the CHOP NICU indicates high *S. aureus* colonization rates among NICU infants, primarily MSSA with a few MRSA cases.

Infection Prevention Measures

Current infection prevention measures are broadly applied, but our findings suggest strain-specific approaches may be more effective. Certain *S. aureus* strains exhibit both high transmissibility and invasive potential, highlighting the need for targeted mitigation strategies such as enhanced environmental cleaning and selective decolonization. Identifying these high-risk strains could refine prevention efforts beyond standard measures. Alternatively, other factors may facilitate transmission, allowing specific strains to colonize a larger pool of vulnerable hosts. When conditions permit (e.g., a breakdown of physical or immunological barriers), these strains can cause invasive infections.

Limitations

Our strain determination method, which integrates SNP thresholds with phylogenetic structure, tends to expand rather than contract transmission clusters. The phylogenetic correction only adds members missed by the clustering algorithm rather than removing them, potentially overestimating cluster sizes. However, our within-cluster SNP distances align with other studies. Since most screening swabs were collected from nasal sites, the available data to assess the correlation between colonized body sites and infection risk is limited.

Supplementary figures and tables

Group	Number of Genomes	Bootstrap Support	SH-aLRT Support	CC	Threshold
1	1	N/A	N/A	Unassigned	N/A
2	163	100	97	CC5	34
3	311	100	100	CC398	72
4	3	100	100	Unassigned	15
5	154	100	100	CC30	40
6	104	100	100	CC45	19
7	4	100	100	CC45	8
8	67	100	100	Unassigned	19
9	22	100	100	CC121	31
10	9	100	100	CC22	19
11	6	100	100	Unassigned	10
12	3	100	100	Unassigned	11
13	245	100	100	CC8	32
14	45	100	100	CC97	15
15	83	100	100	CC15	56
16	3	100	100	CC1	14
17	14	100	100	CC1	18
18	35	100	100	CC1	21
19	1	N/A	N/A	CC1	N/A
20	30	100	100	Unassigned	9
21	4	100	100	Unassigned	6
22	46	100	100	CC5	22
23	11	100	100	CC8	12
24	73	100	100	CC8	26
25	2	100	100	Unassigned	6
26	1	N/A	N/A	Unassigned	N/A
27	1	N/A	N/A	Unassigned	N/A
28	5	100	100	CC1	3

Table S1: Phylogenetically corrected SNP Thresholds for Hierarchical Groups

This table summarizes the hierarchical clustering groups. The table provides the number of genomes, bootstrap and SH-aLRT supports for the grouping, clonal complex assignment, and the SNP thresholds. Groups with only one genome lack bootstrap and SH-aLRT supports (N/A), as these require at least 2 genomes to form a clade in the phylogeny. Nine groups were designated as "Unassigned" as they were not assigned any sequence types.

Cluster	Invasive Status	Observed Persistence (Days)	Molecular Clock Persistence (Days)
1	Invasive	975	1201
2	Invasive	708	998
3	Colonizing	944	1710
4	Invasive	583	879
5	Invasive	291	542
6	Invasive	848	2905
7	Invasive	449	610
8	Colonizing	947	1526
9	Colonizing	399	449
10	Invasive	840	1854
11	Colonizing	149	267
12	Colonizing	645	776
13	Invasive	449	661
14	Colonizing	358	393
15	Colonizing	949	1099
16	Colonizing	903	1045
17	Invasive	421	497
18	Colonizing	406	957
19	Invasive	557	909
20	Colonizing	155	159
21	Colonizing	99	183
22	Colonizing	309	373
23	Colonizing	603	1501
24	Colonizing	128	157
25	Colonizing	406	466
26	Colonizing	29	46
28	Colonizing	91	250
29	Invasive	422	535
30	Invasive	360	465
31	Colonizing	428	479
32	Colonizing	352	507
33	Colonizing	128	189
34	Invasive	27	80

35	Colonizing	22	143
37	Invasive	191	238
38	Colonizing	29	345
46	Colonizing	35	163
48	Colonizing	120	215
49	Colonizing	128	189
51	Invasive	216	251
56	Colonizing	495	673
57	Colonizing	379	430
58	Invasive	231	902
59	Colonizing	190	234
63	Colonizing	29	220
65	Invasive	69	161
67	Colonizing	595	602
68	Invasive	208	214

Table S2: Characteristics of 69 Identified Transmission Clusters

This table presents invasive status and temporal metrics for identified transmission clusters. For each cluster, the table reports invasive status, earliest and most recent detection dates, time to most recent common ancestor (tMRCA), observed persistence period in days, and molecular clock-estimated persistence in days.

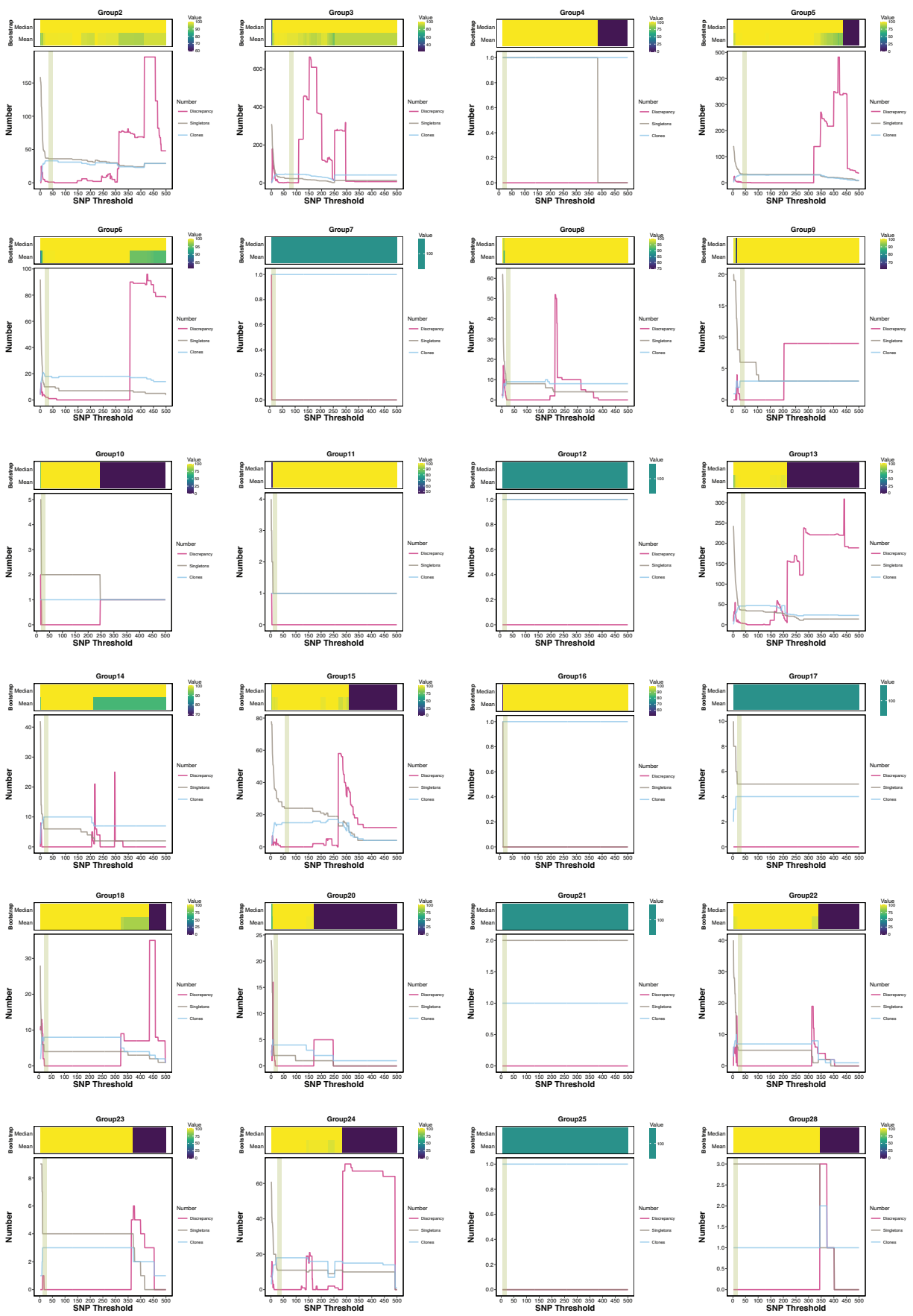


Figure S1: Multi-SNP Threshold Plots for Hierarchical Groups containing at Least Two Genomes

For each plot, the x-axis represents tested SNP threshold values. The upper panel displays average and median bootstrap support values for clones identified at each threshold. The lower panel quantifies discrepancy genomes (purple), clones (blue), and singletons (gray) across the tested threshold range. The green bar indicates the plateau where the strain compositions were finalized.

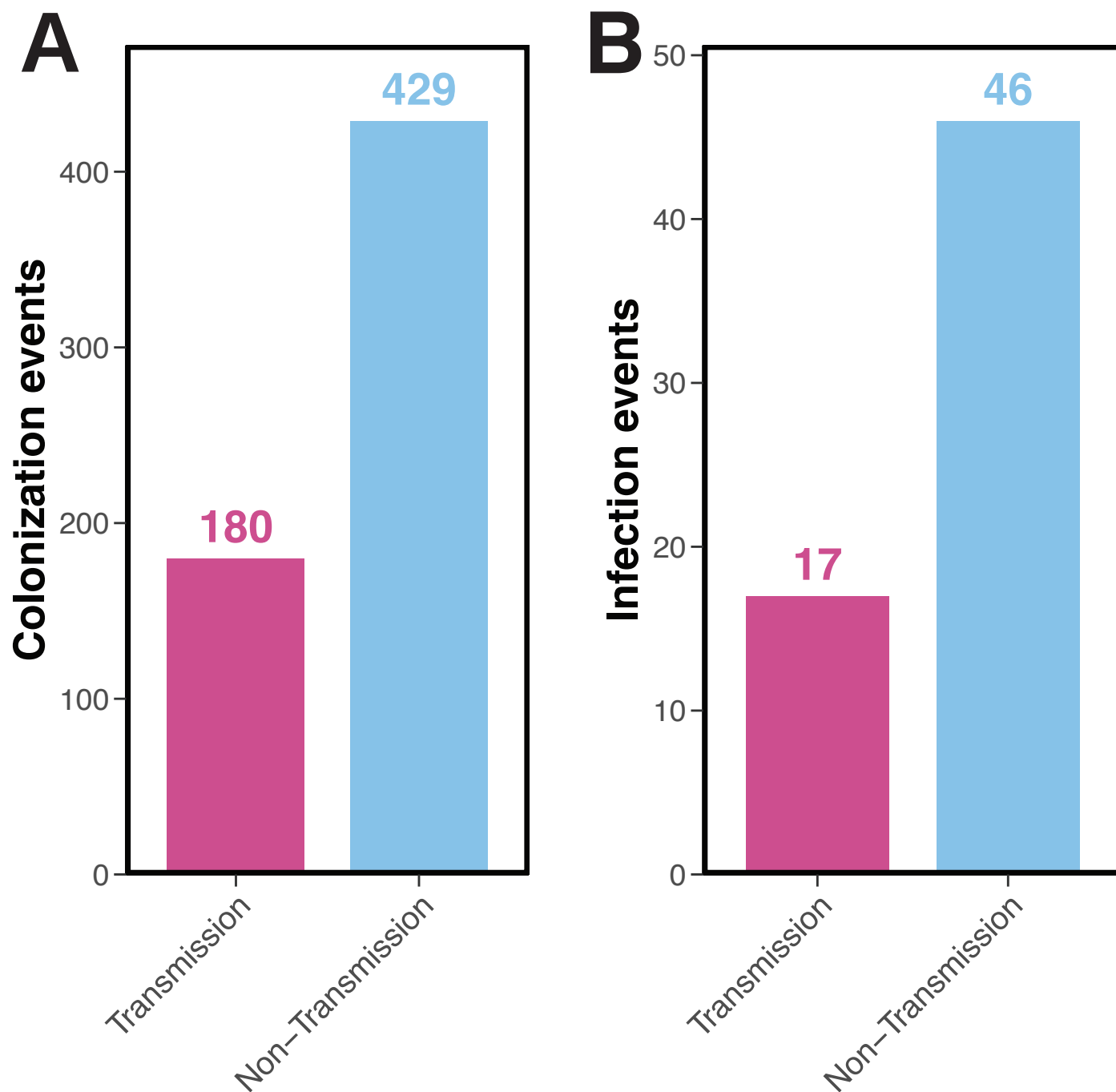


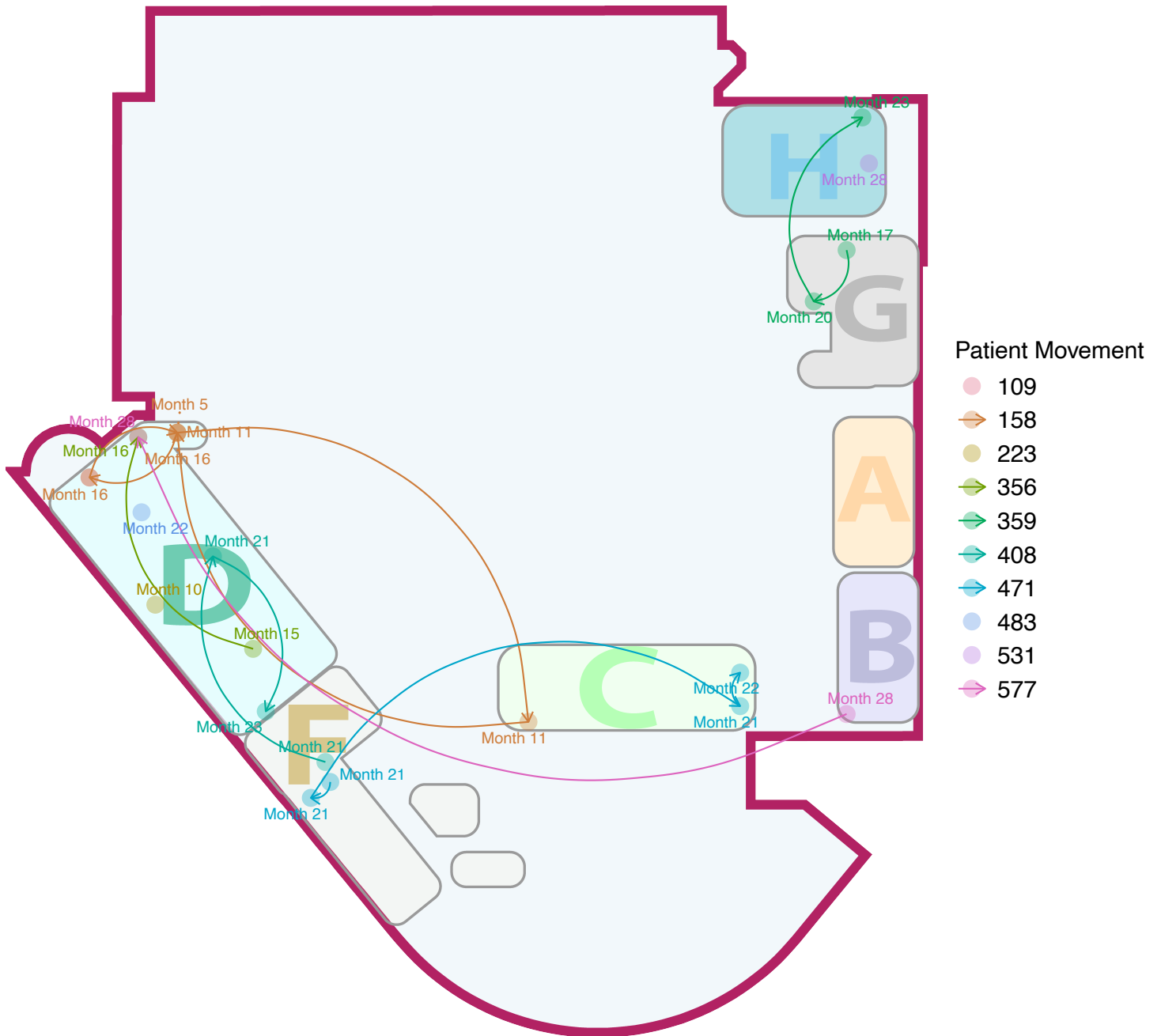
Figure S2: Distribution of Transmission-acquired Colonization and Infection Events

(A) Distribution of colonization events in the NICU (n=609), showing 180 cases (29.56%) identified as transmission-acquired after excluding index patients. (B) Distribution of infection events across CHOP departments (n=63), with 17 cases (26.98%) identified as transmission-acquired.

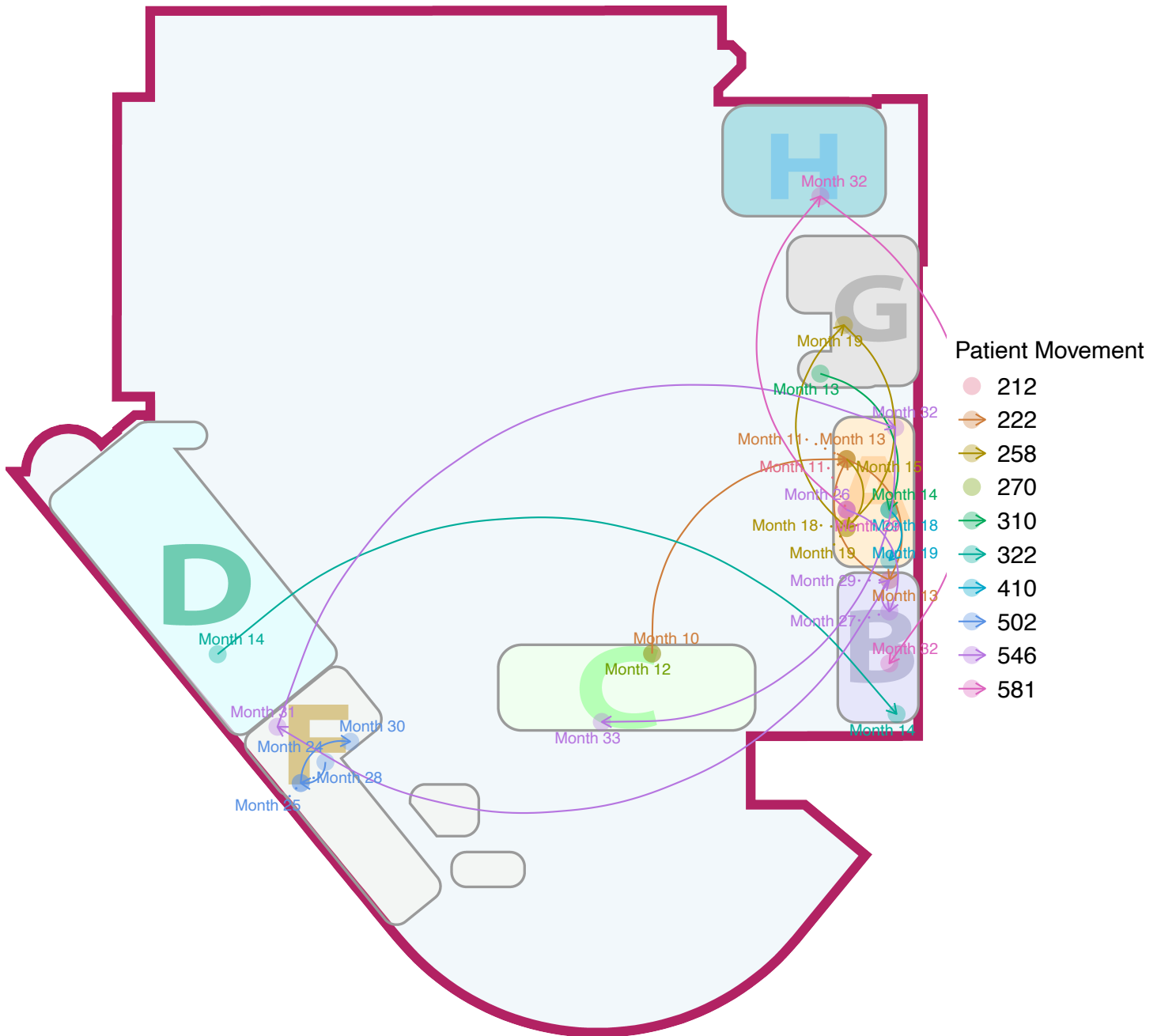
Visualization of patient movements across the NICU for each identified transmission cluster. Patient locations are mapped chronologically, with connecting lines indicating the patient movements.



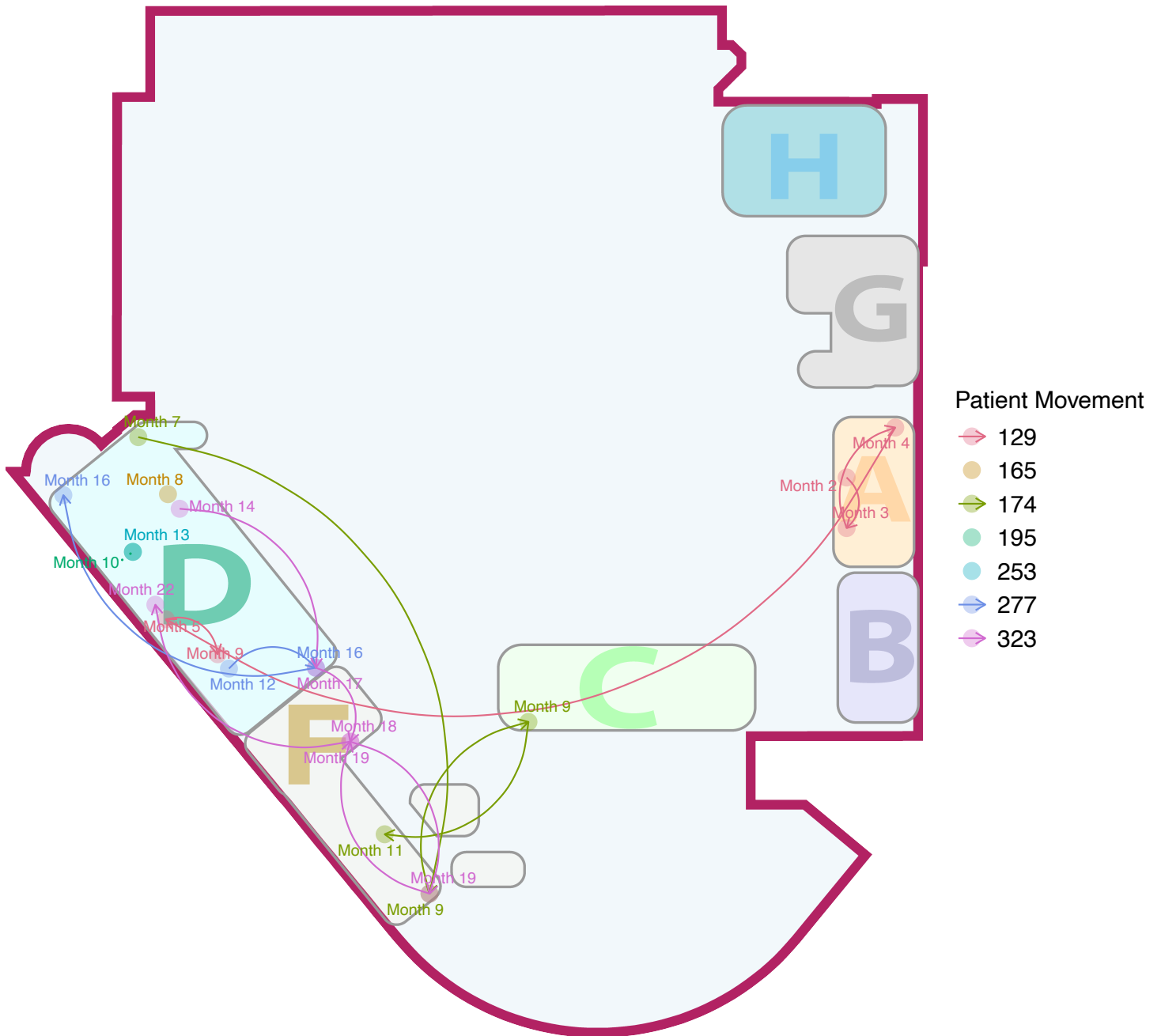
Cluster3



Cluster4



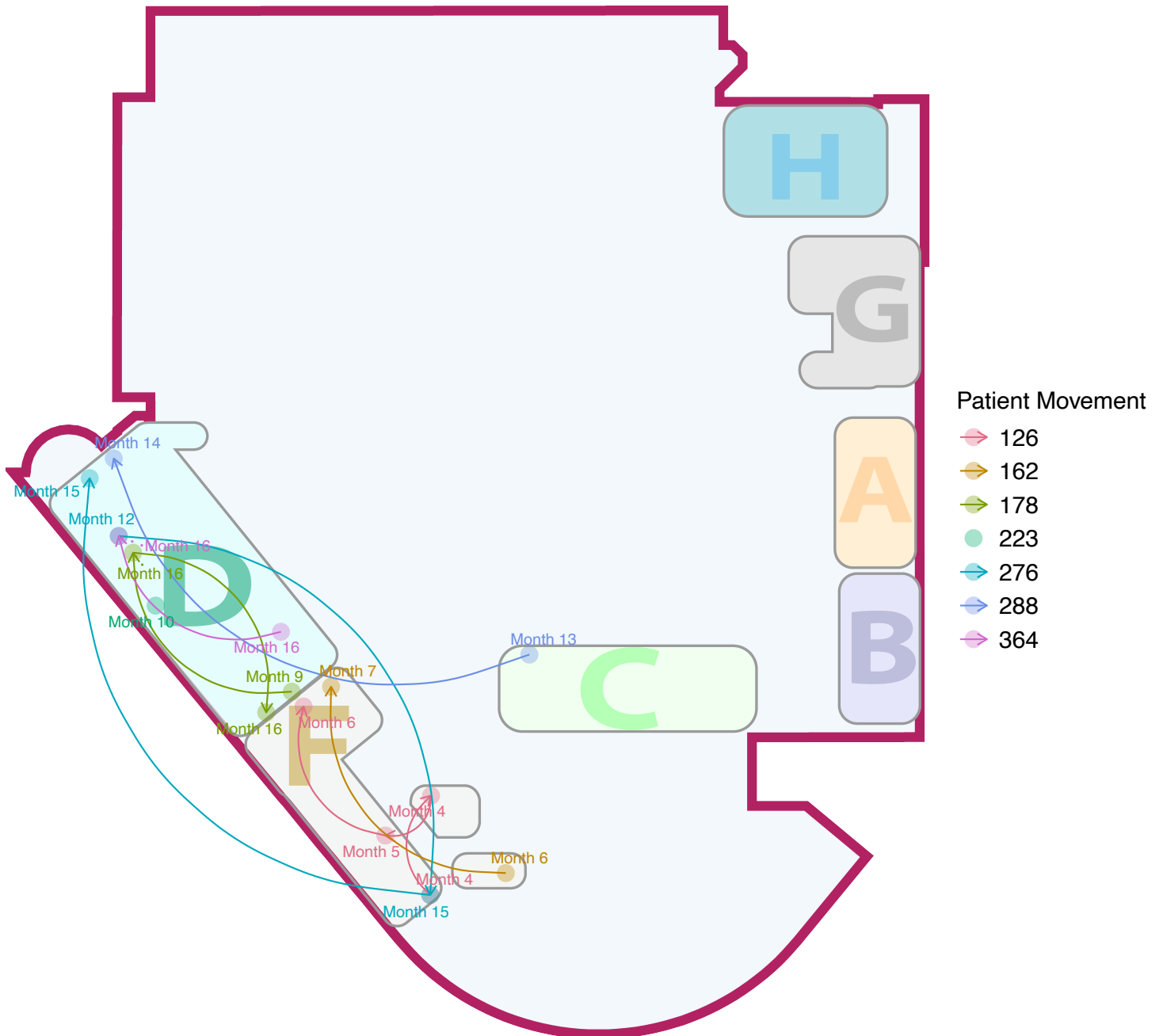
Cluster5



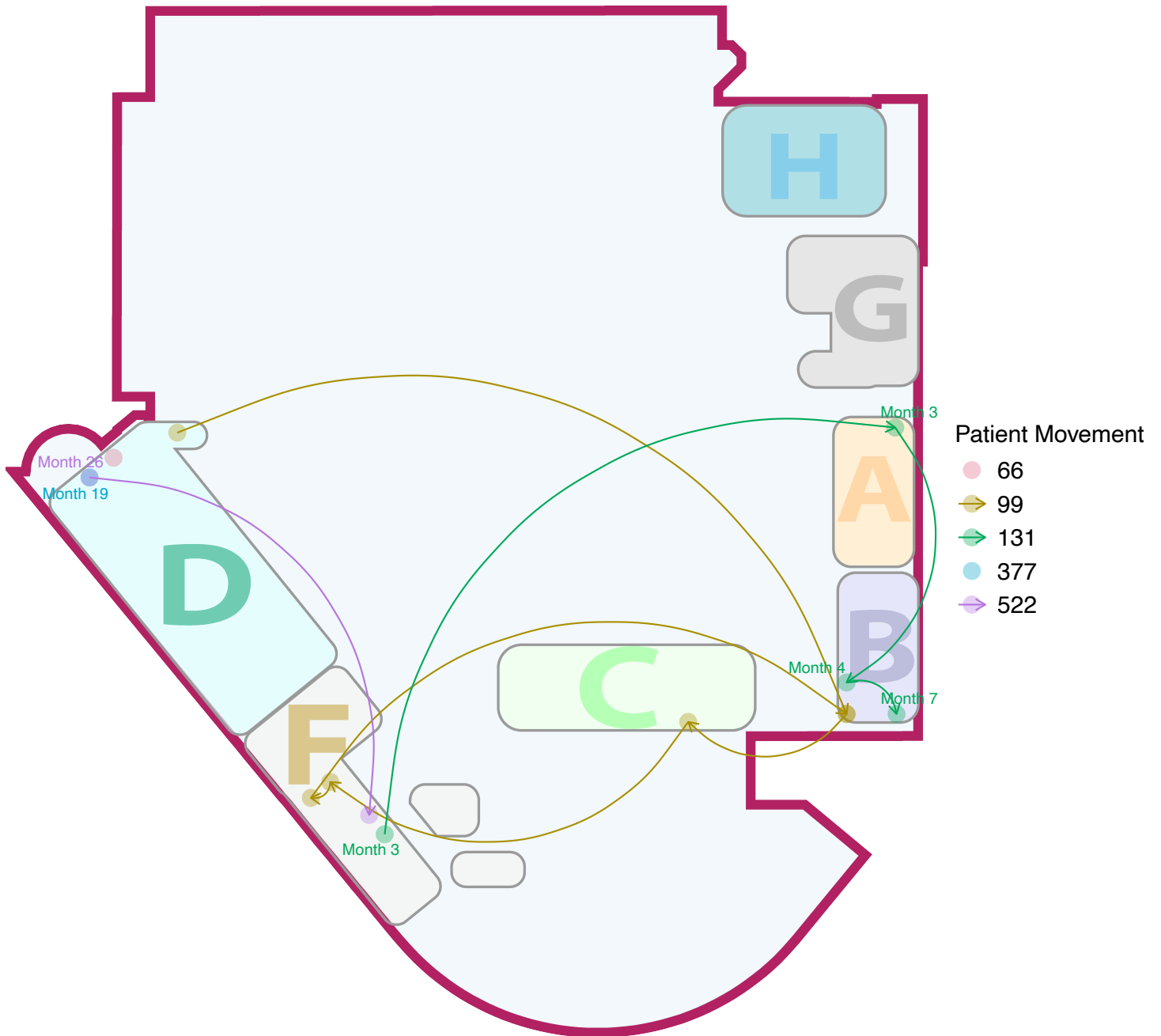
Cluster6



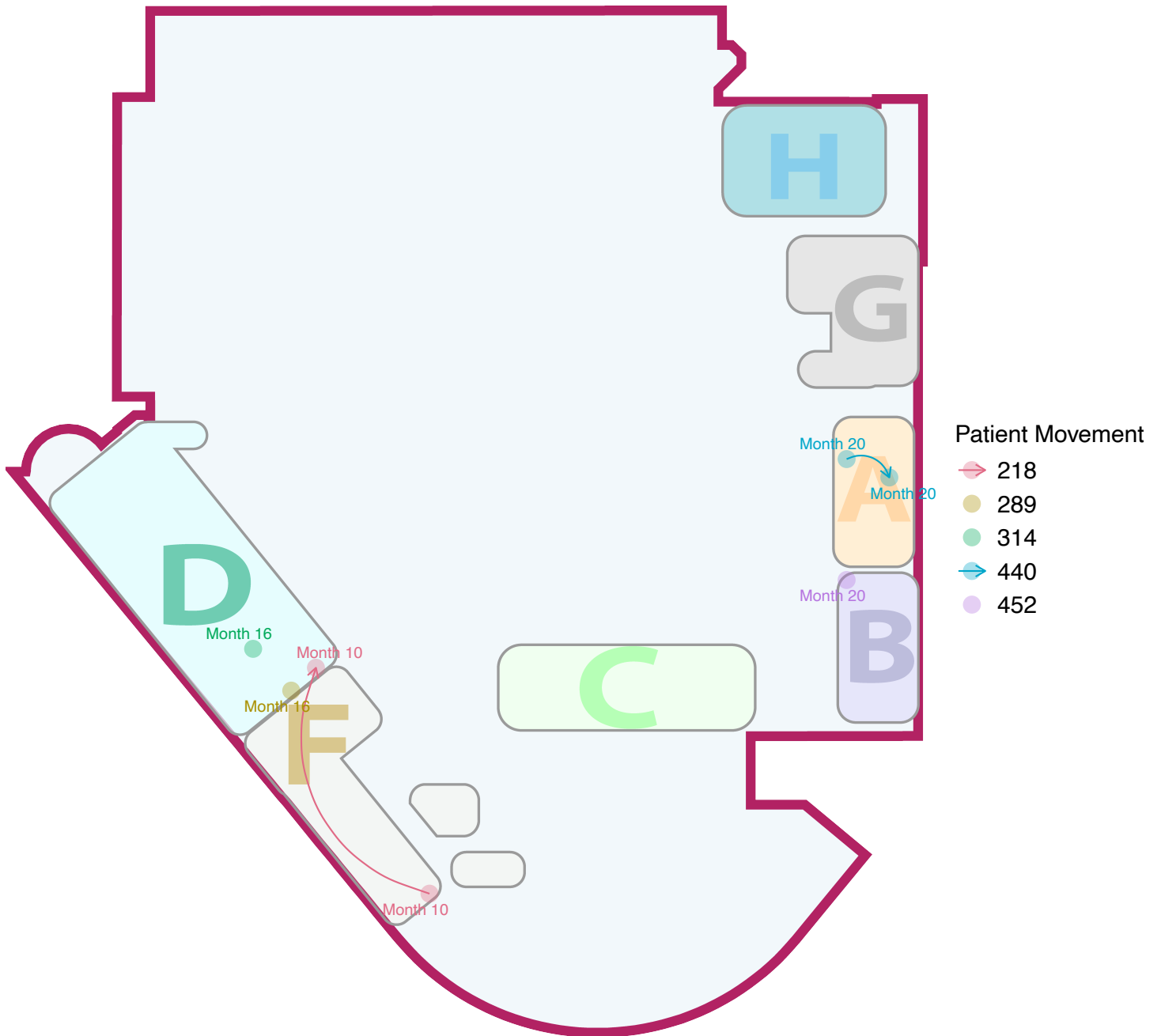
Cluster7



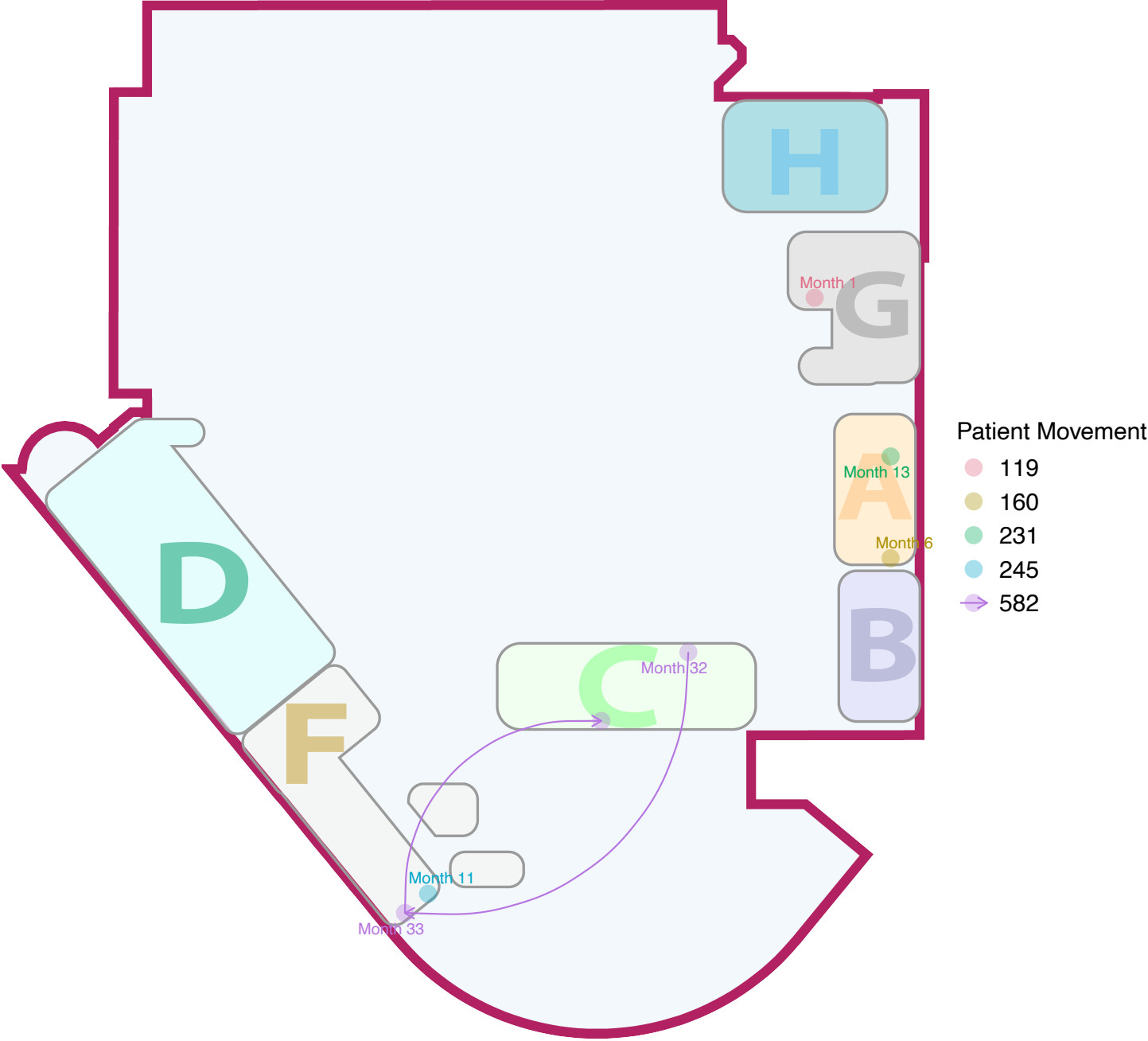
Cluster8



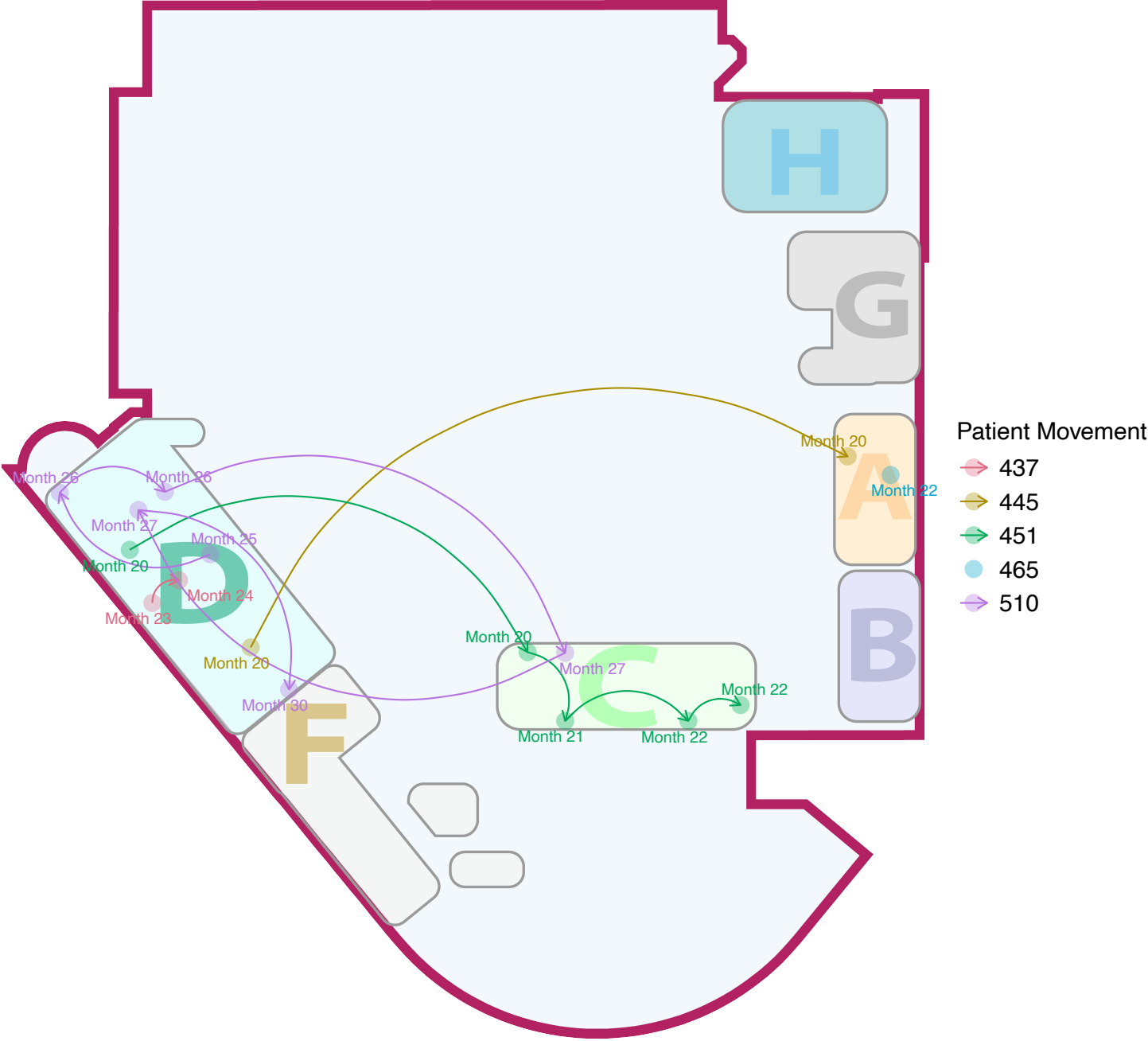
Cluster9



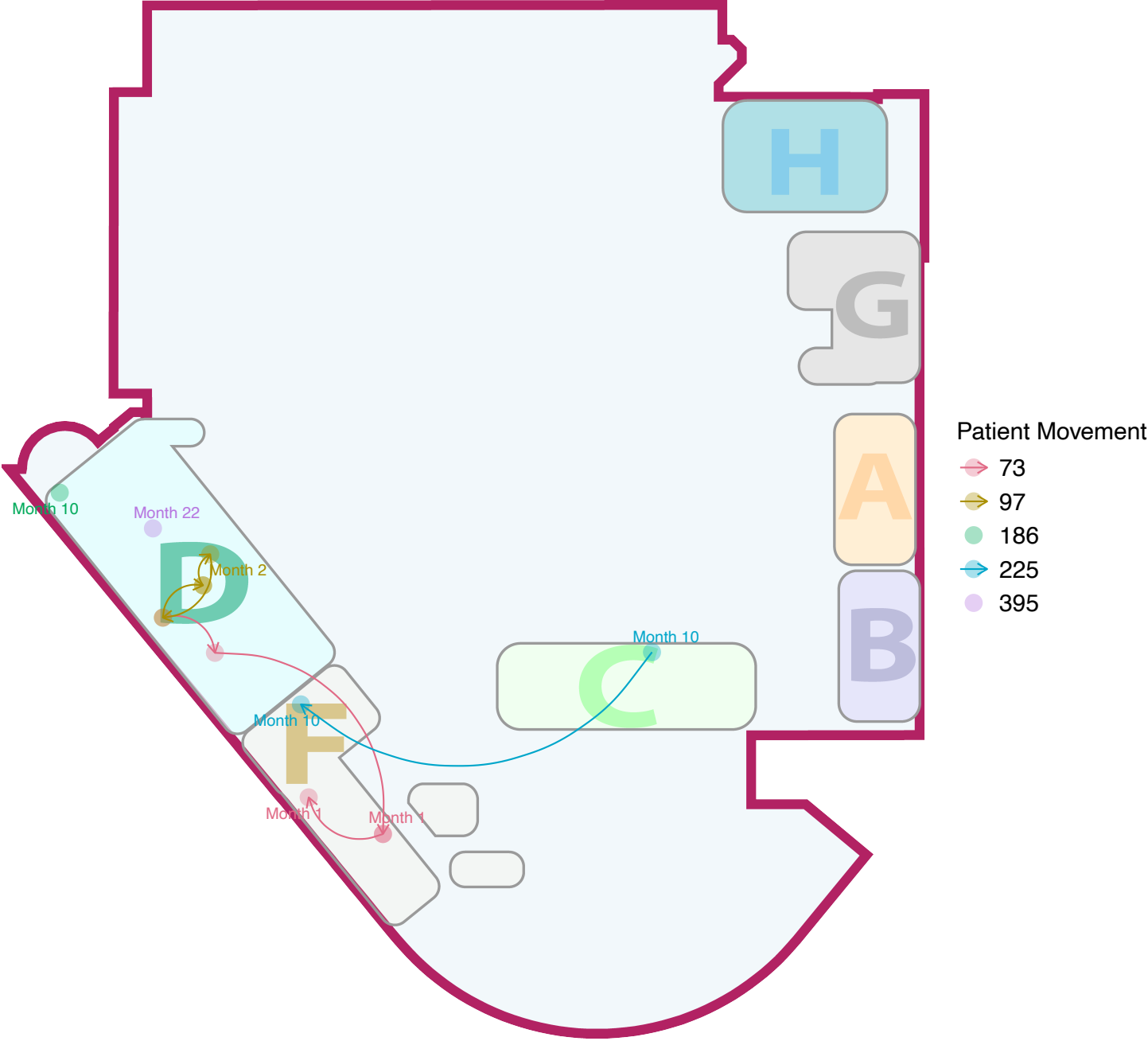
Cluster10



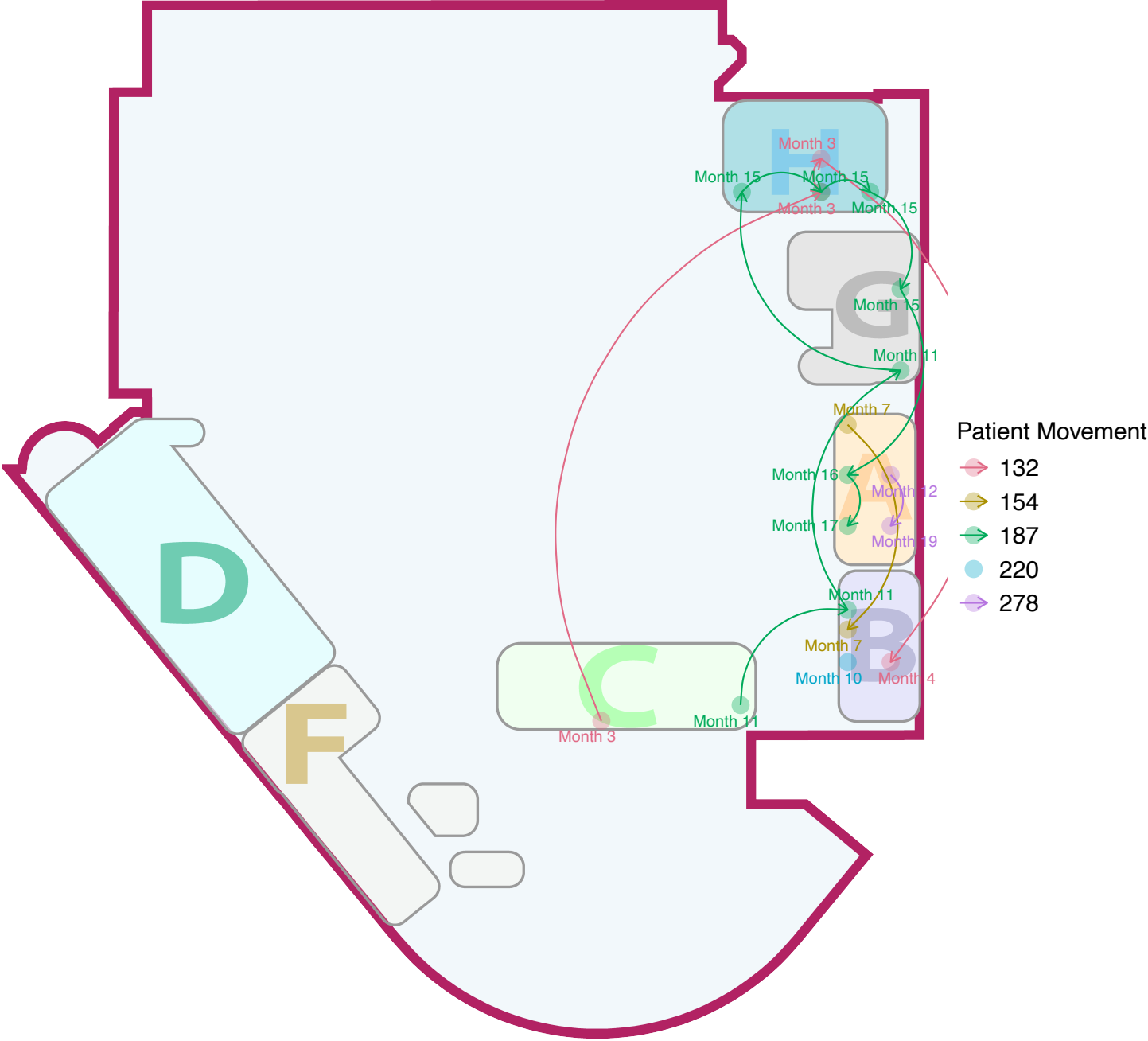
Cluster11



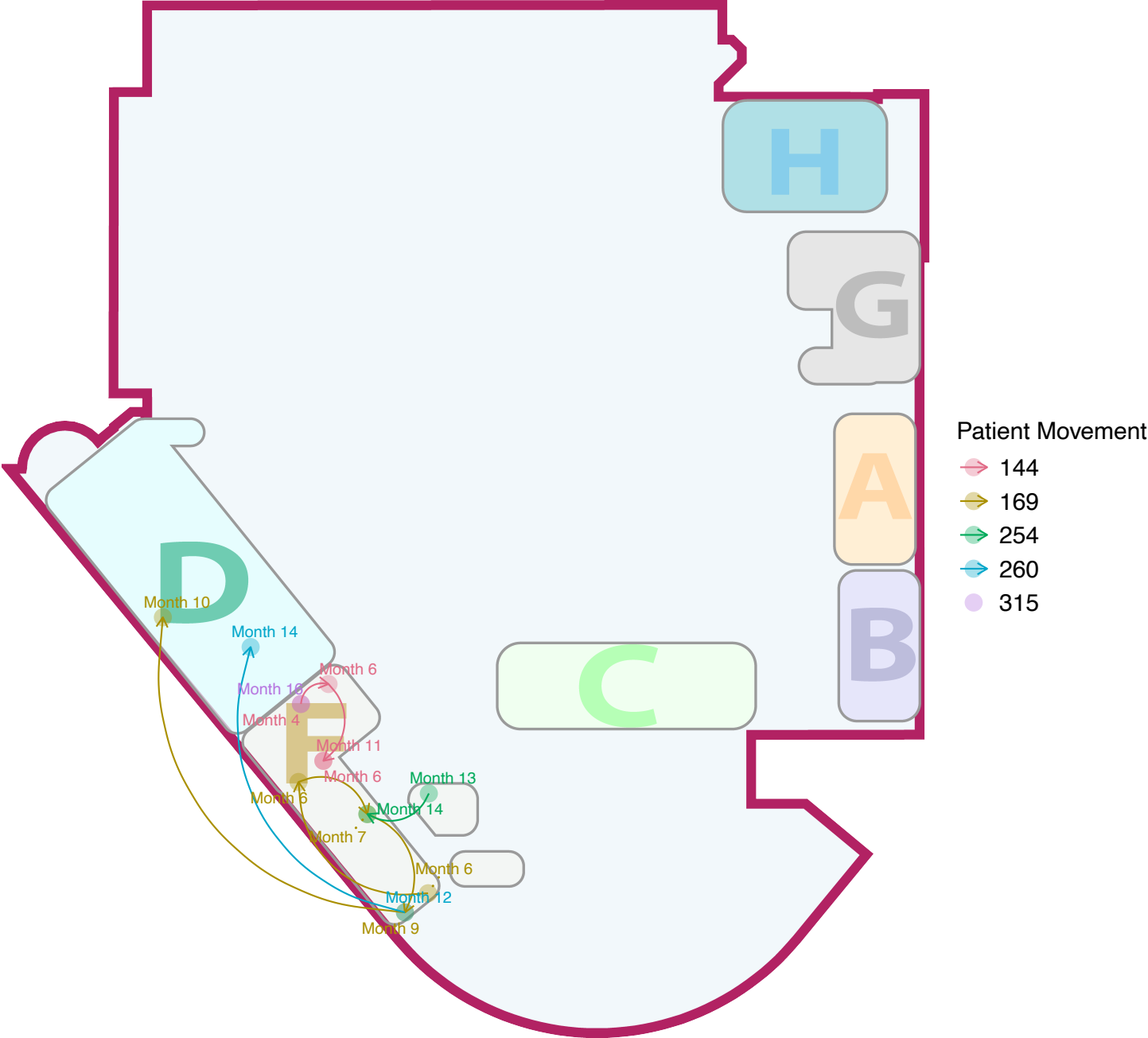
Cluster12



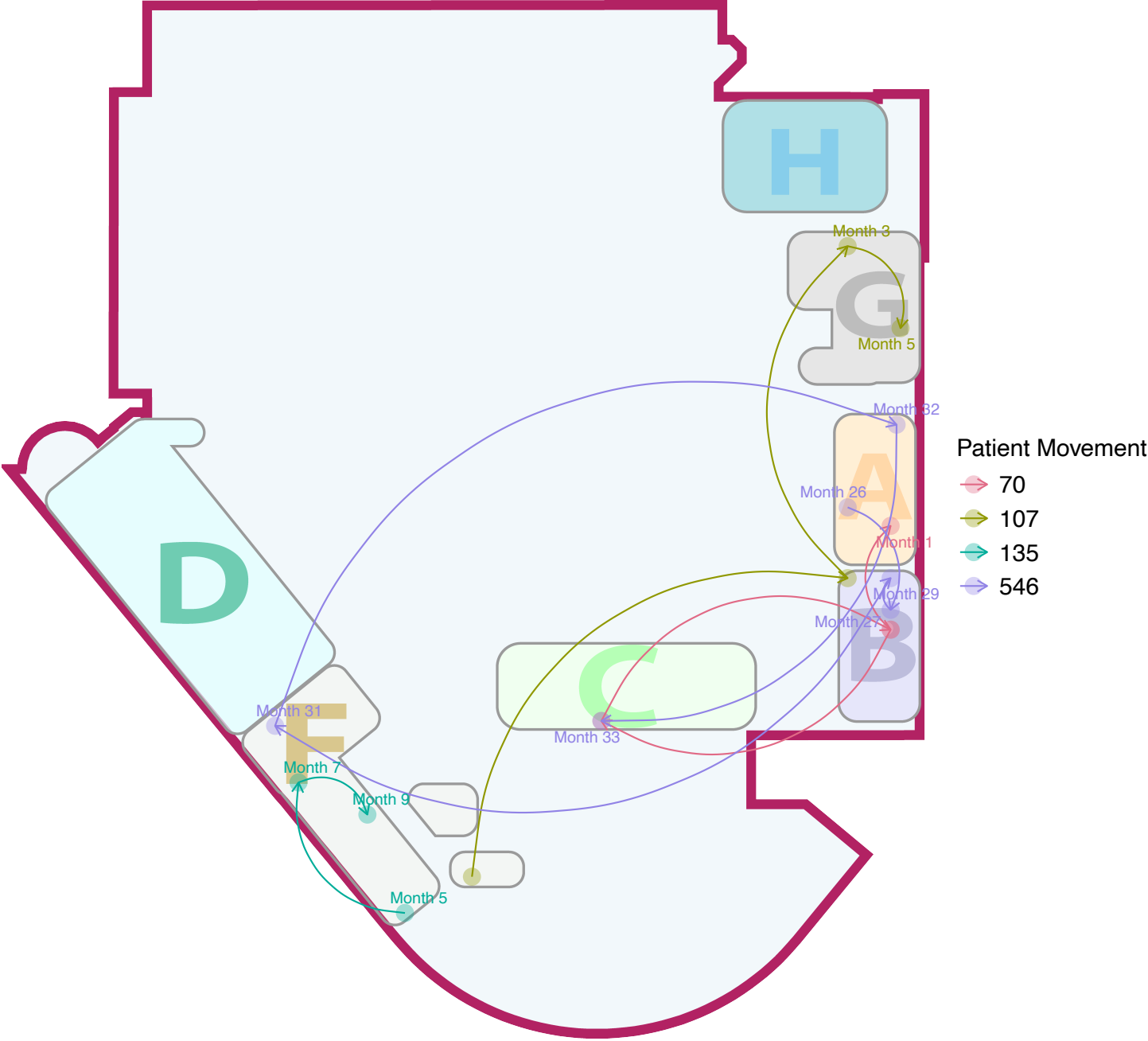
Cluster13



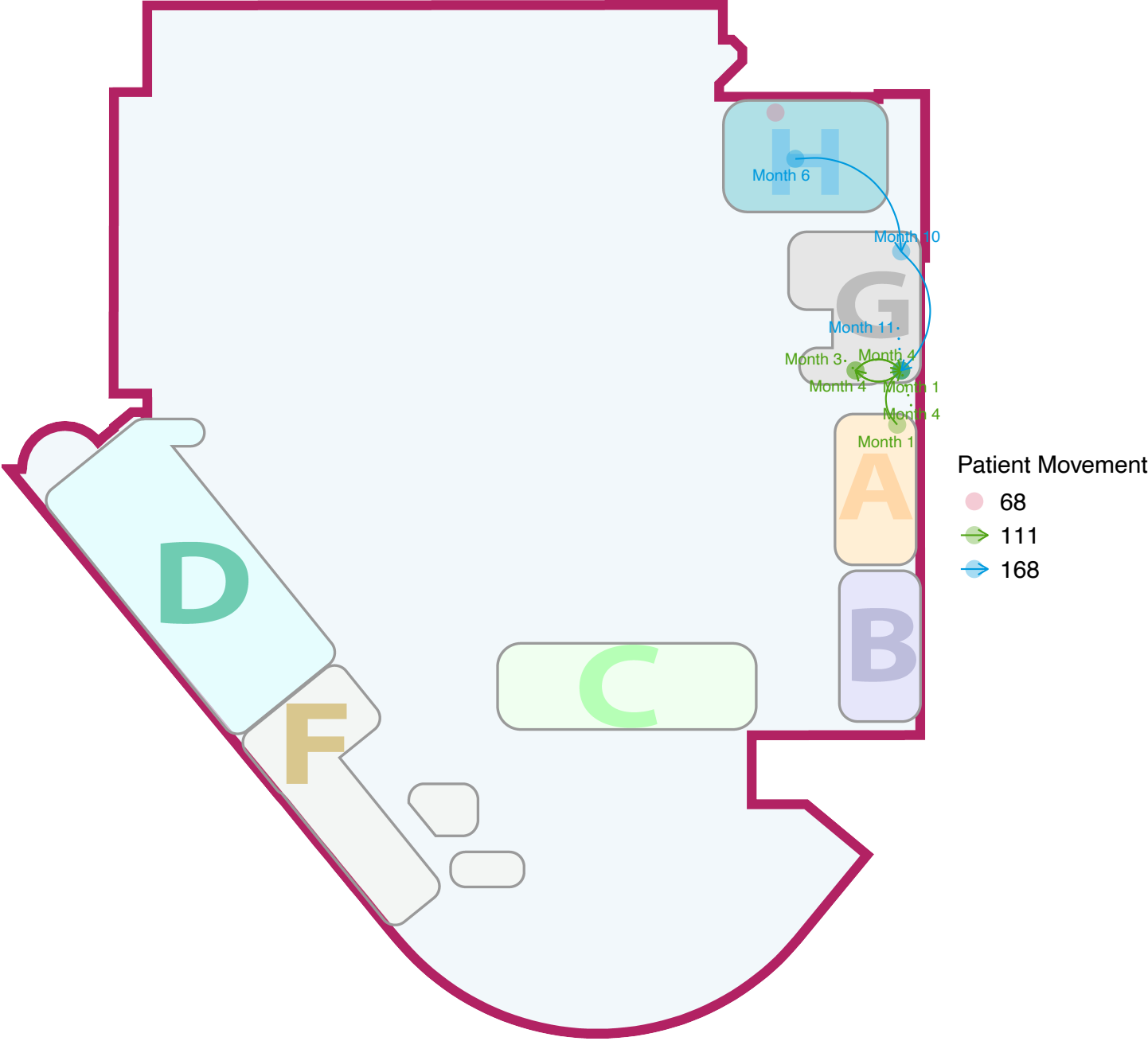
Cluster14



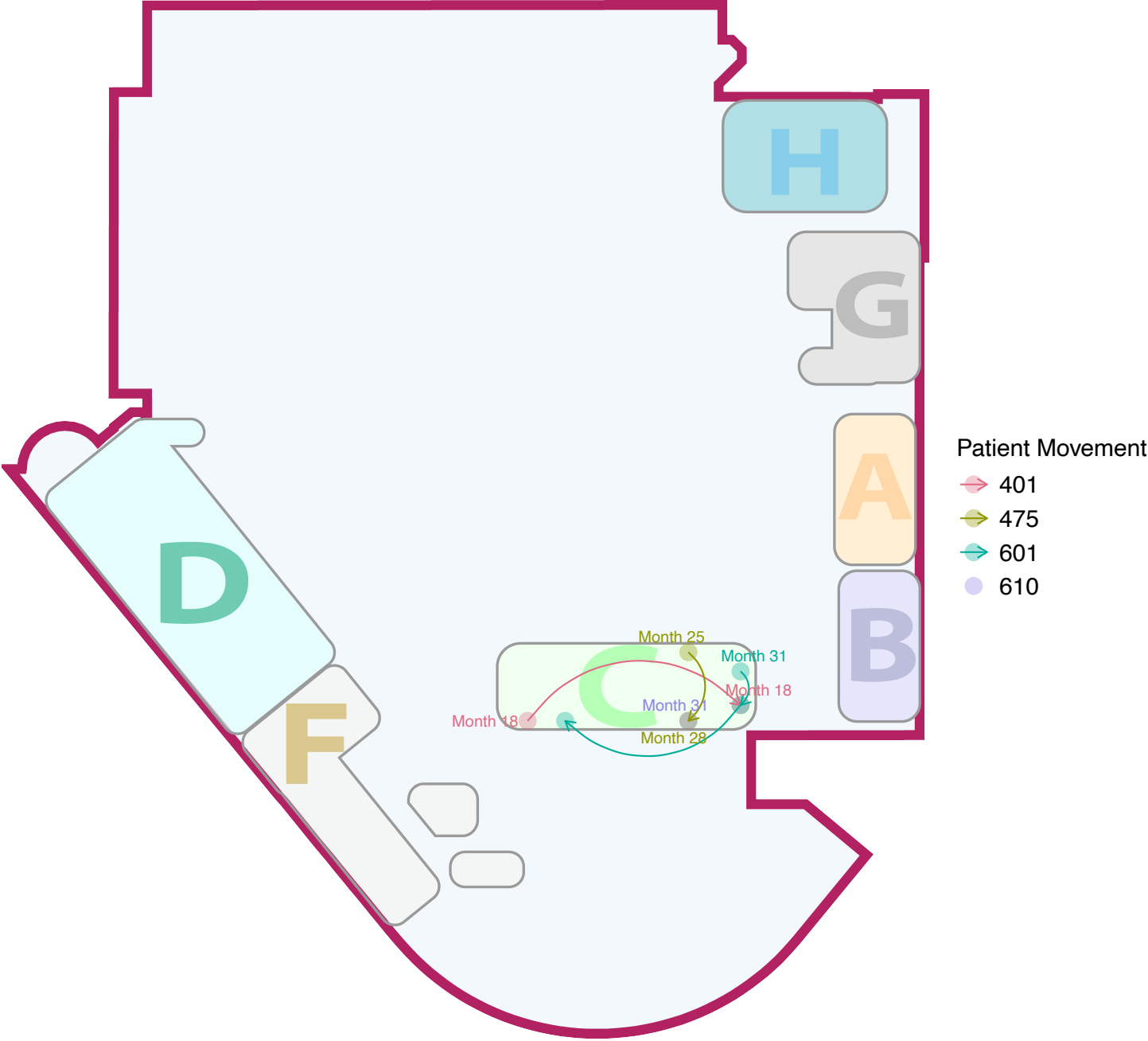
Cluster15



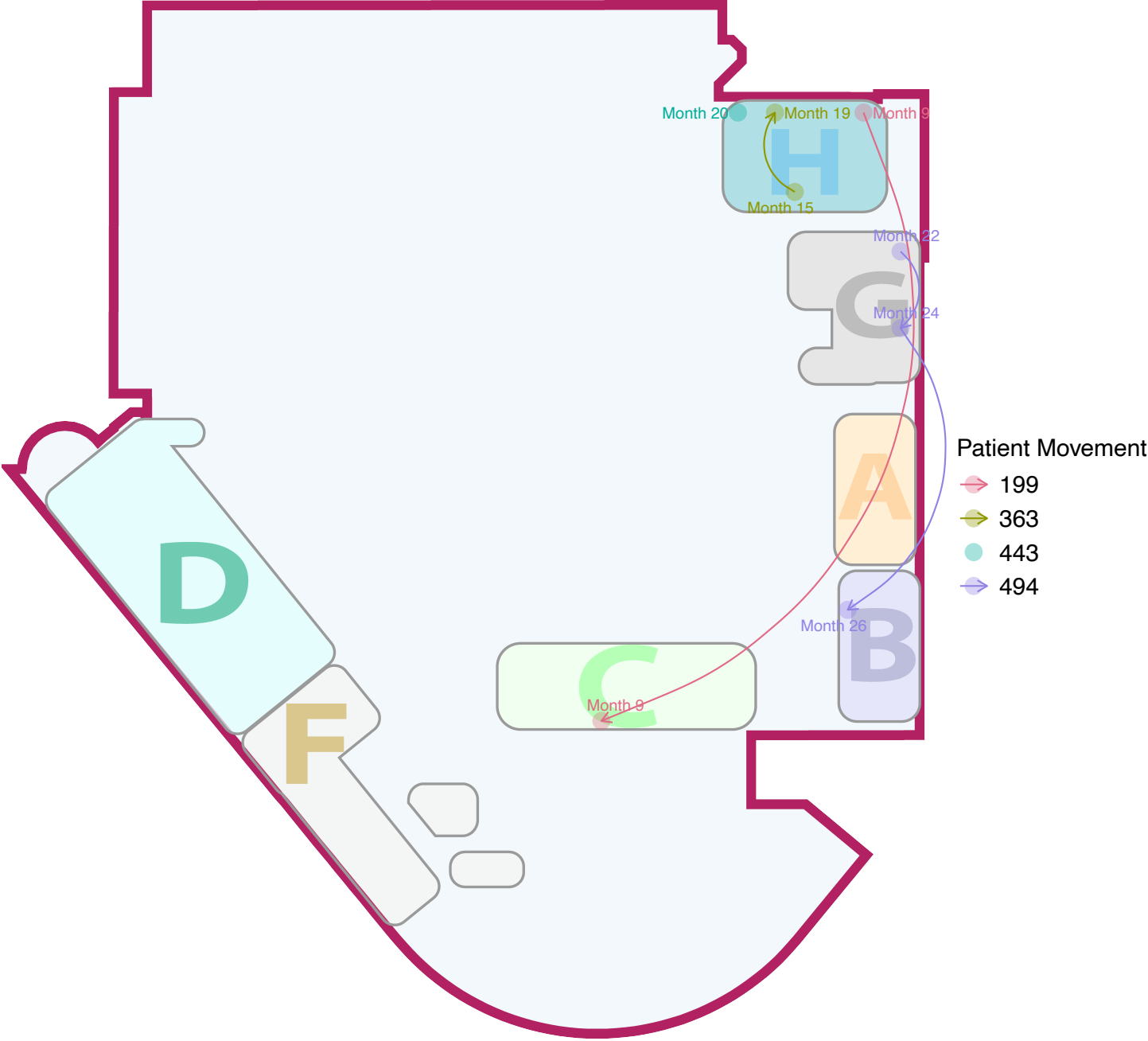
Cluster16



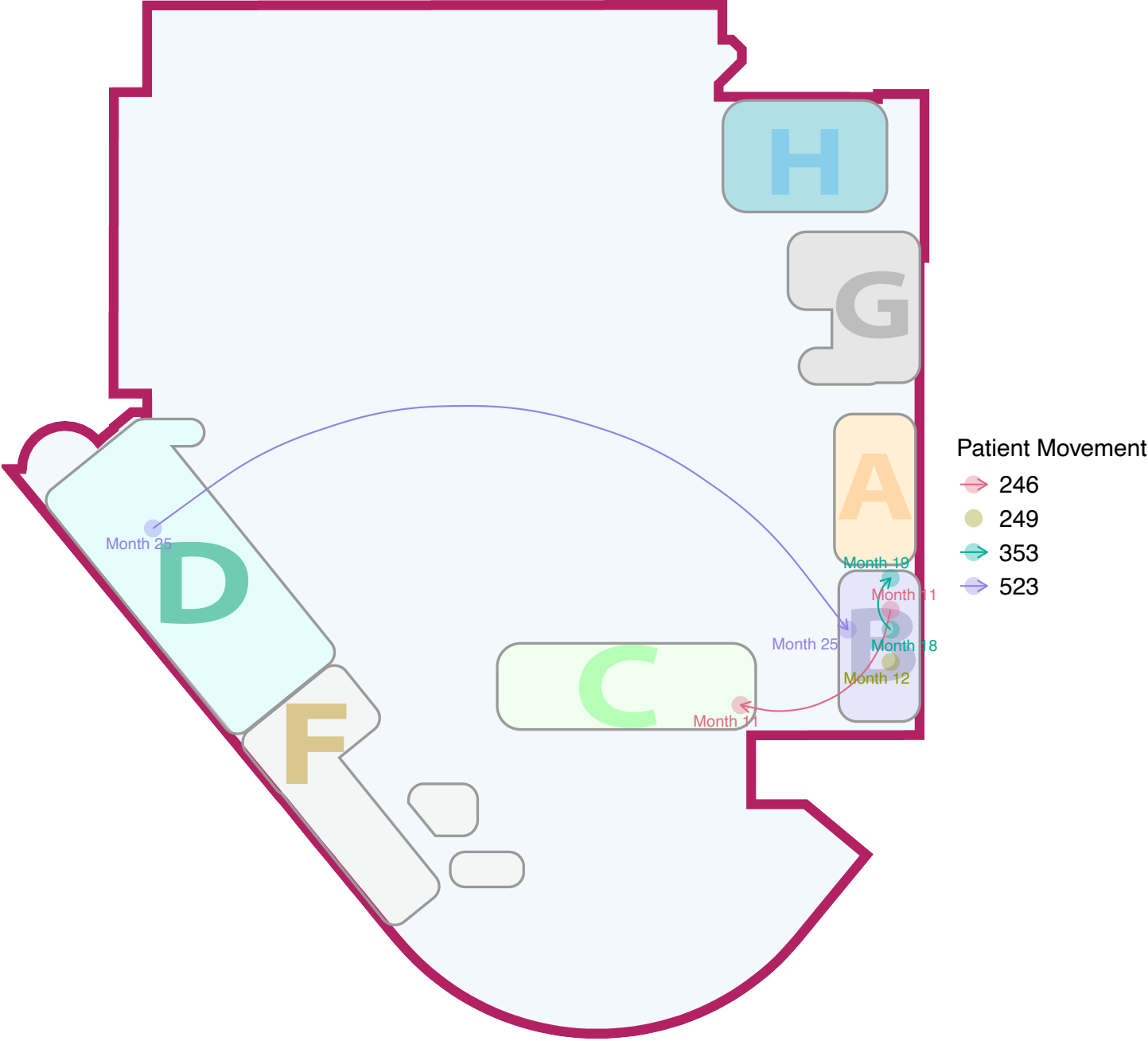
Cluster17



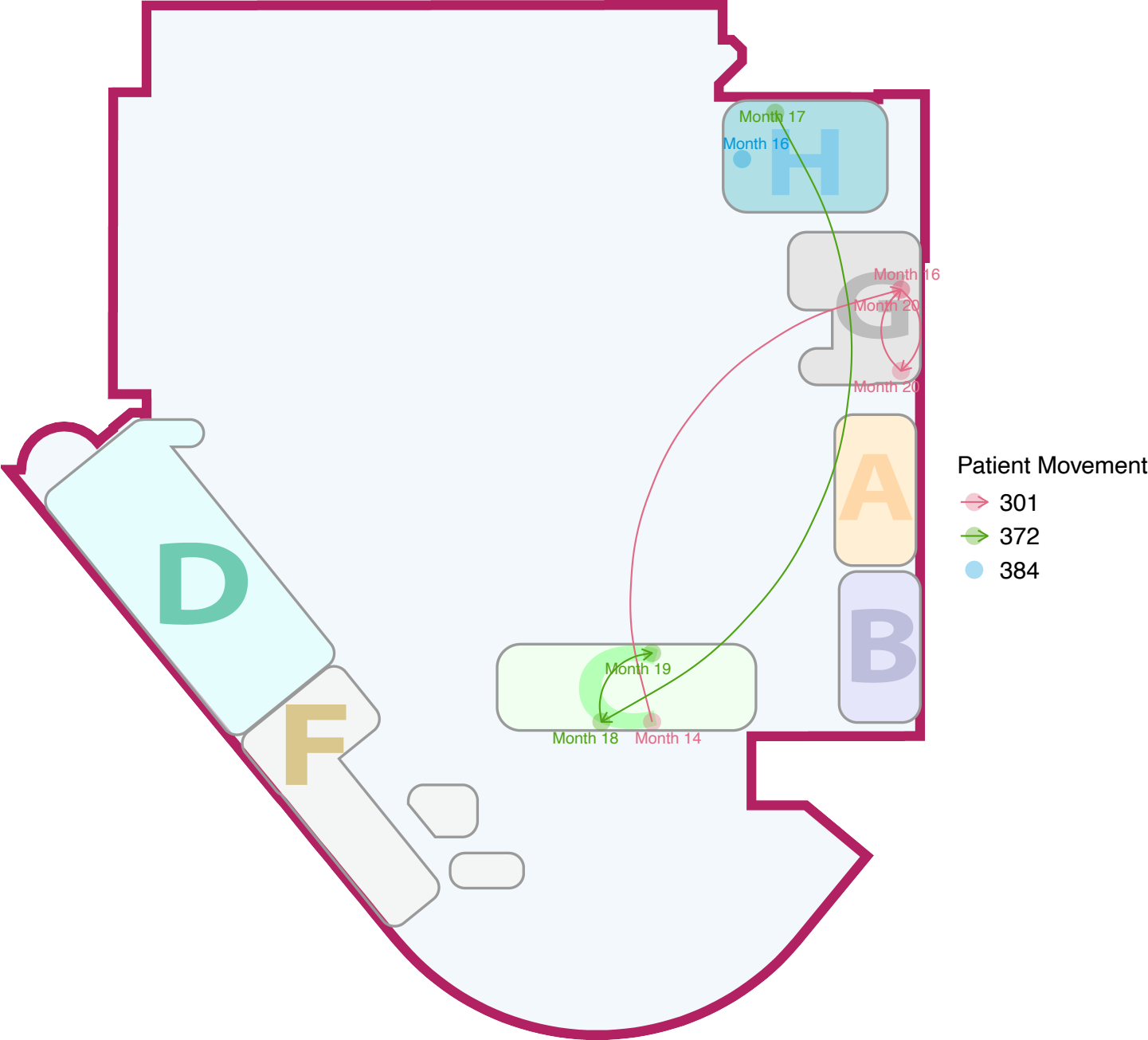
Cluster18



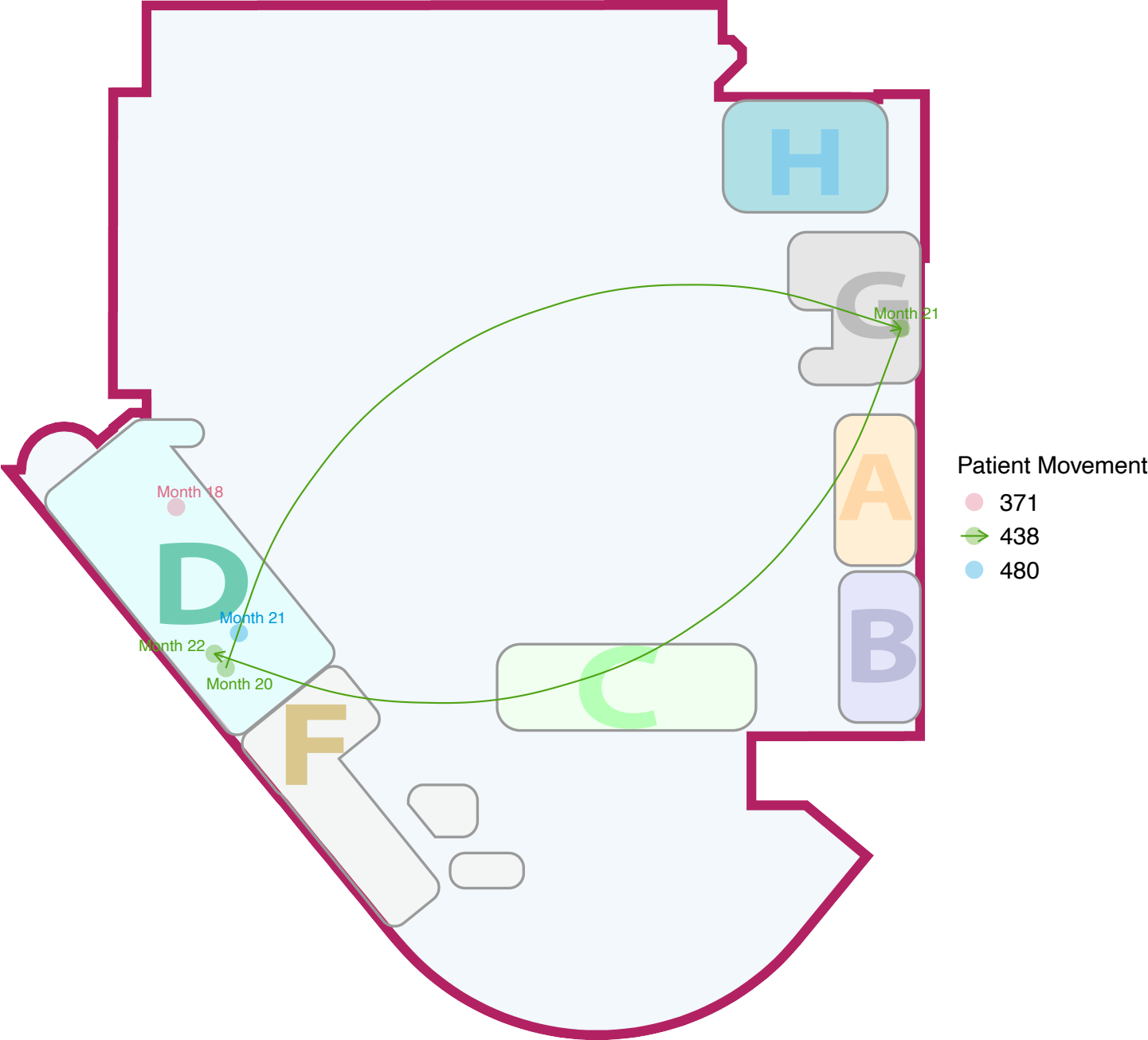
Cluster19



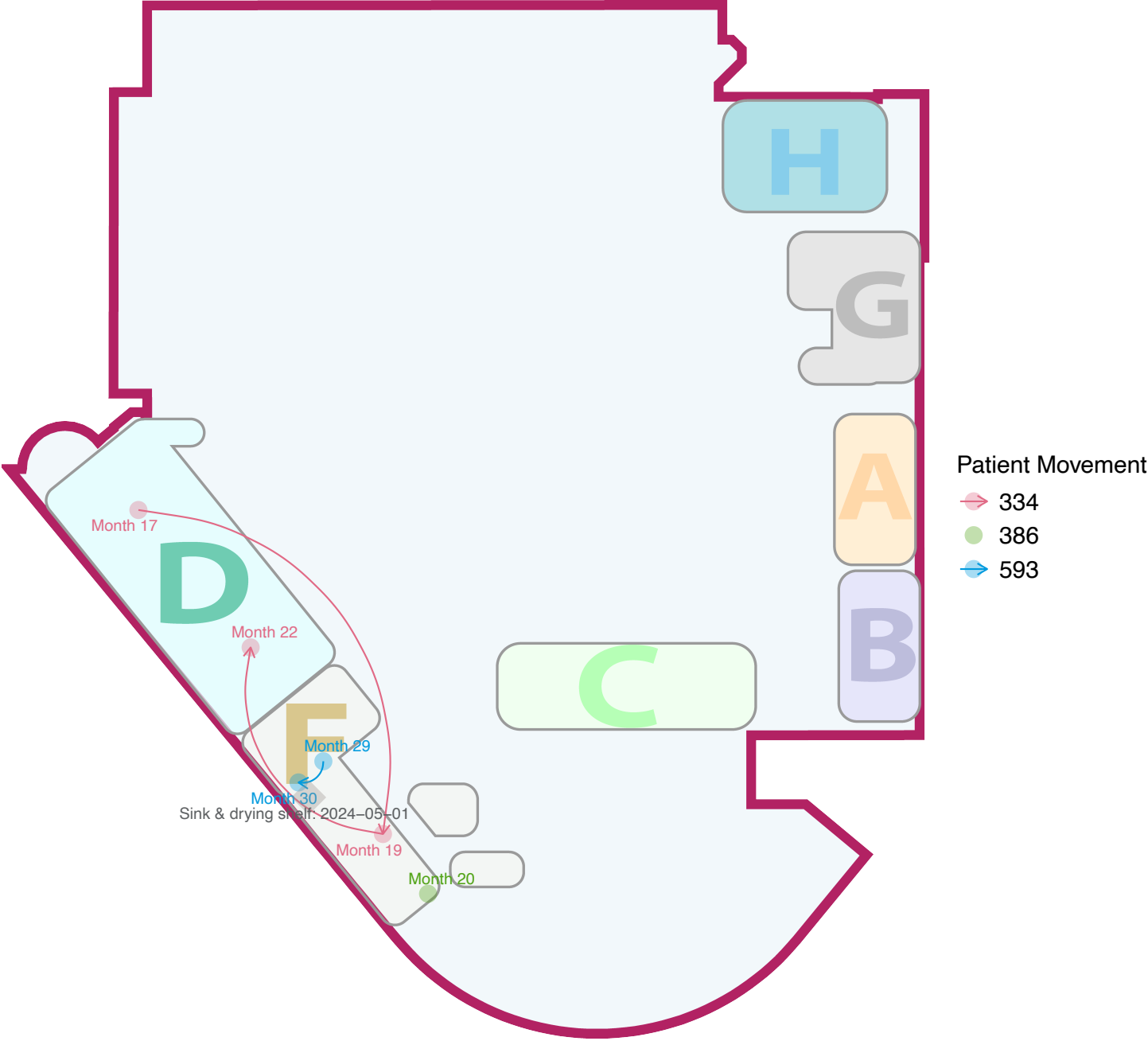
Cluster20



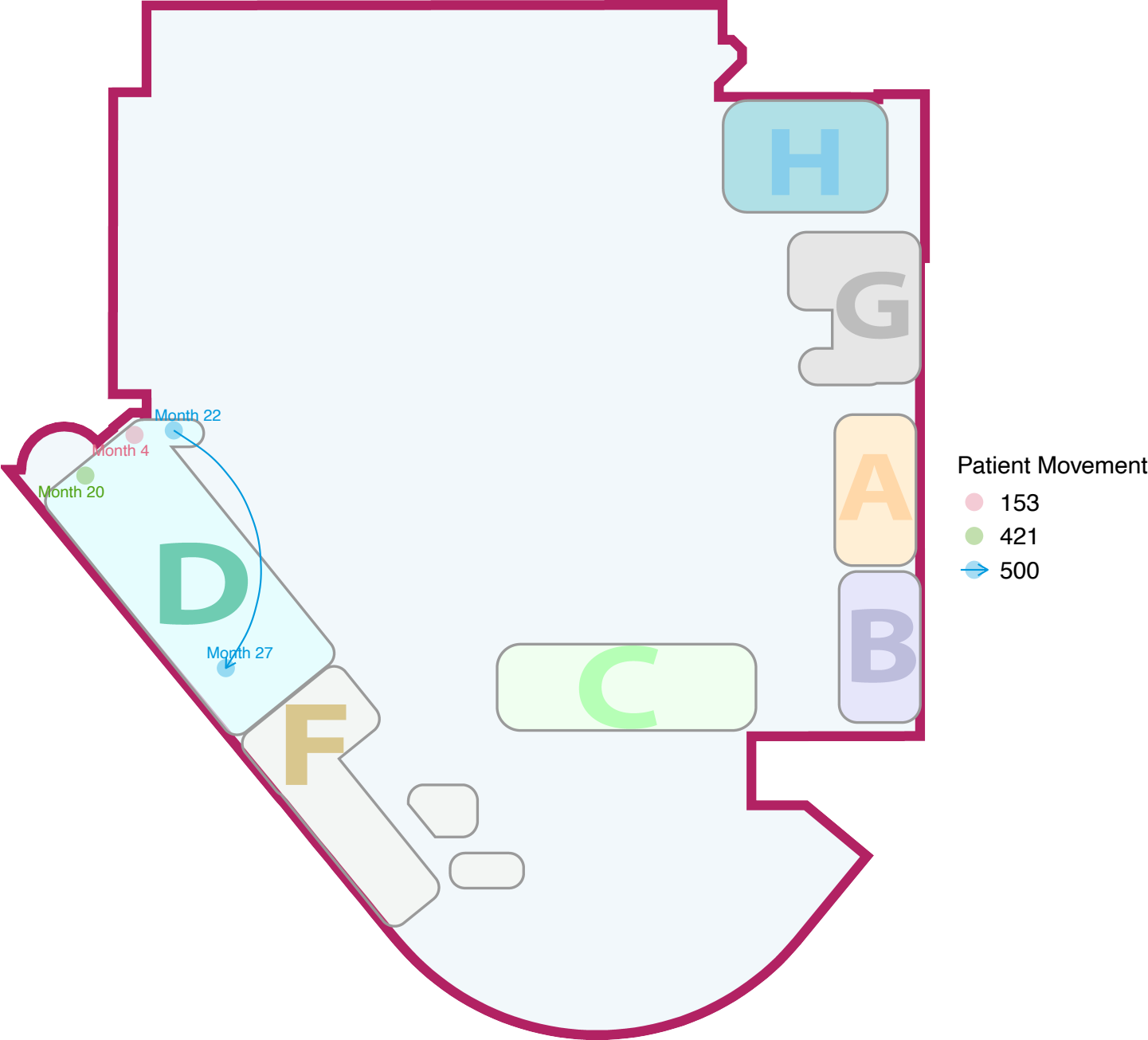
Cluster21



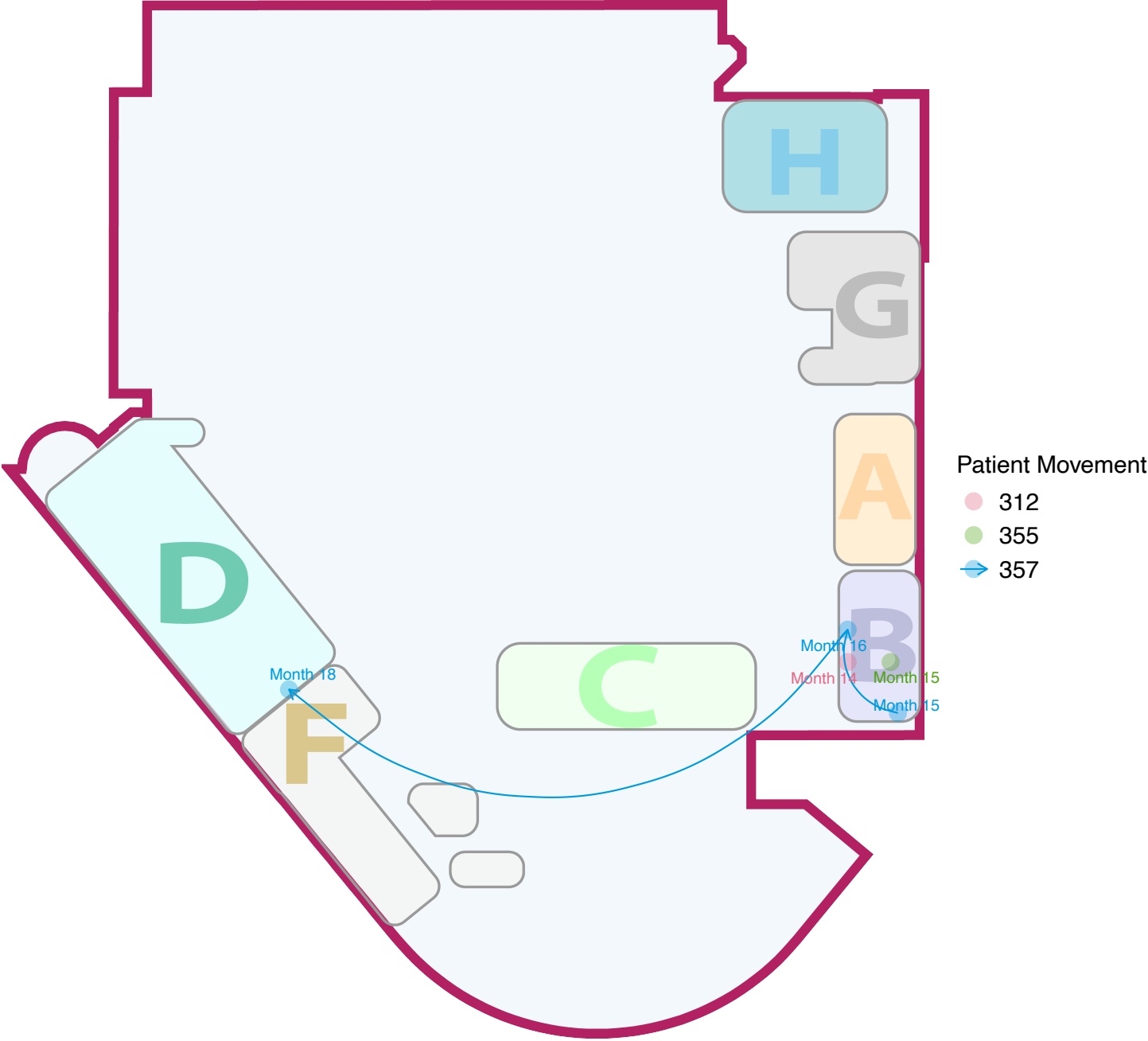
Cluster22



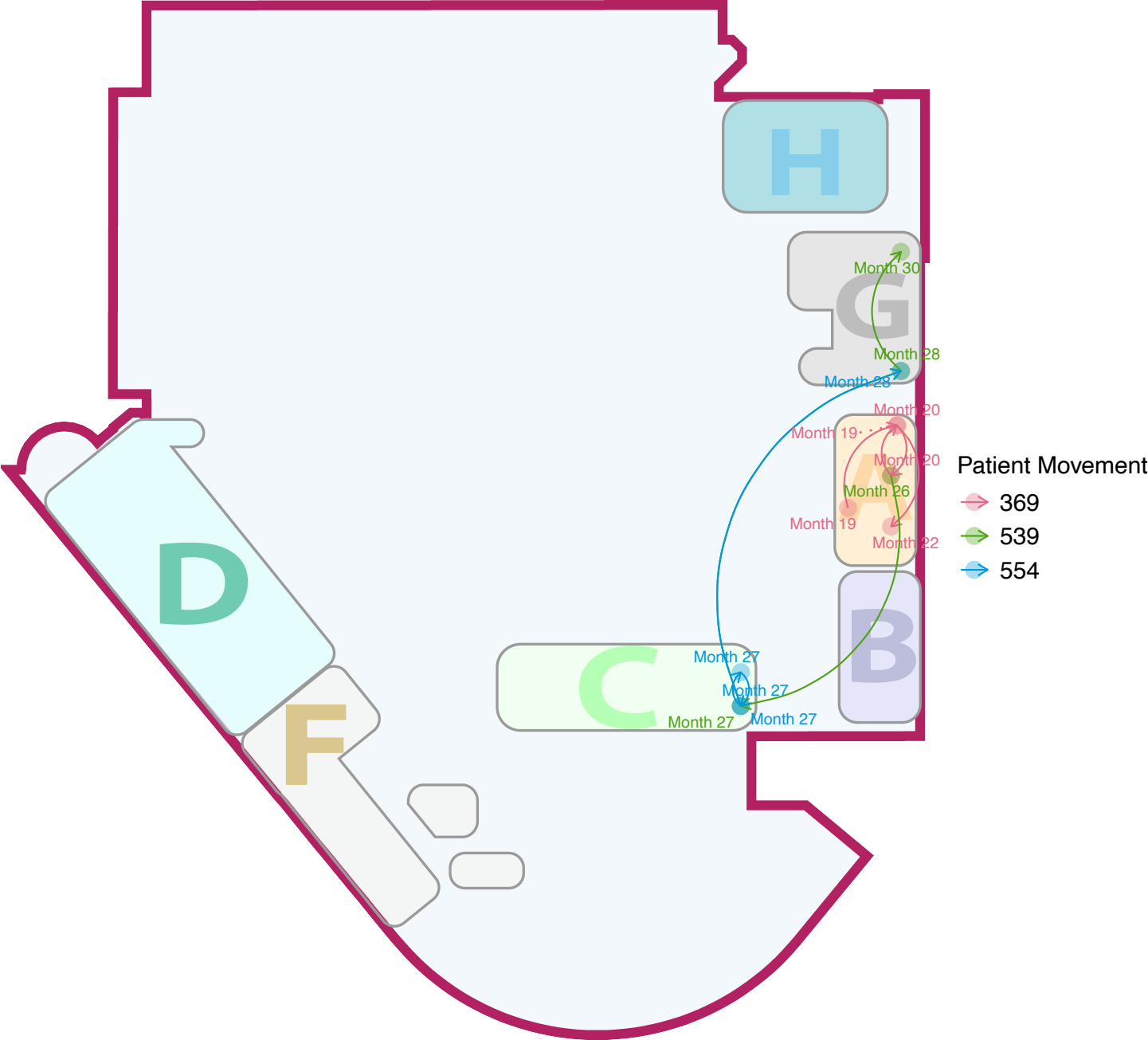
Cluster23



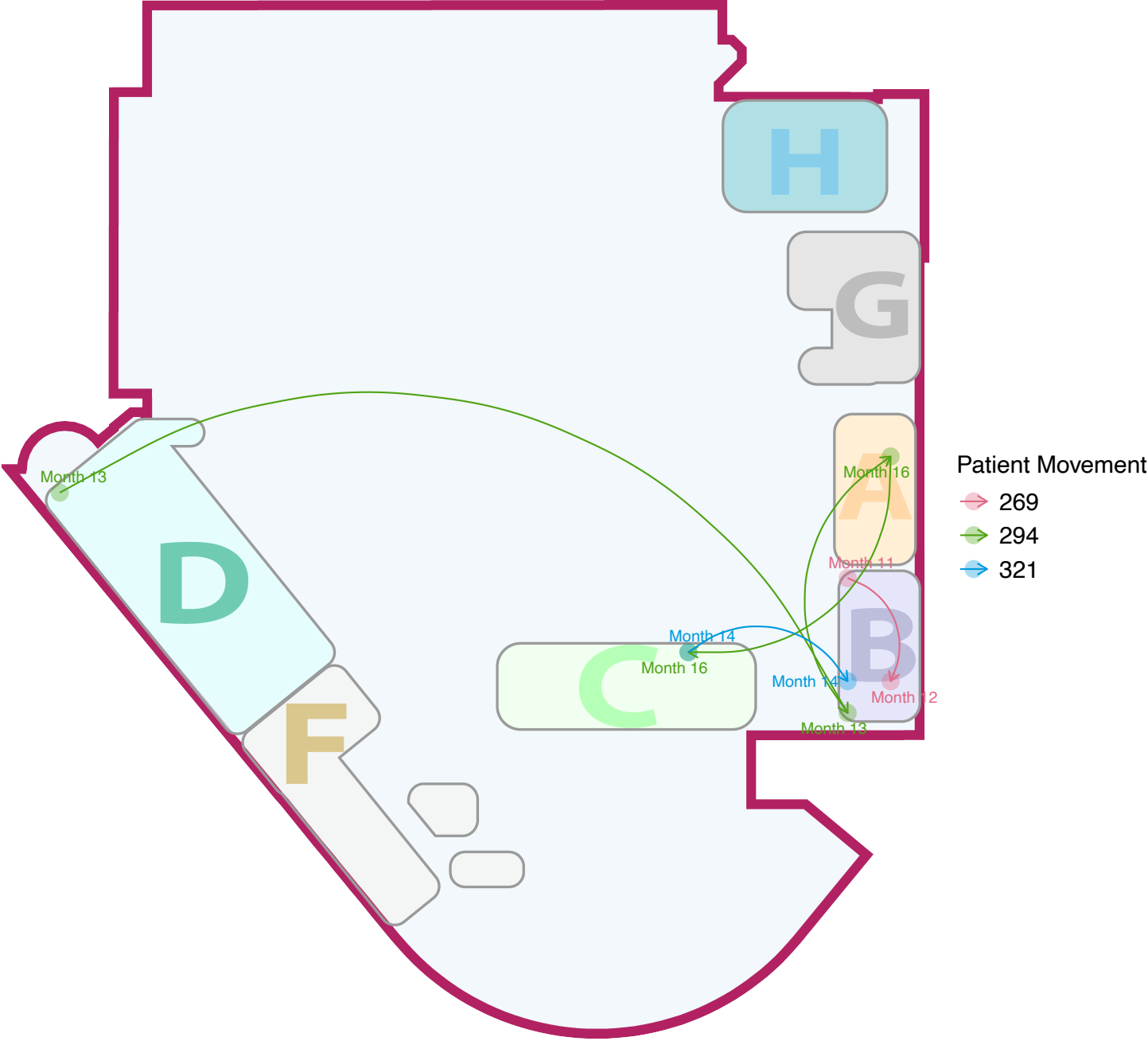
Cluster24



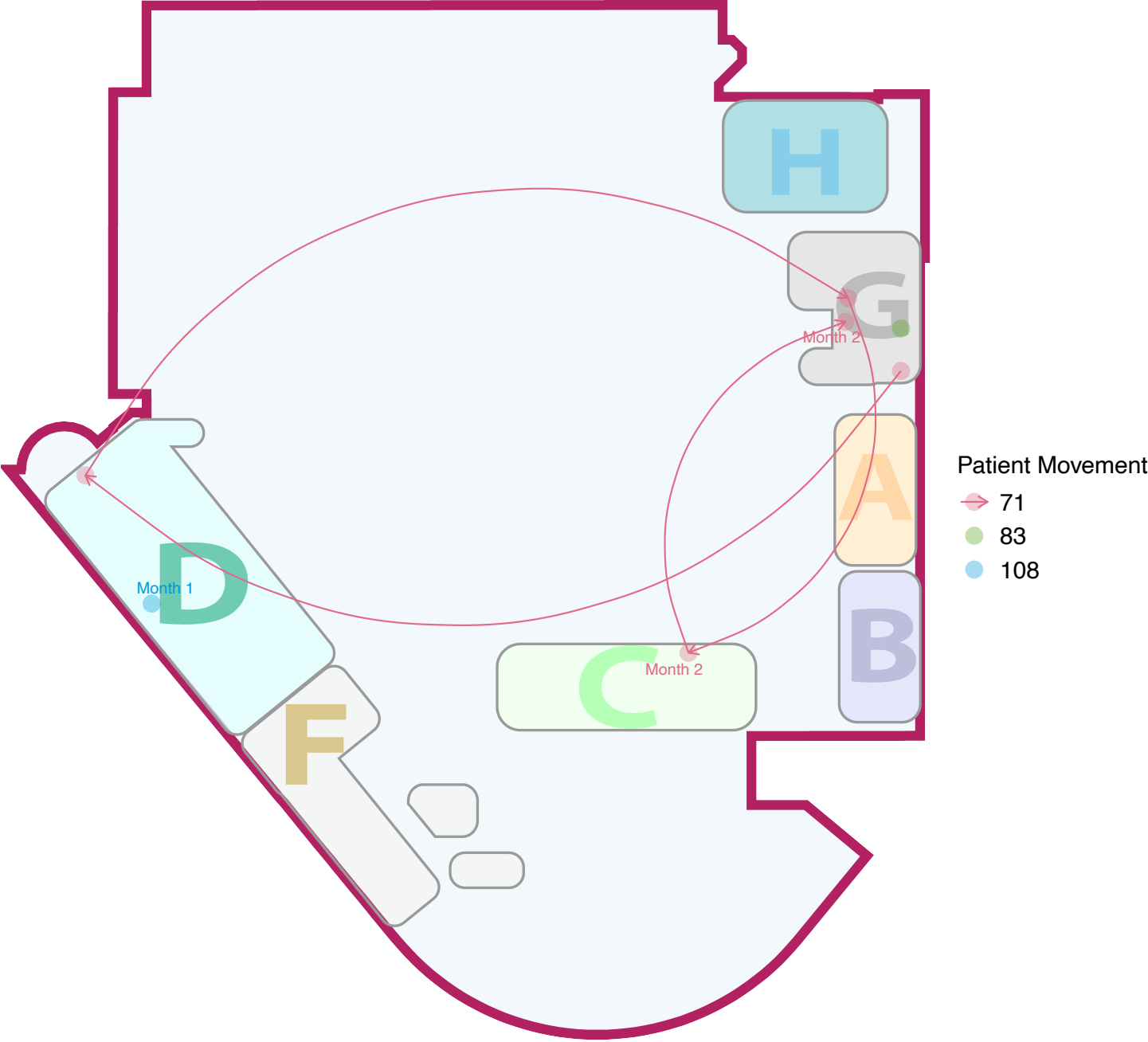
Cluster25



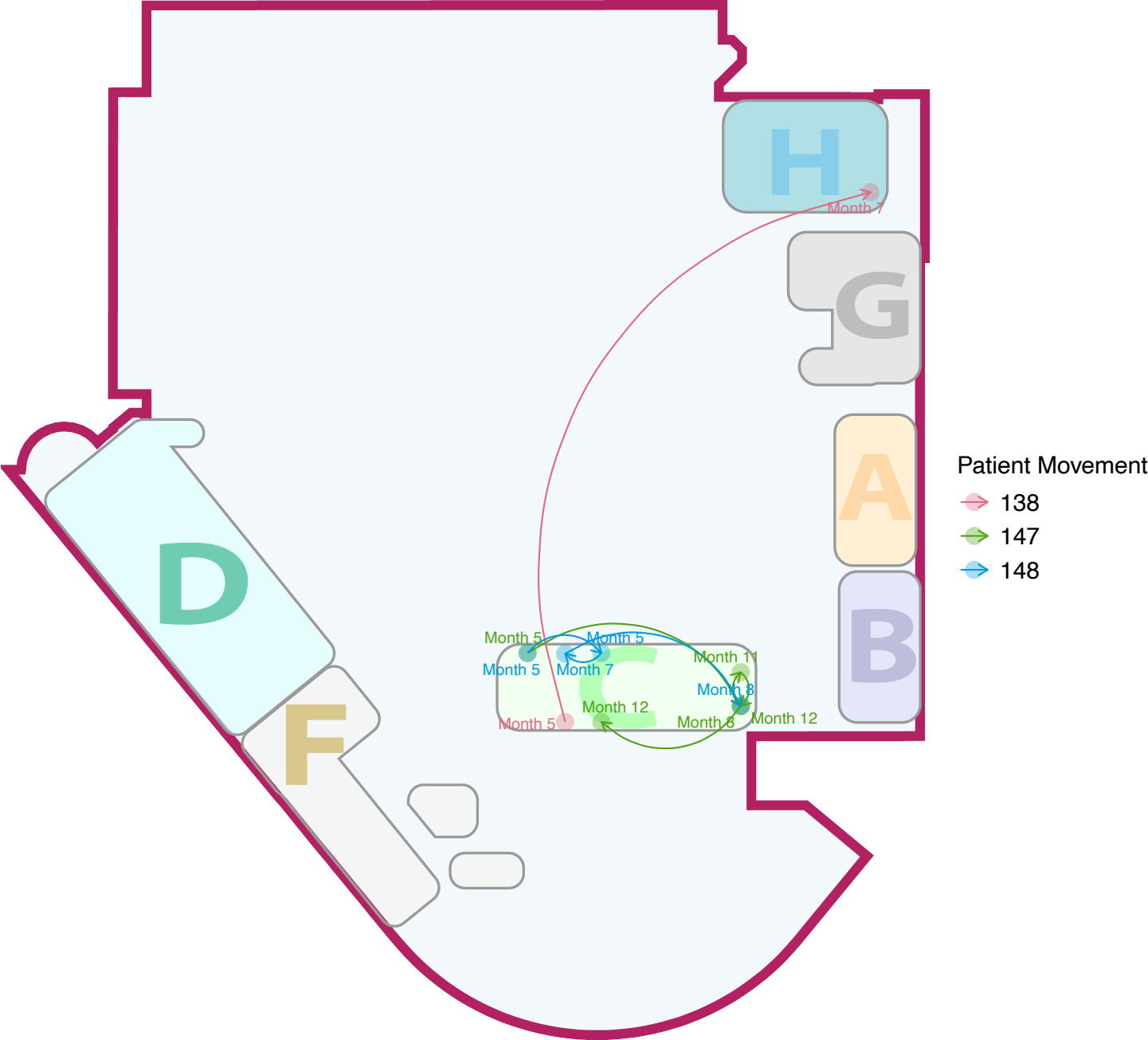
Cluster26



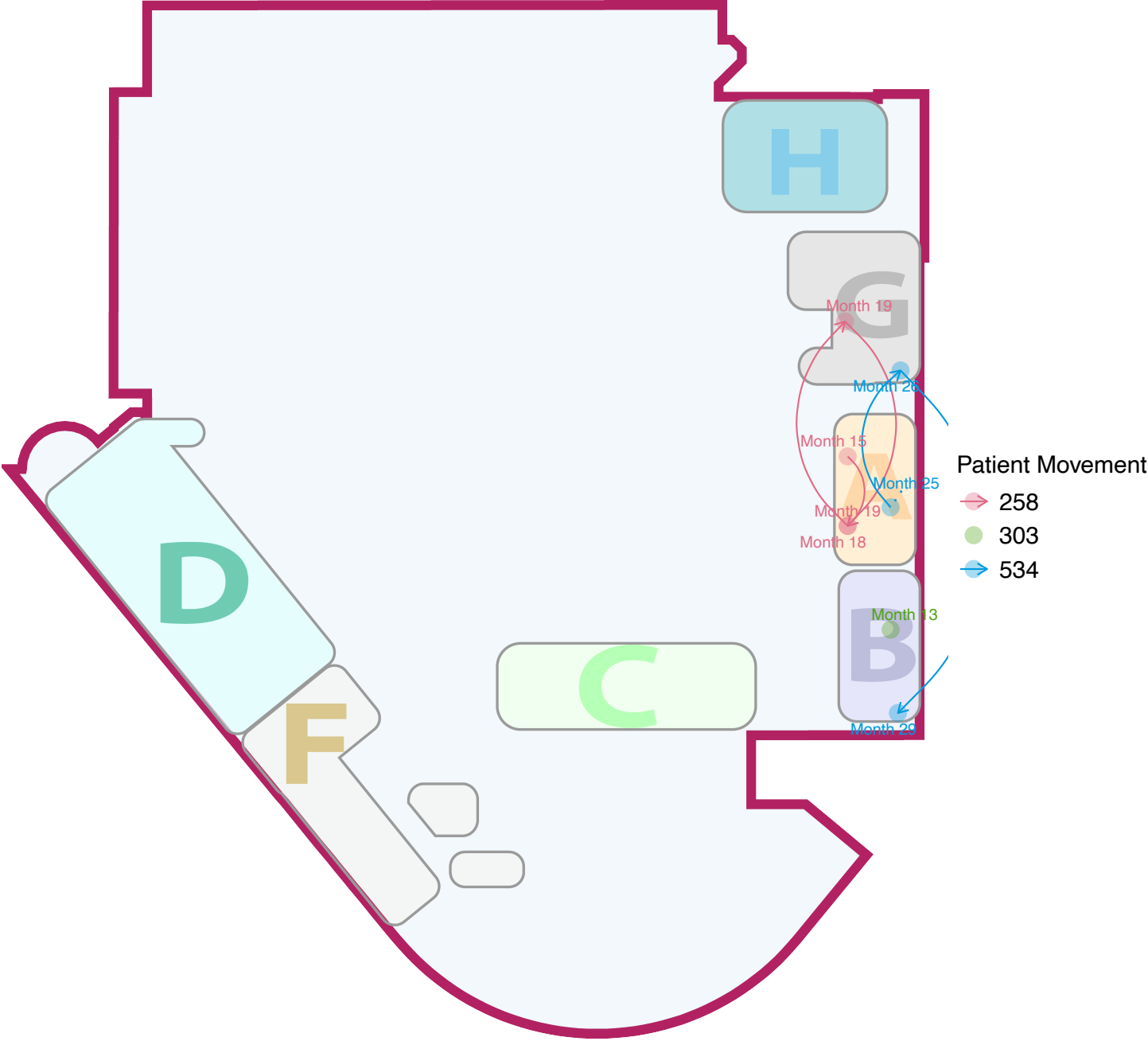
Cluster27



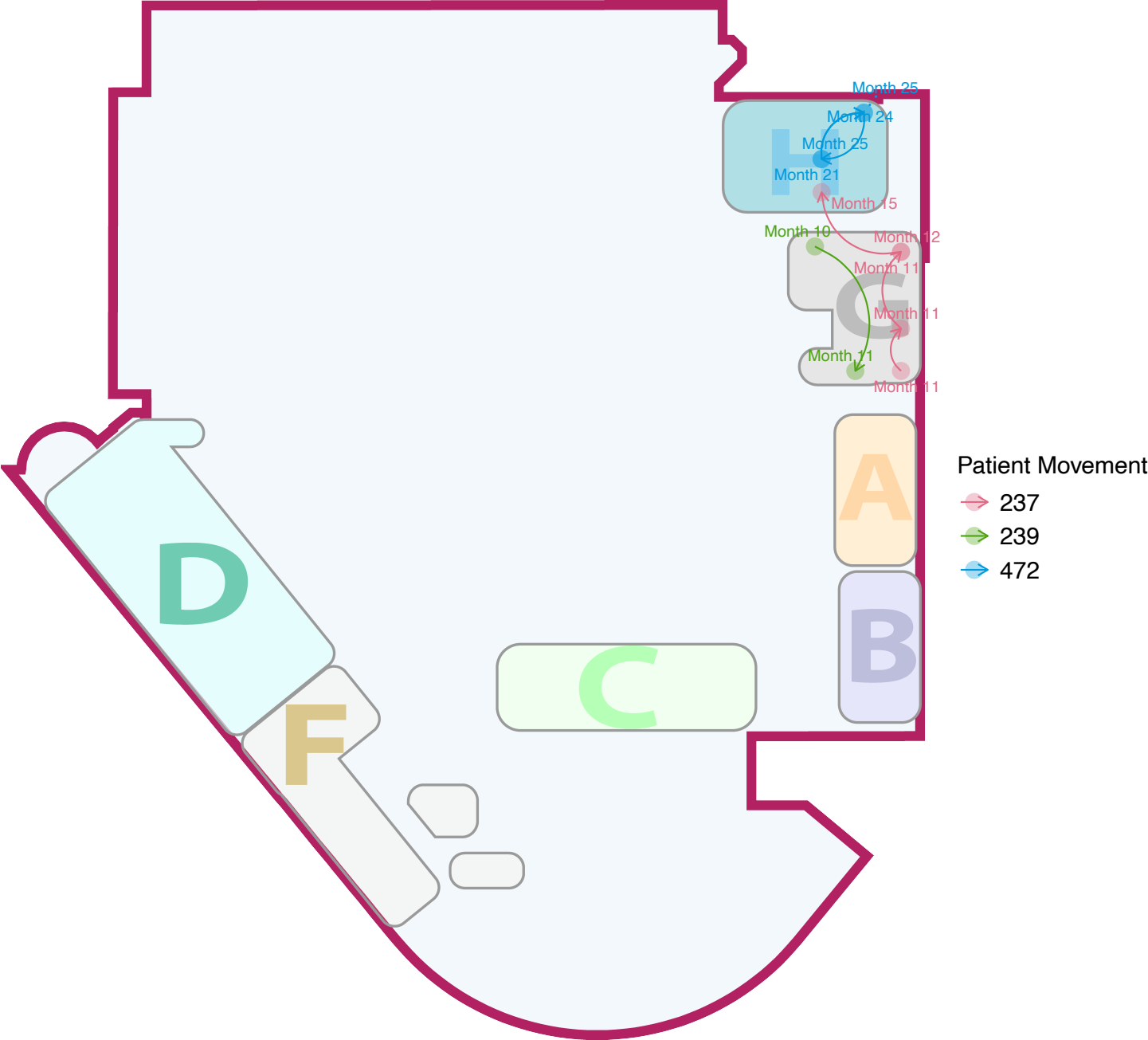
Cluster28



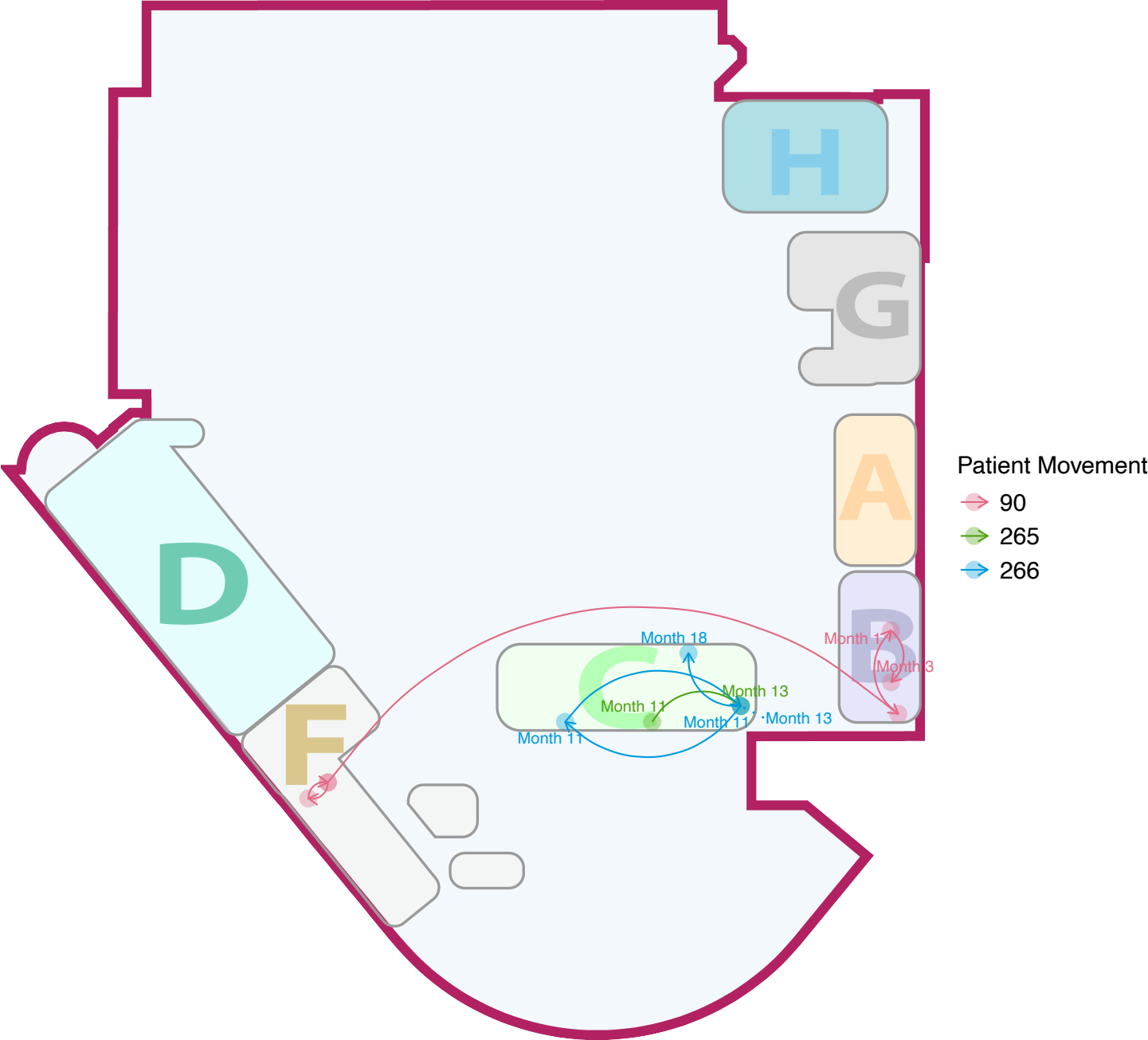
Cluster29



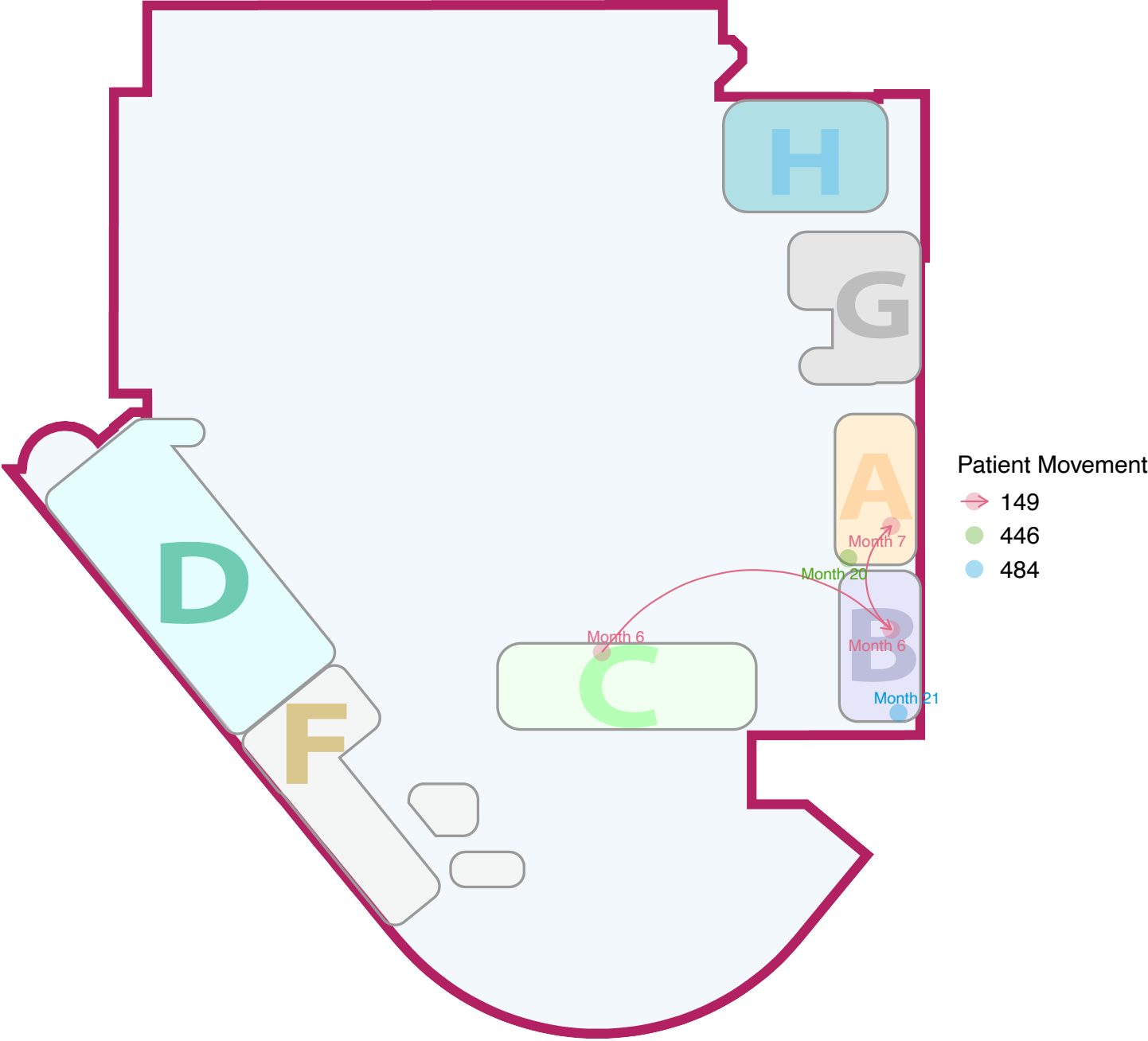
Cluster30



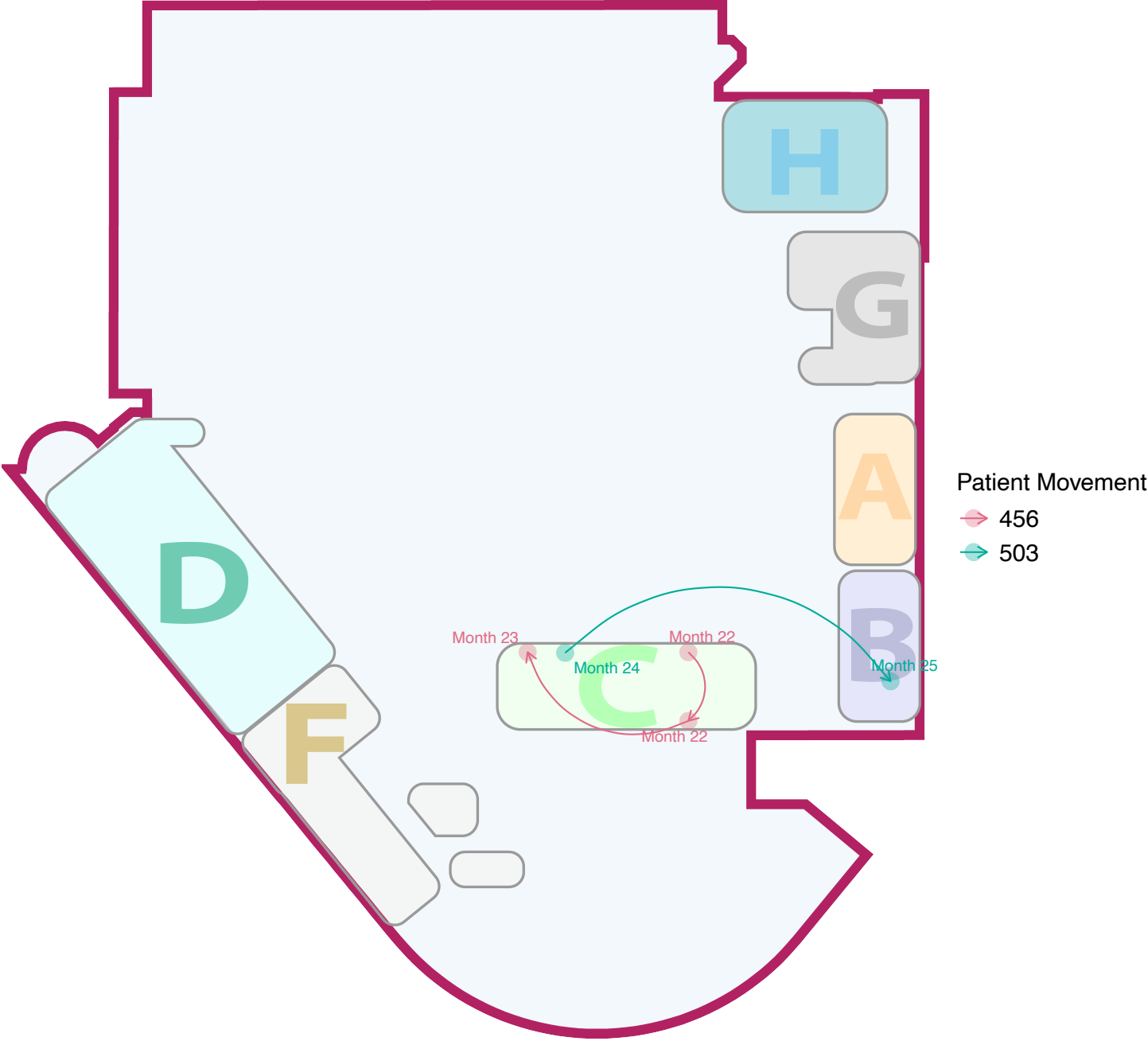
Cluster31



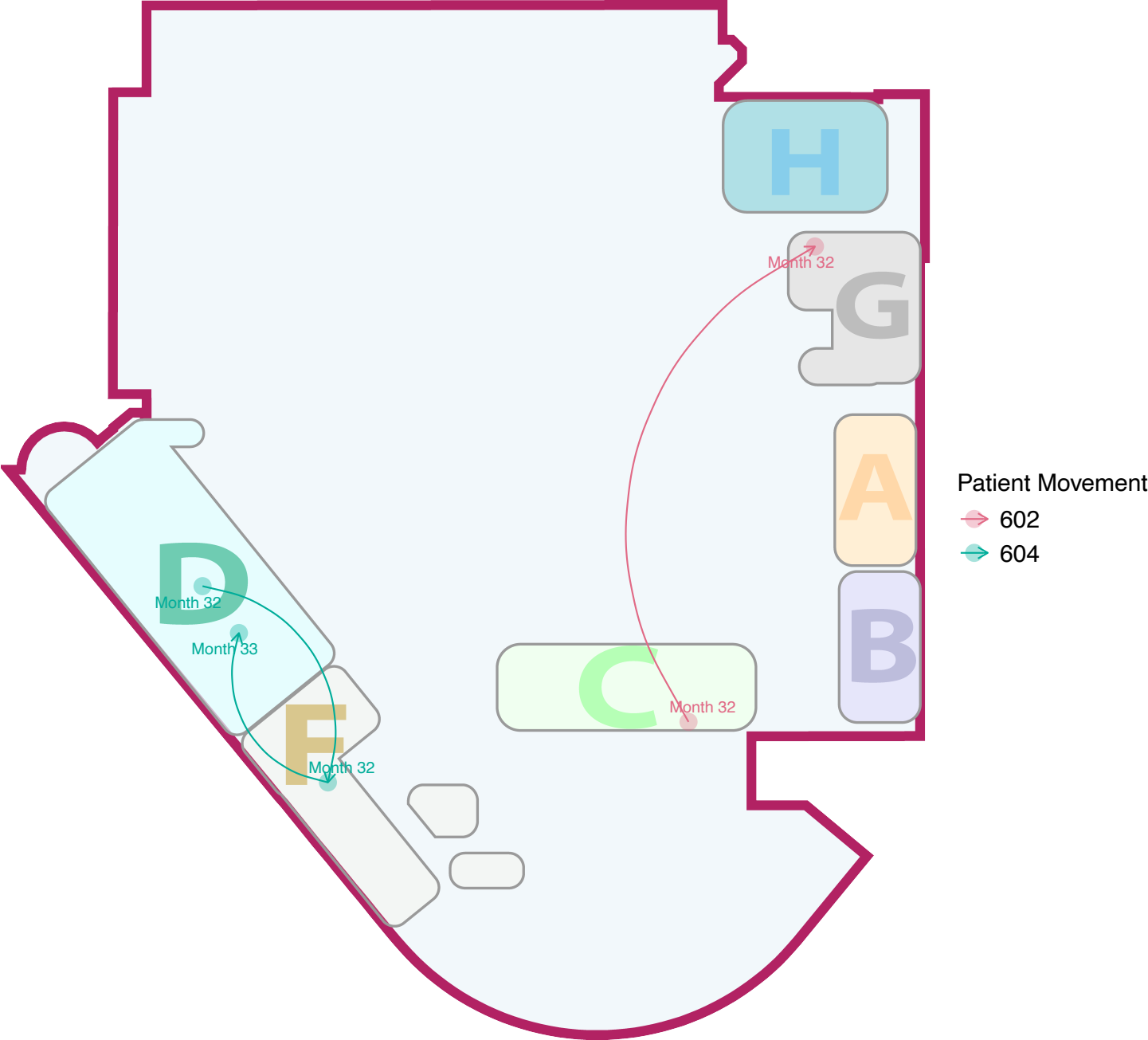
Cluster32



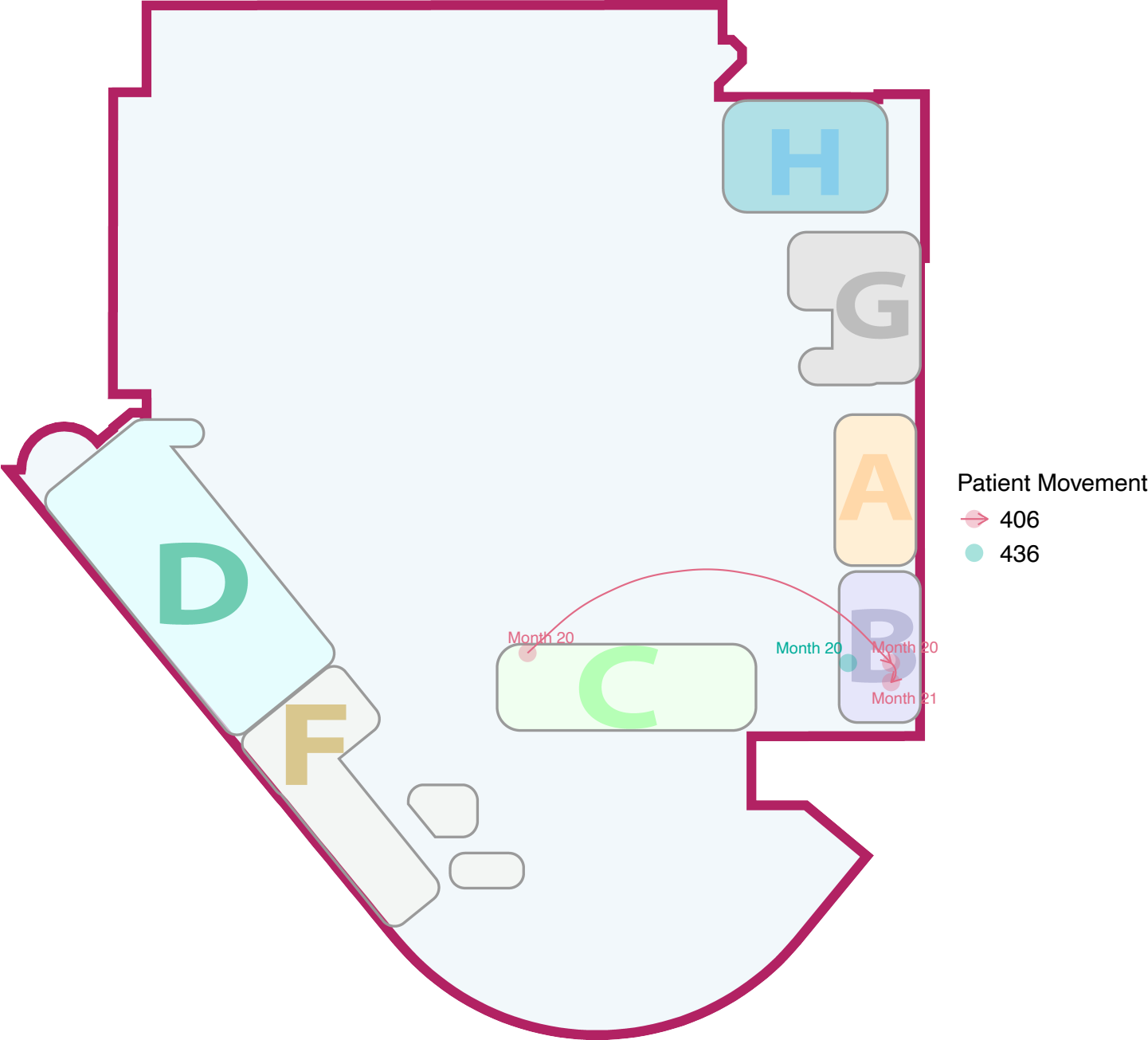
Cluster33



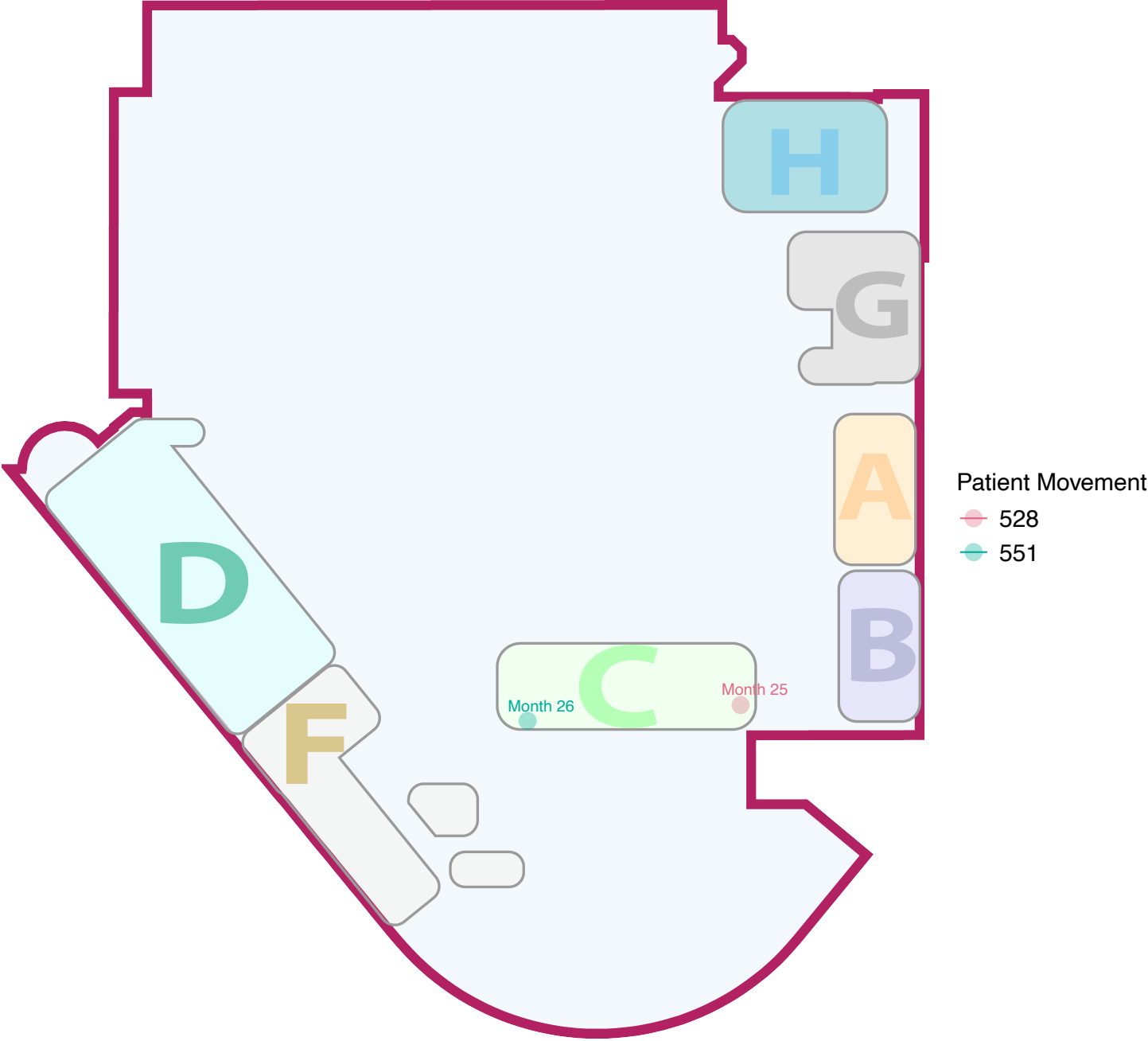
Cluster34



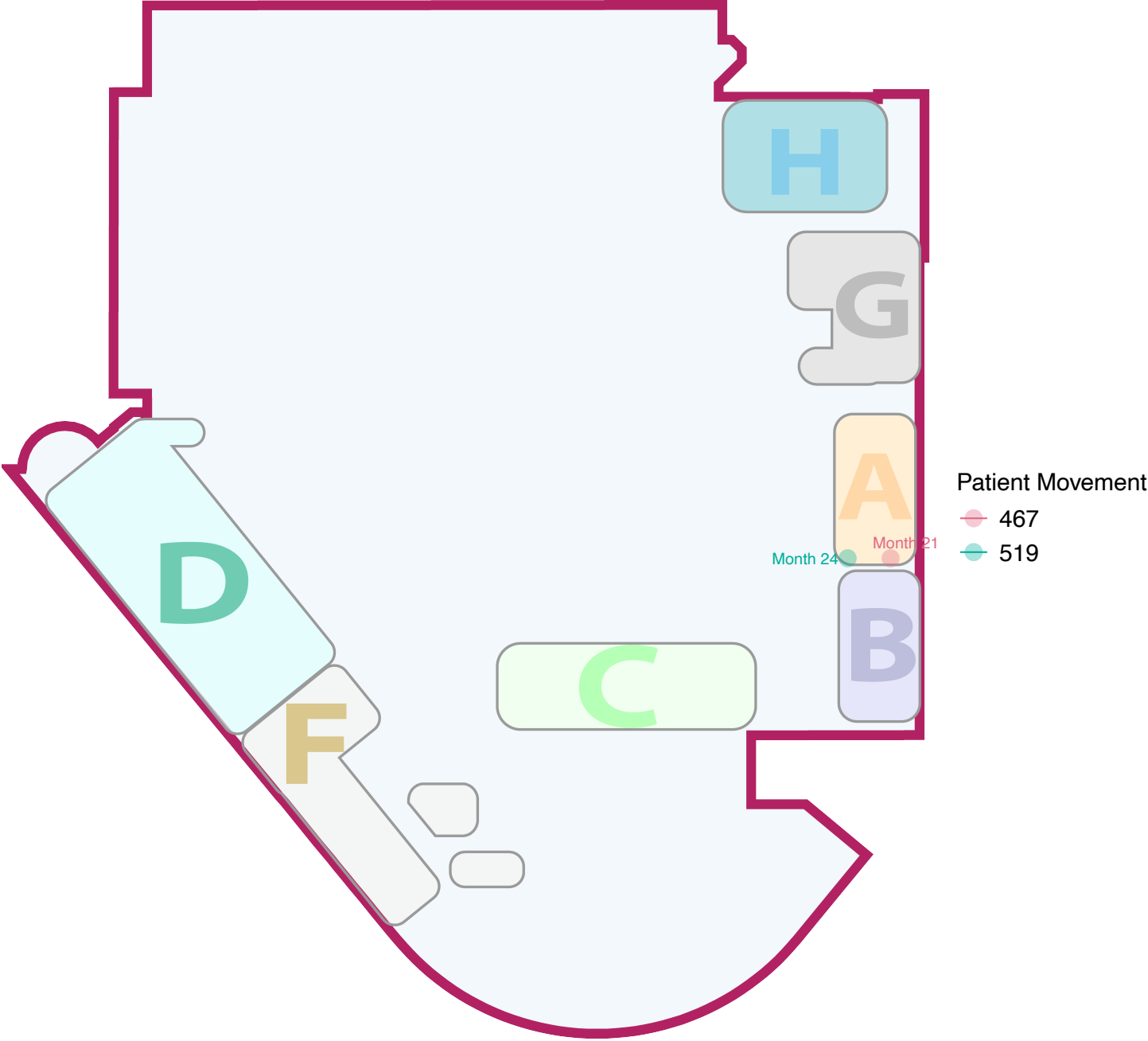
Cluster35



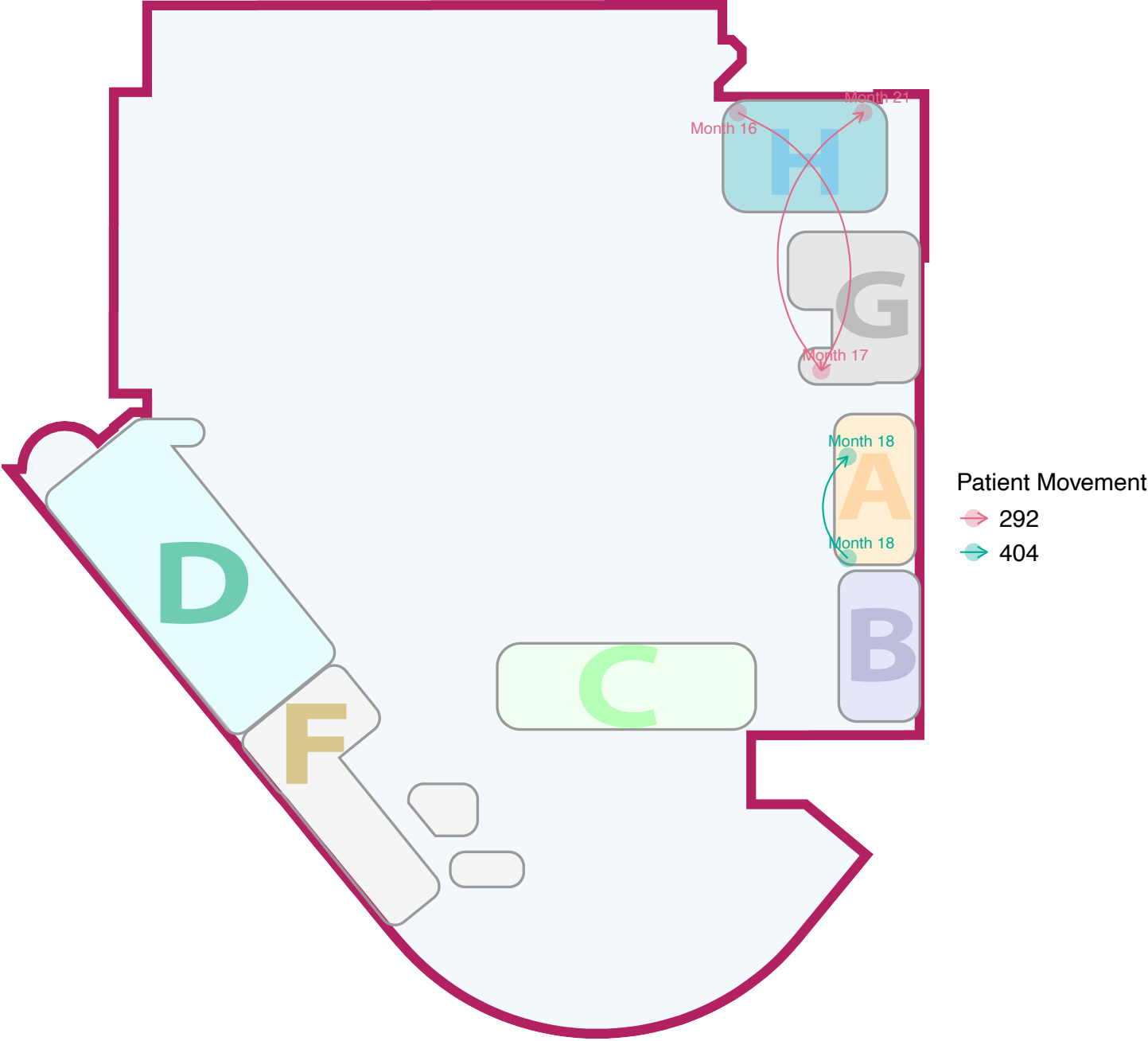
Cluster36



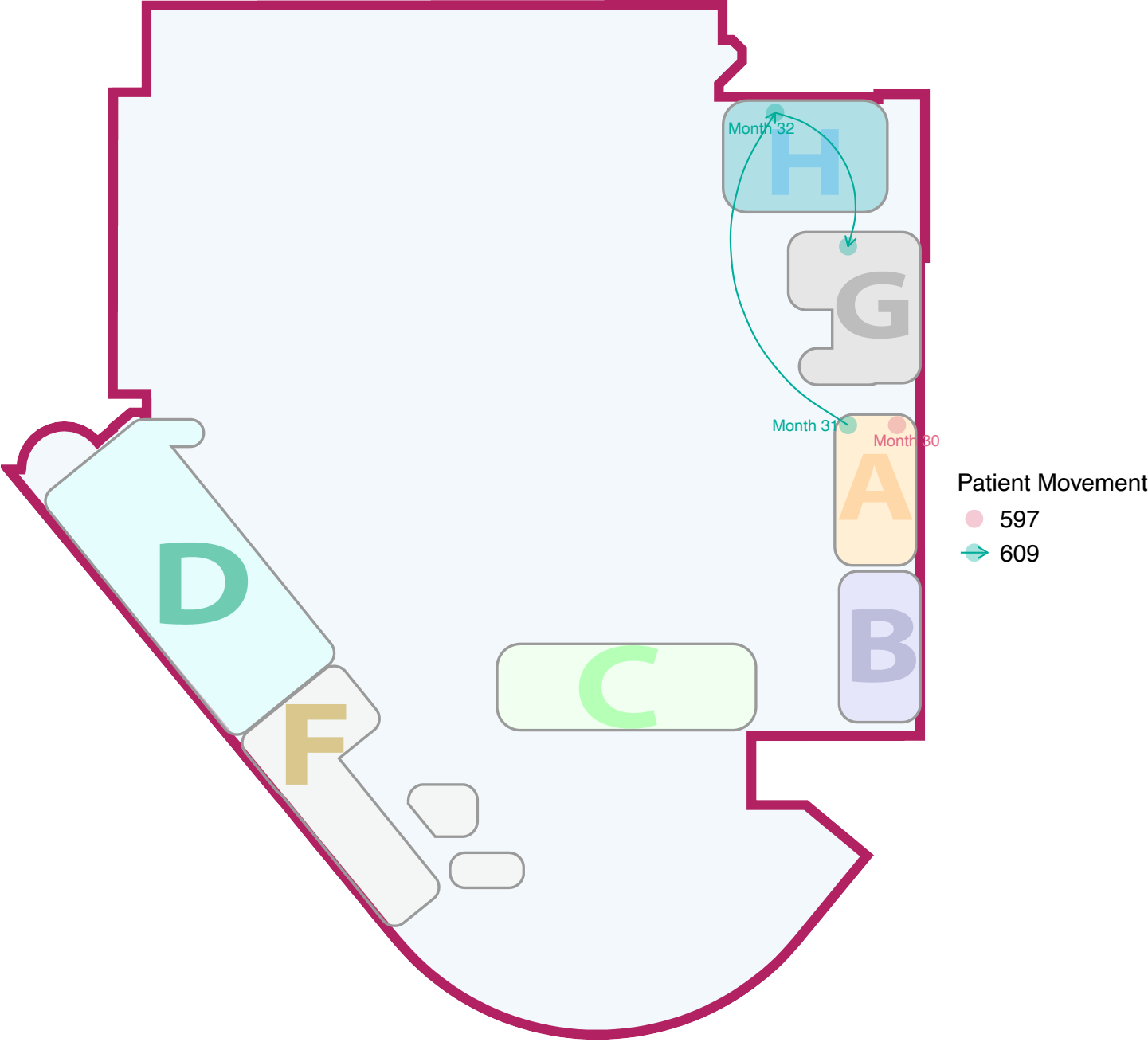
Cluster37



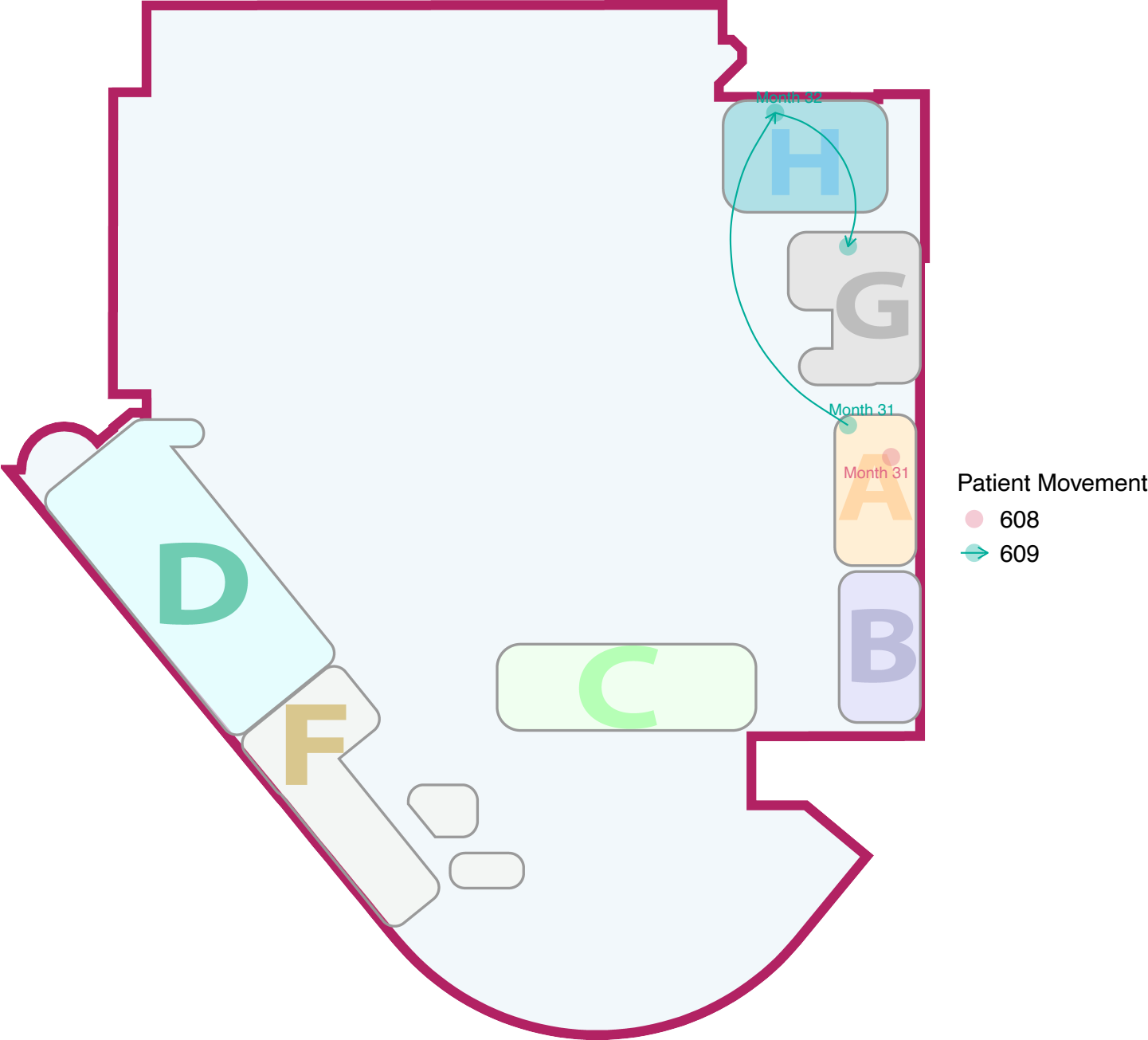
Cluster38



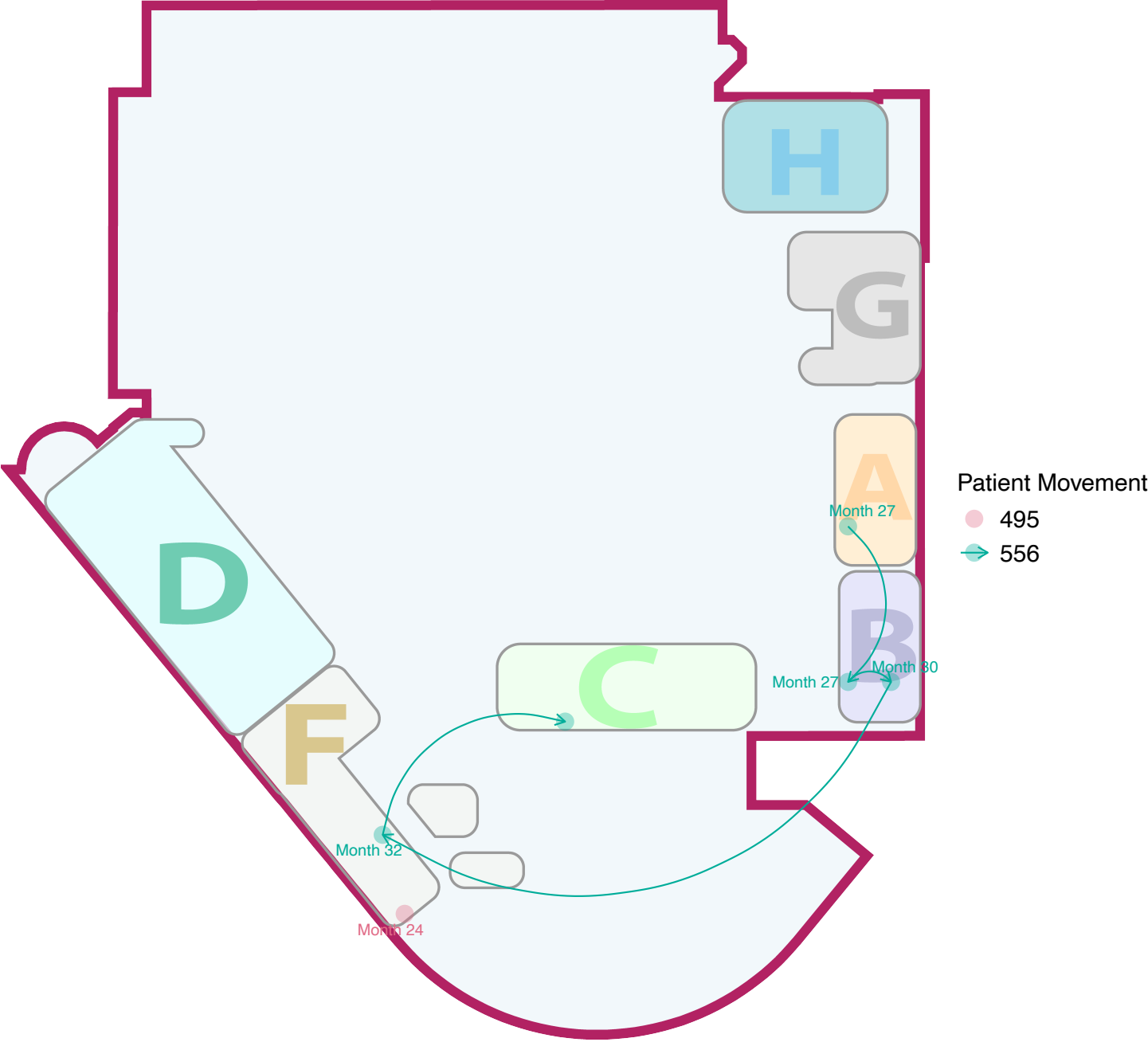
Cluster39



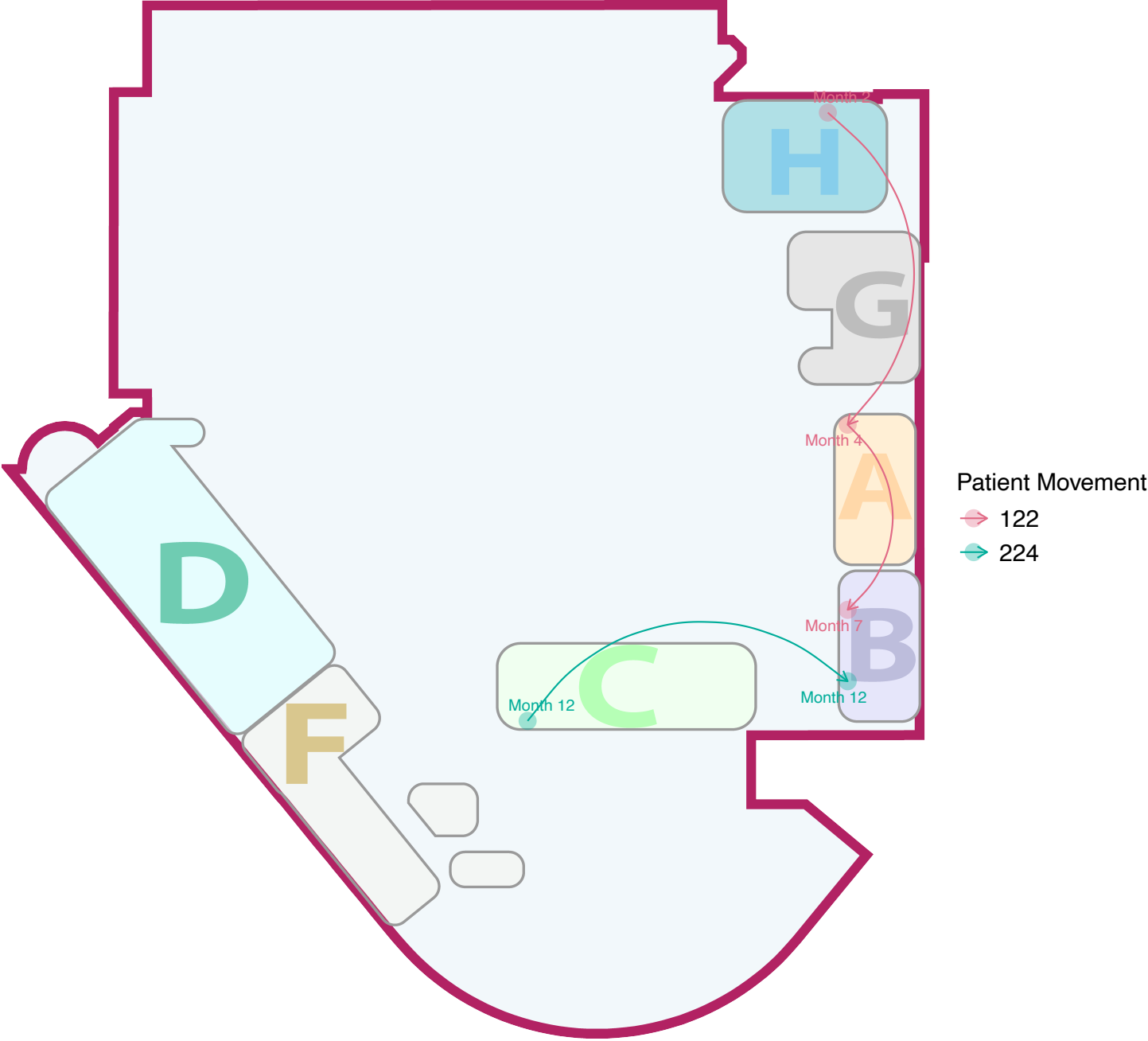
Cluster40



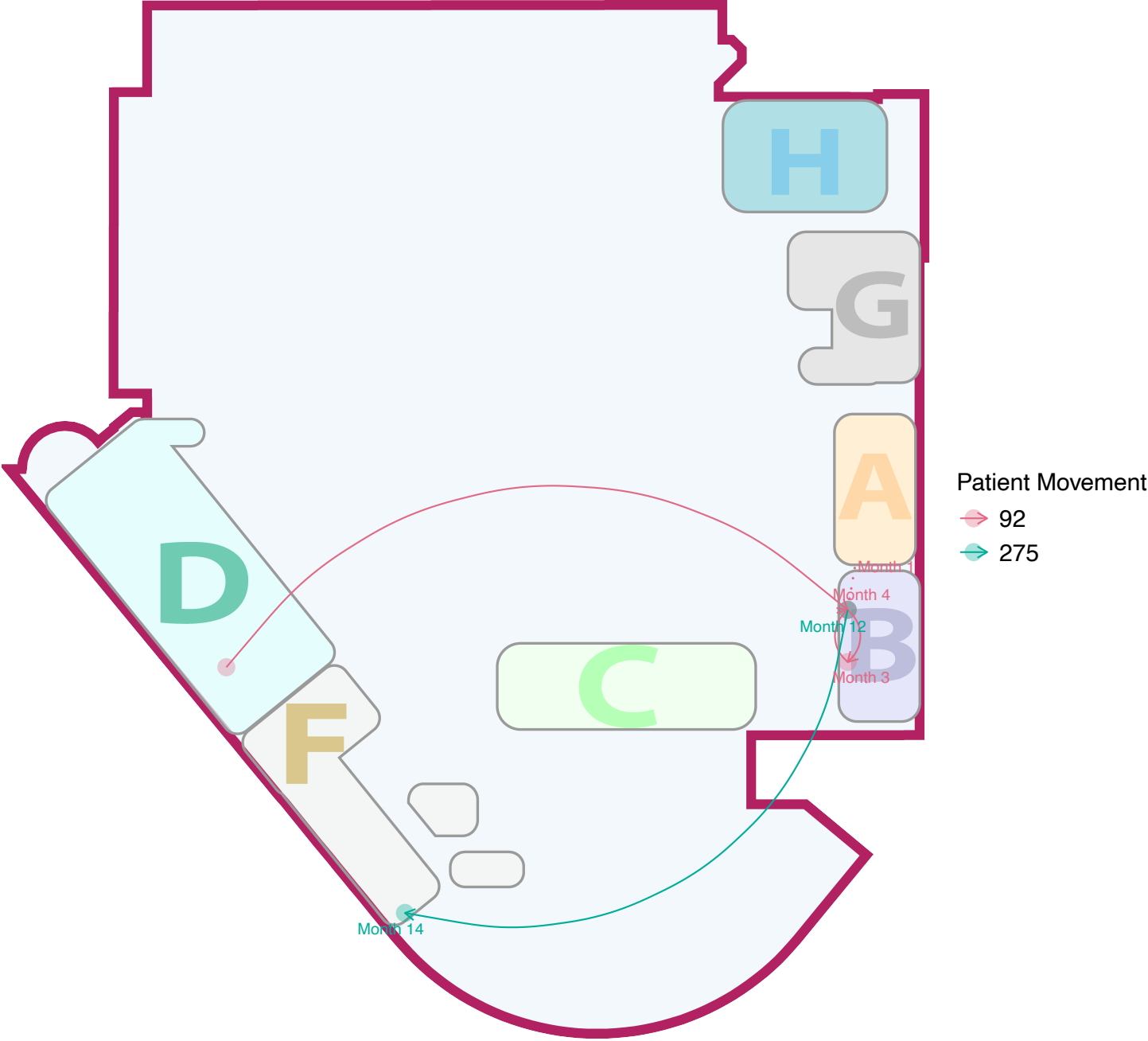
Cluster41



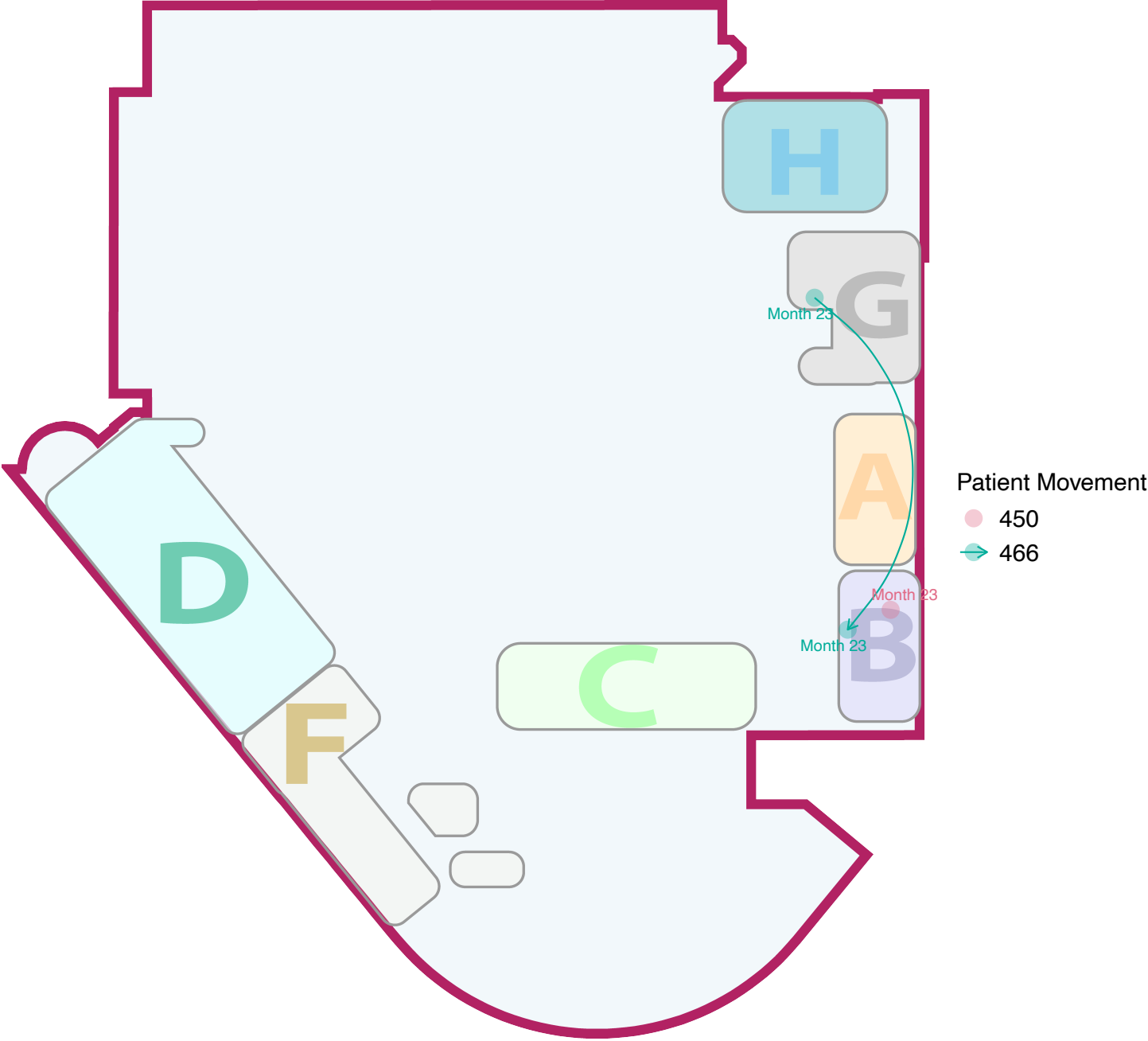
Cluster42



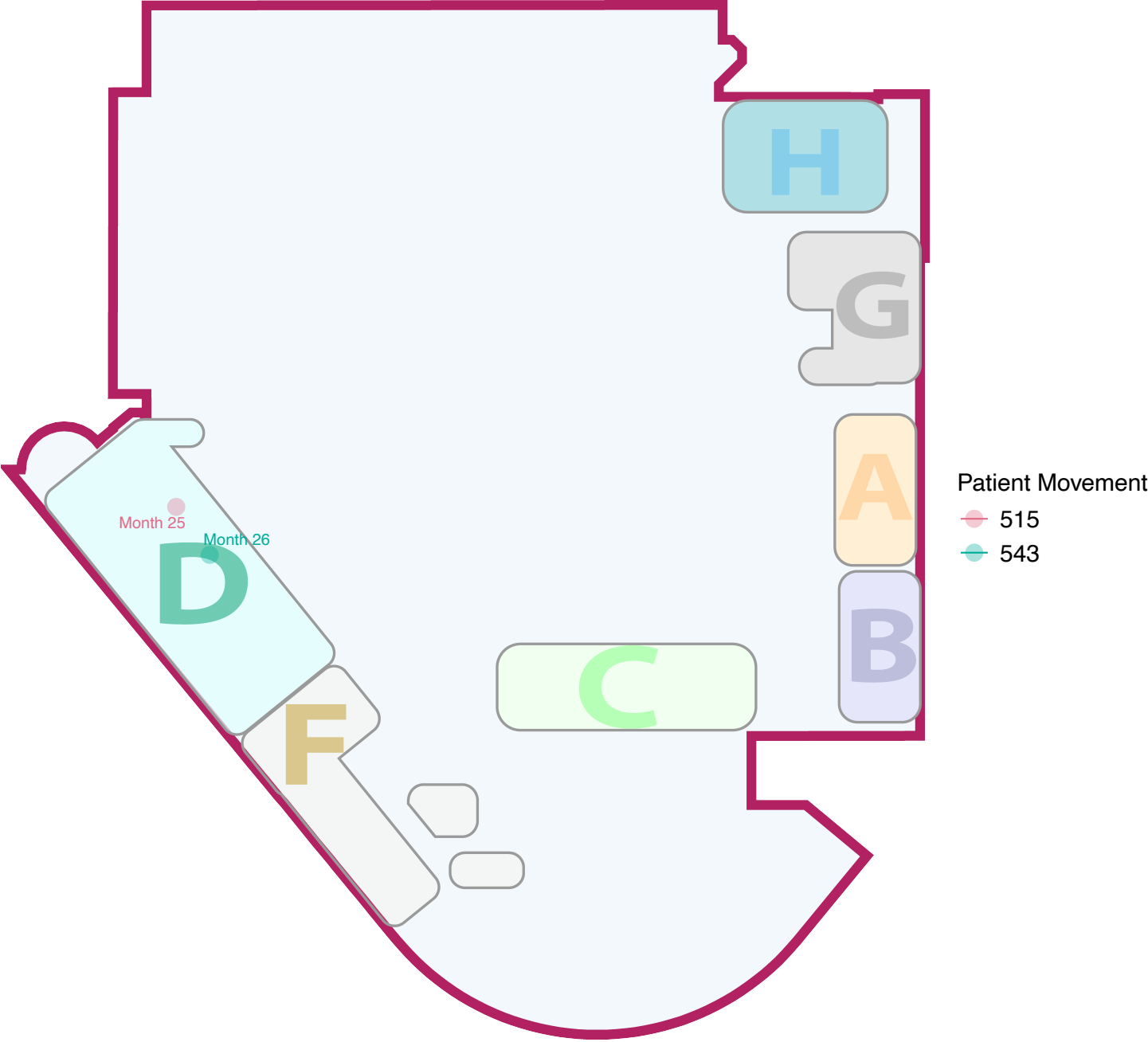
Cluster43



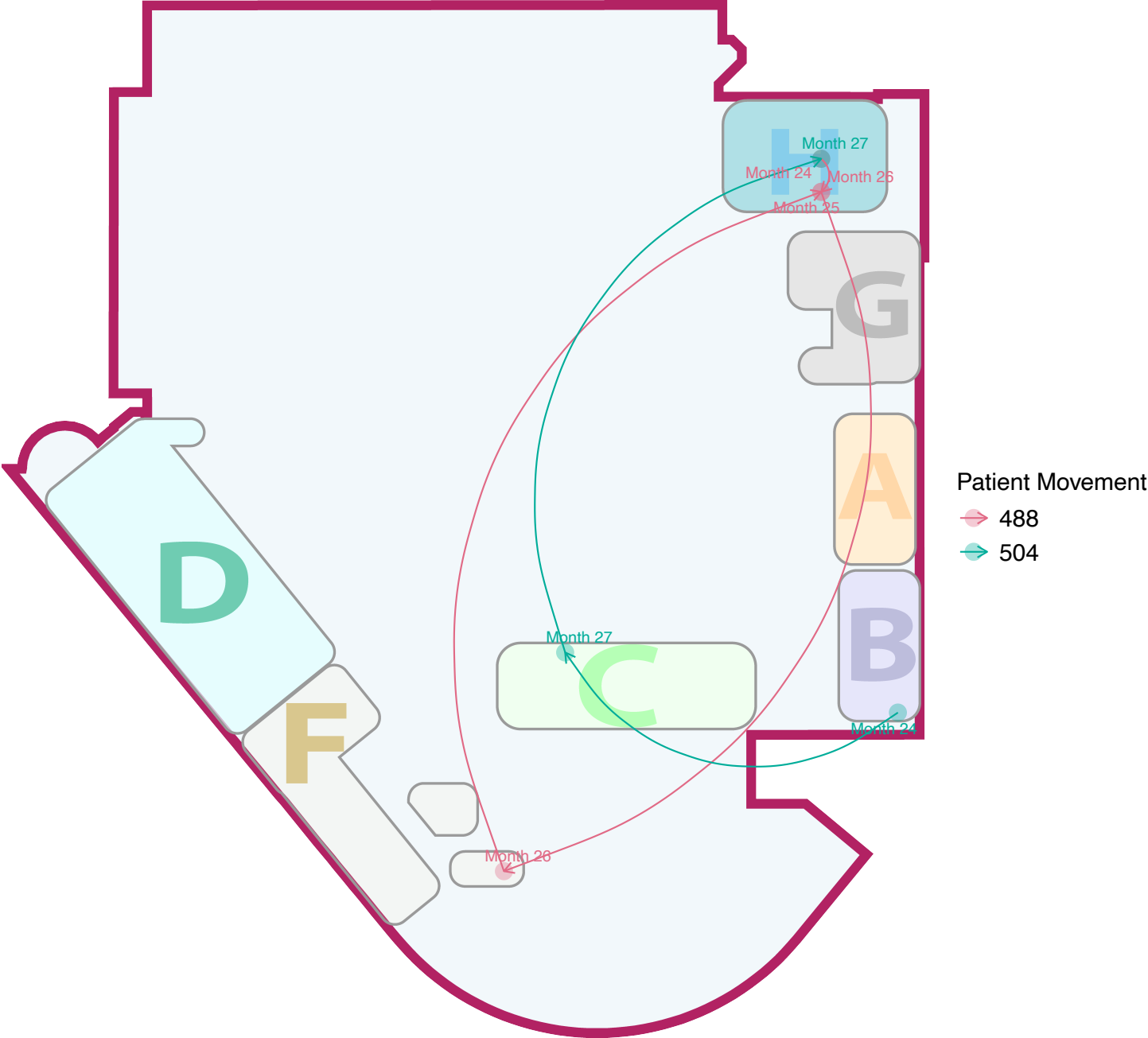
Cluster44



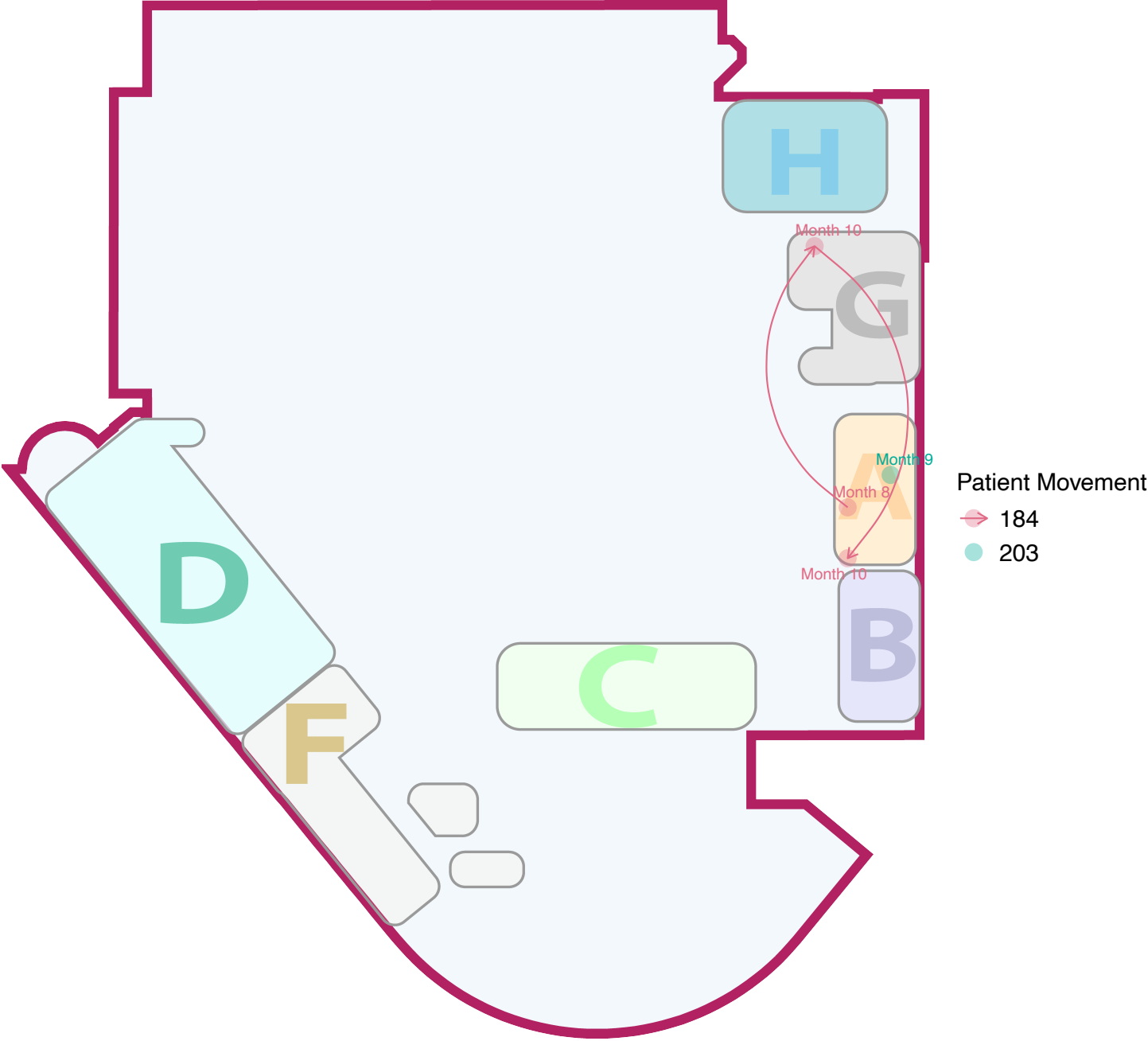
Cluster45



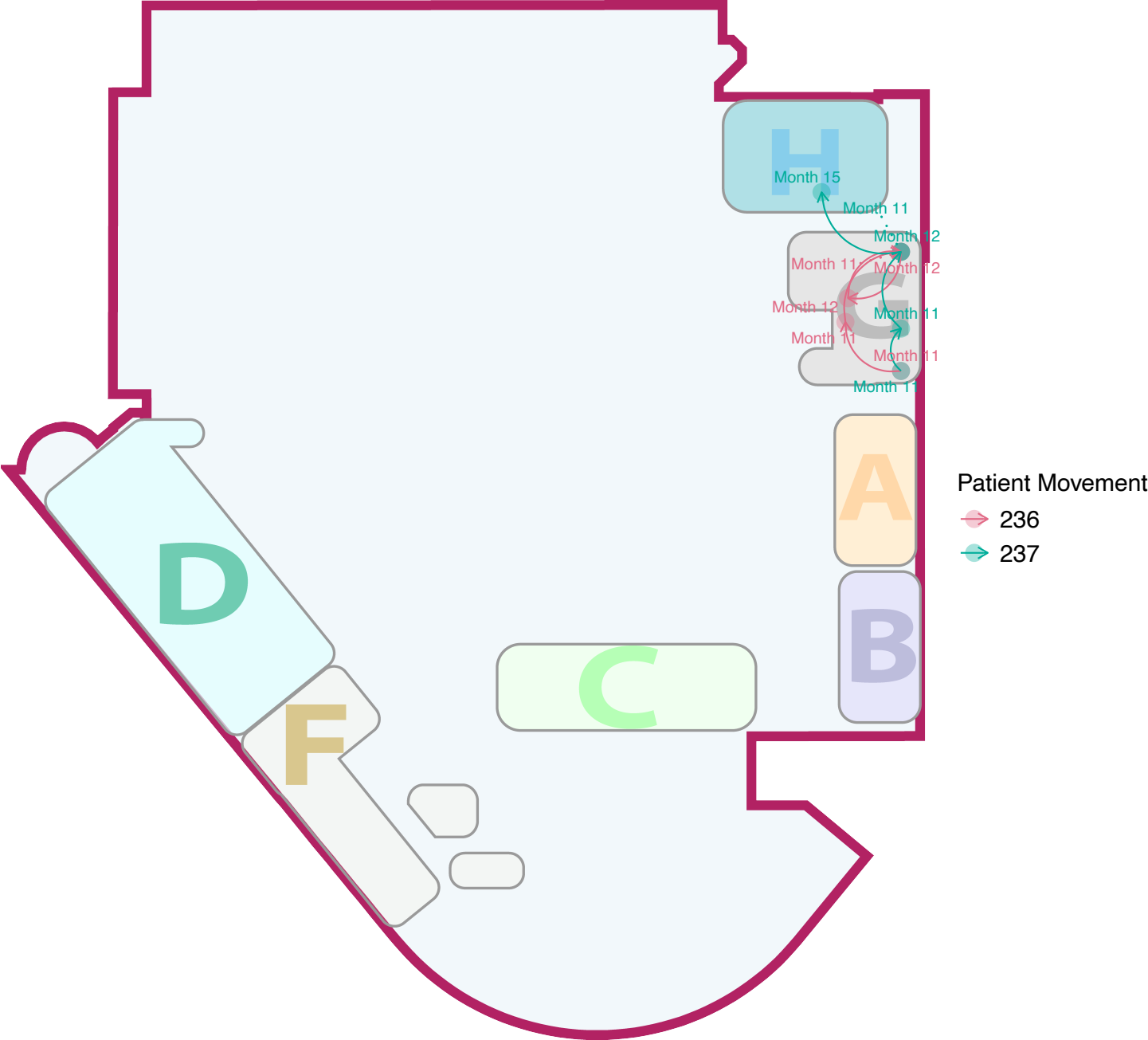
Cluster46



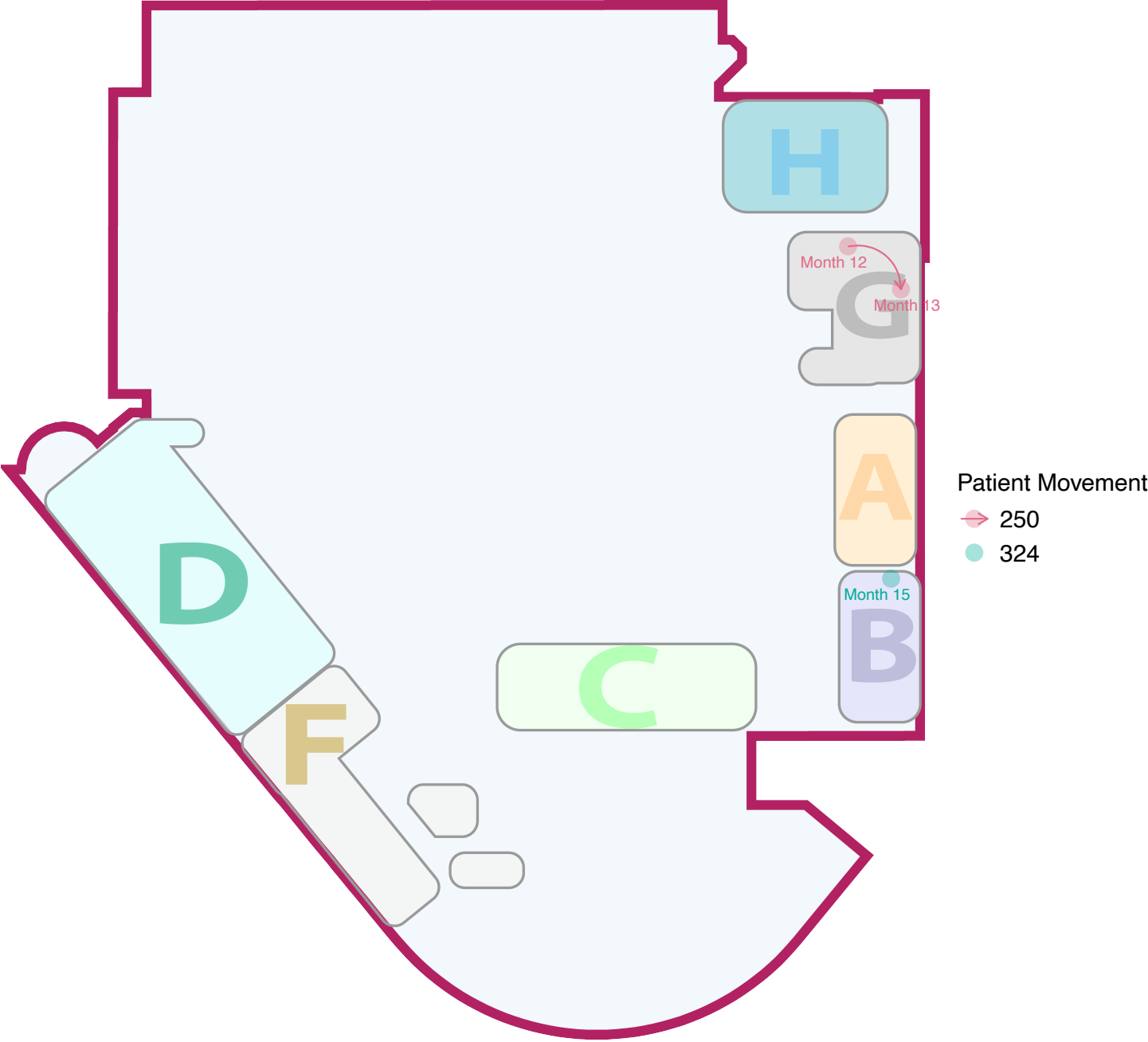
Cluster47



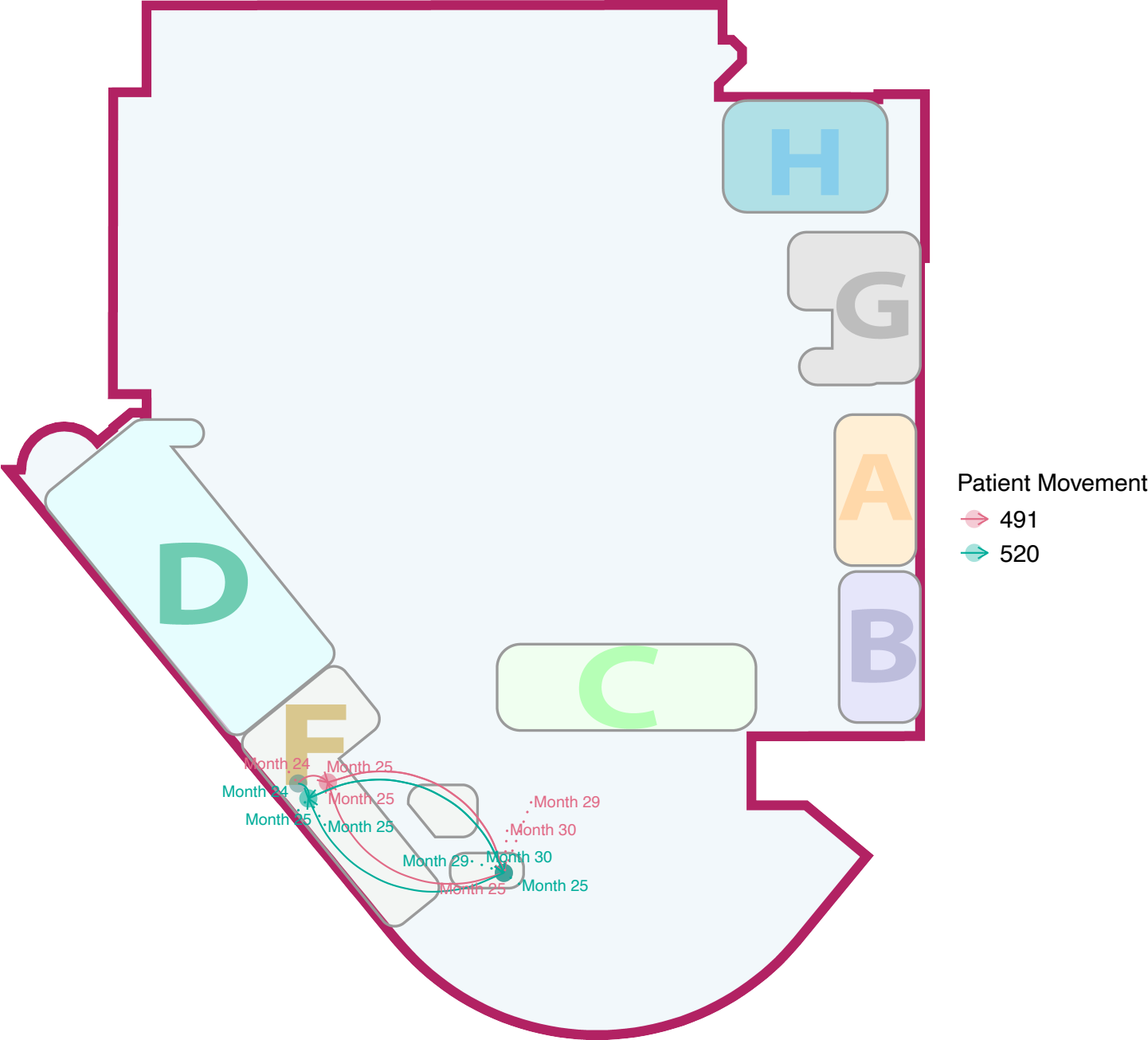
Cluster48



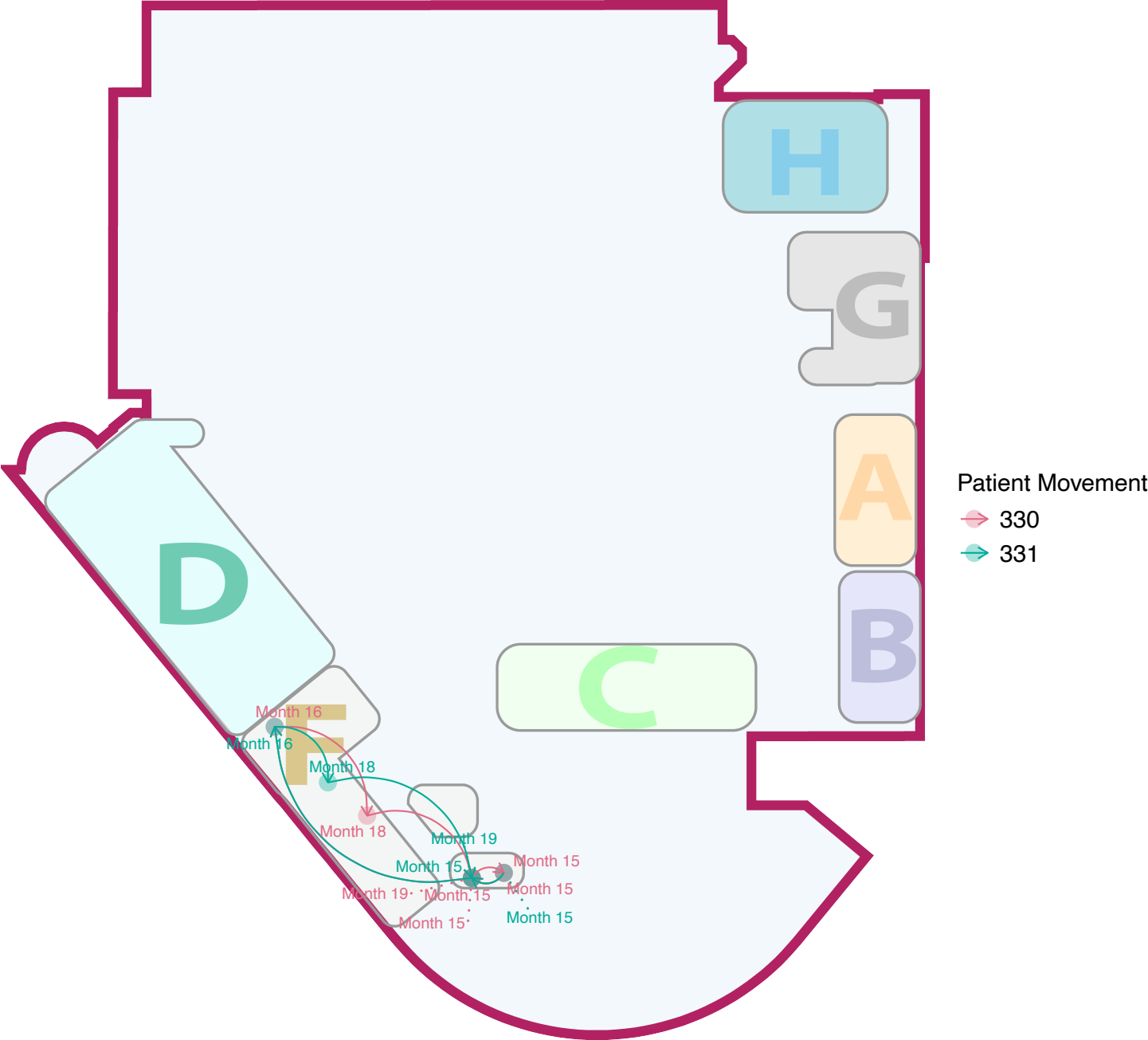
Cluster49



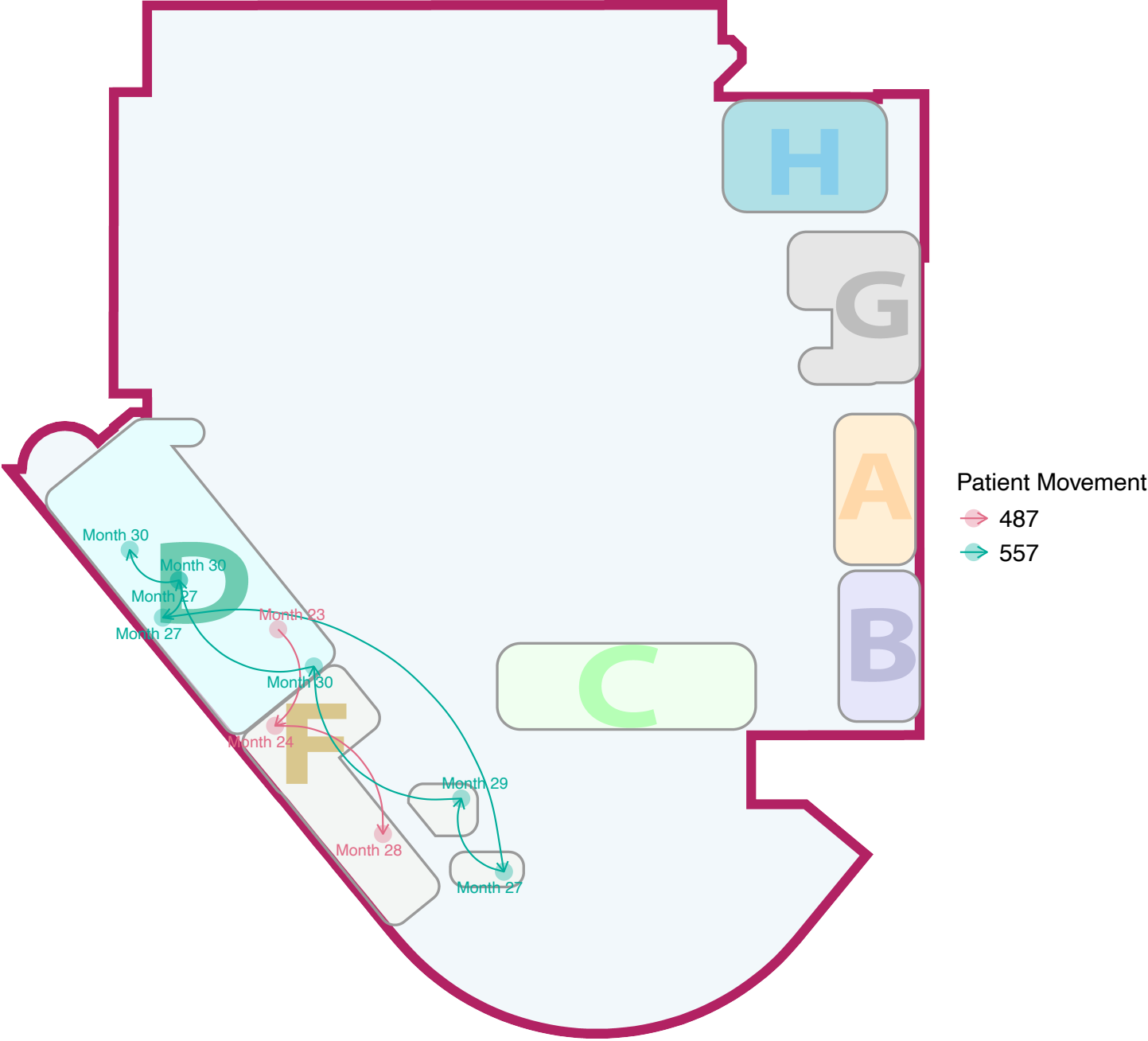
Cluster50



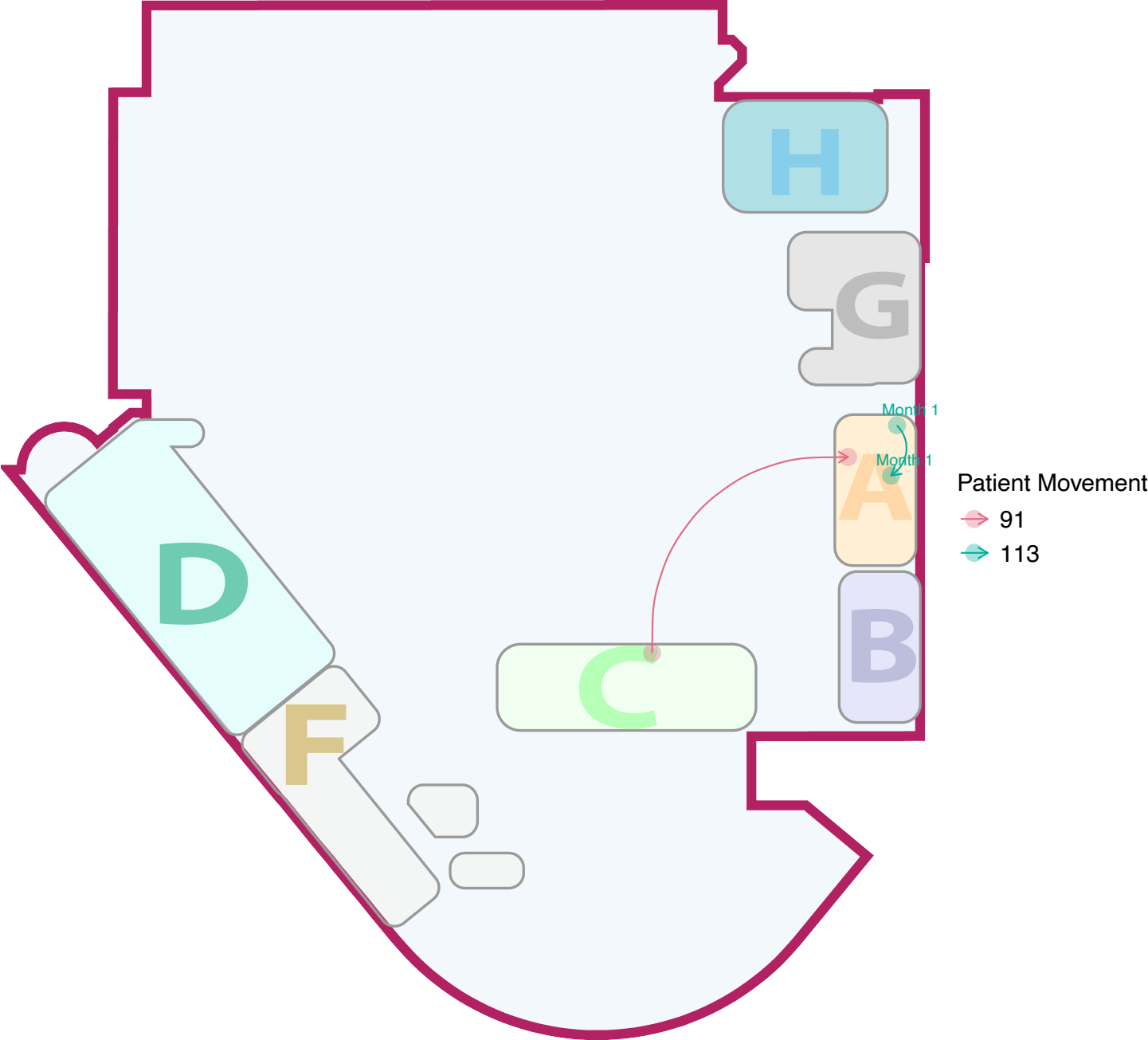
Cluster51



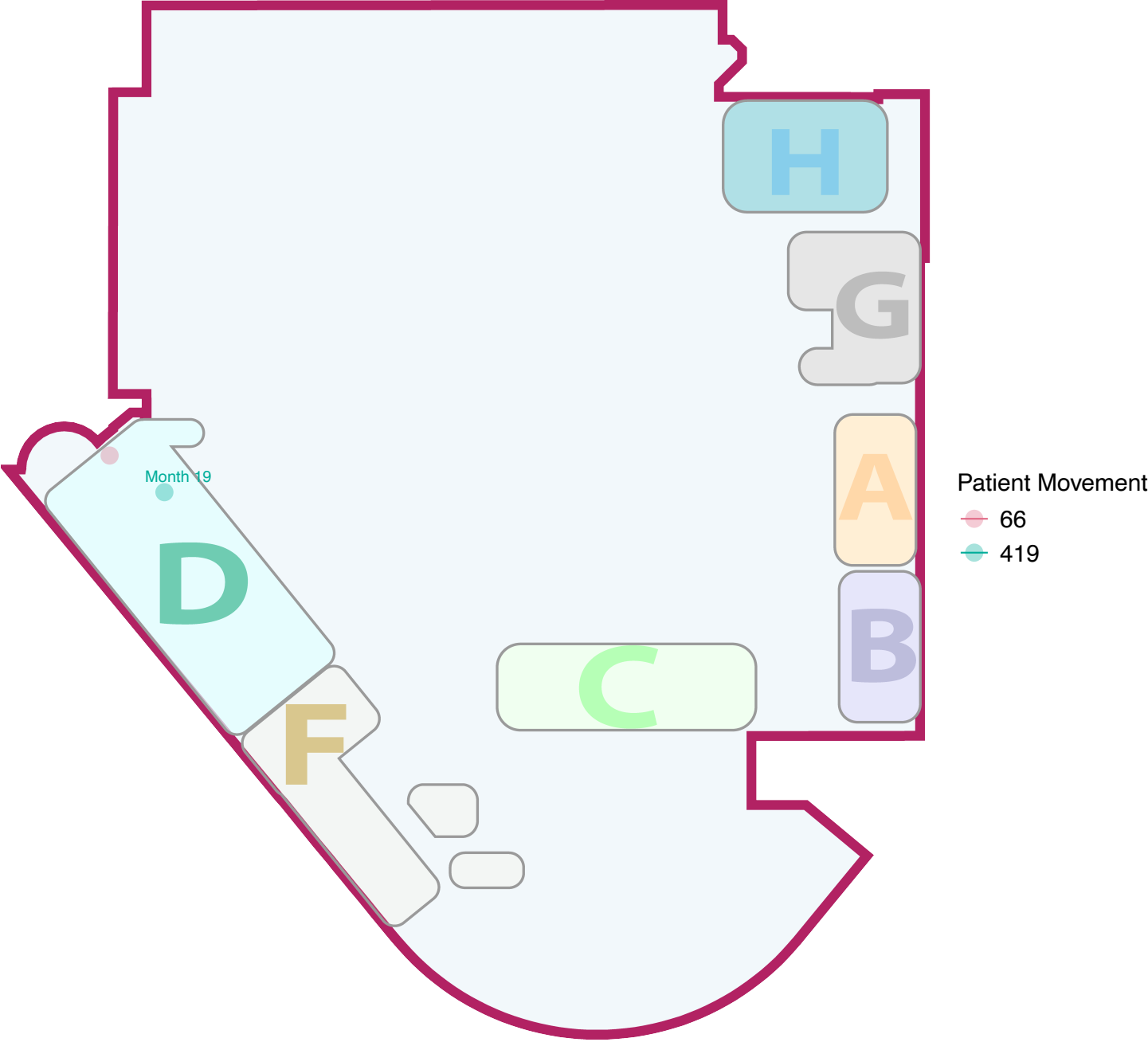
Cluster52



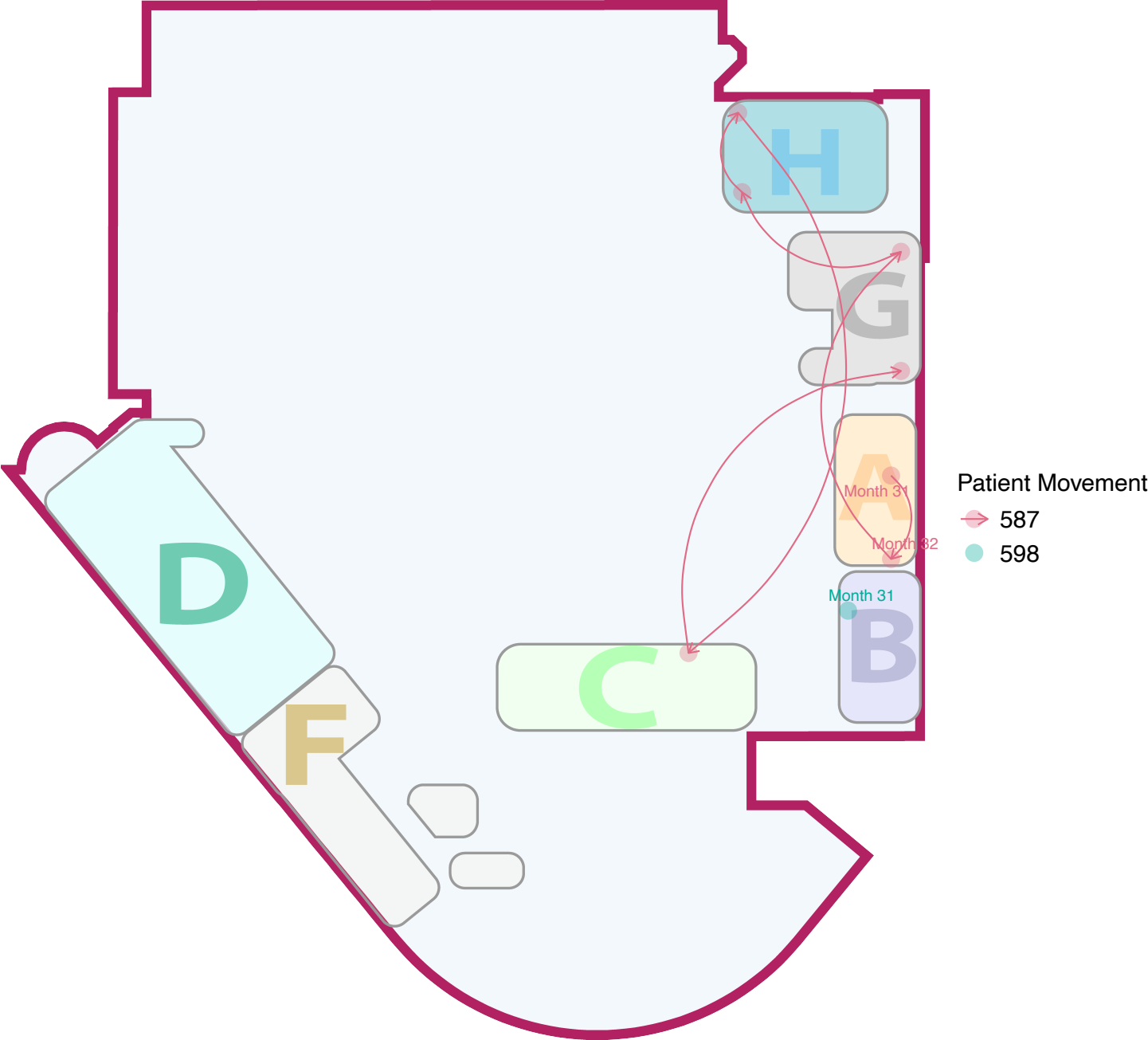
Cluster53



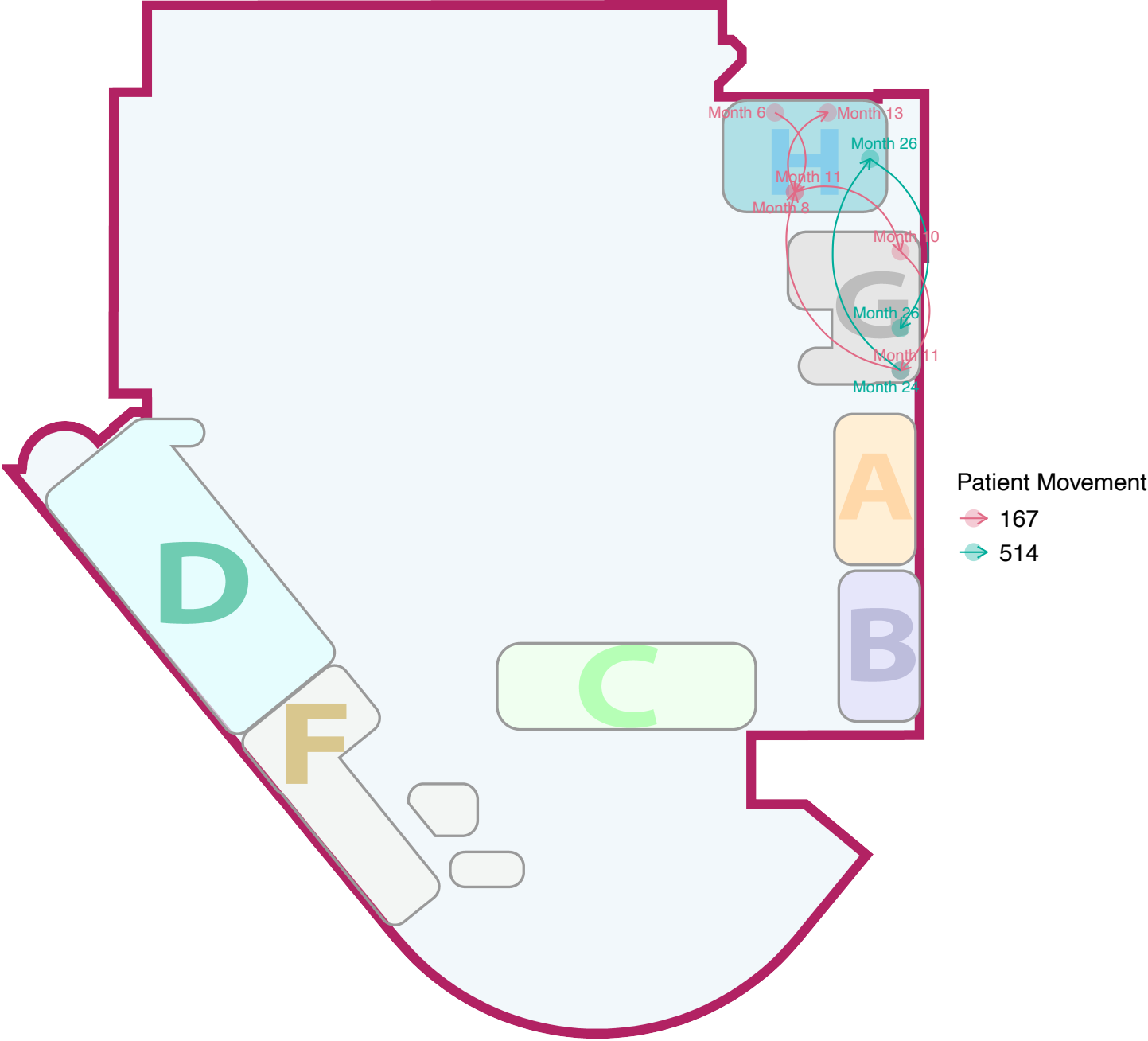
Cluster54



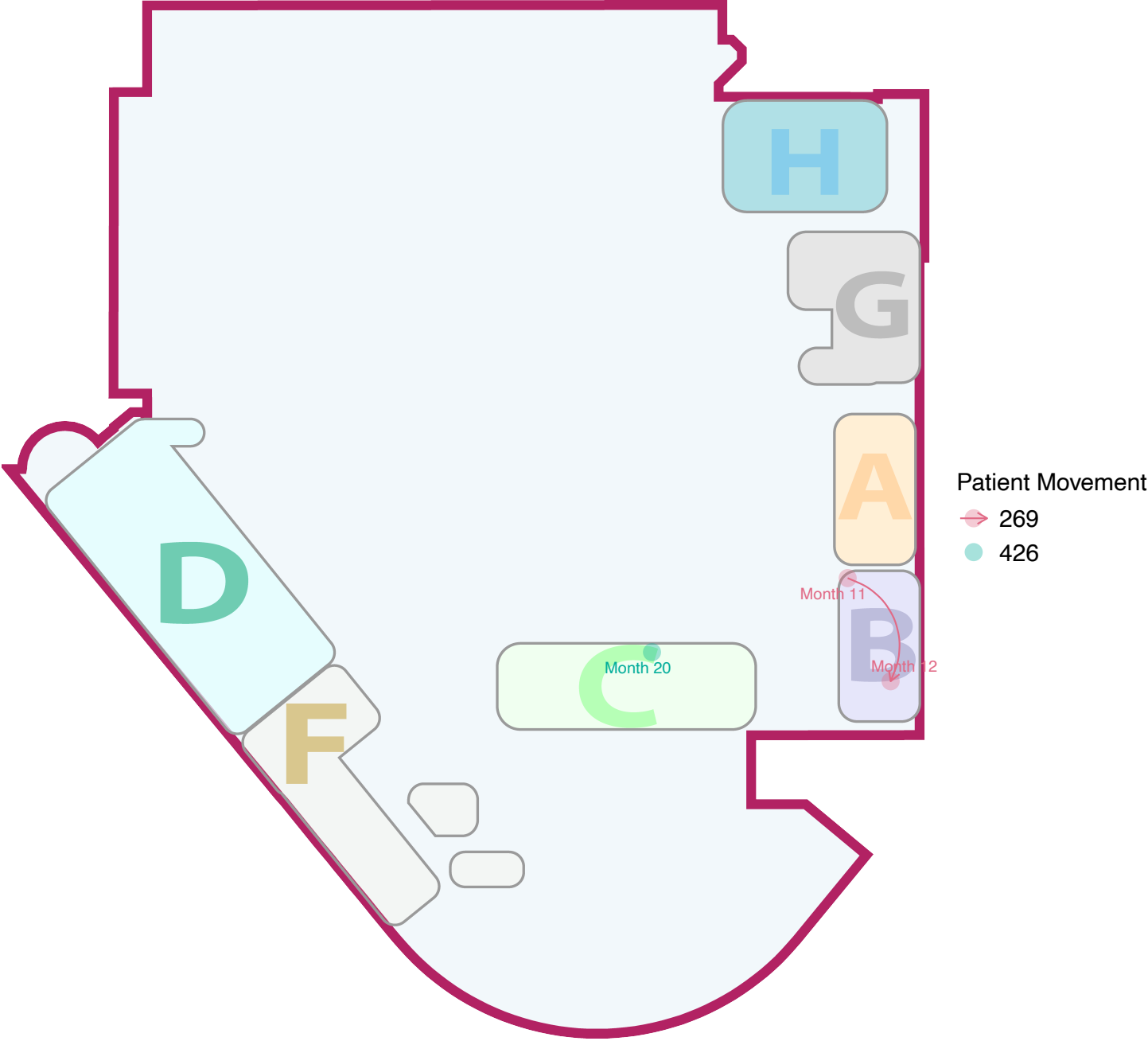
Cluster55



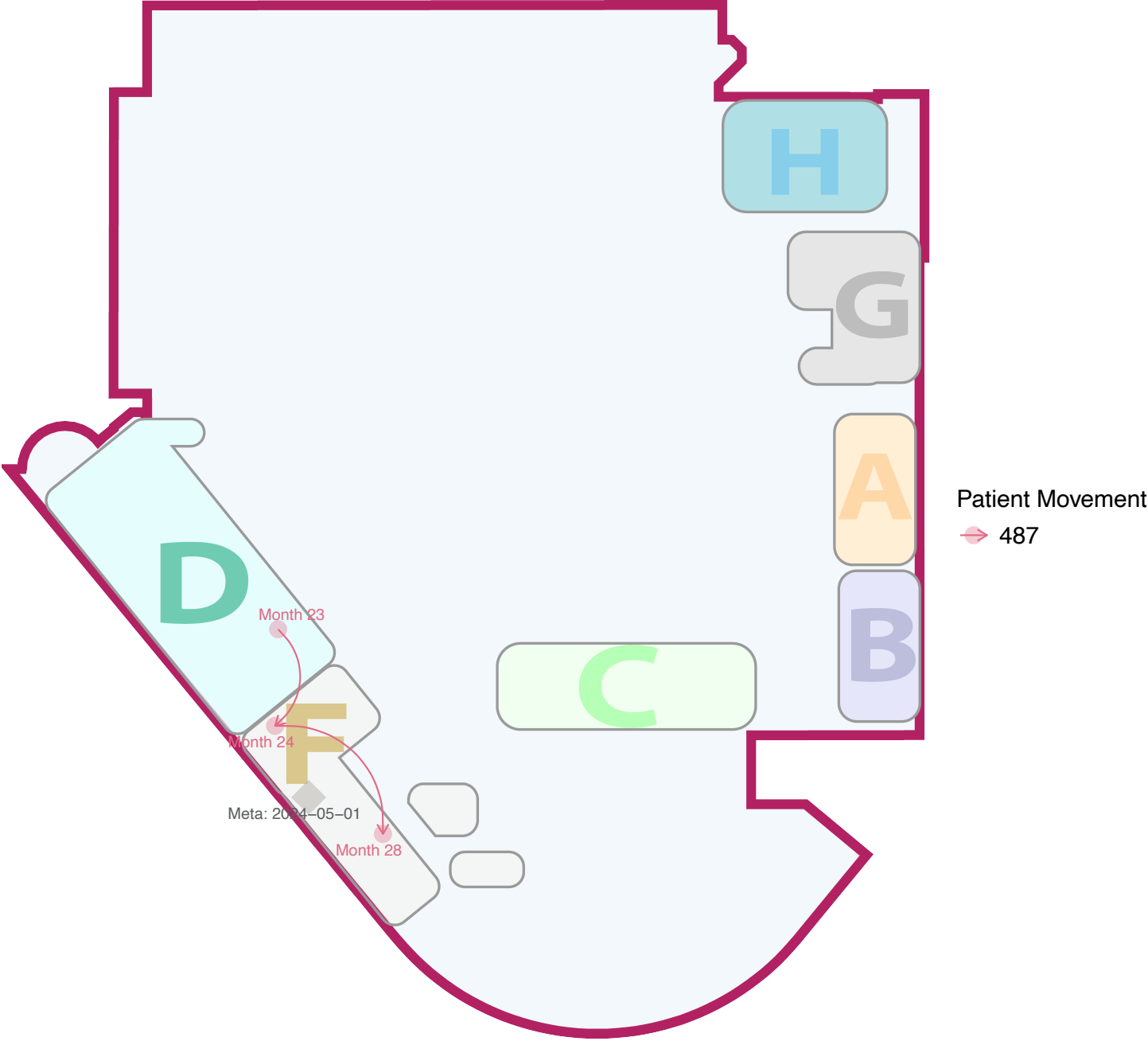
Cluster56



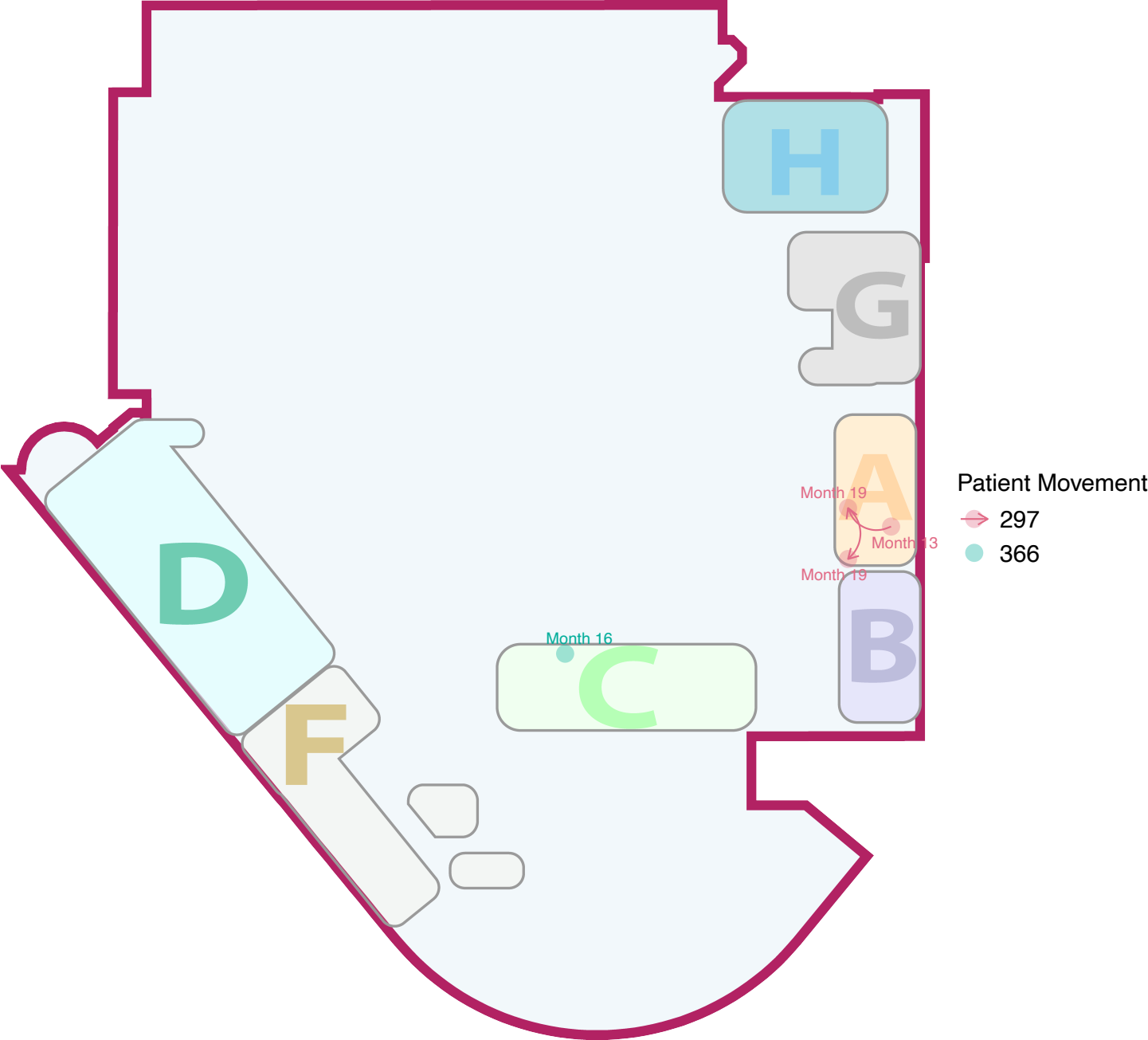
Cluster57



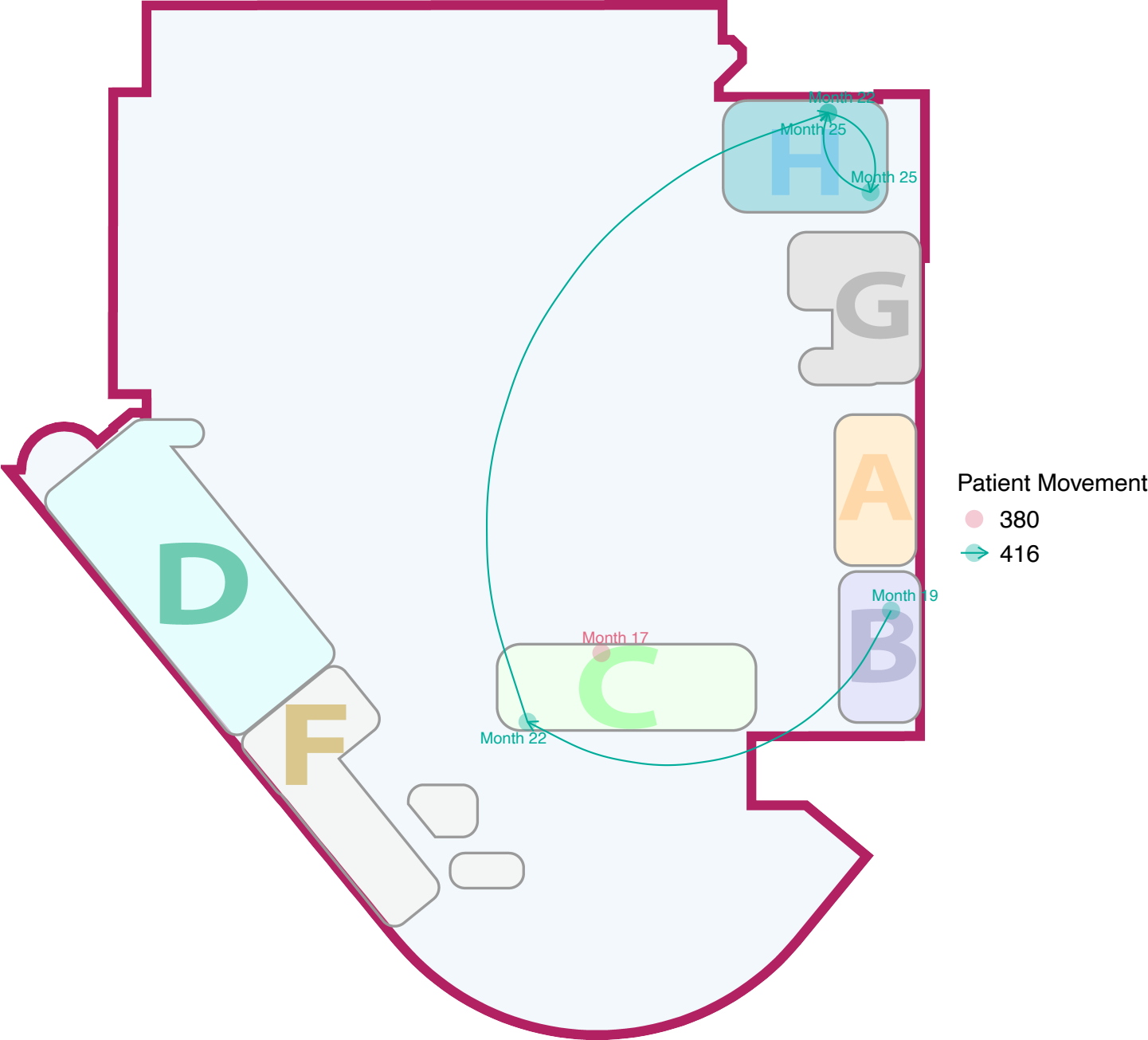
Cluster58



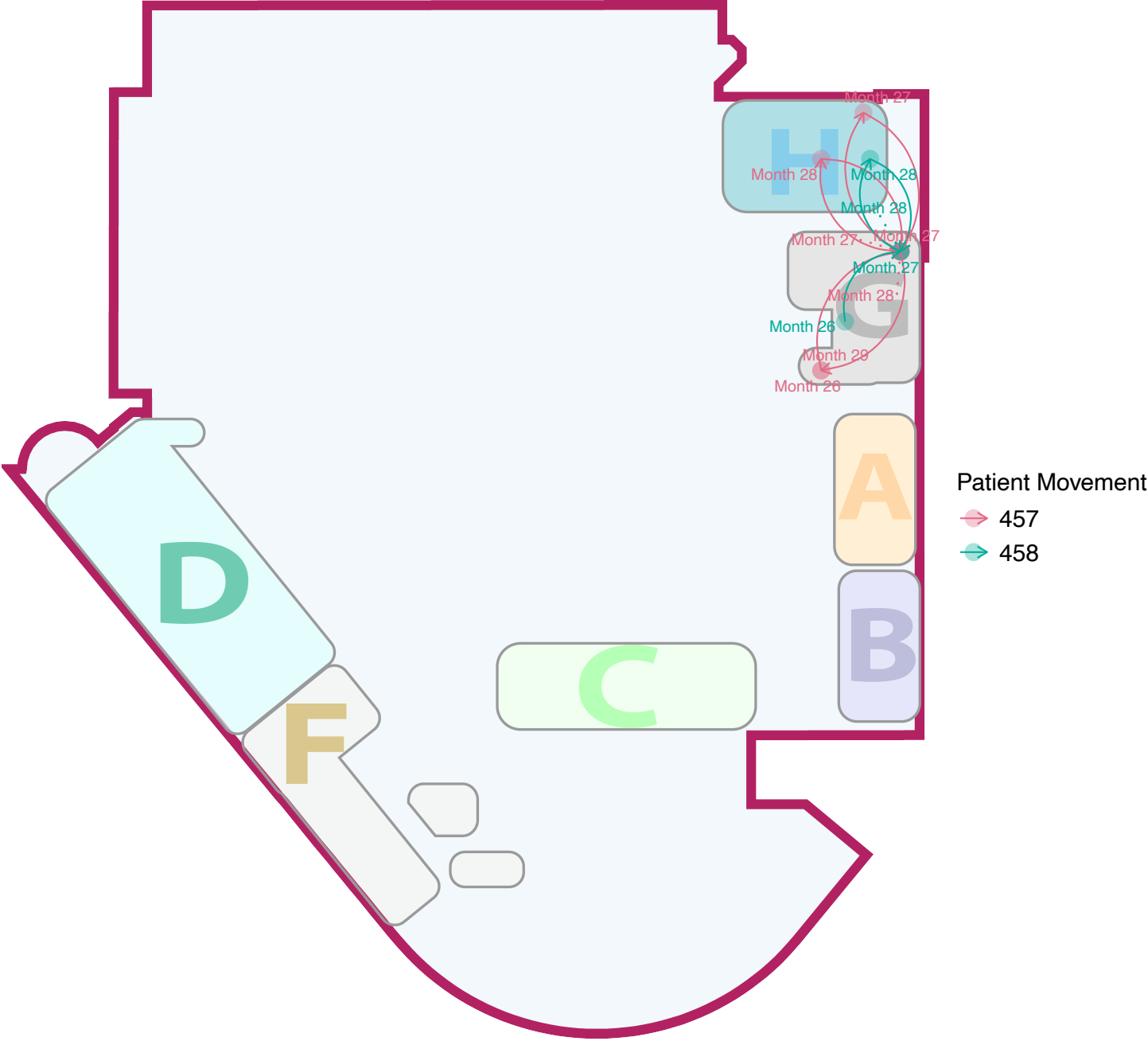
Cluster59



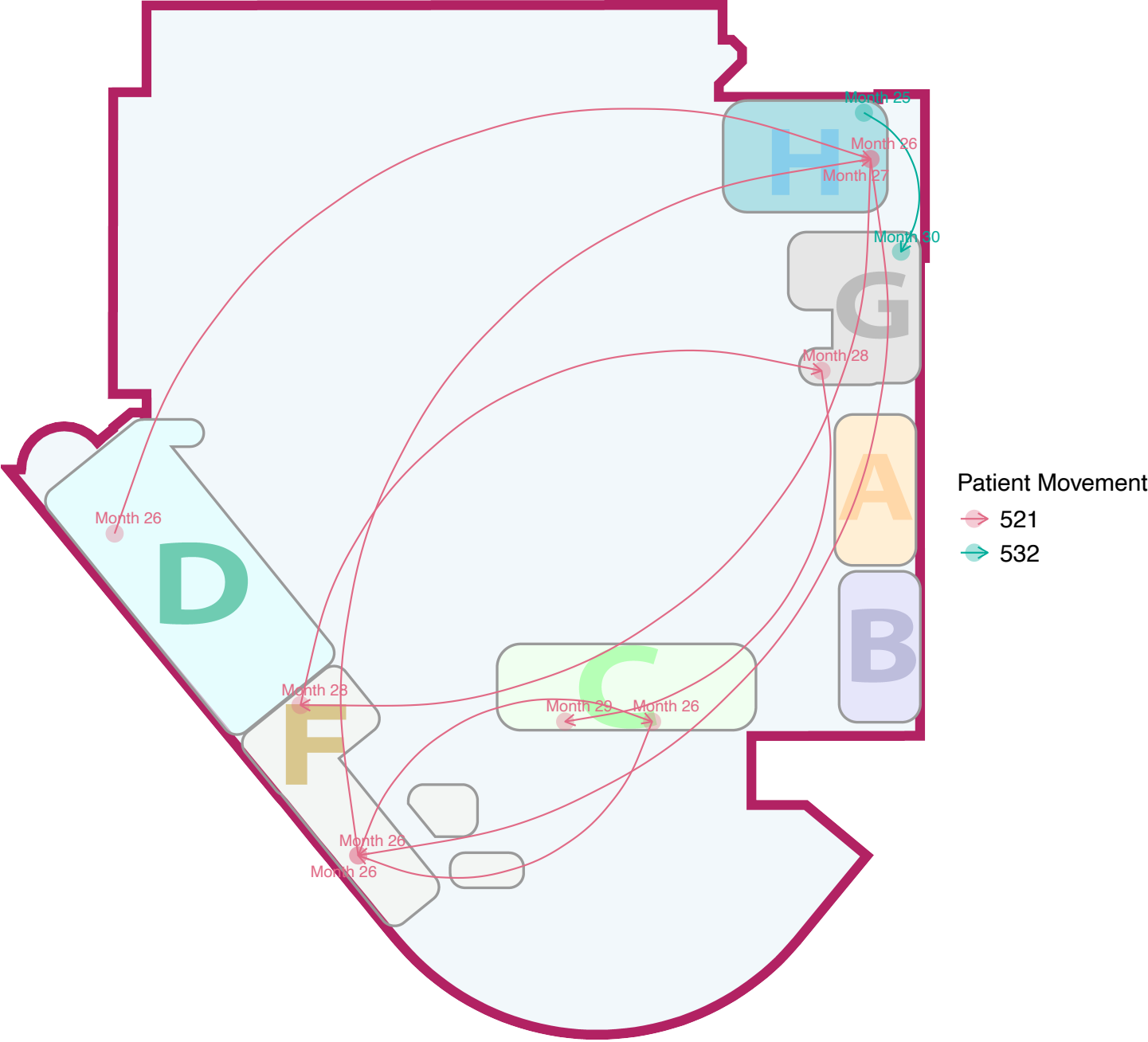
Cluster60



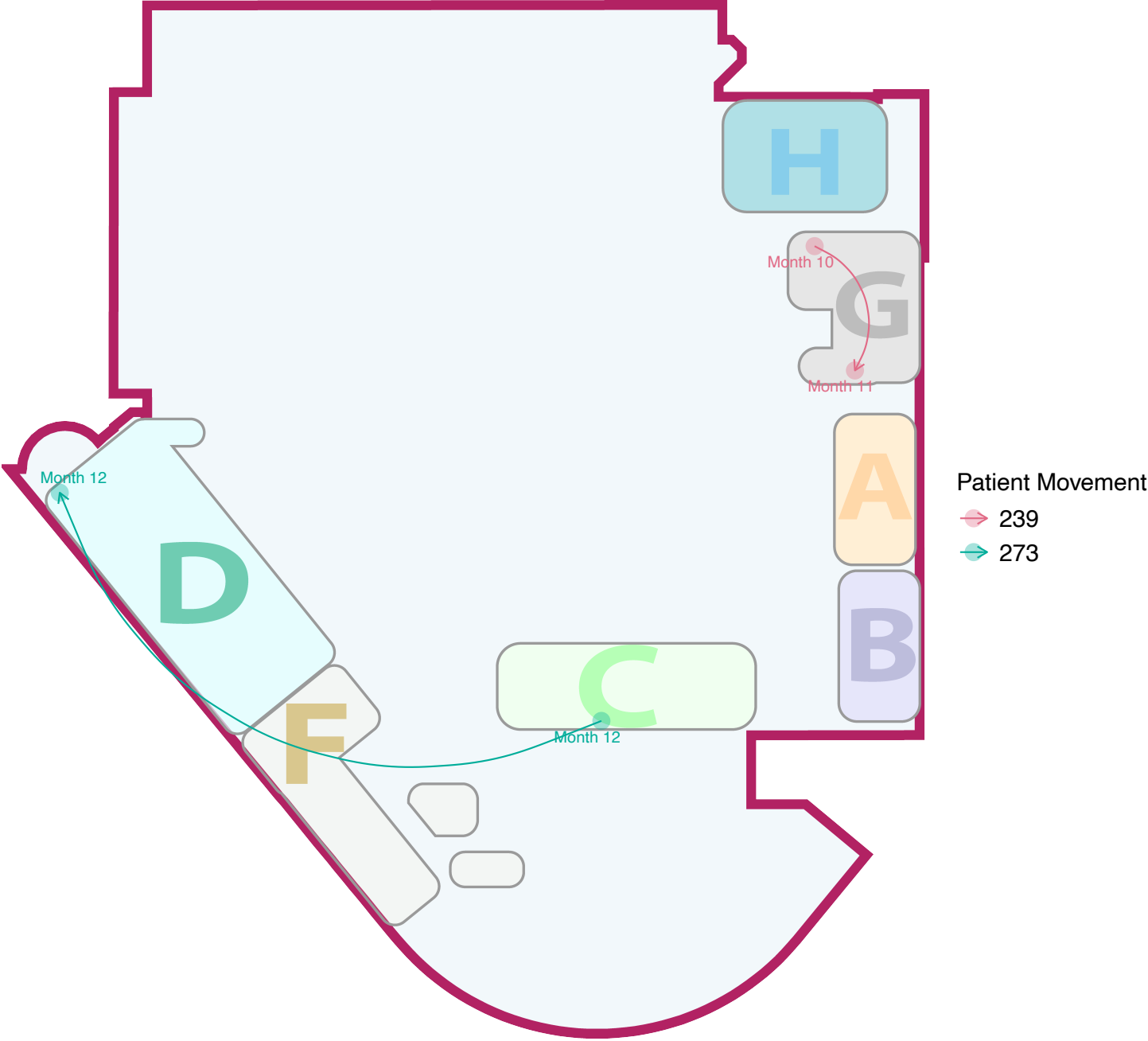
Cluster61



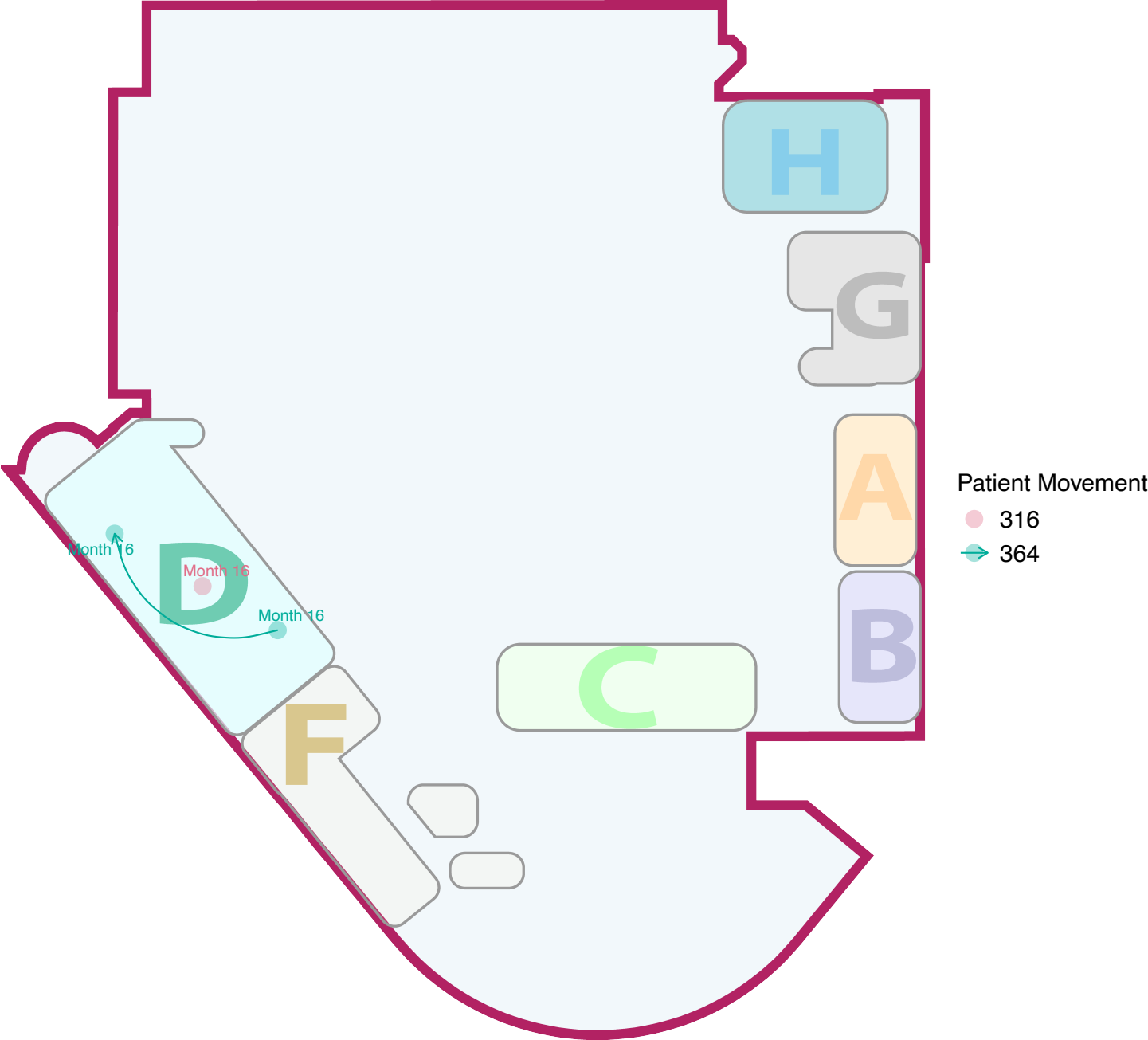
Cluster62



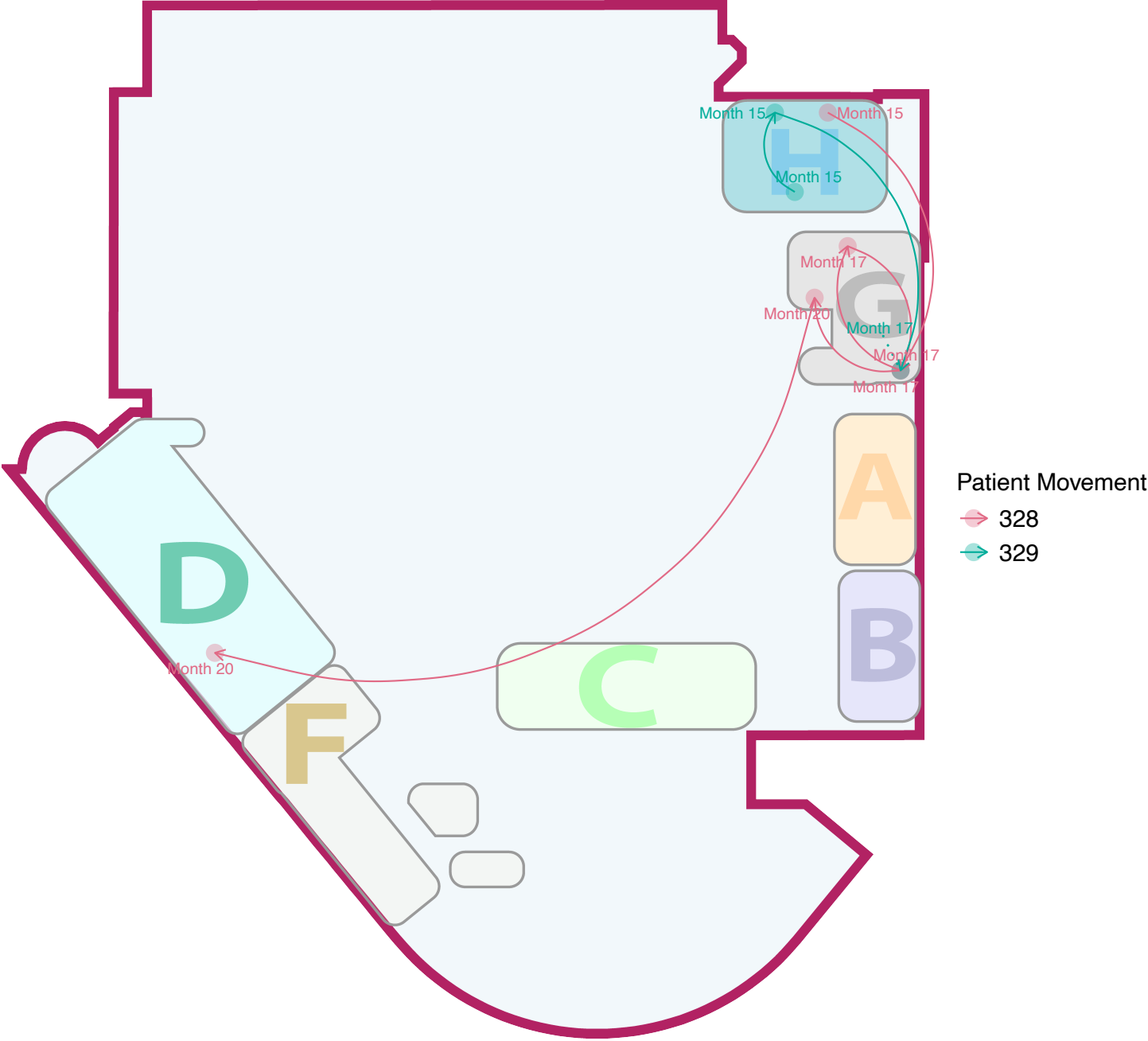
Cluster63



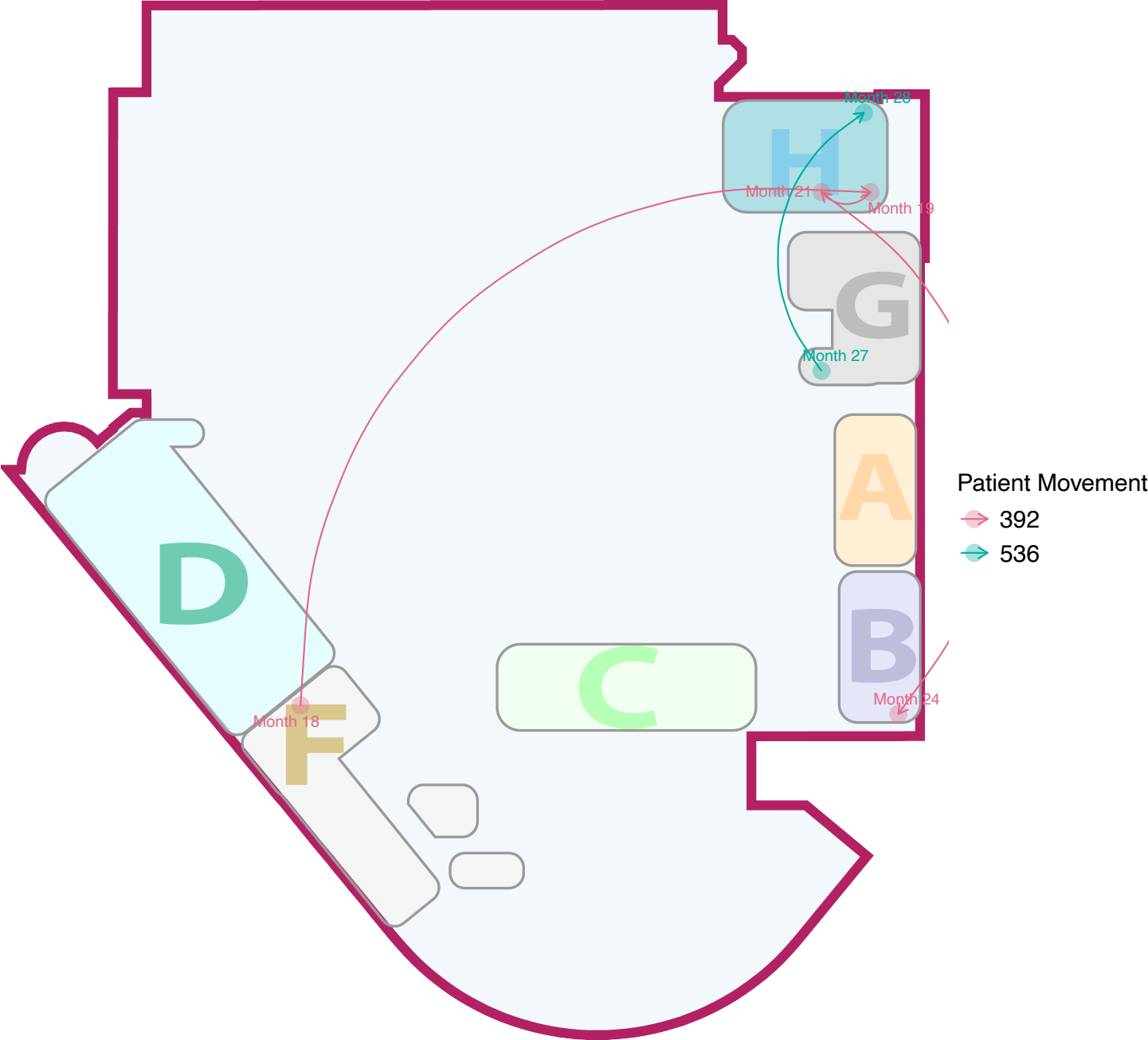
Cluster64



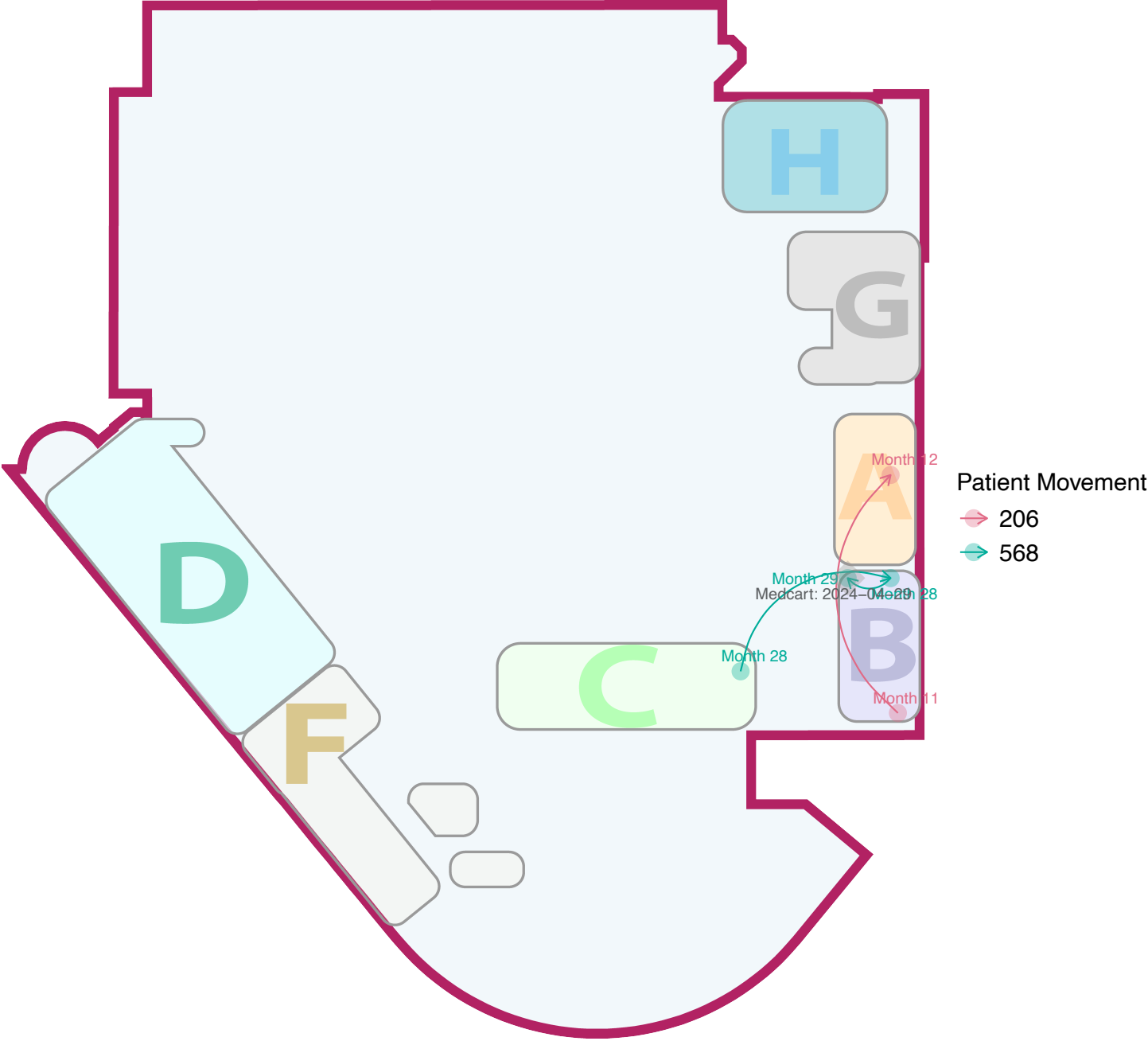
Cluster65



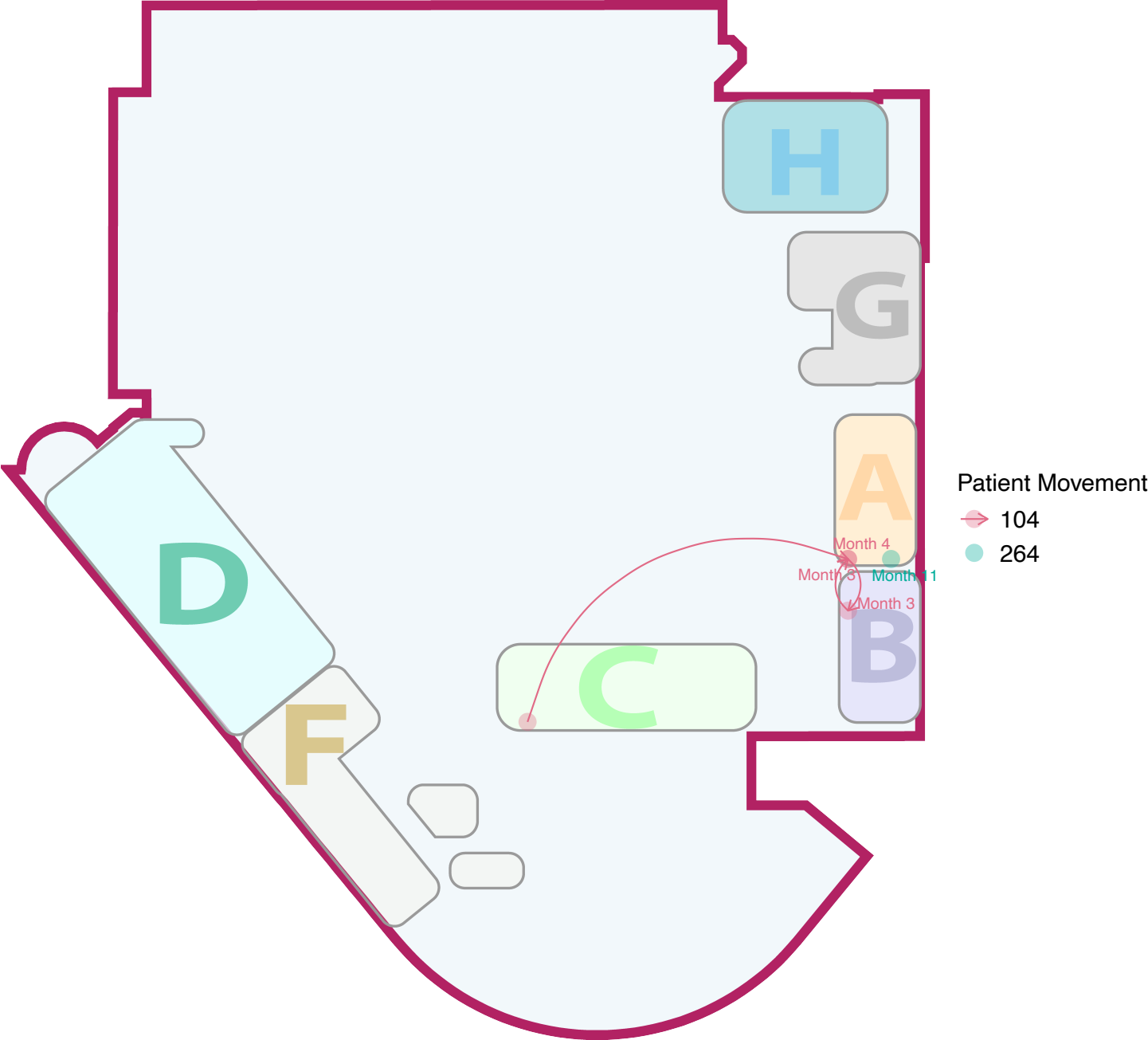
Cluster66



Cluster67



Cluster68



Cluster69

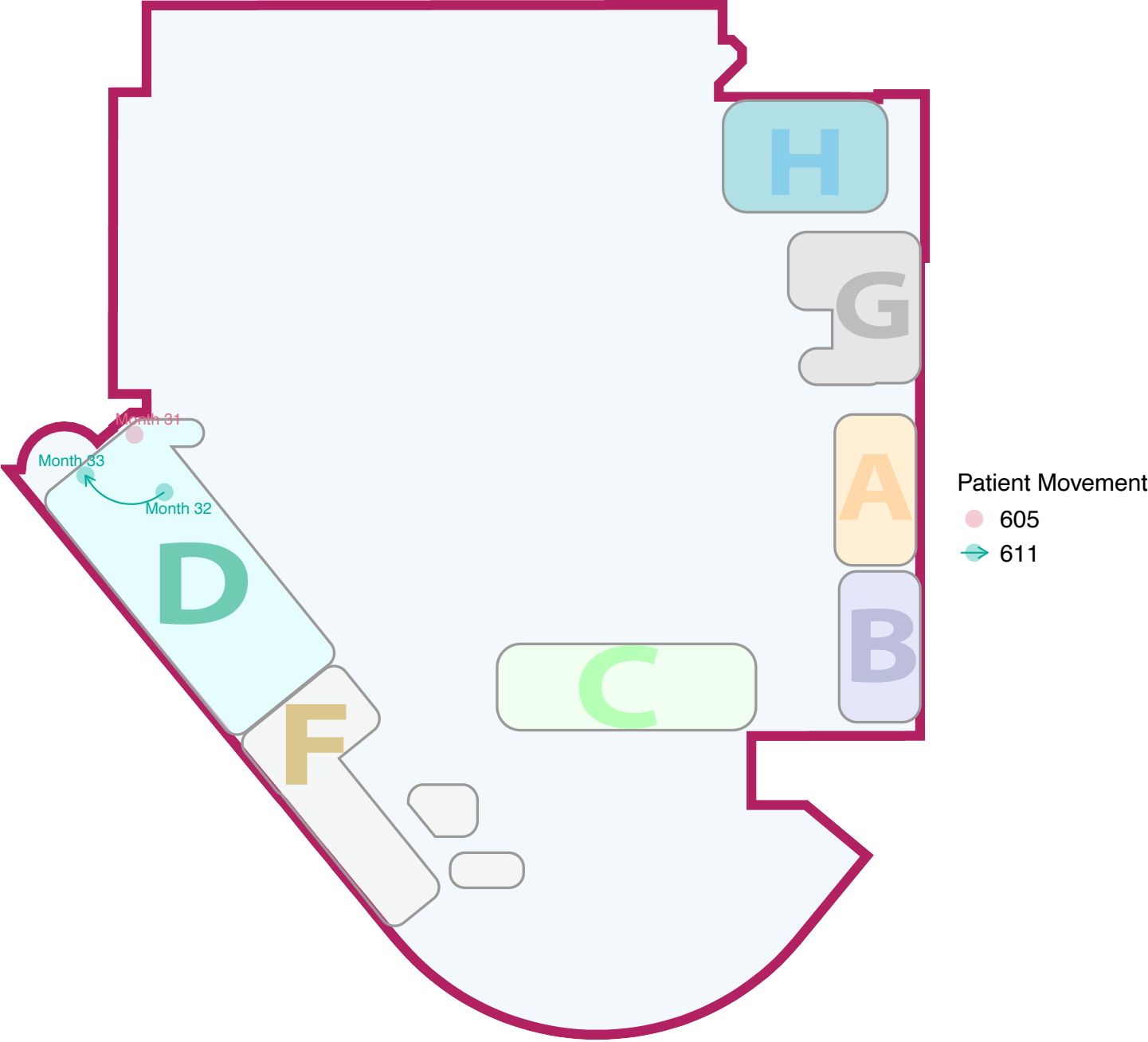
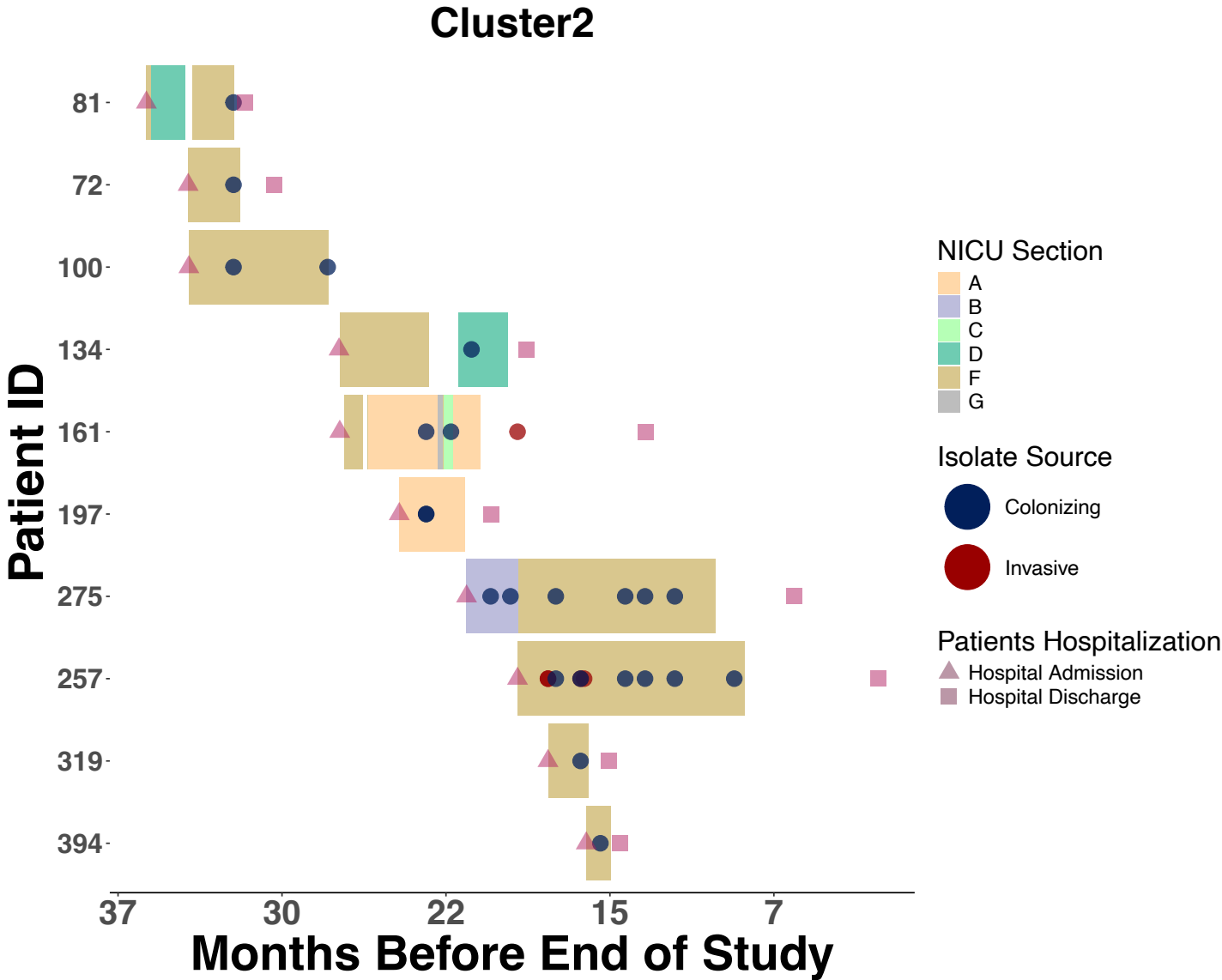
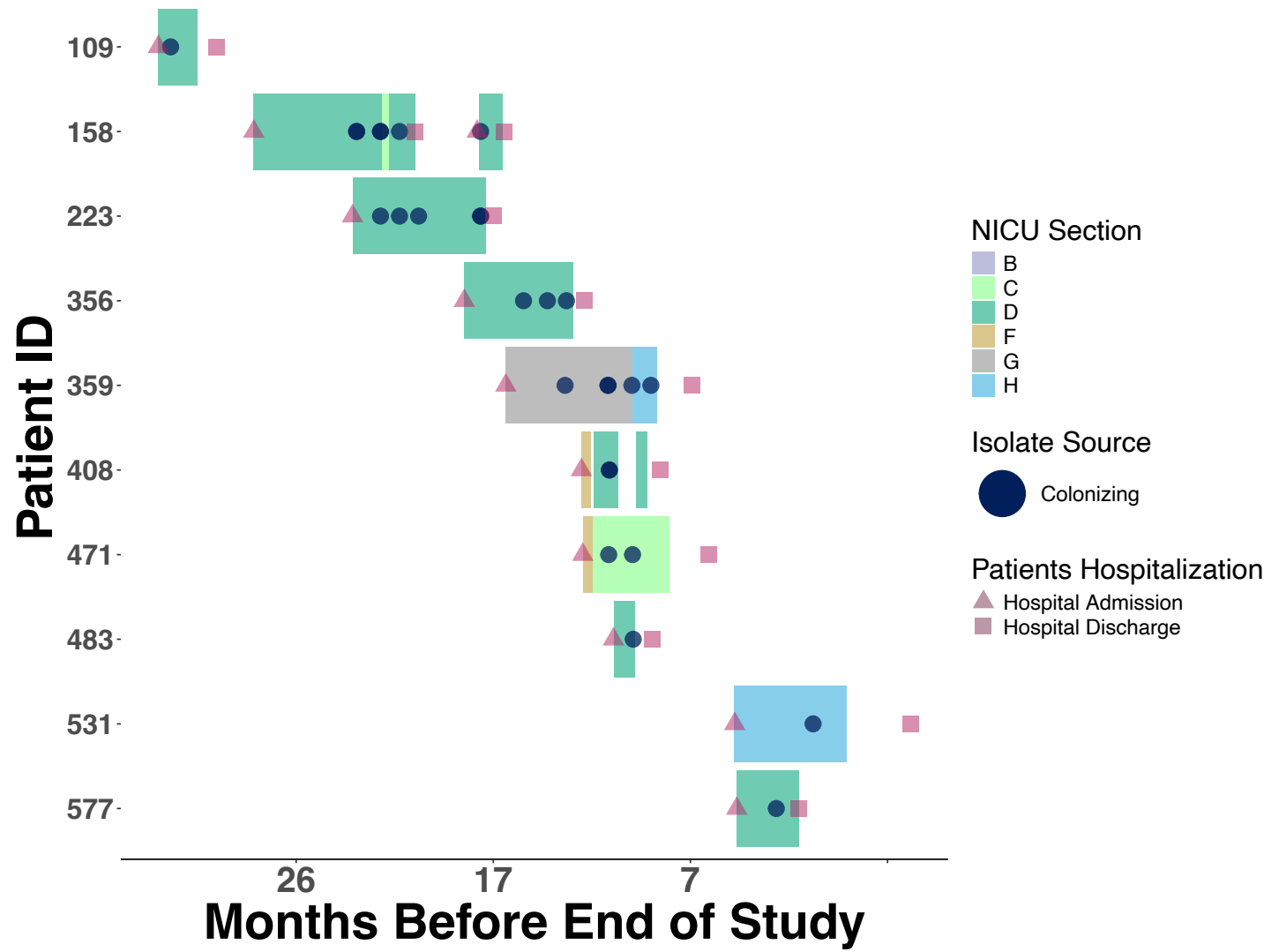


Figure S4: Temporal analysis for each identified transmission cluster by location

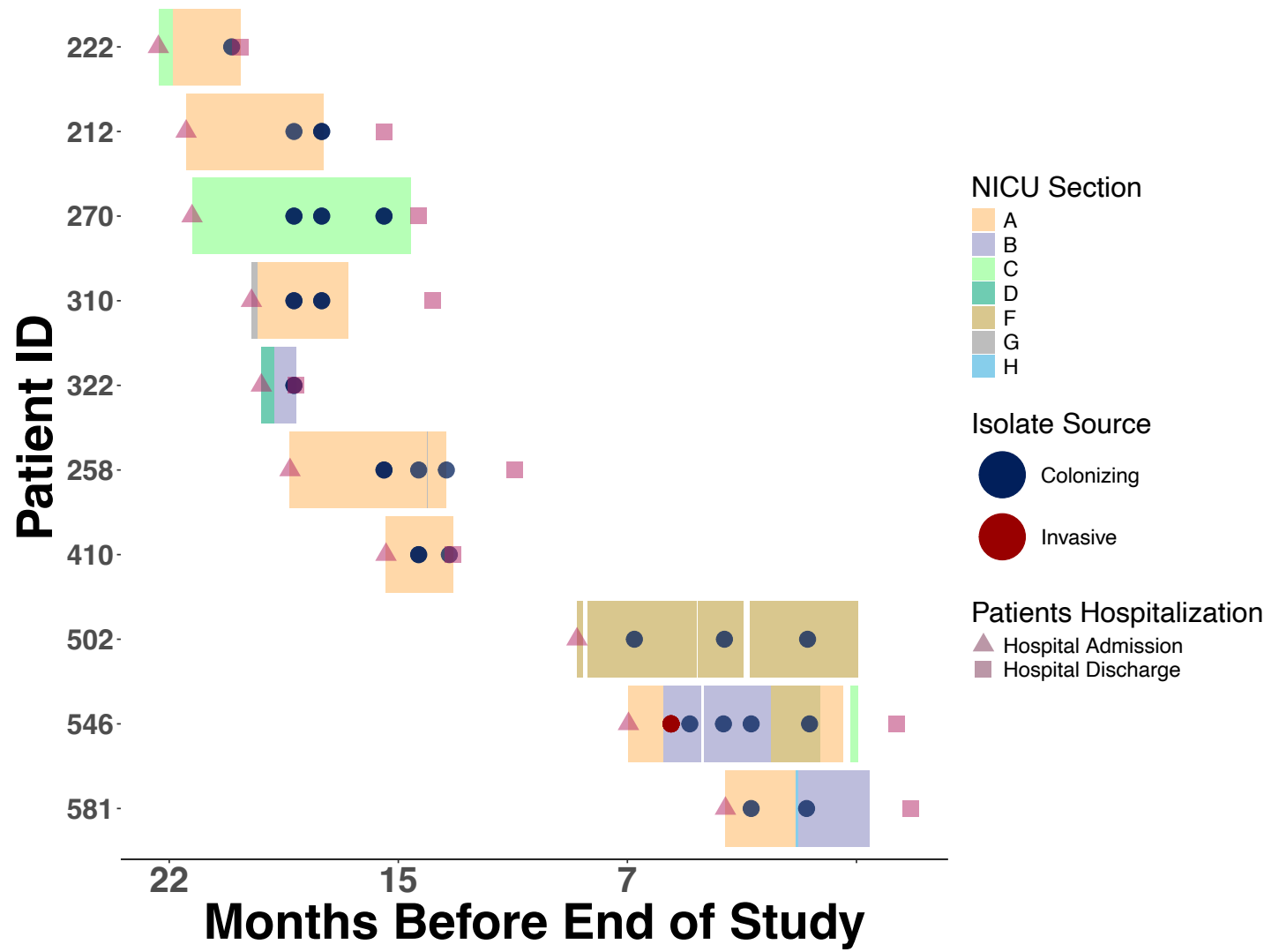
Each bar represents a single patient's NICU stay, with colored rectangles indicating different NICU sections. Admission dates are marked with triangles and discharge dates with squares. X-axis (Months Before End of Study) represents the number of months prior to completion of study period.



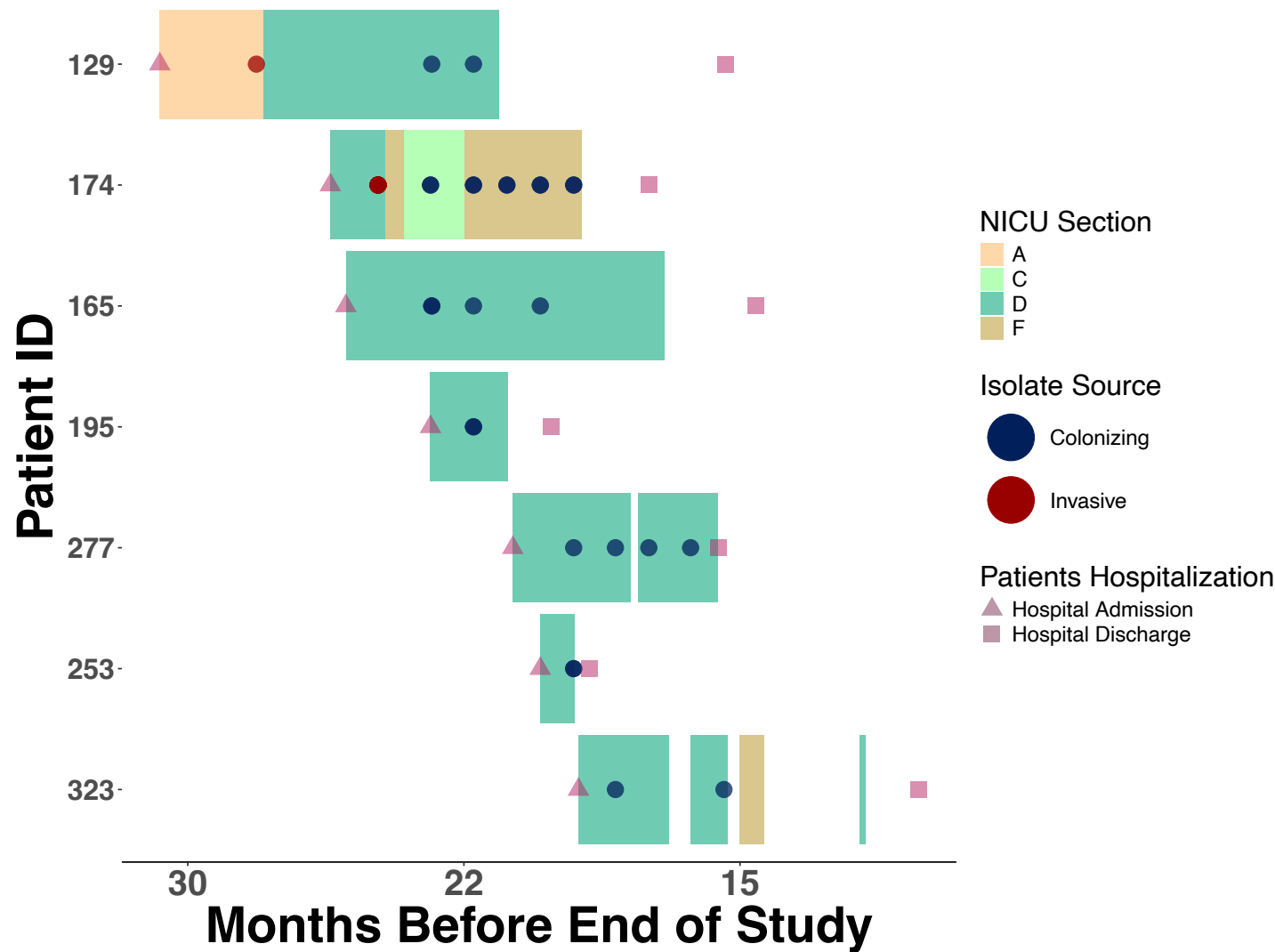
Cluster3



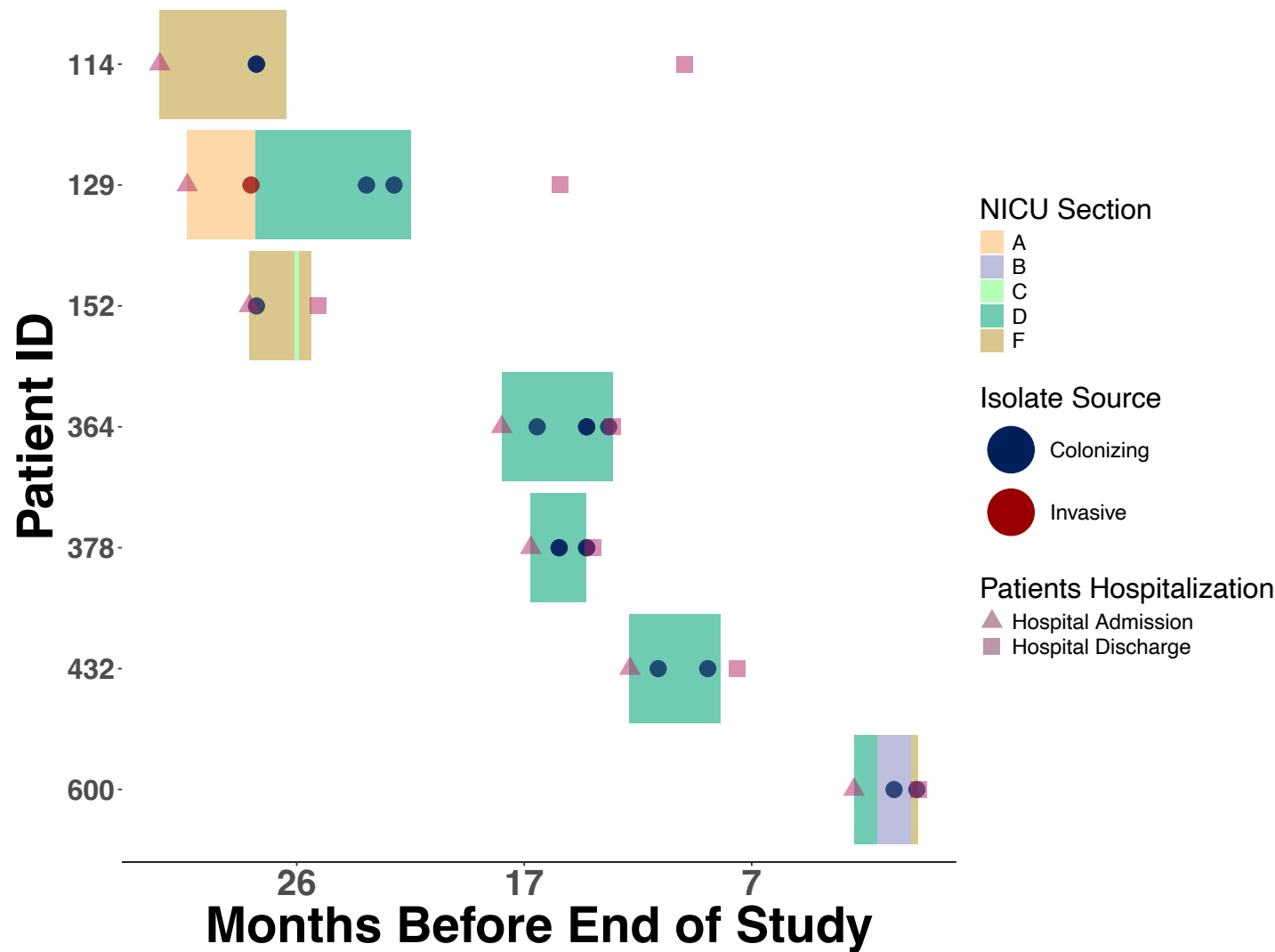
Cluster4



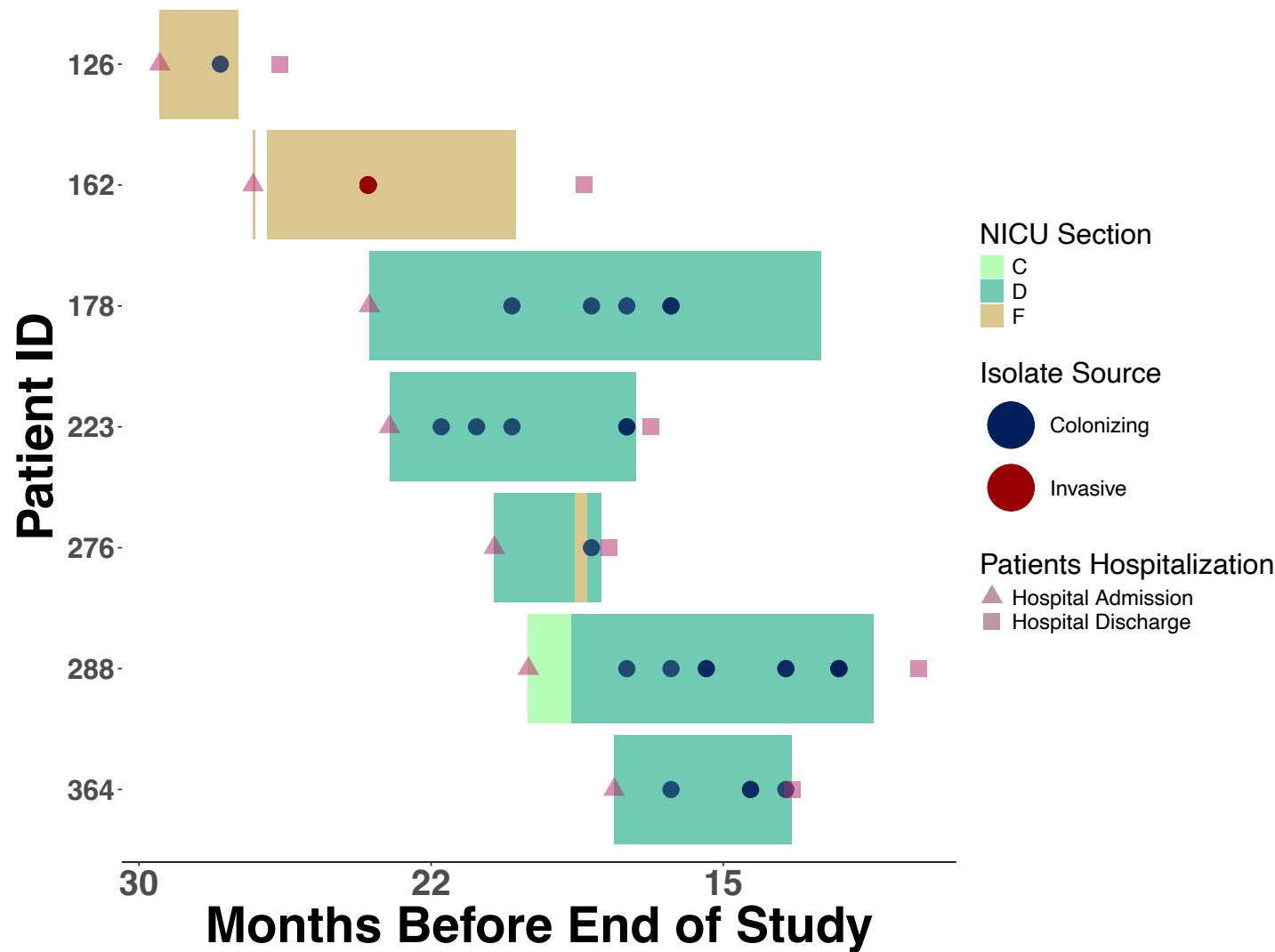
Cluster5



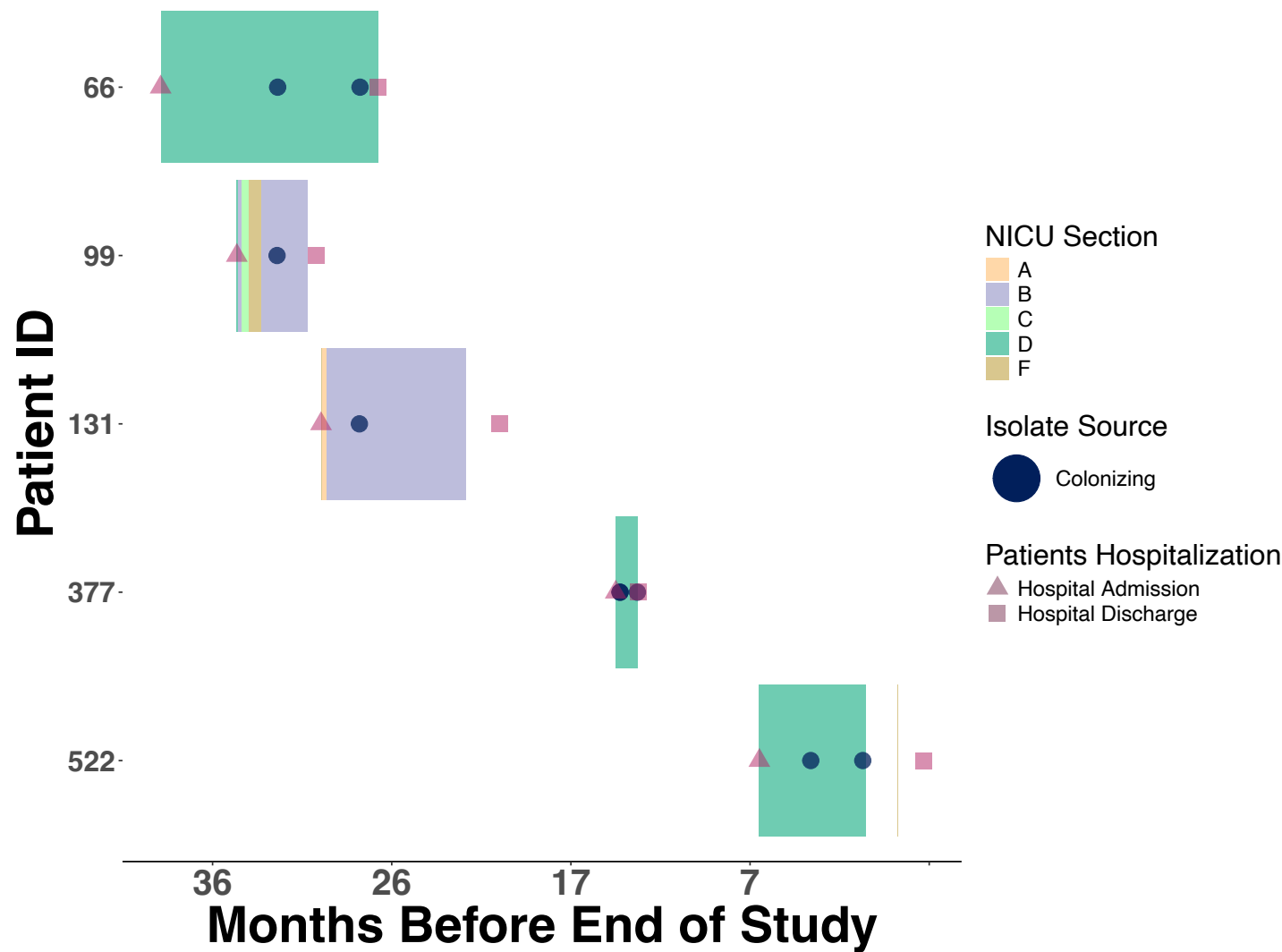
Cluster6



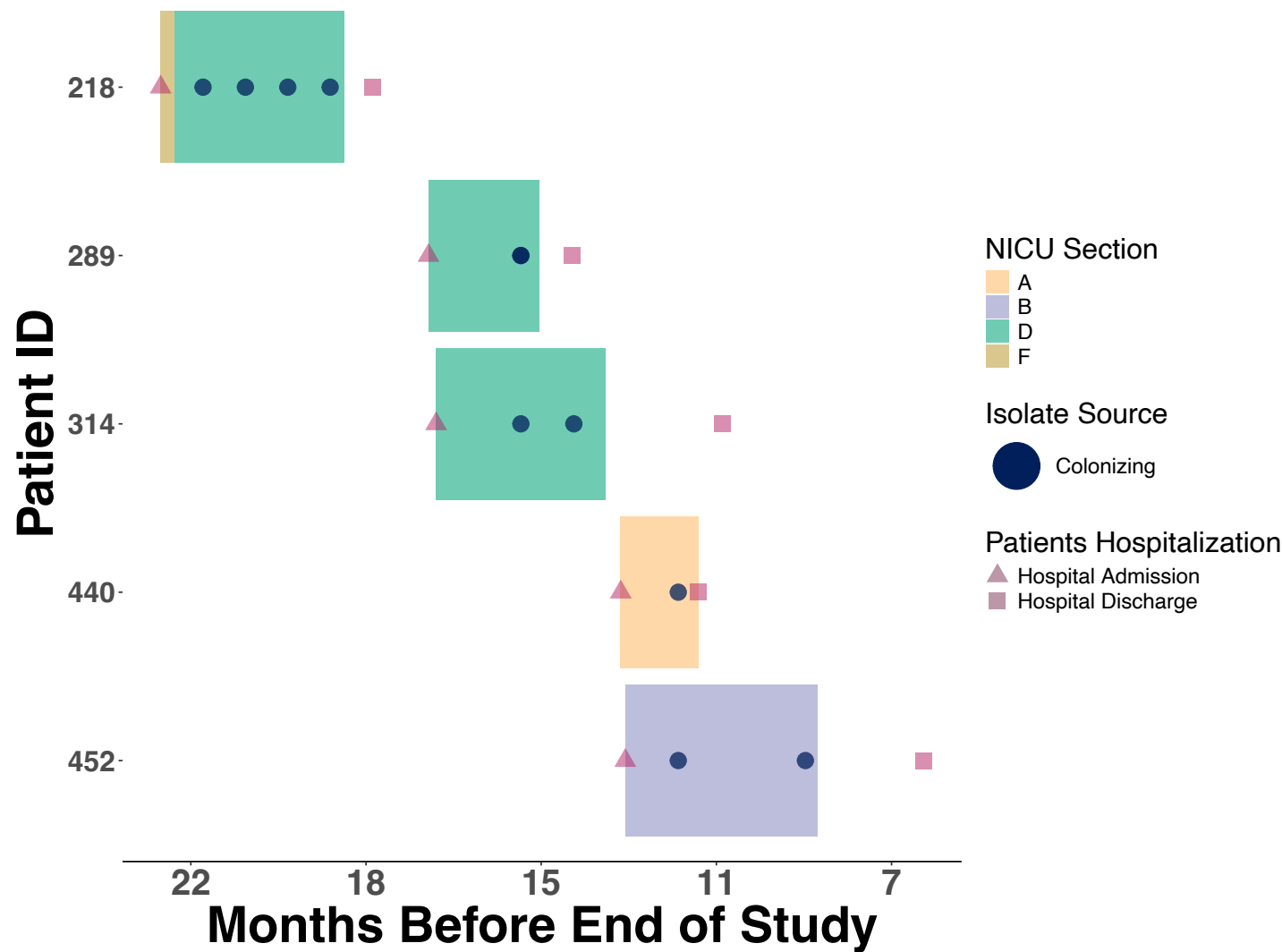
Cluster7



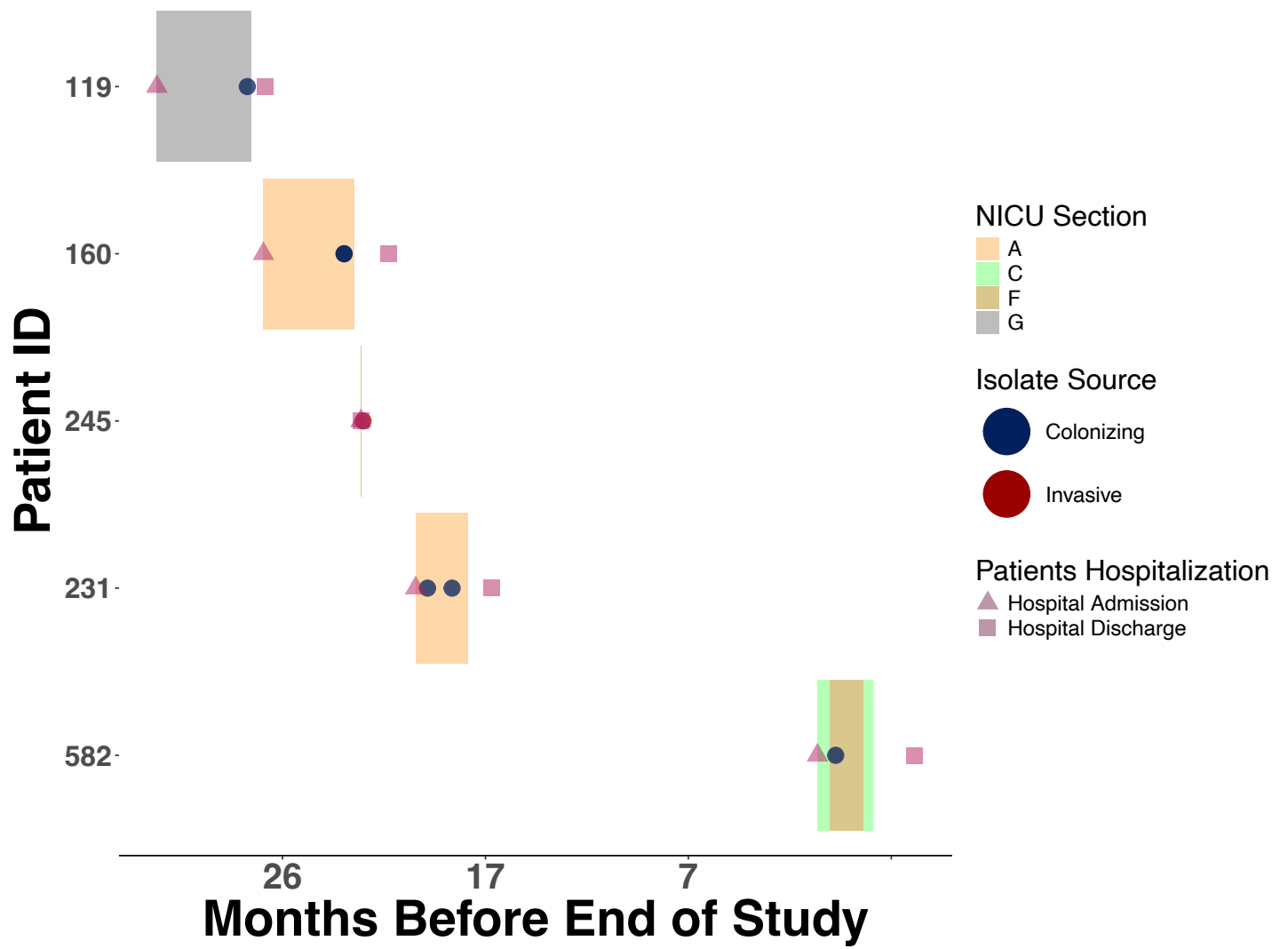
Cluster8



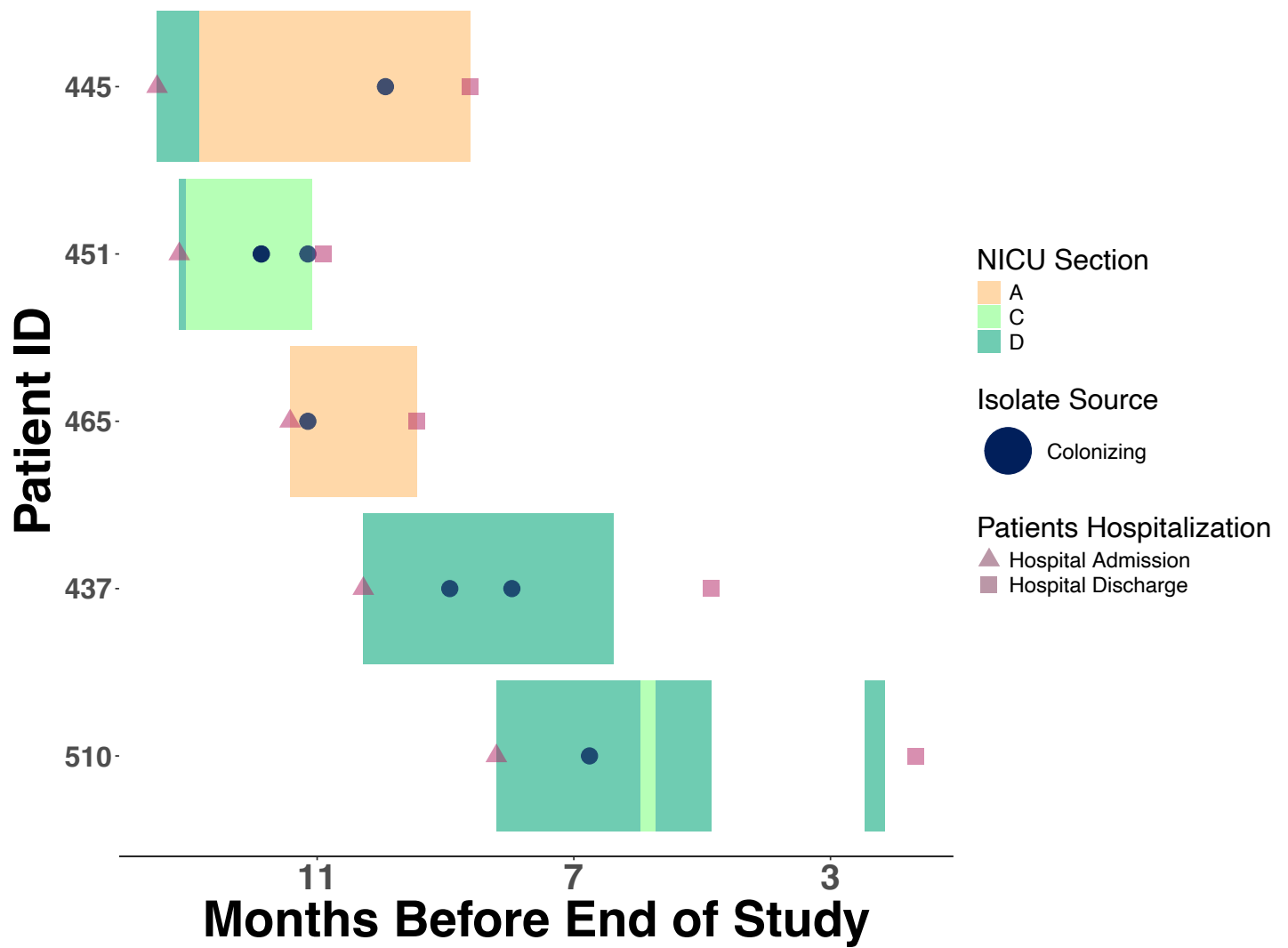
Cluster9



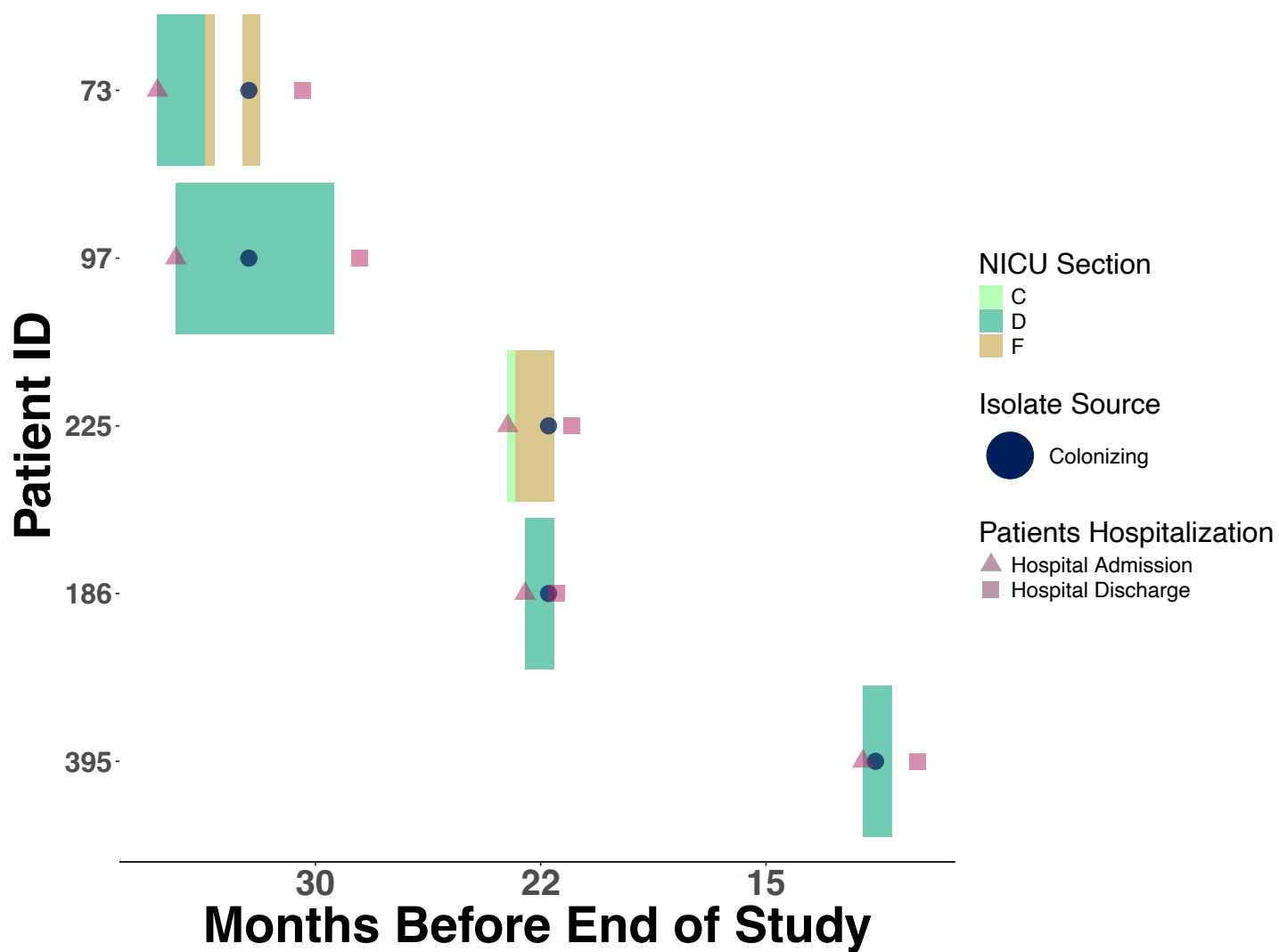
Cluster10



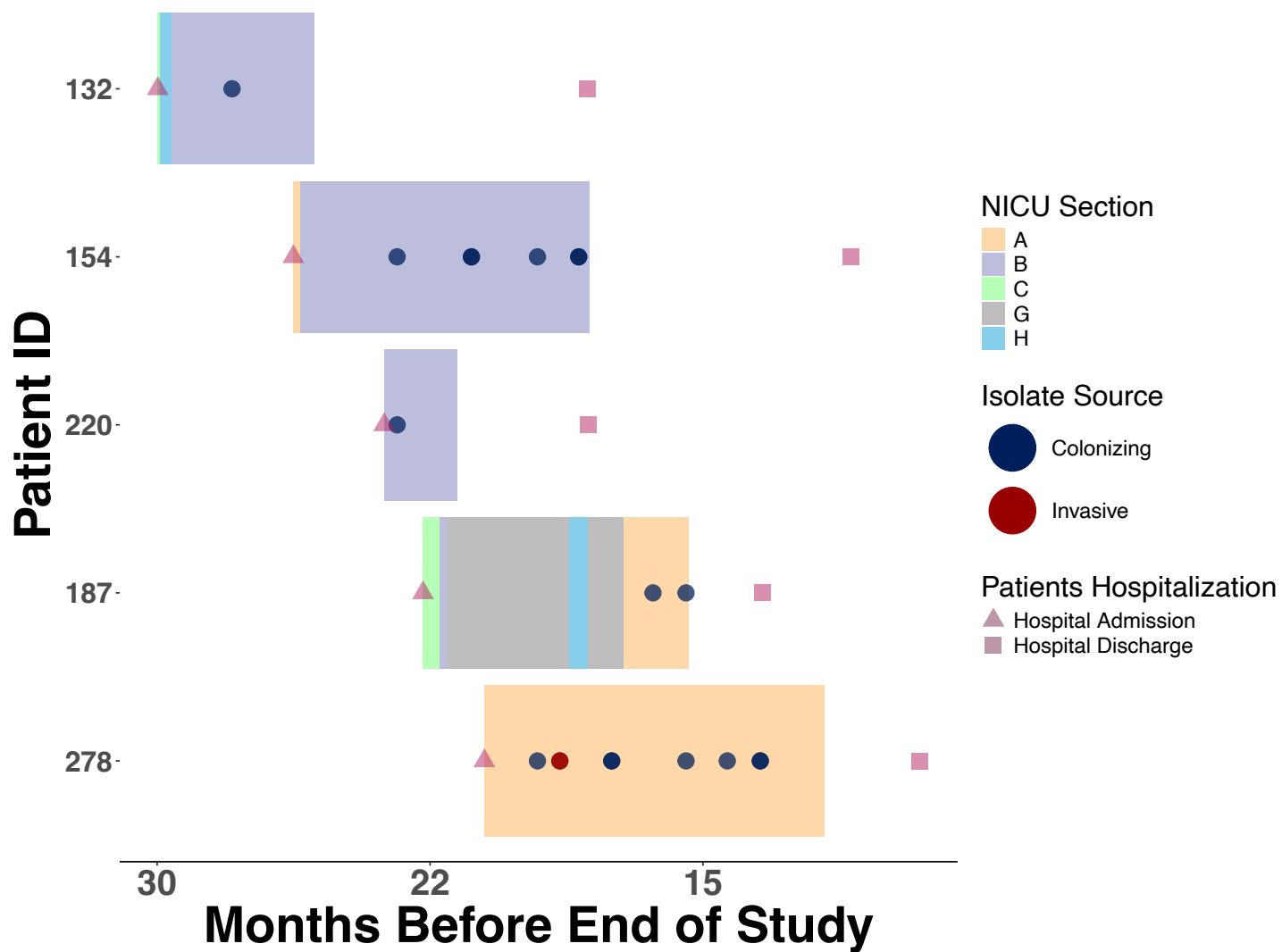
Cluster11



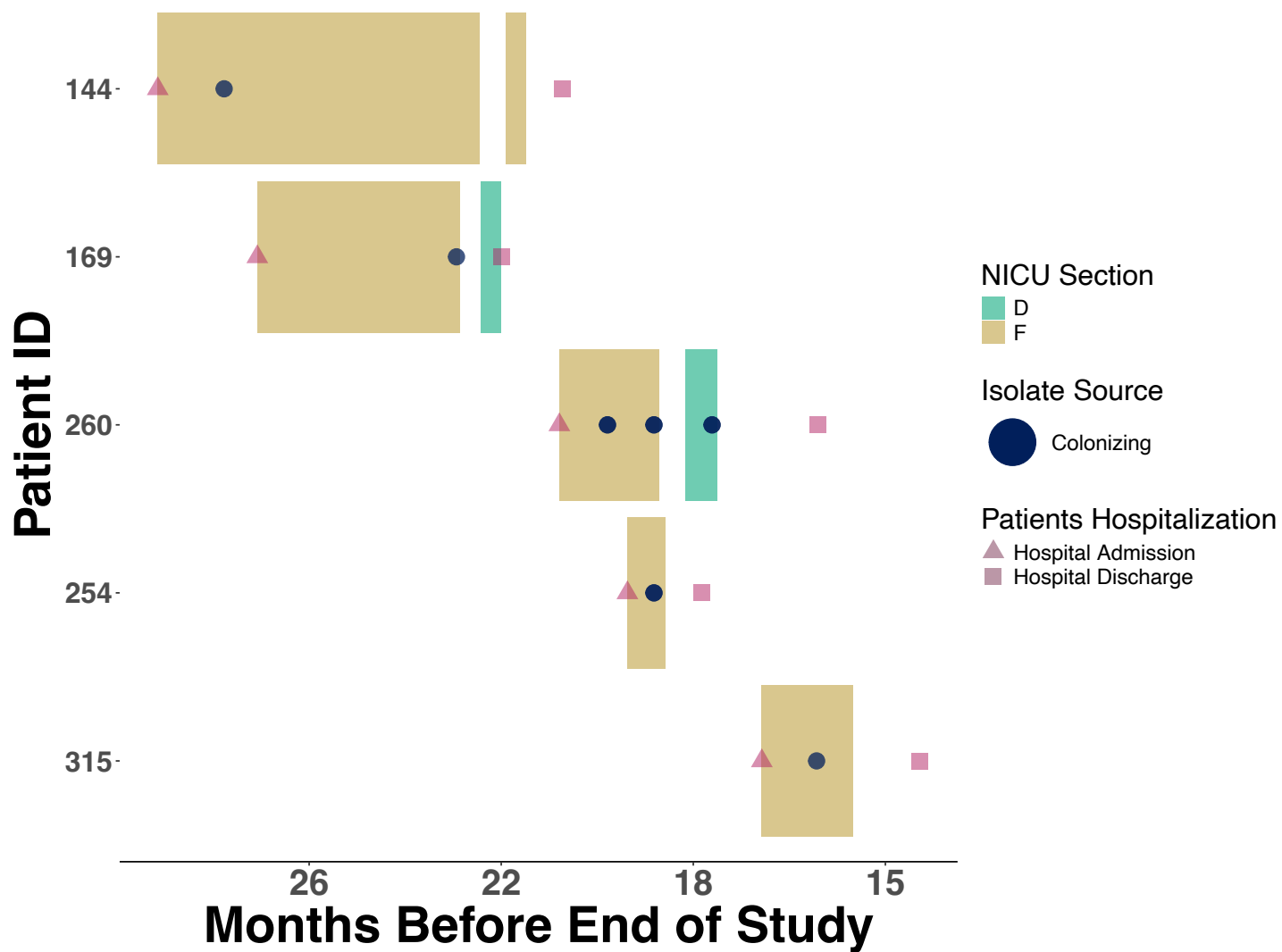
Cluster12



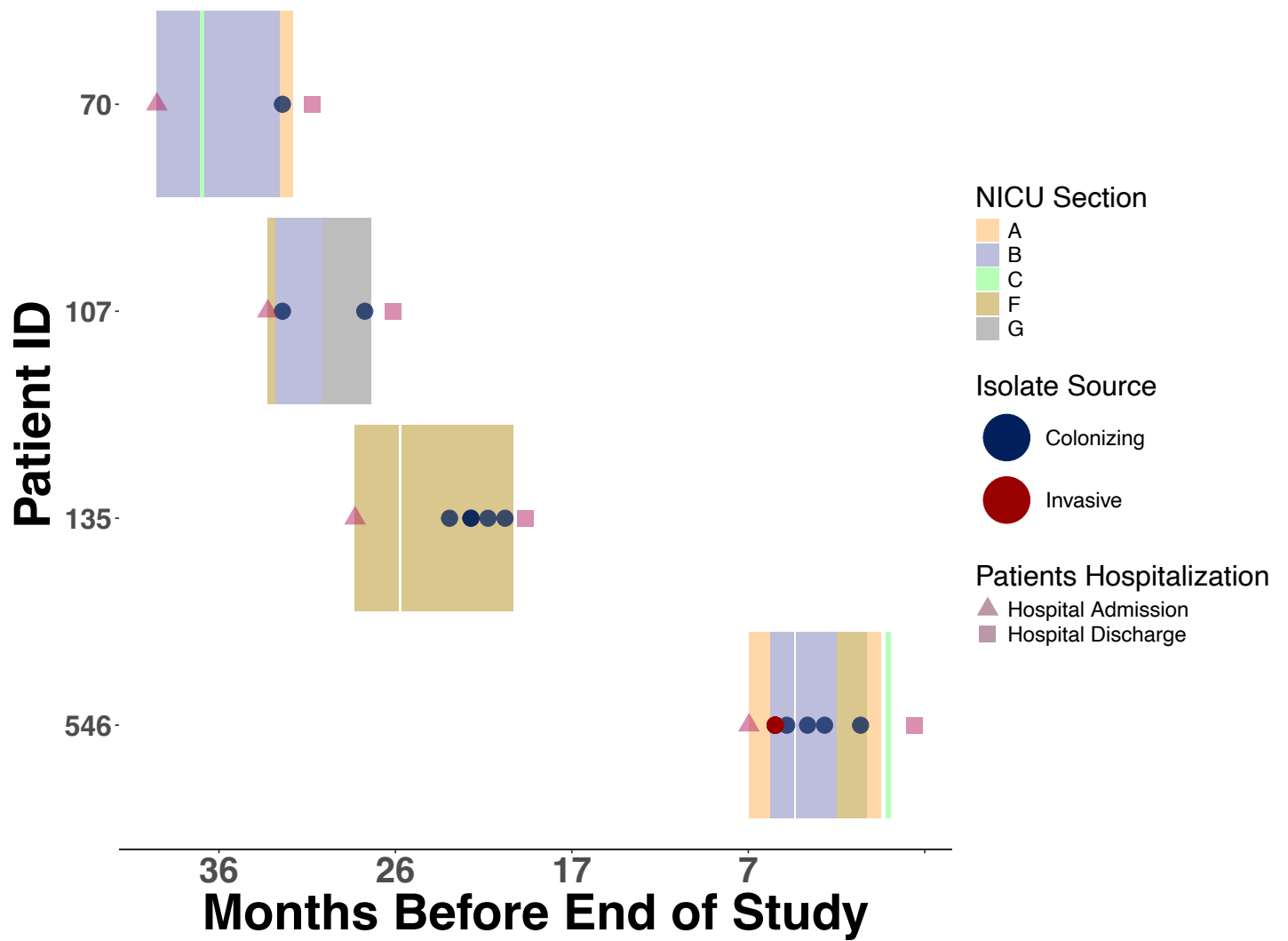
Cluster13



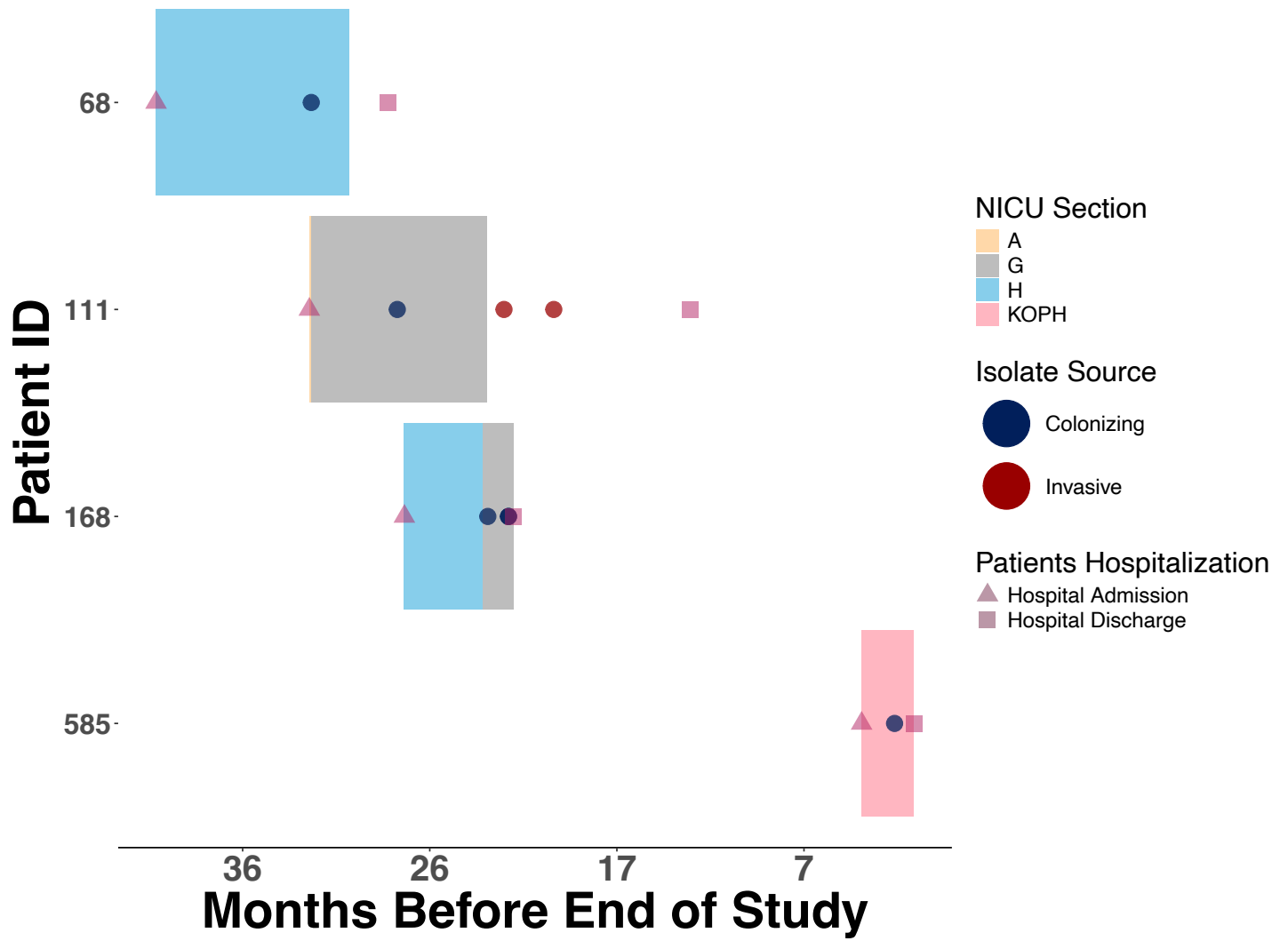
Cluster14



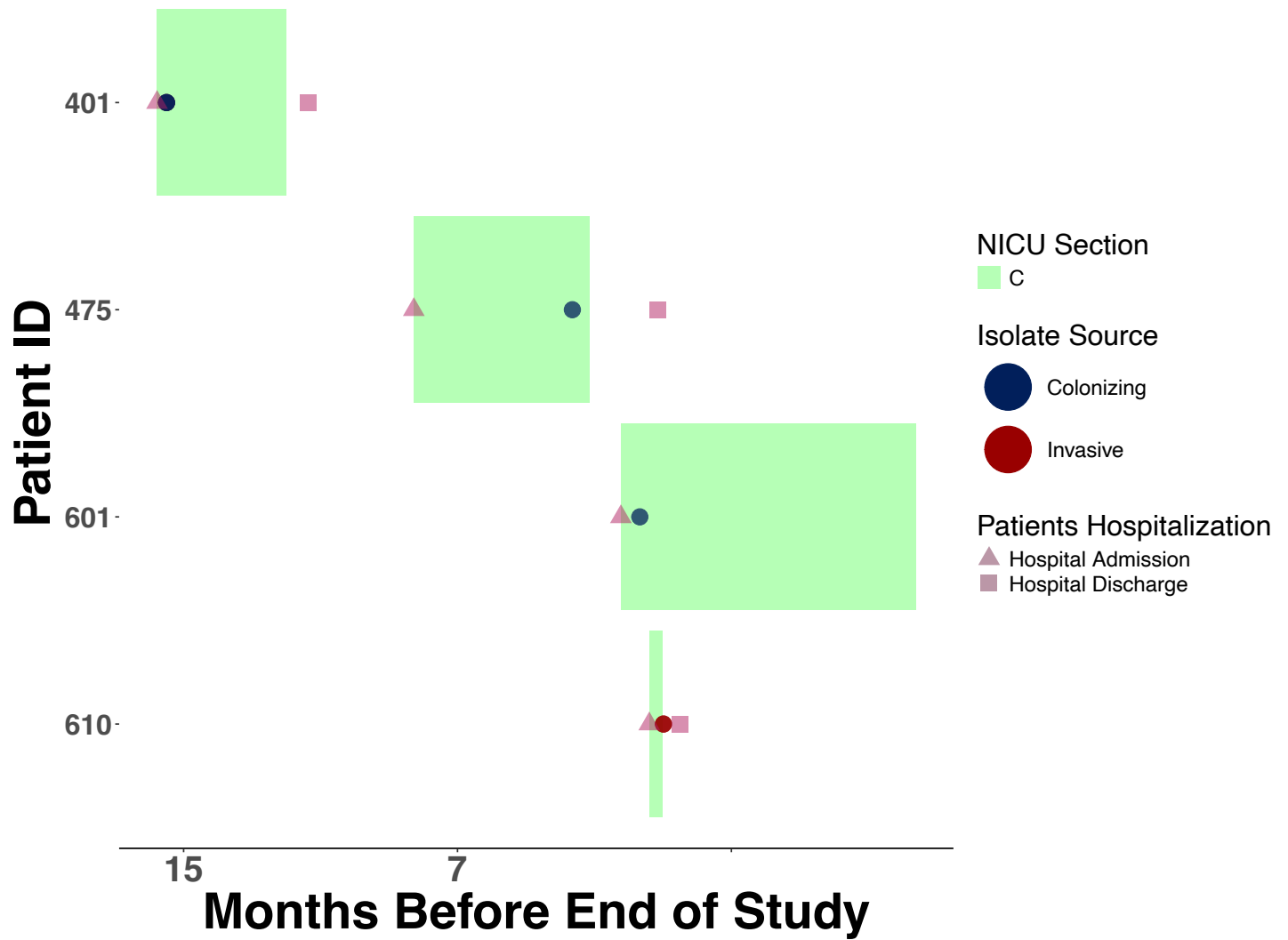
Cluster15



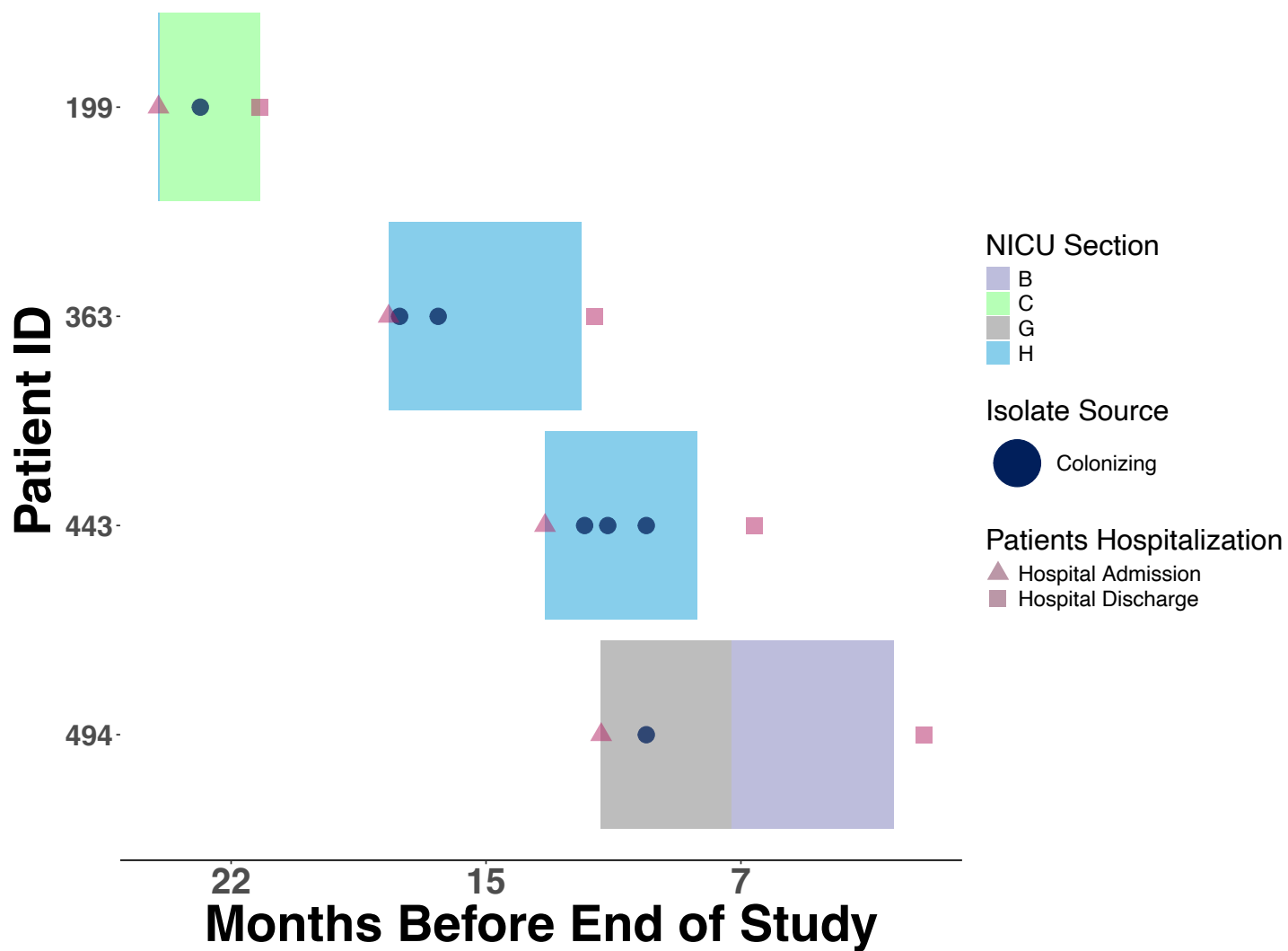
Cluster16



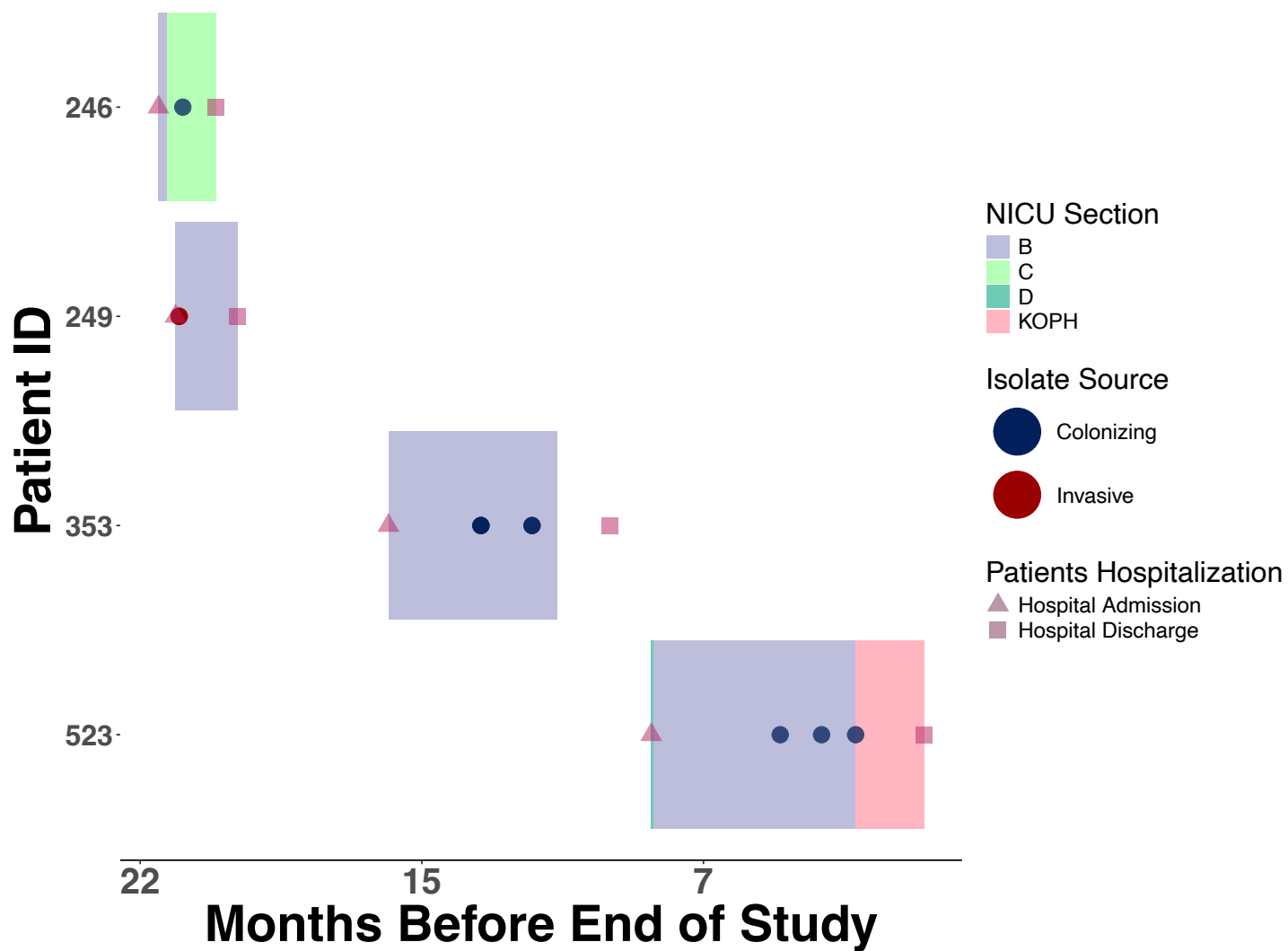
Cluster17



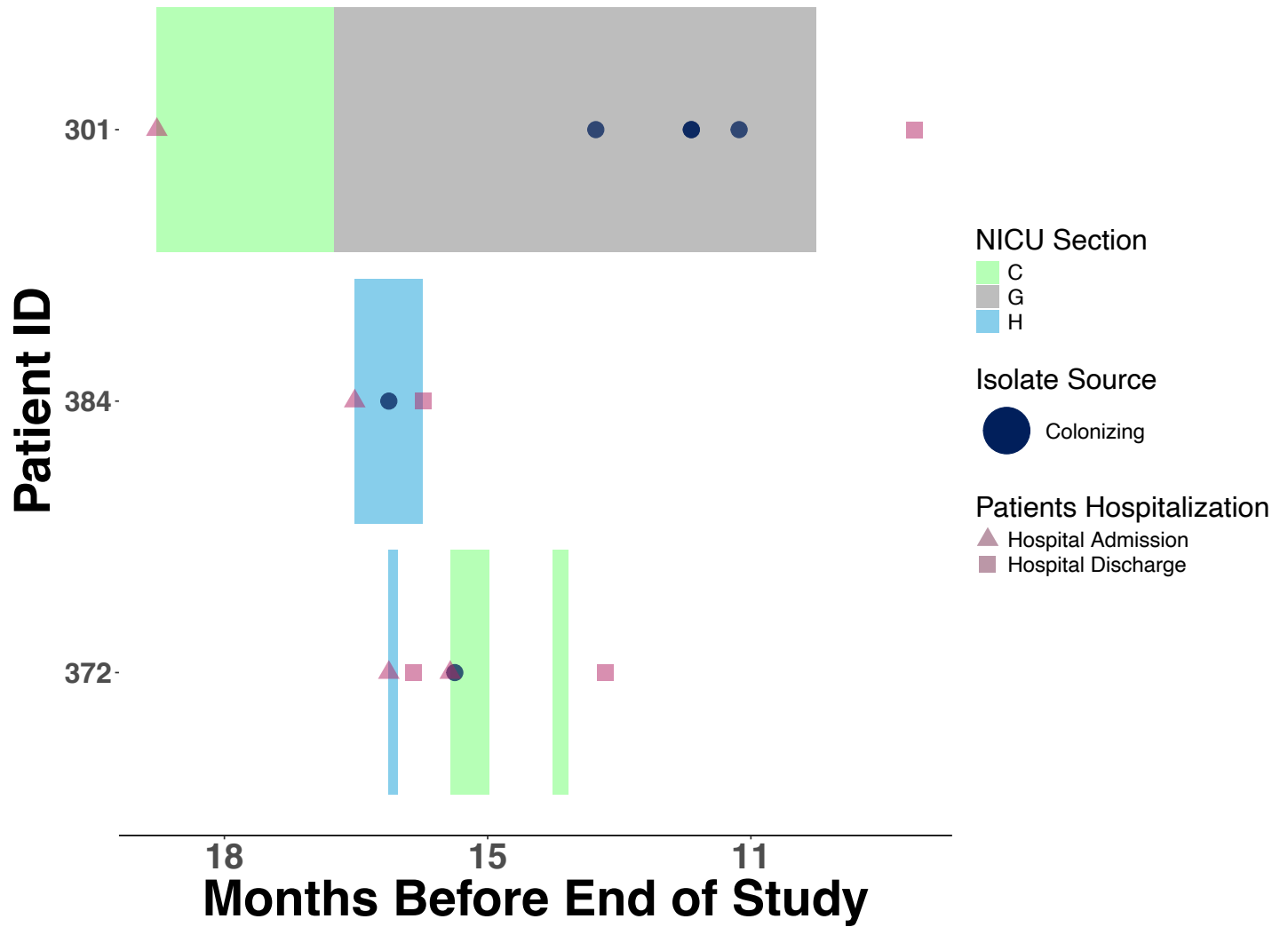
Cluster18



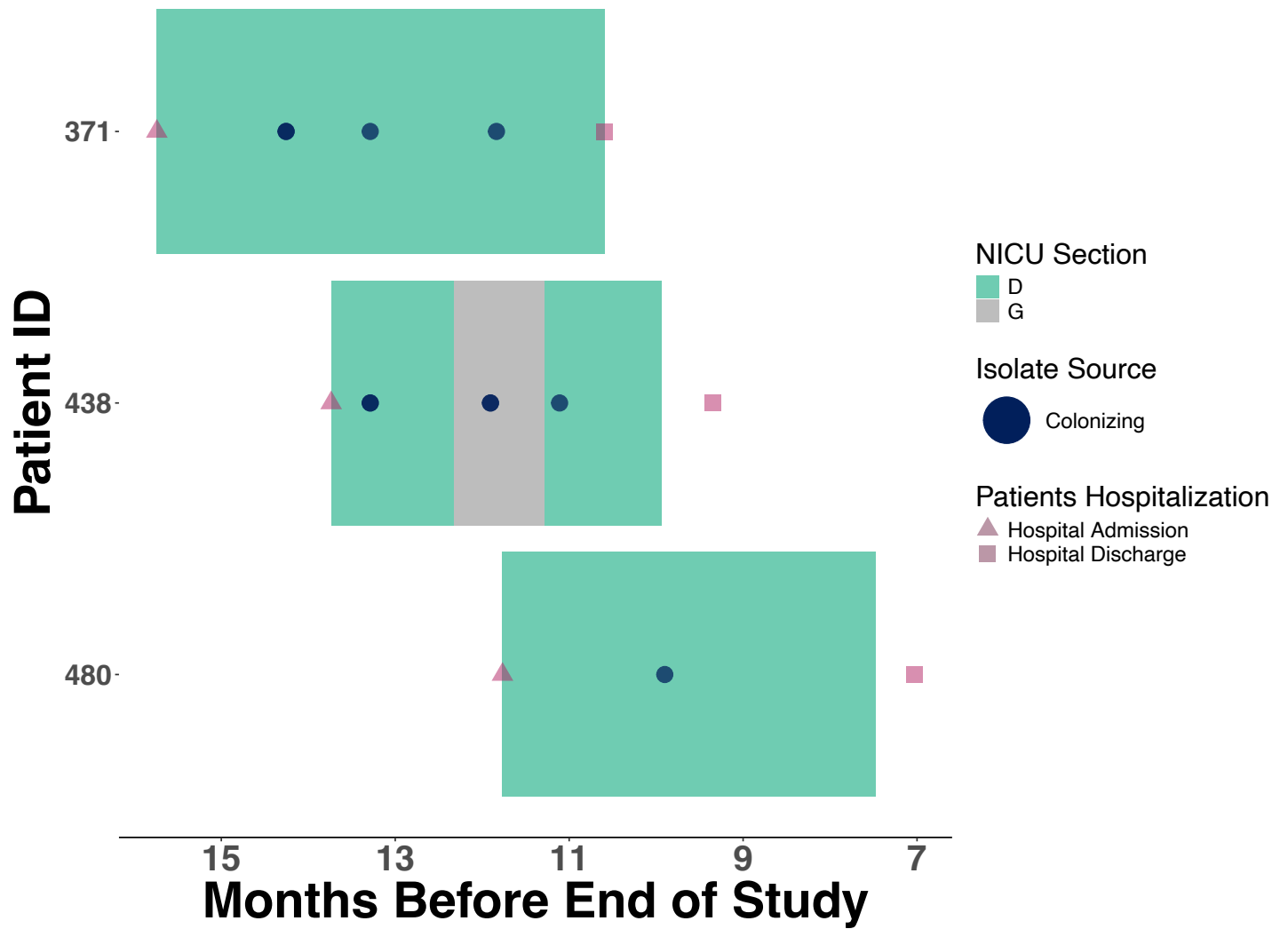
Cluster19



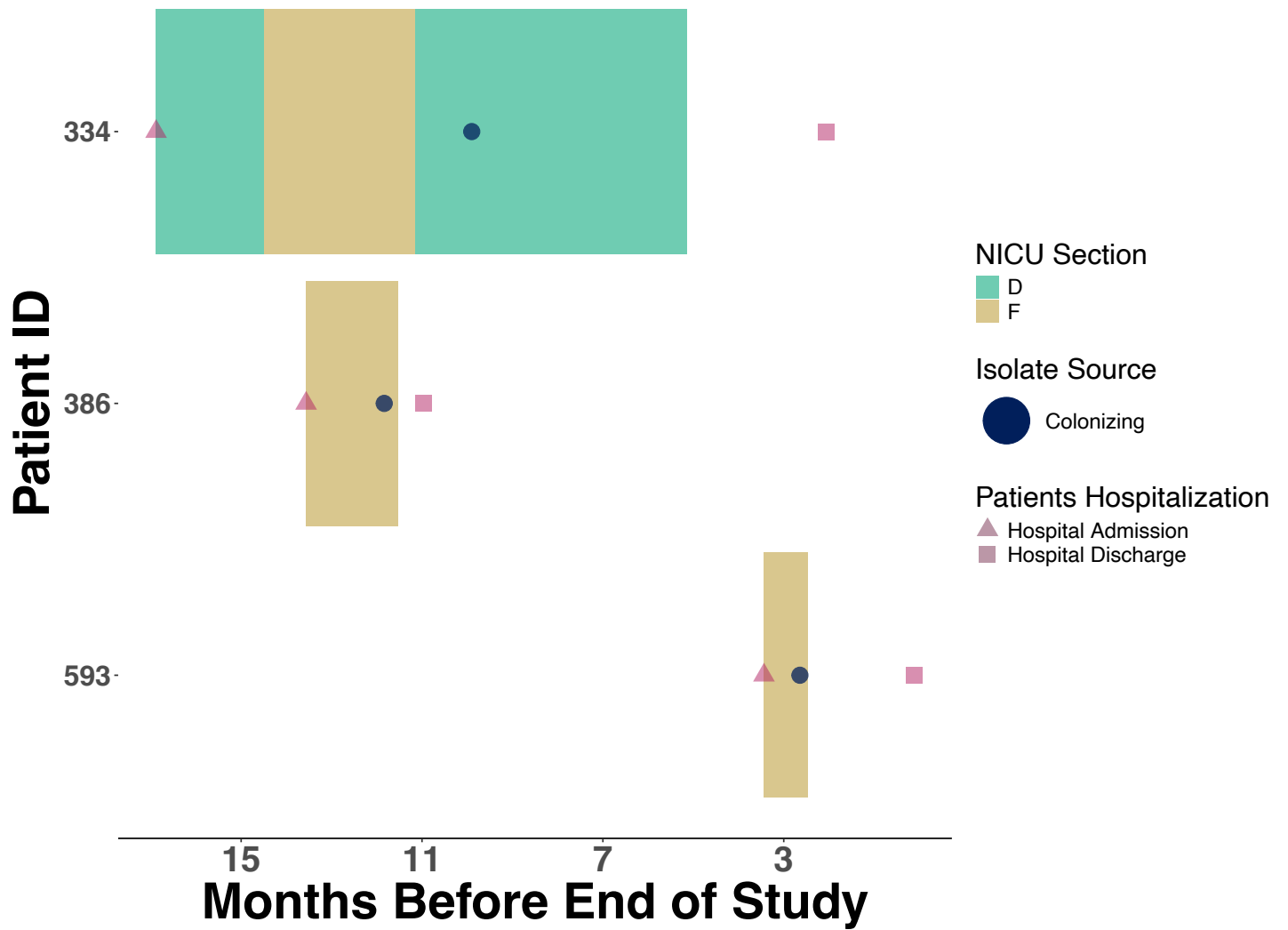
Cluster20



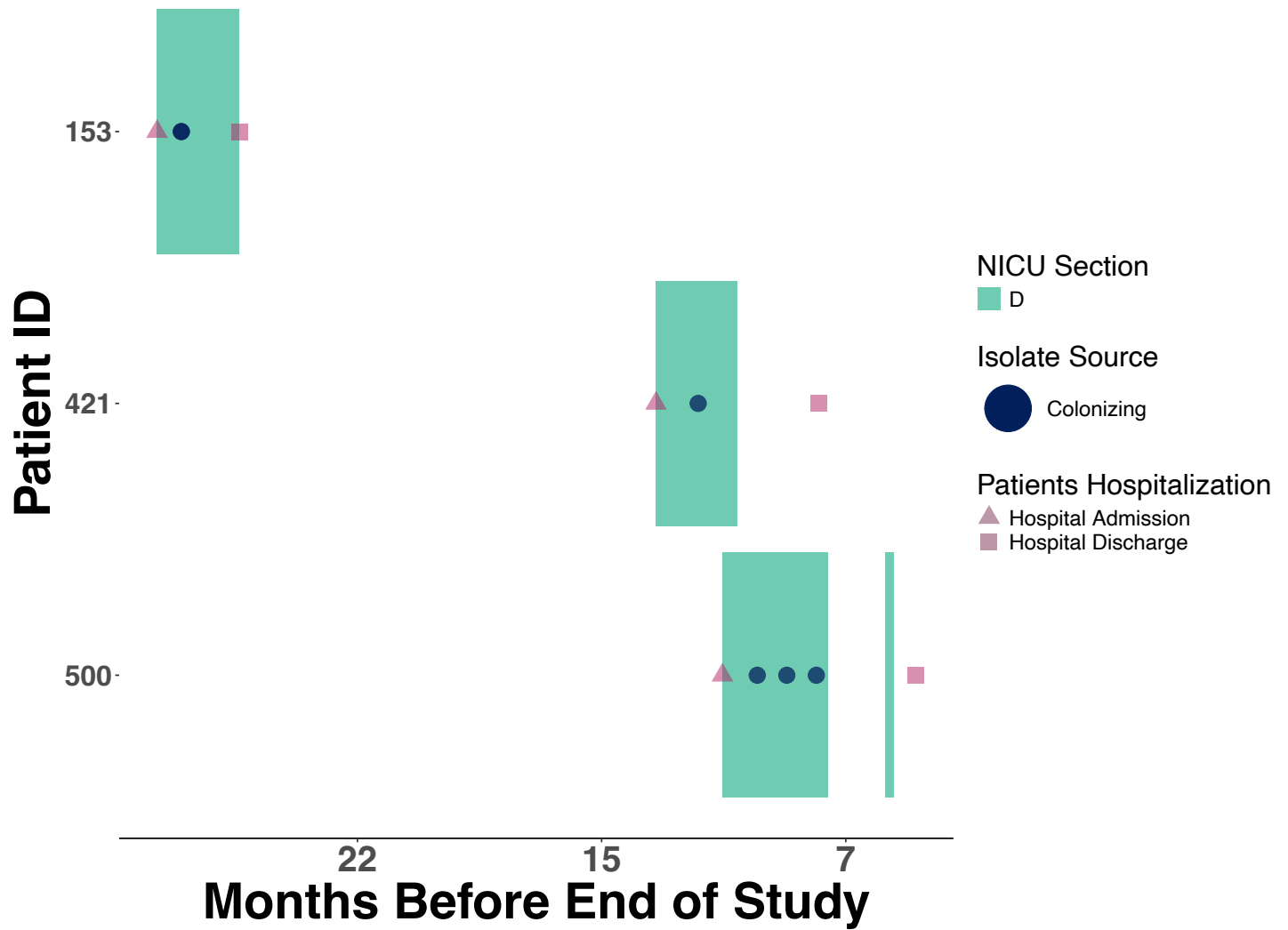
Cluster21



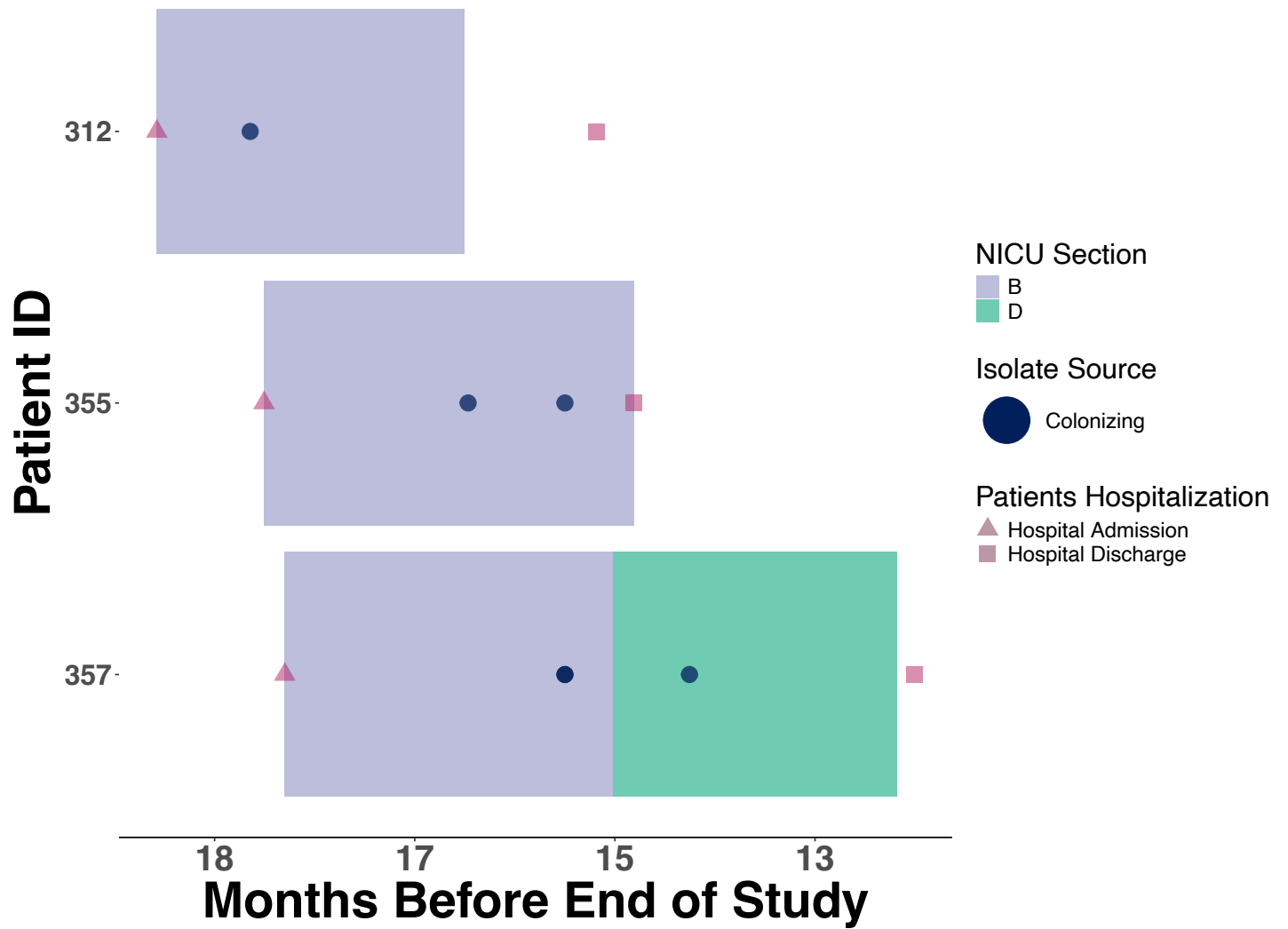
Cluster22



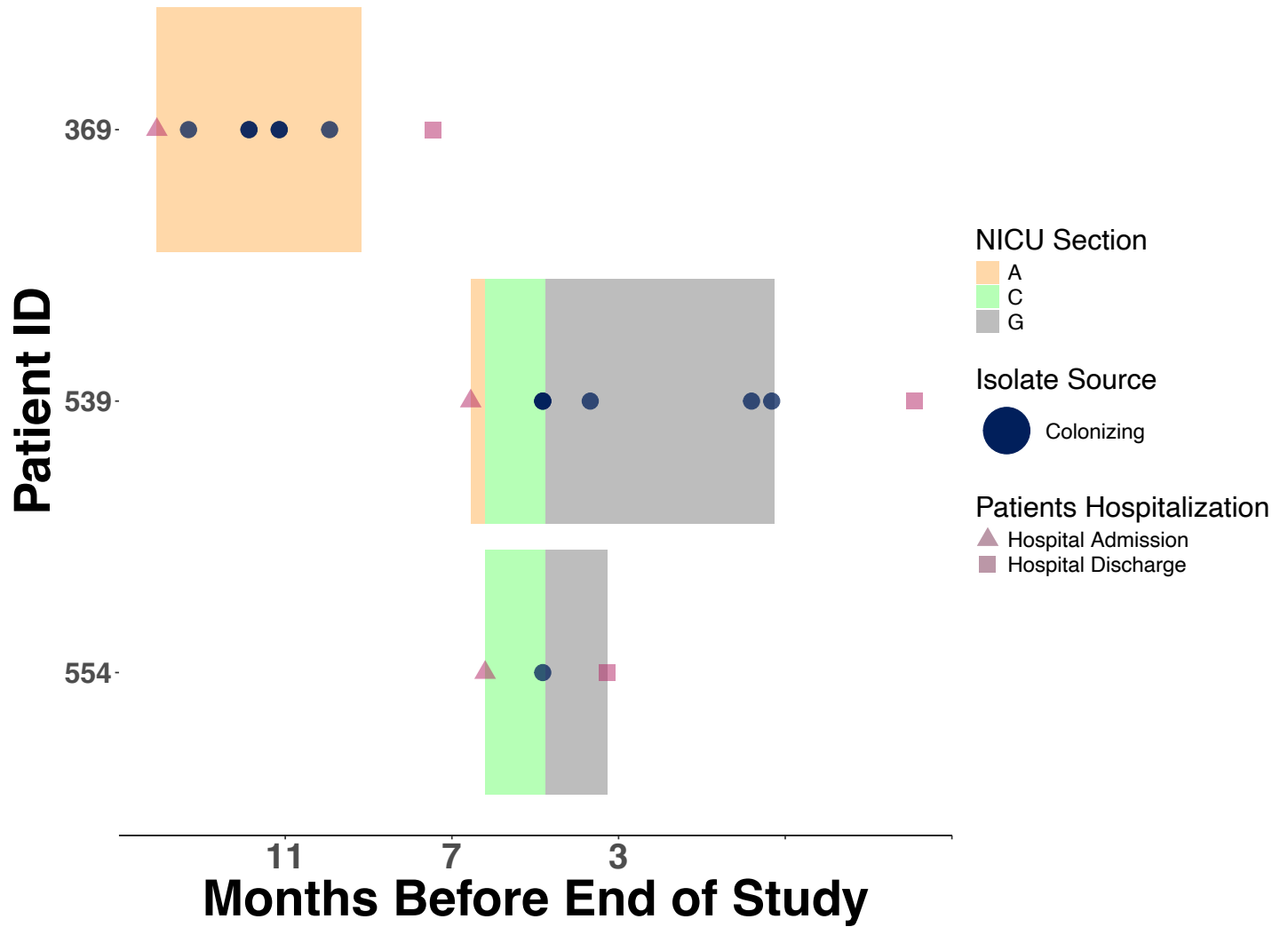
Cluster23



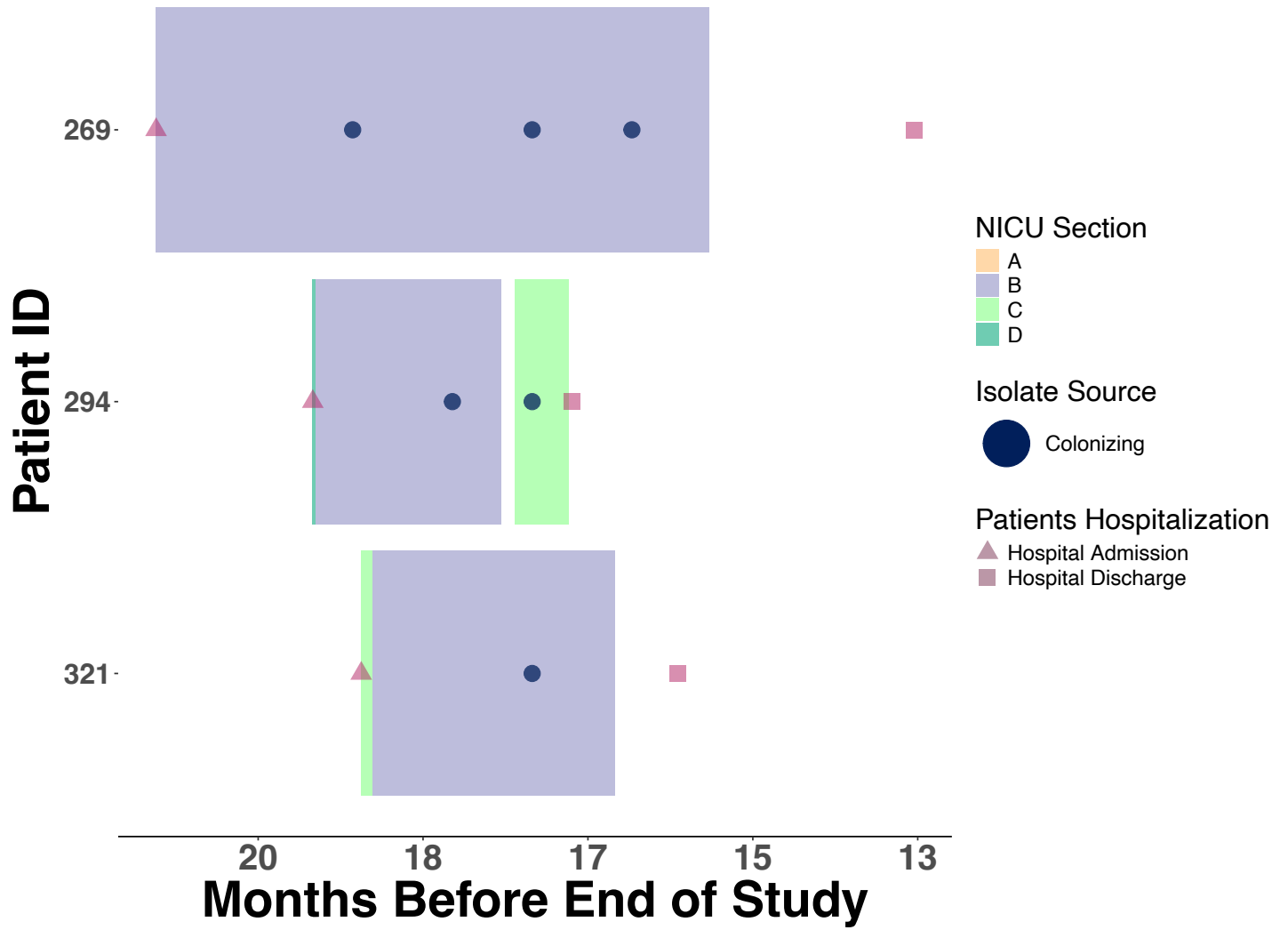
Cluster24



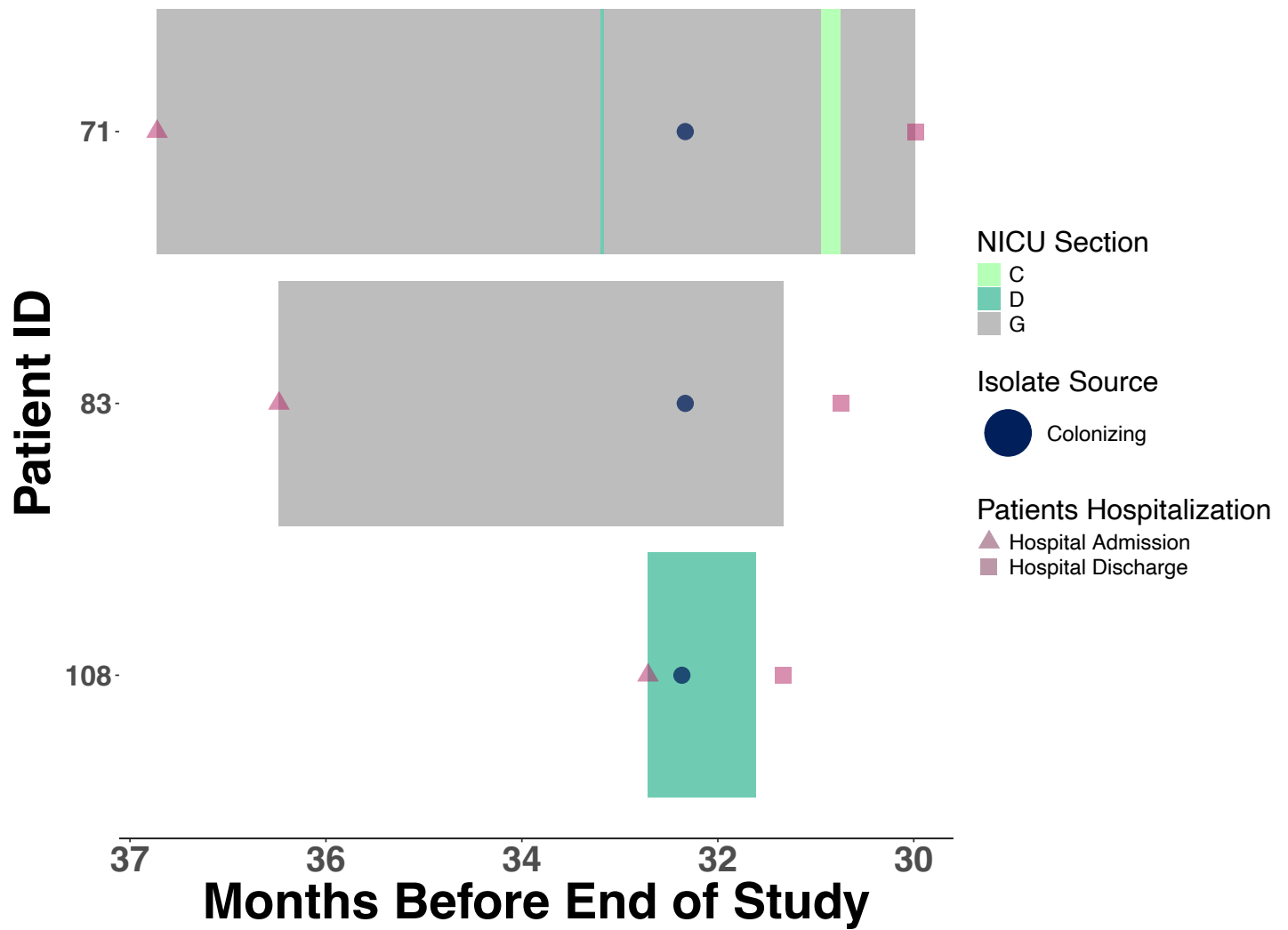
Cluster25



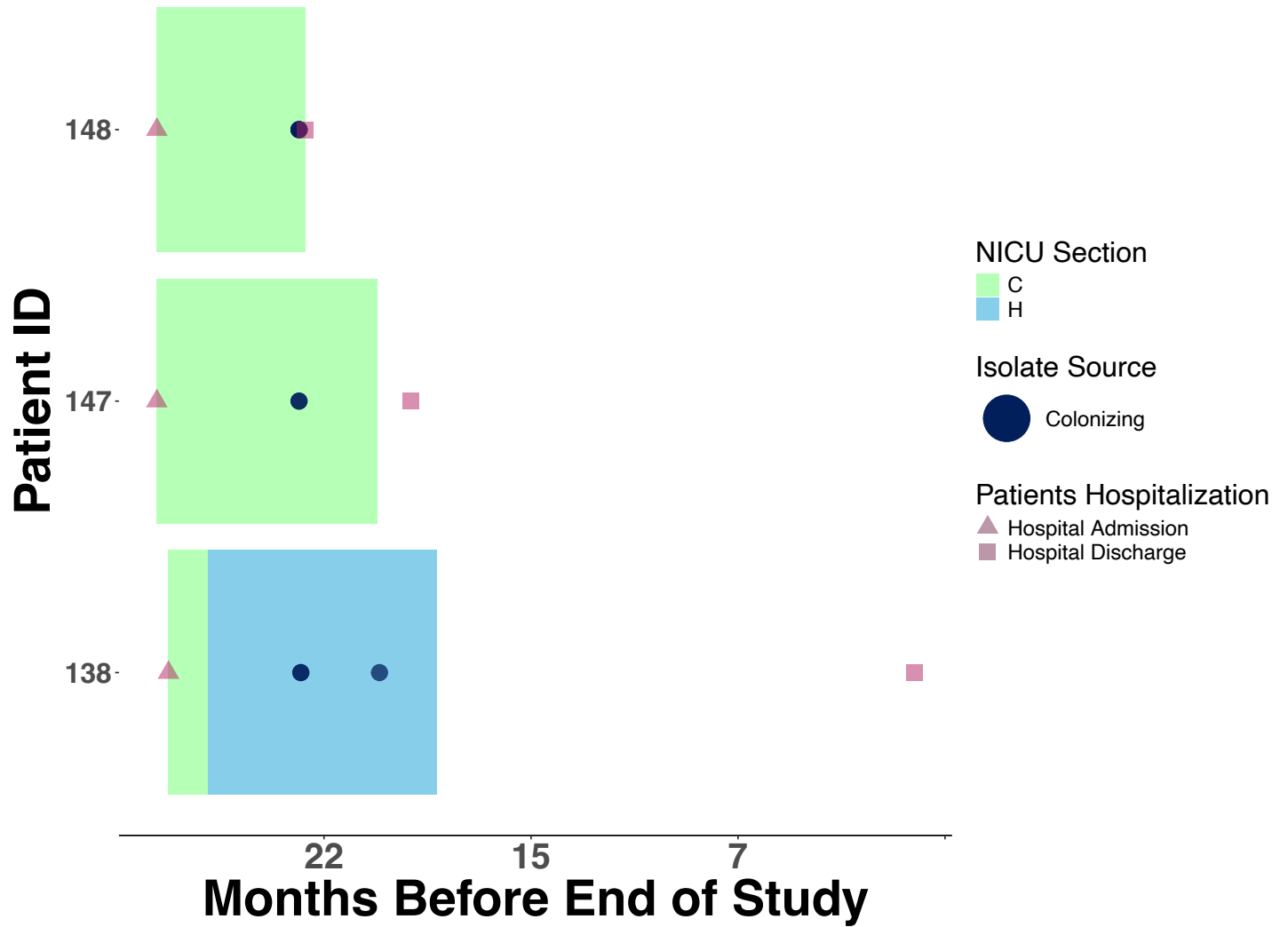
Cluster26



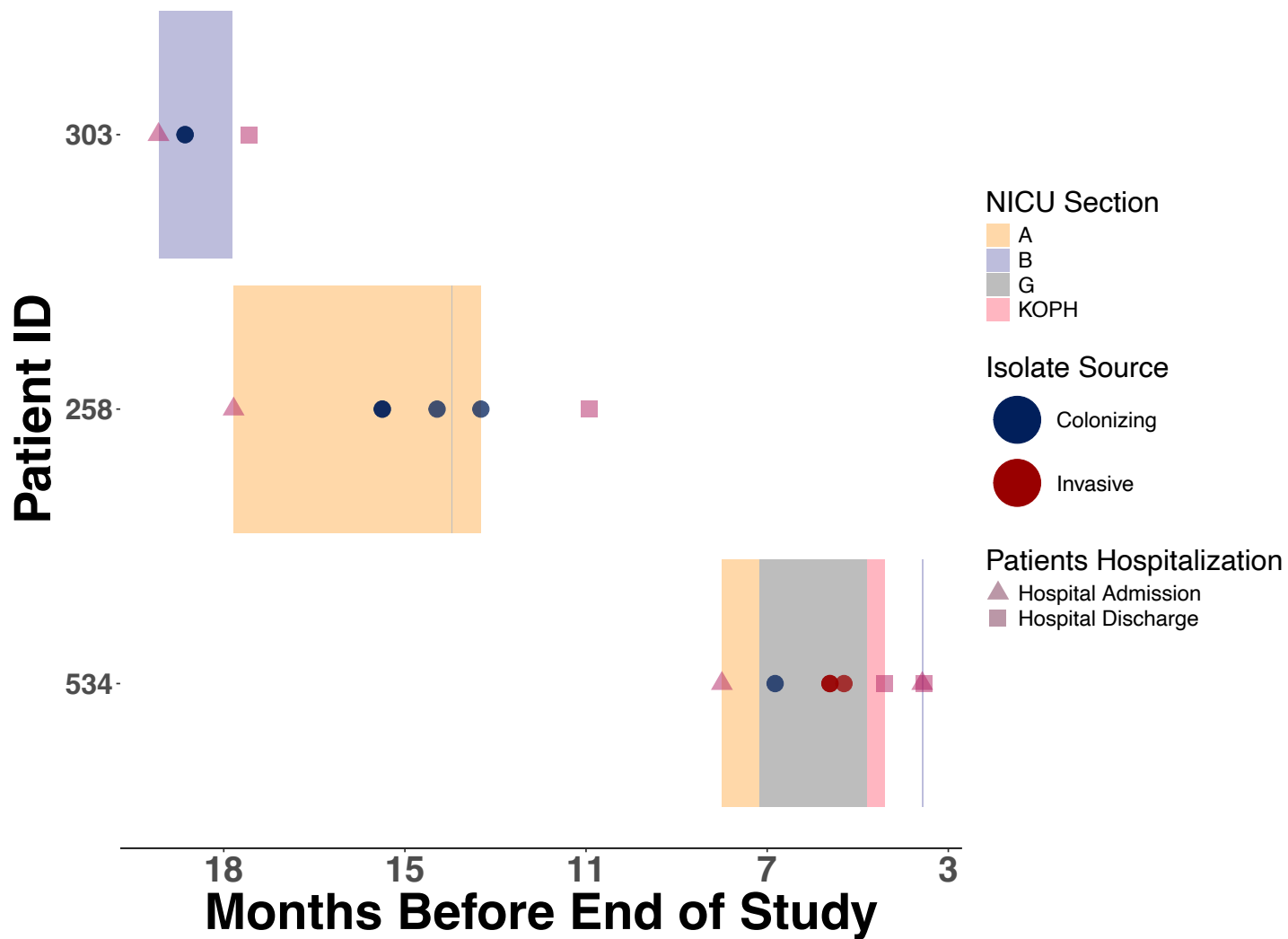
Cluster27



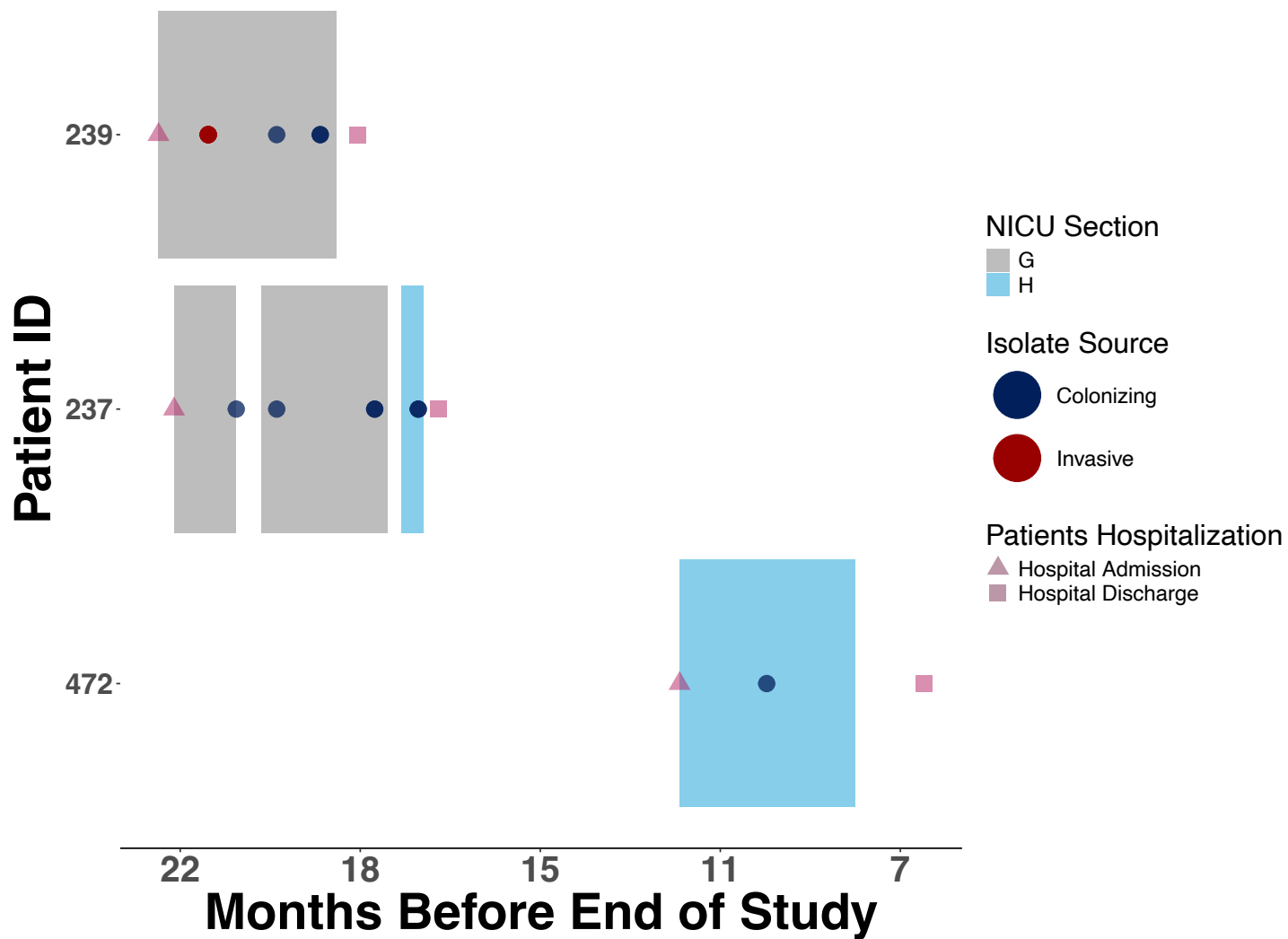
Cluster28



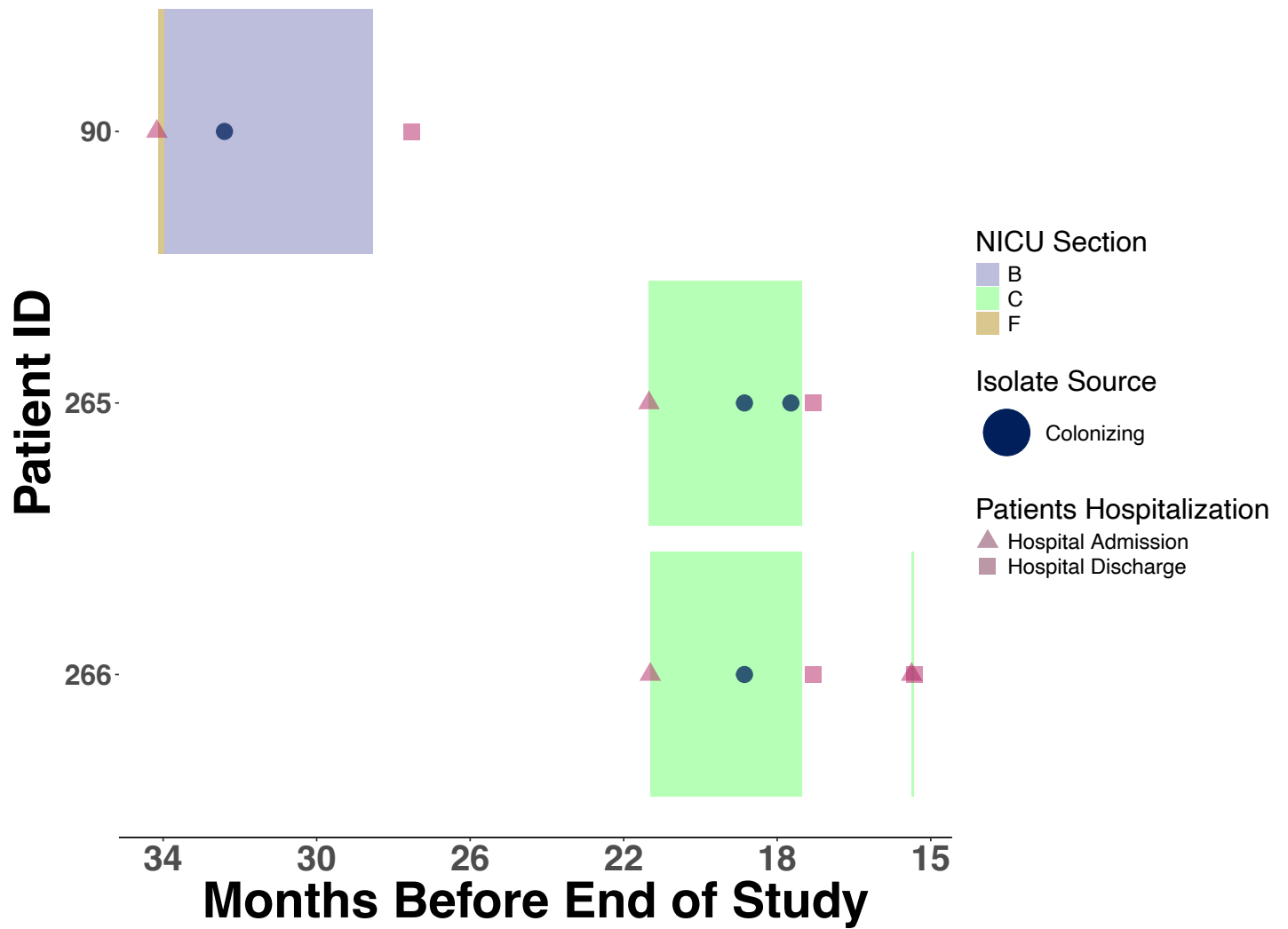
Cluster29



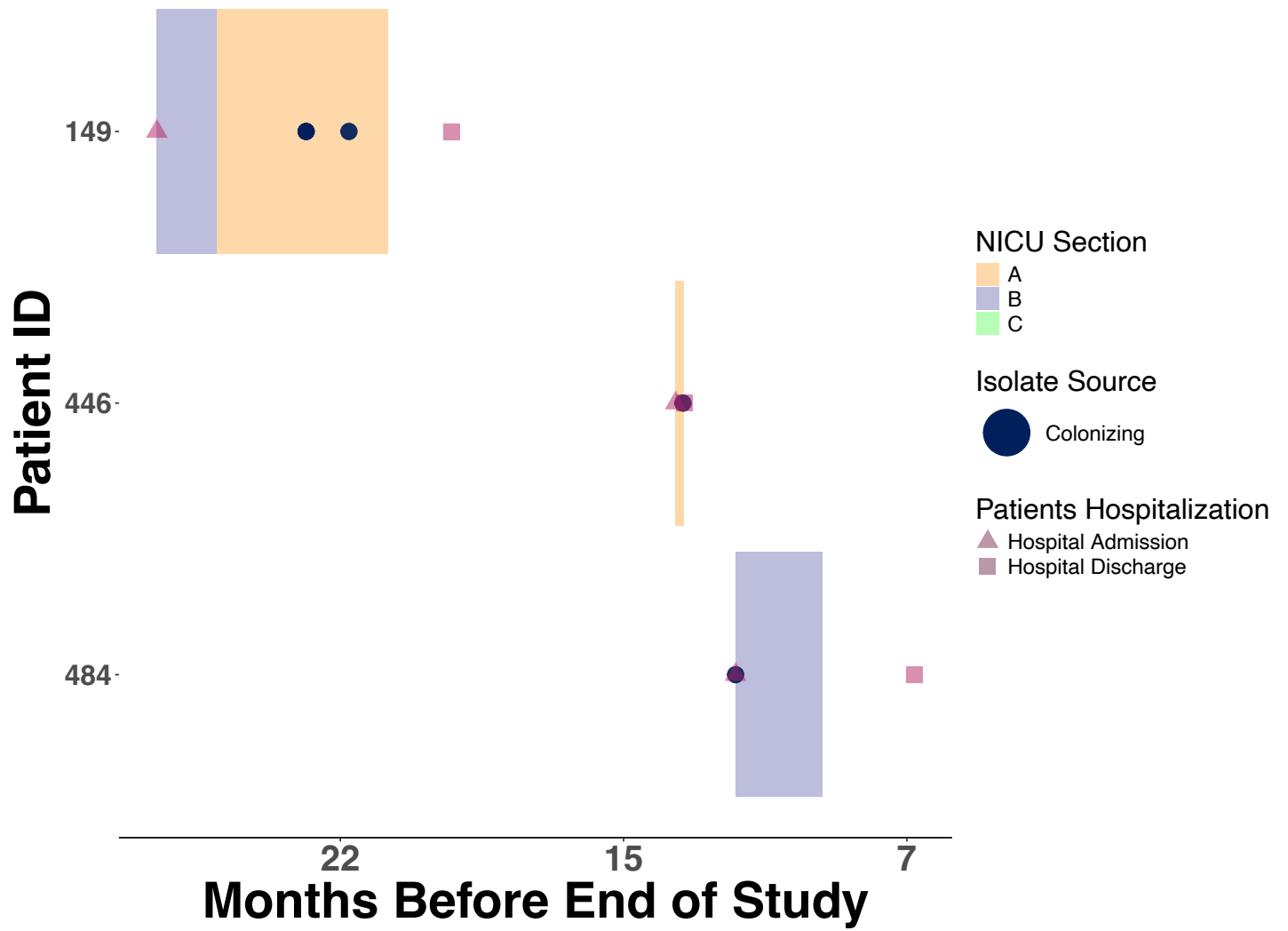
Cluster30



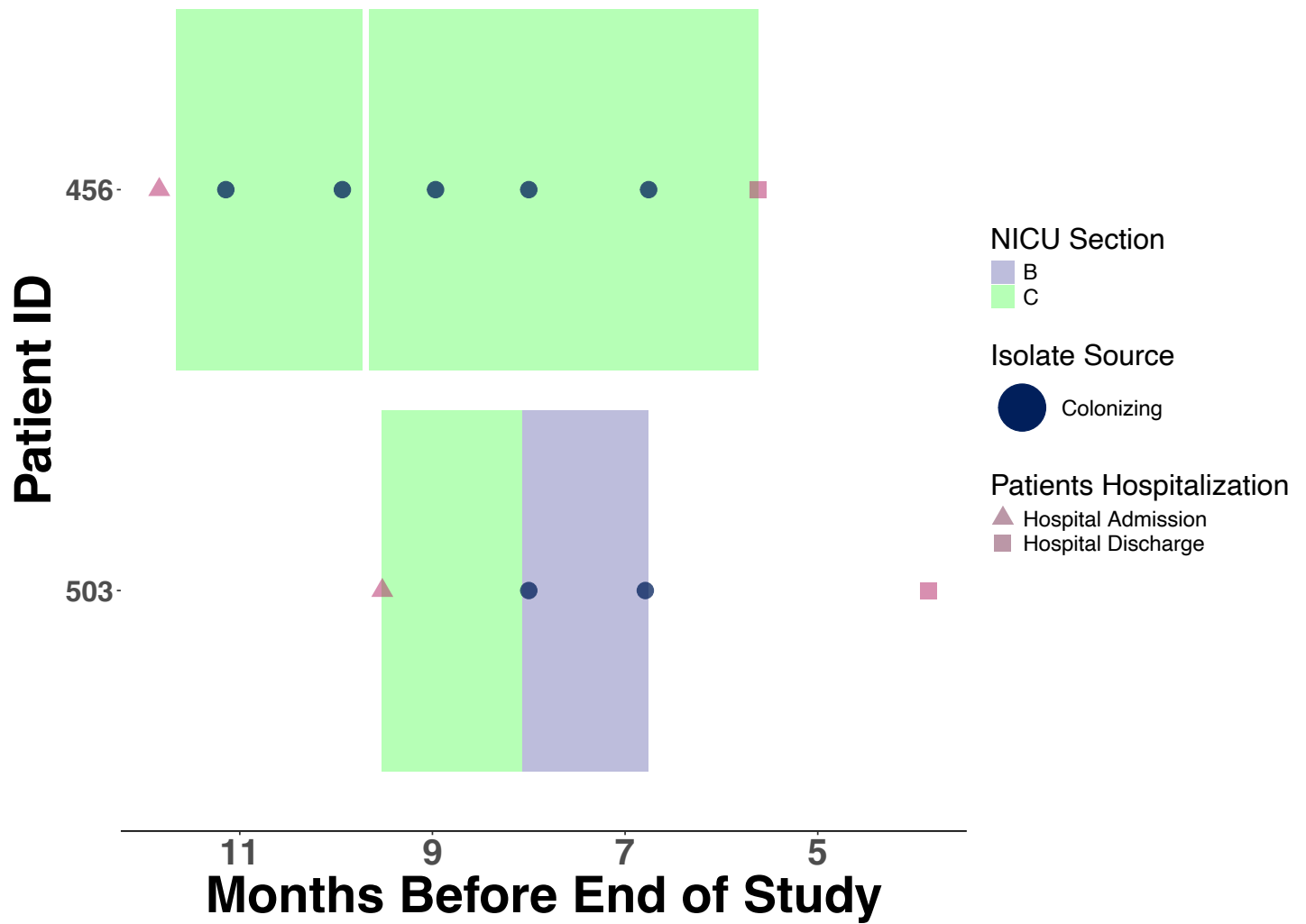
Cluster31



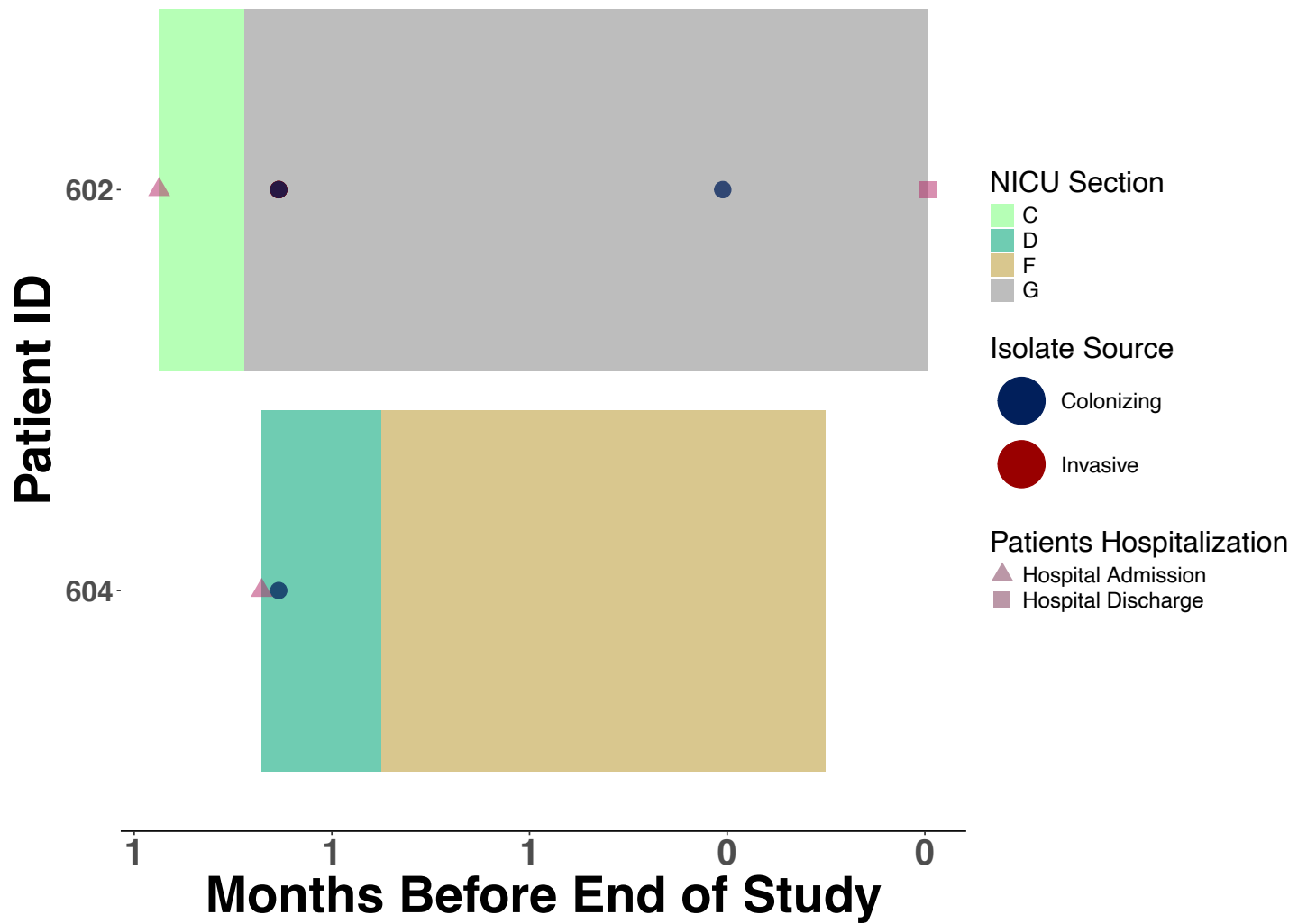
Cluster32



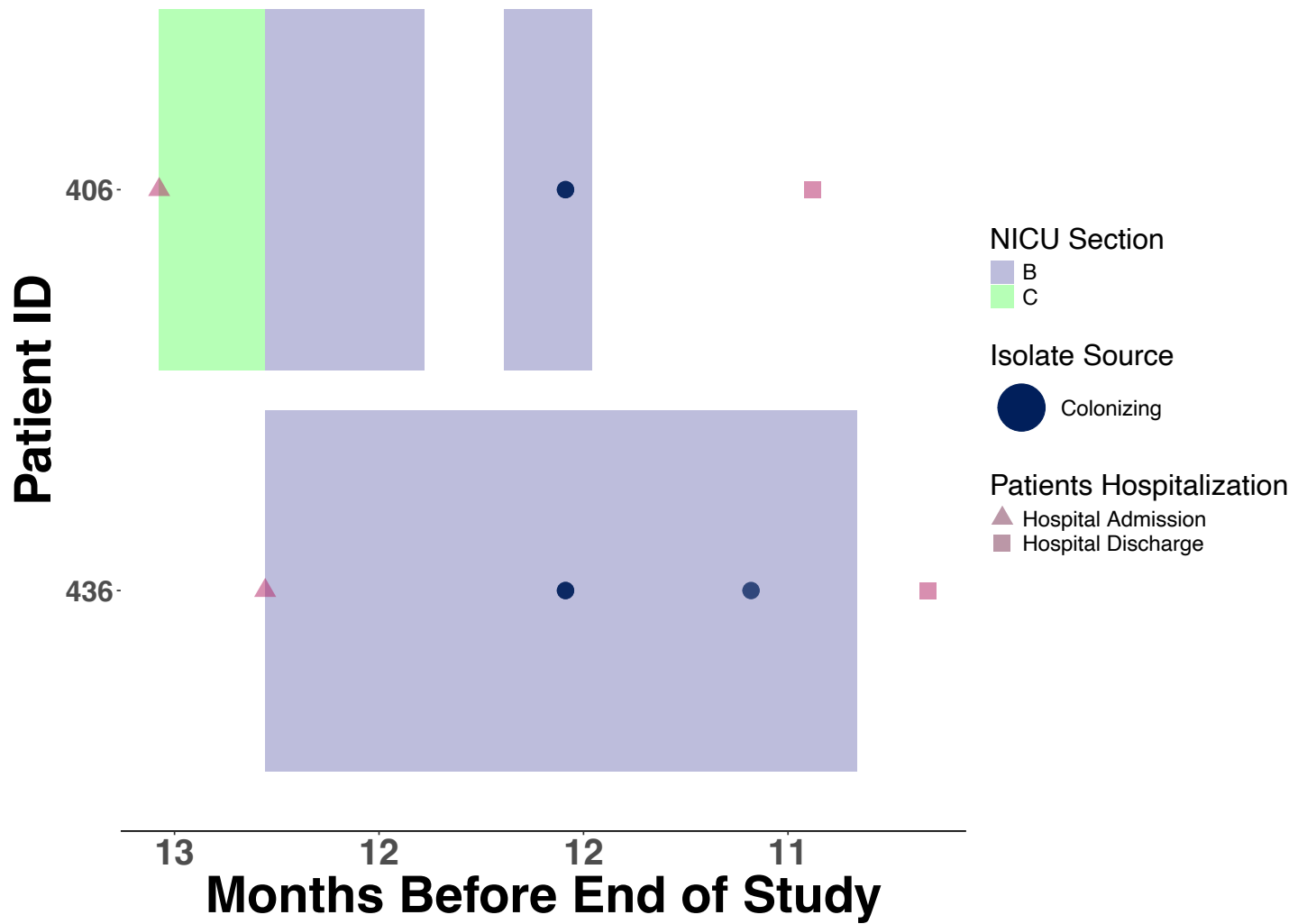
Cluster33



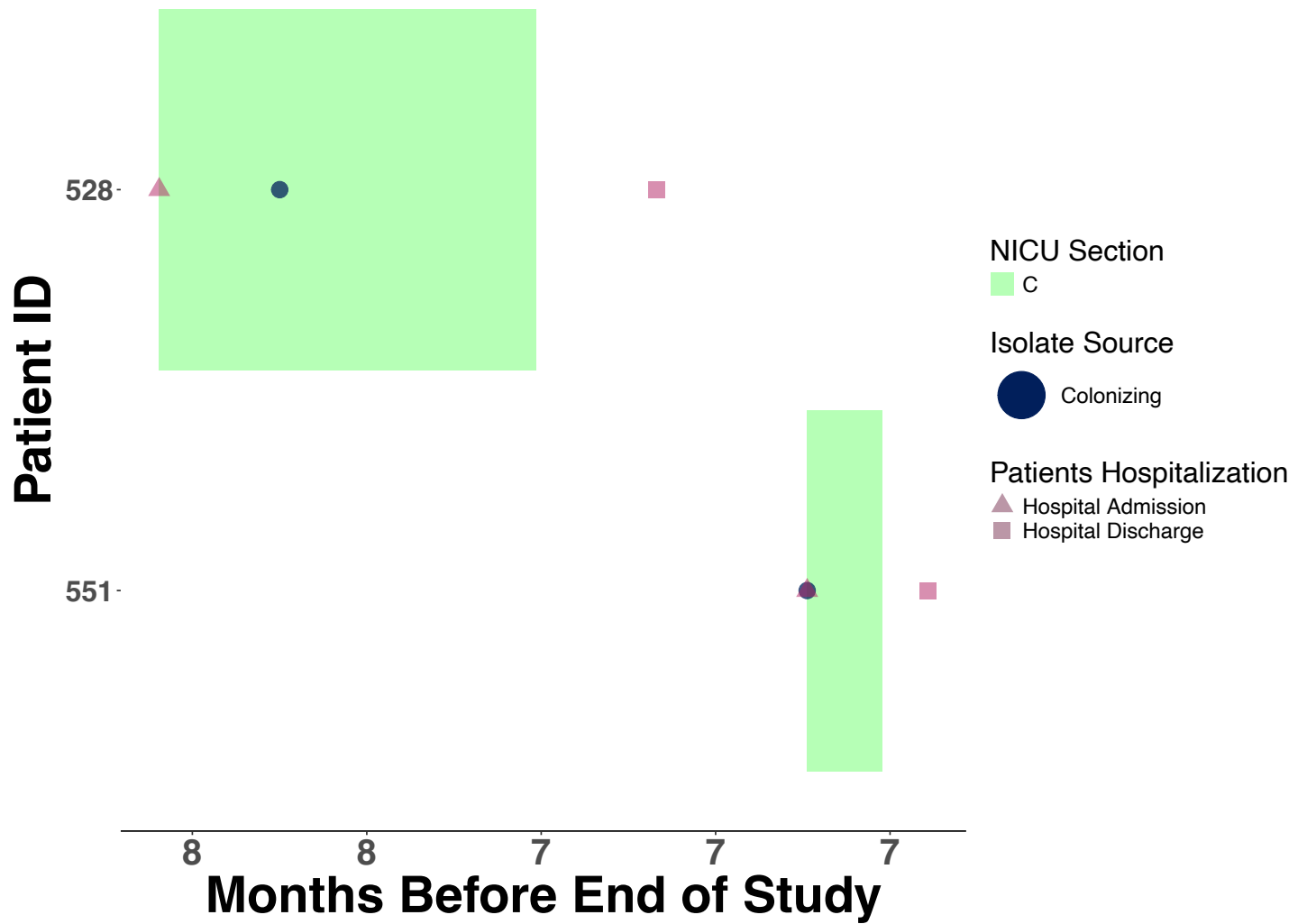
Cluster34



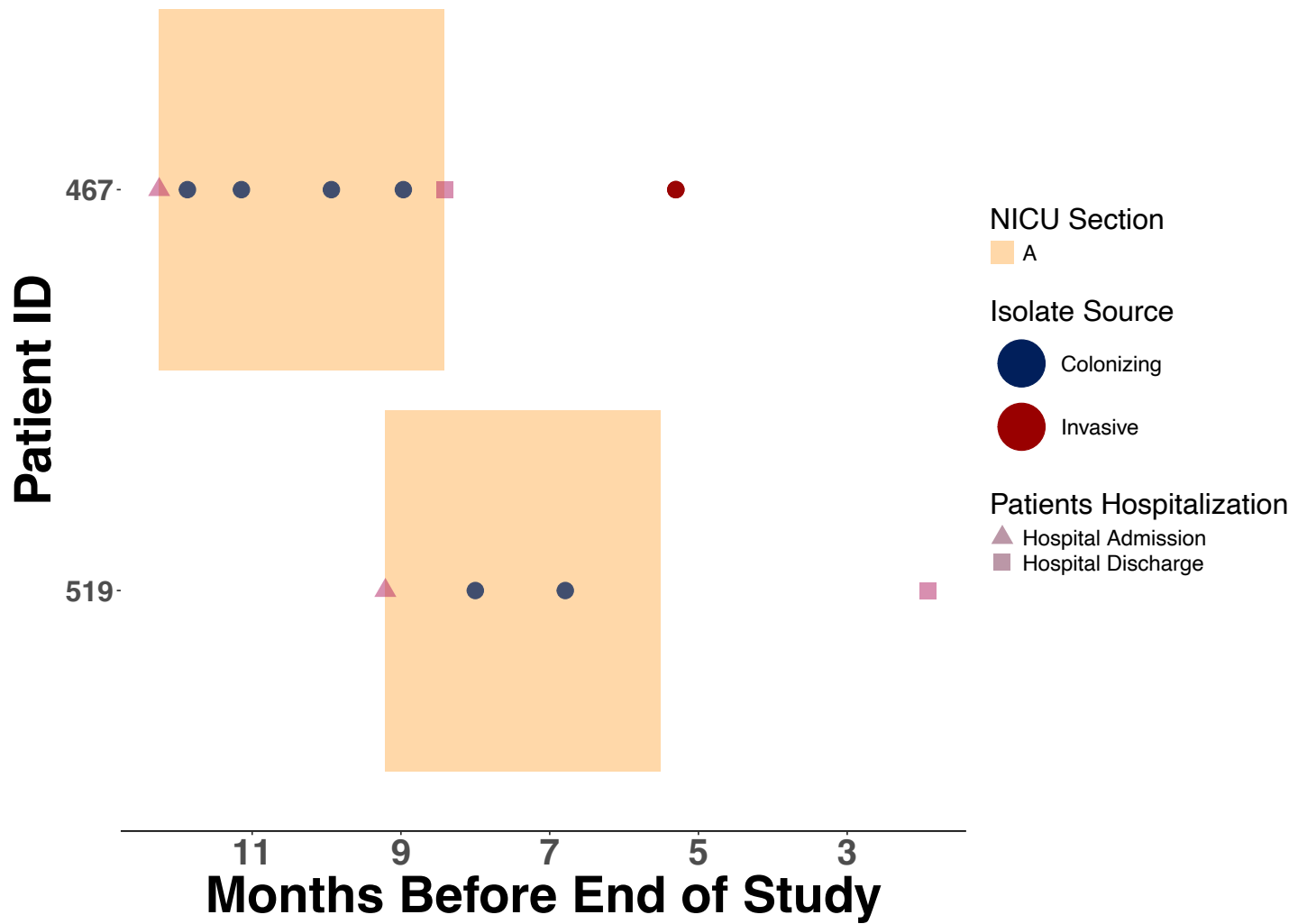
Cluster35



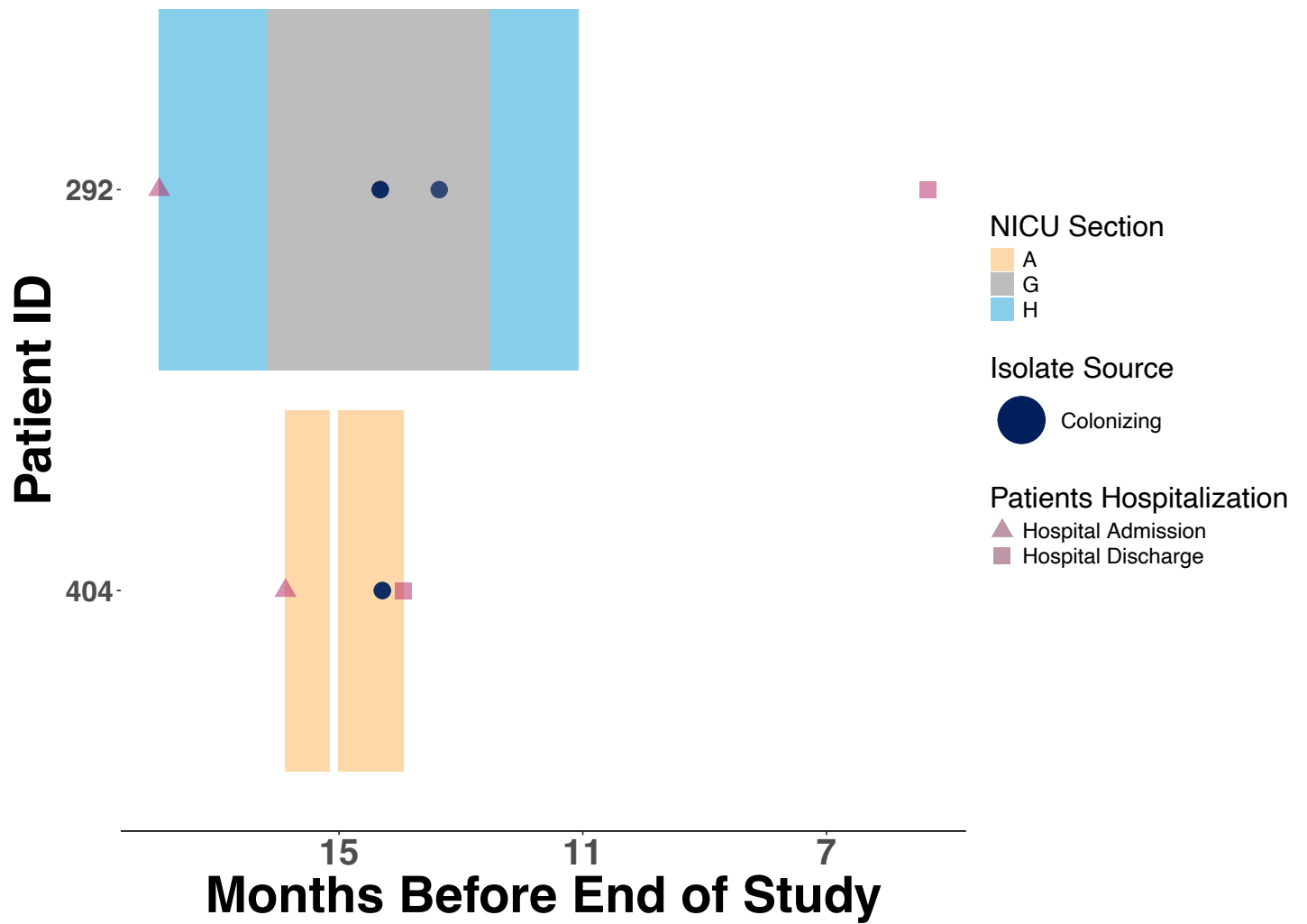
Cluster36



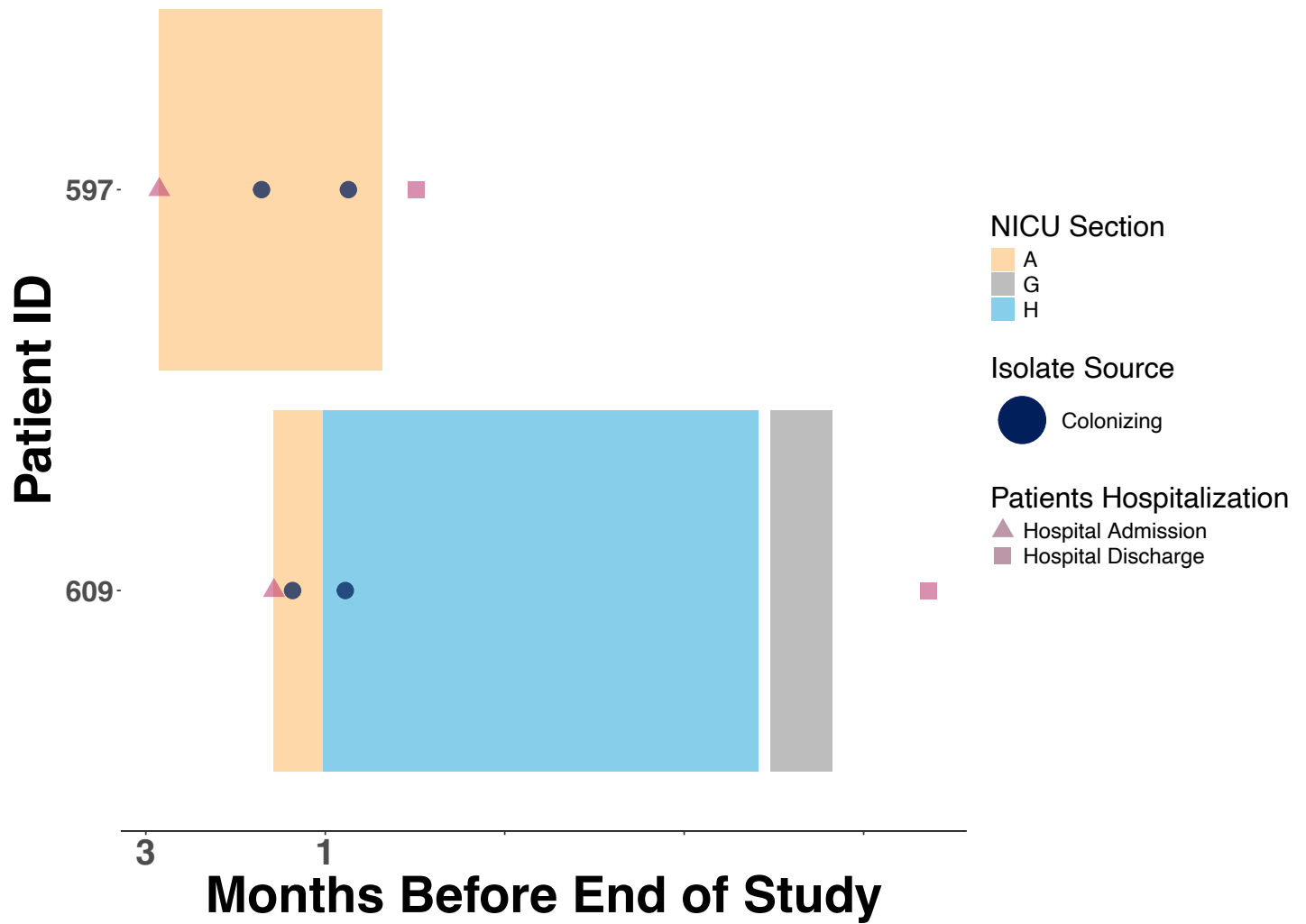
Cluster37



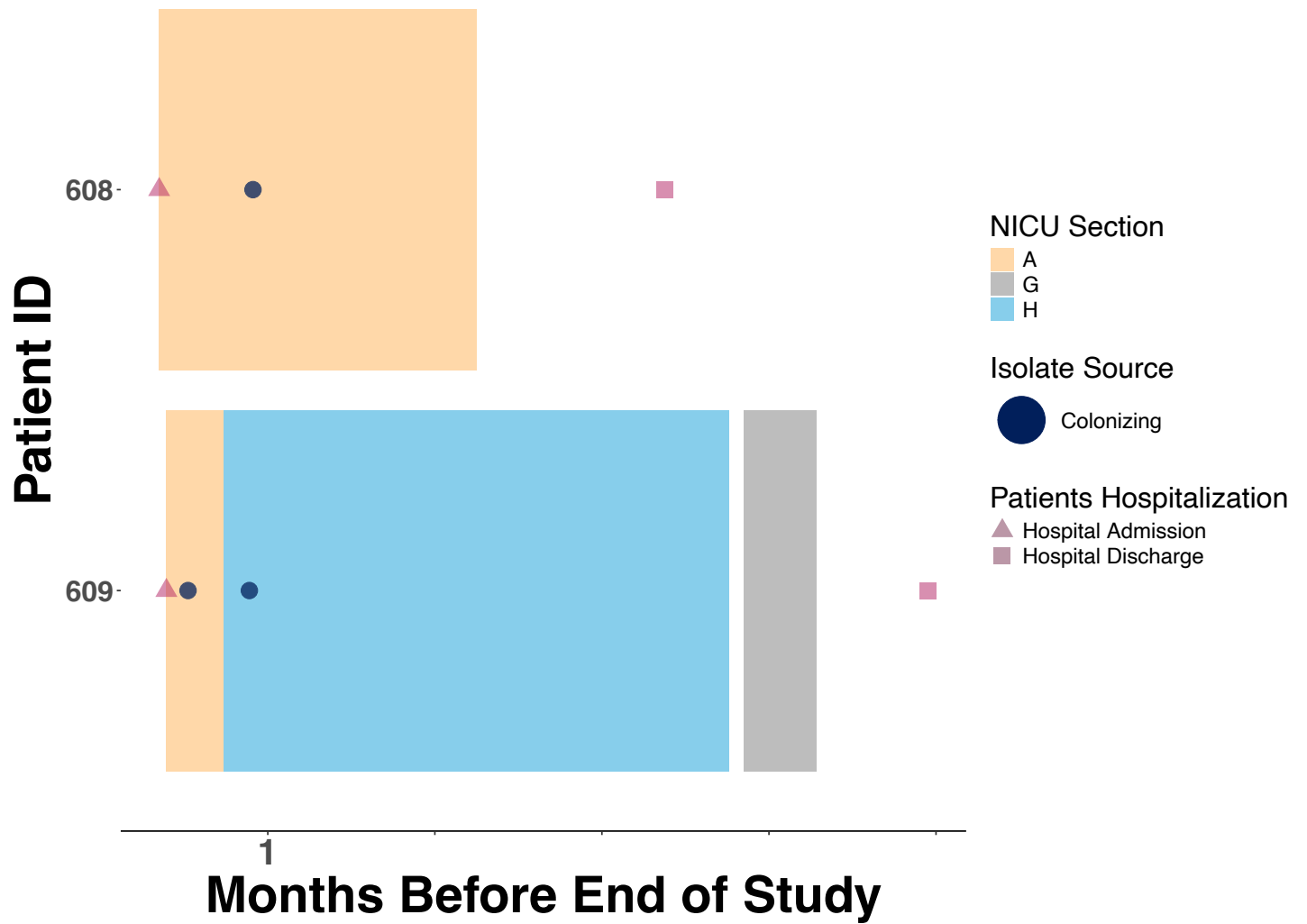
Cluster38



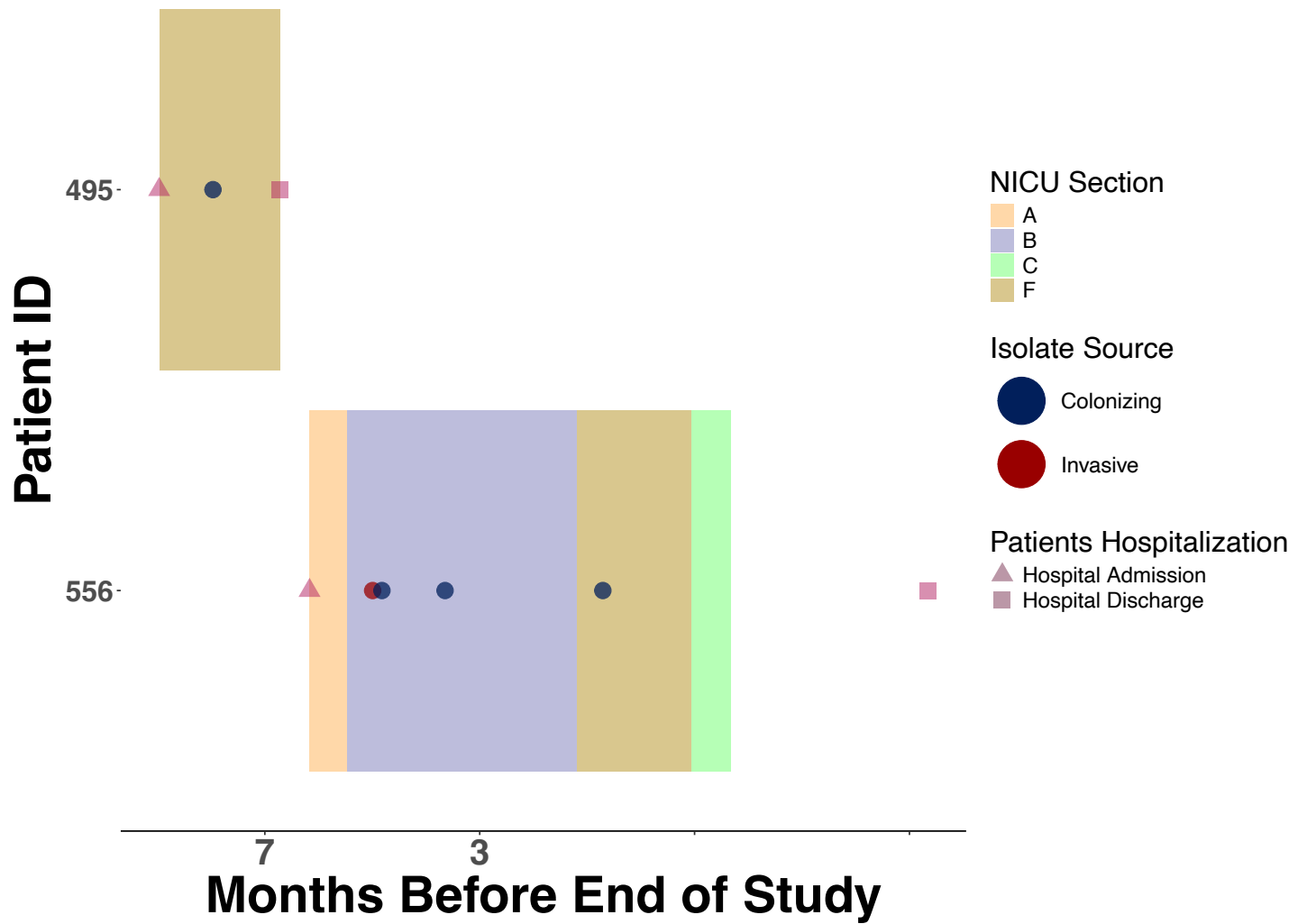
Cluster39



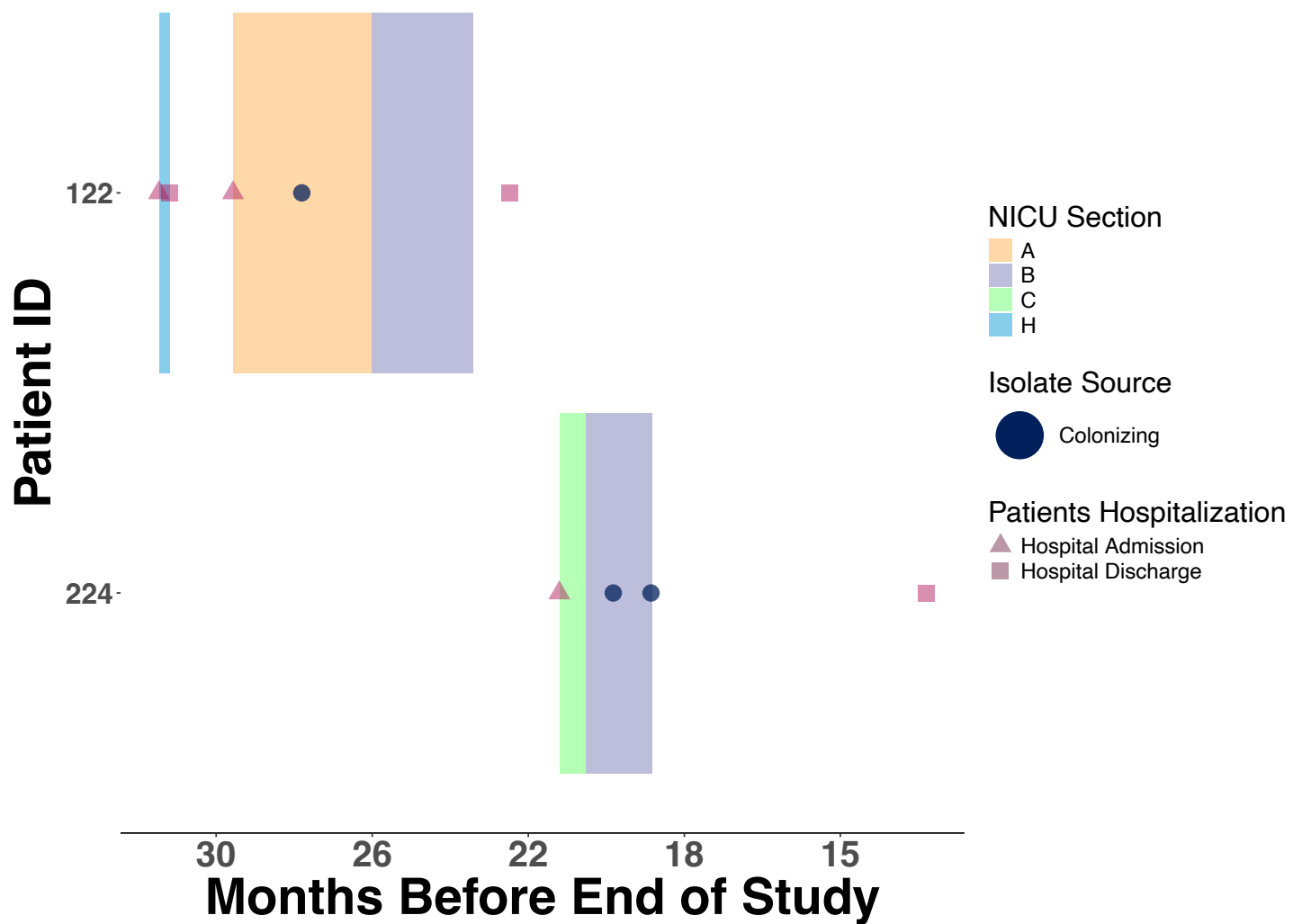
Cluster40



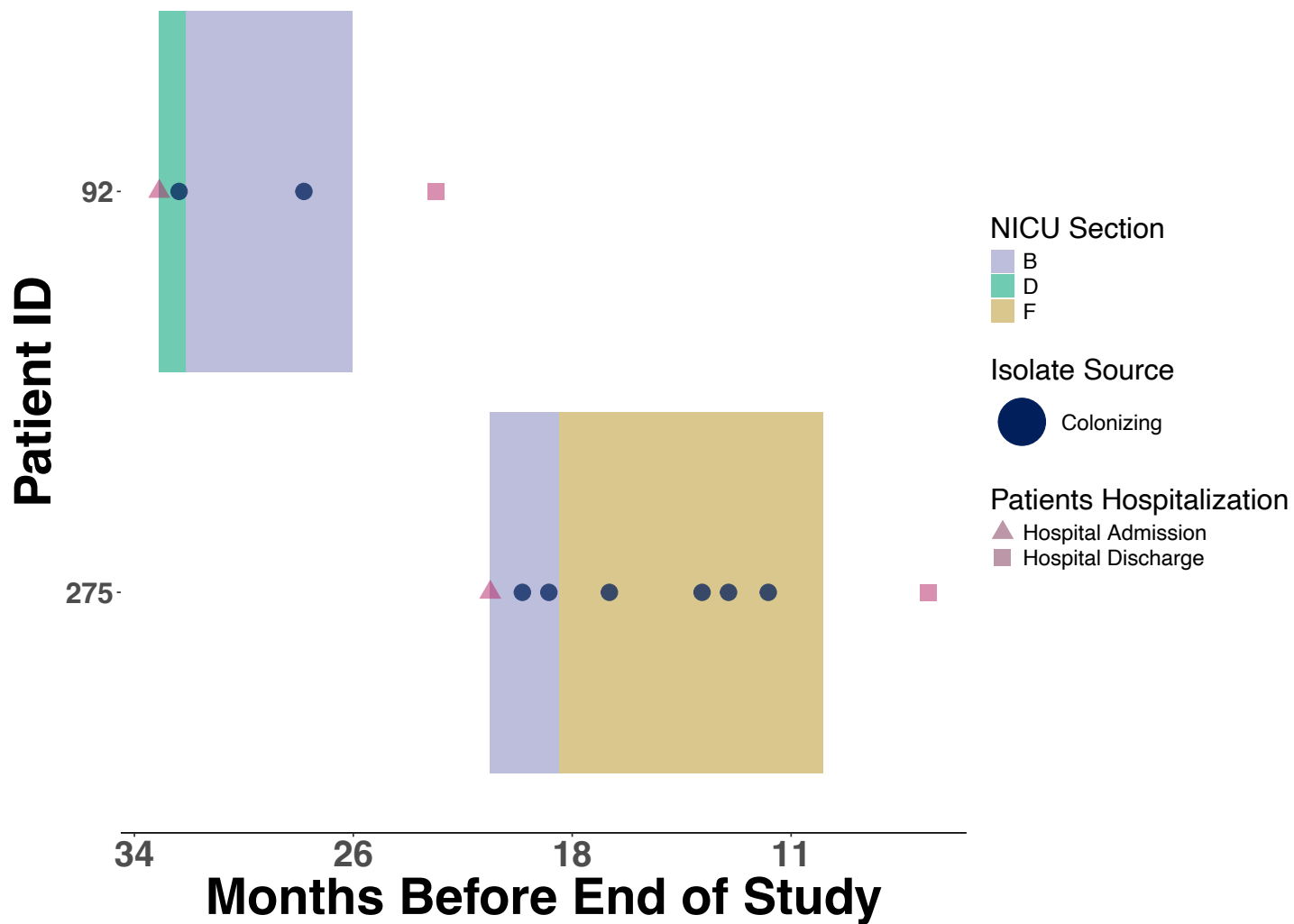
Cluster41



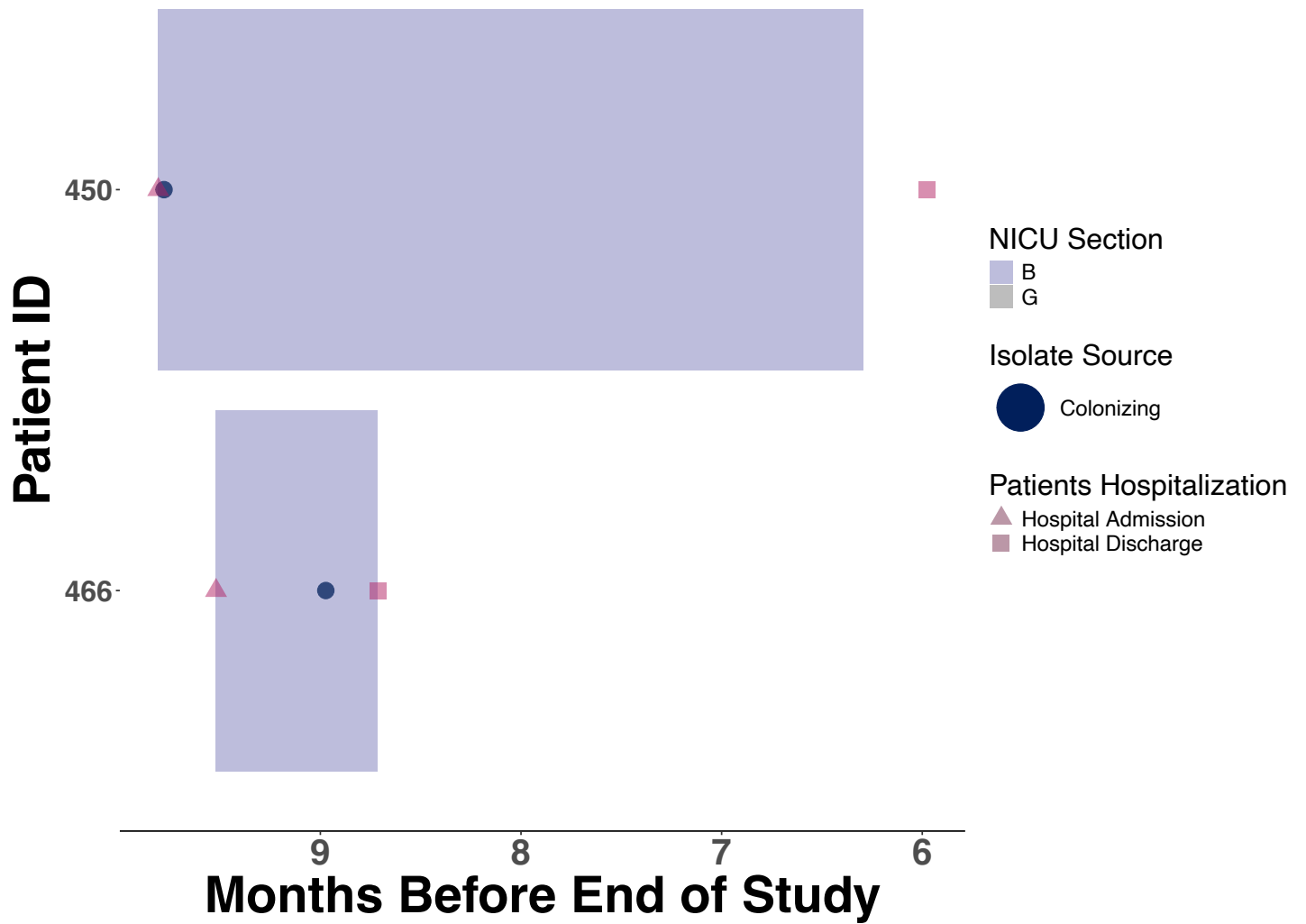
Cluster42



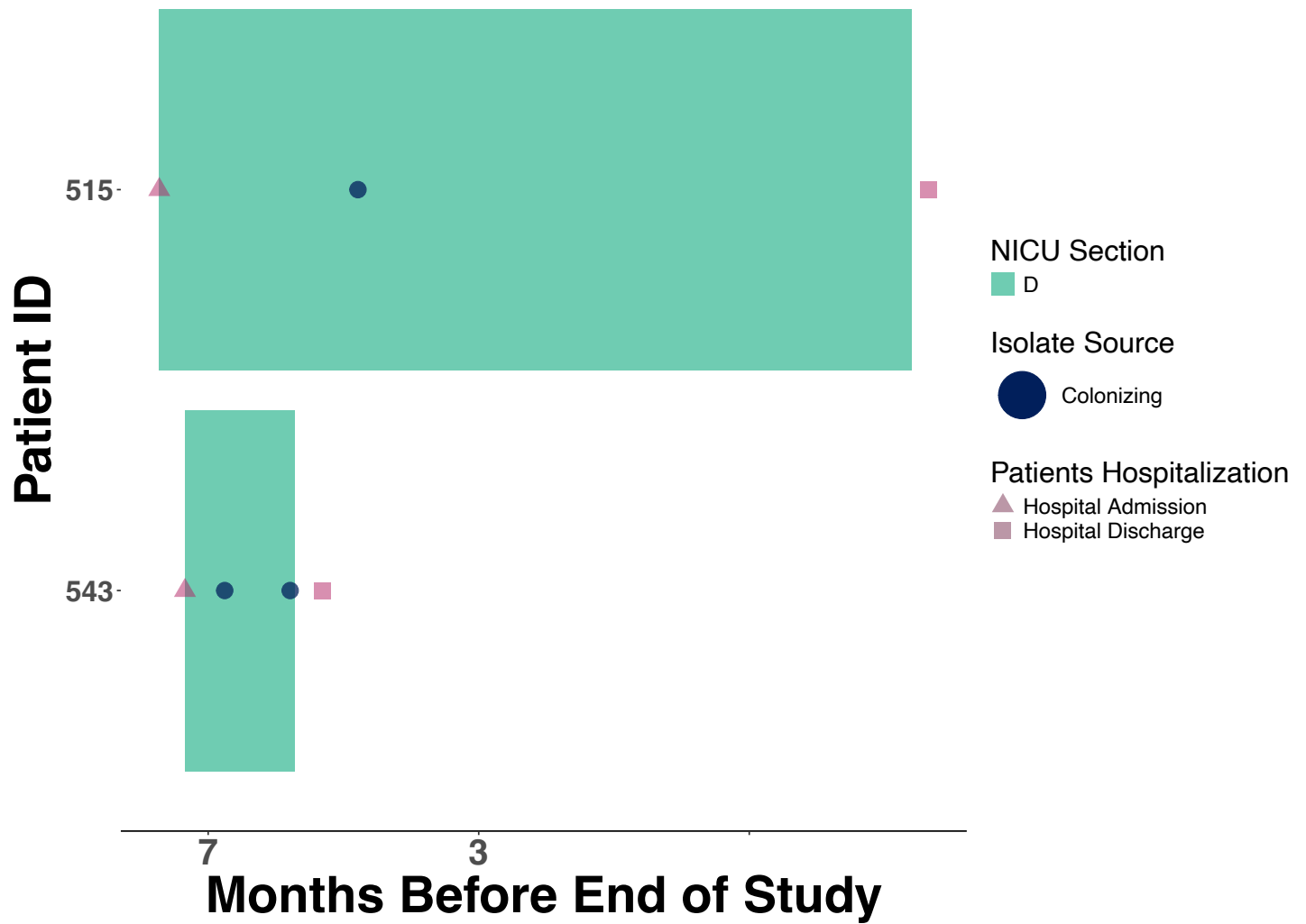
Cluster43



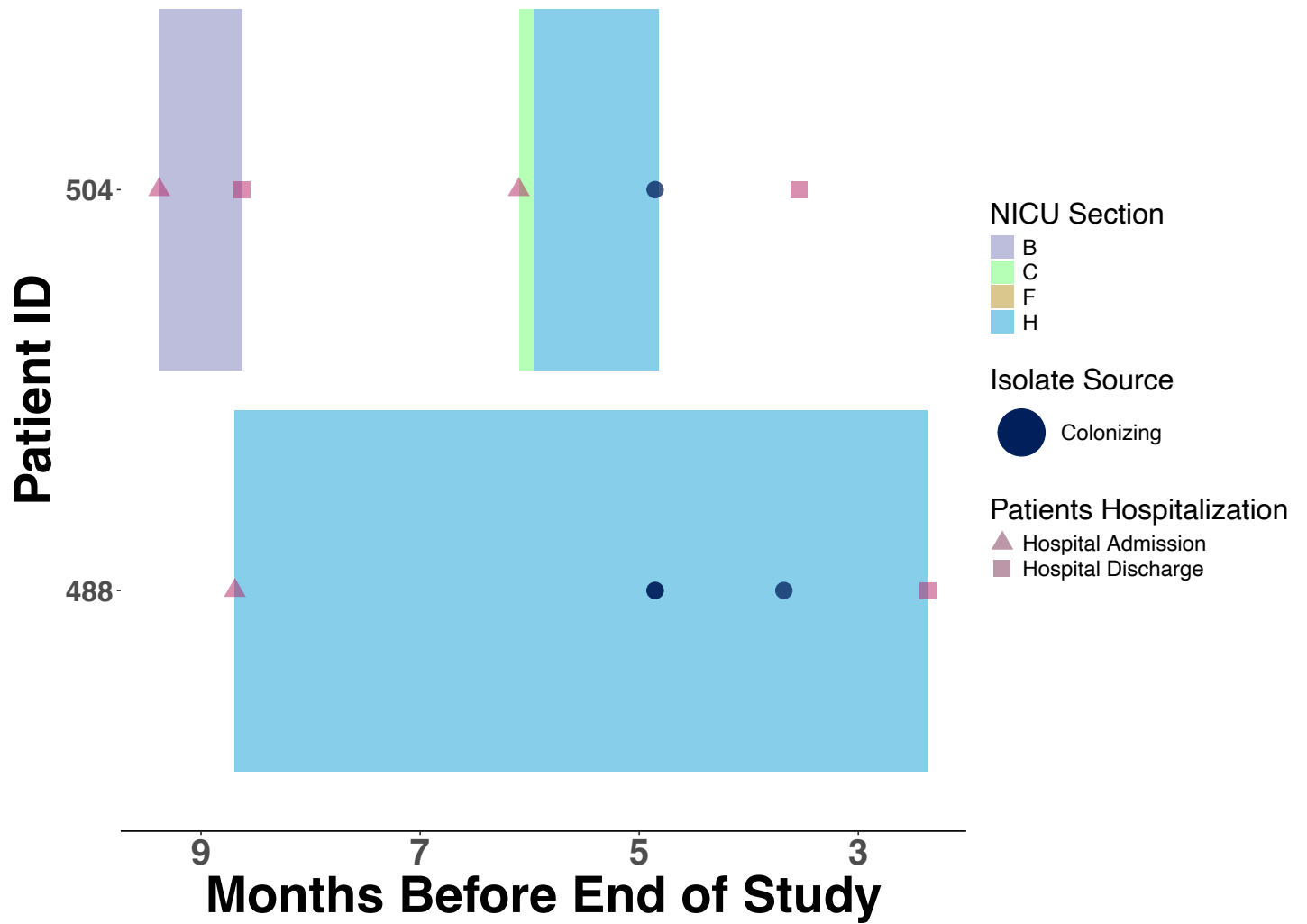
Cluster44



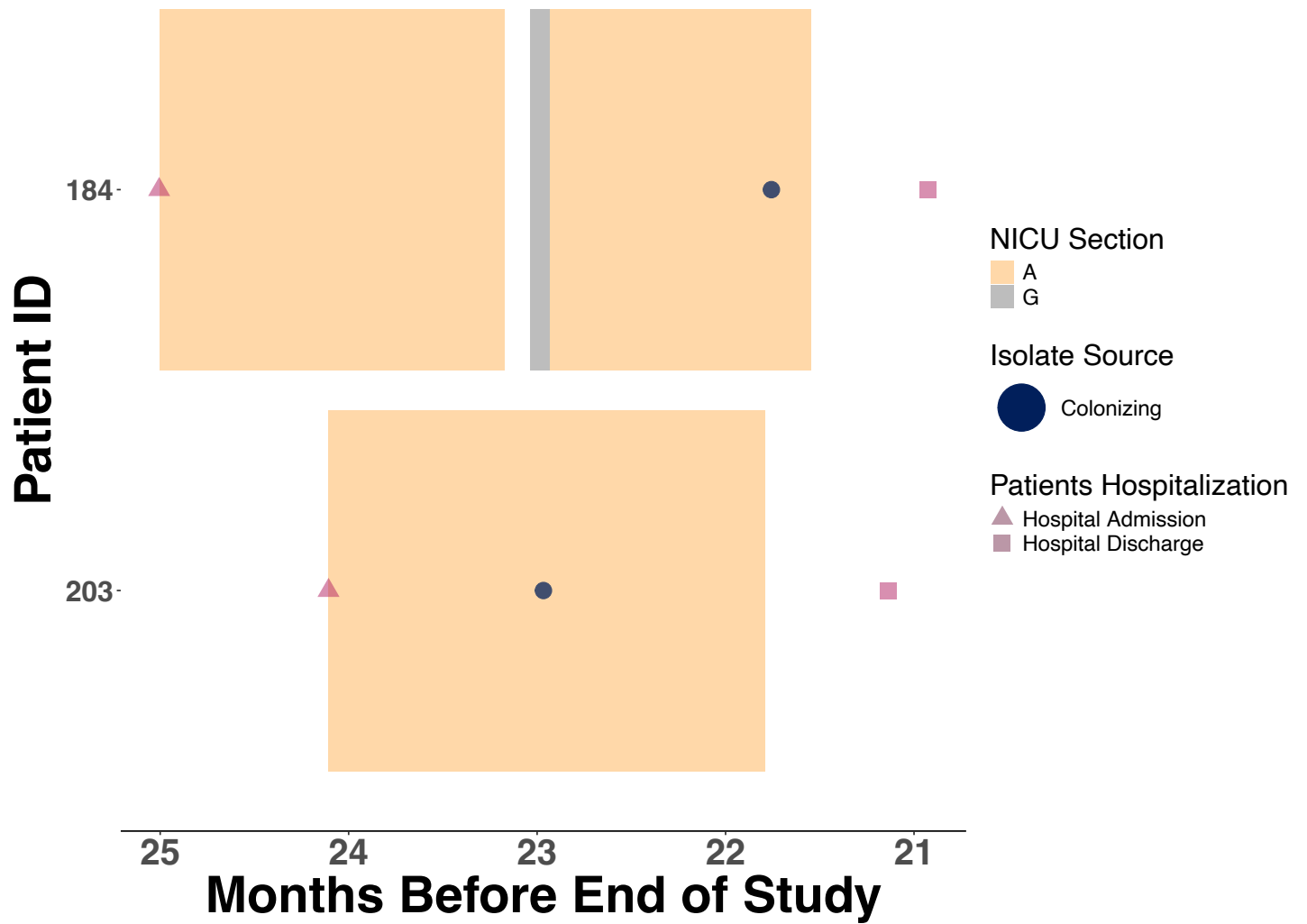
Cluster45



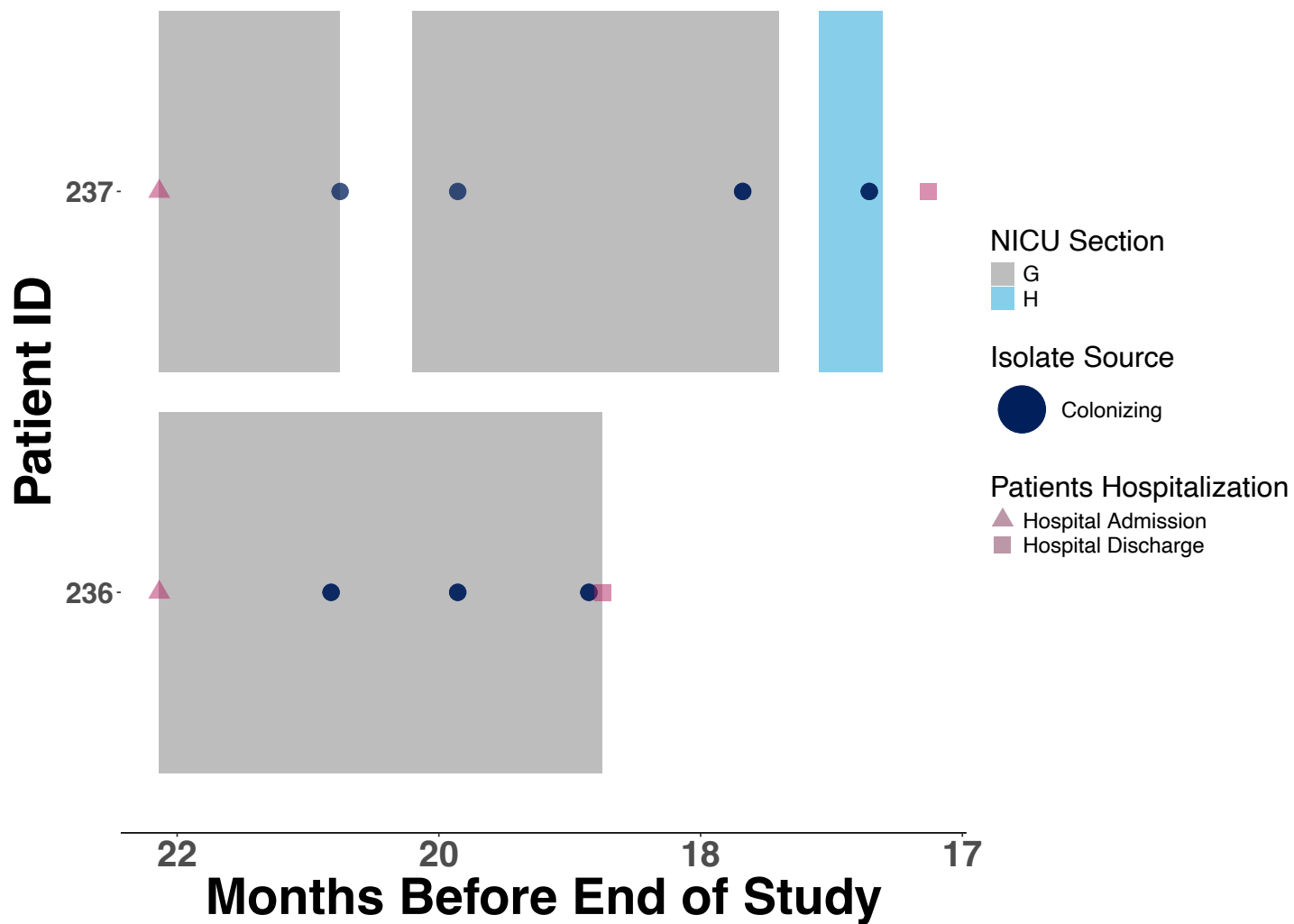
Cluster46



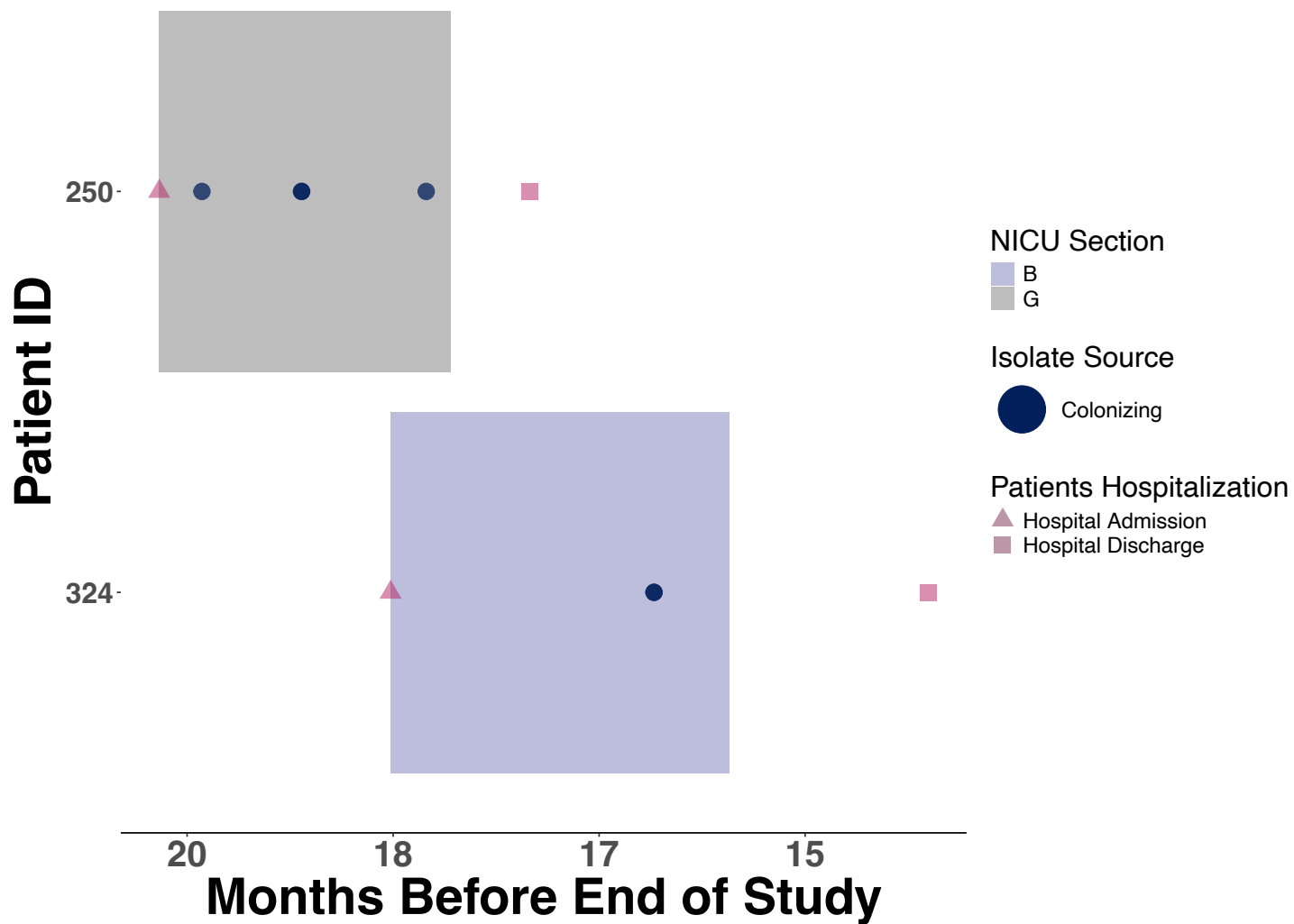
Cluster47



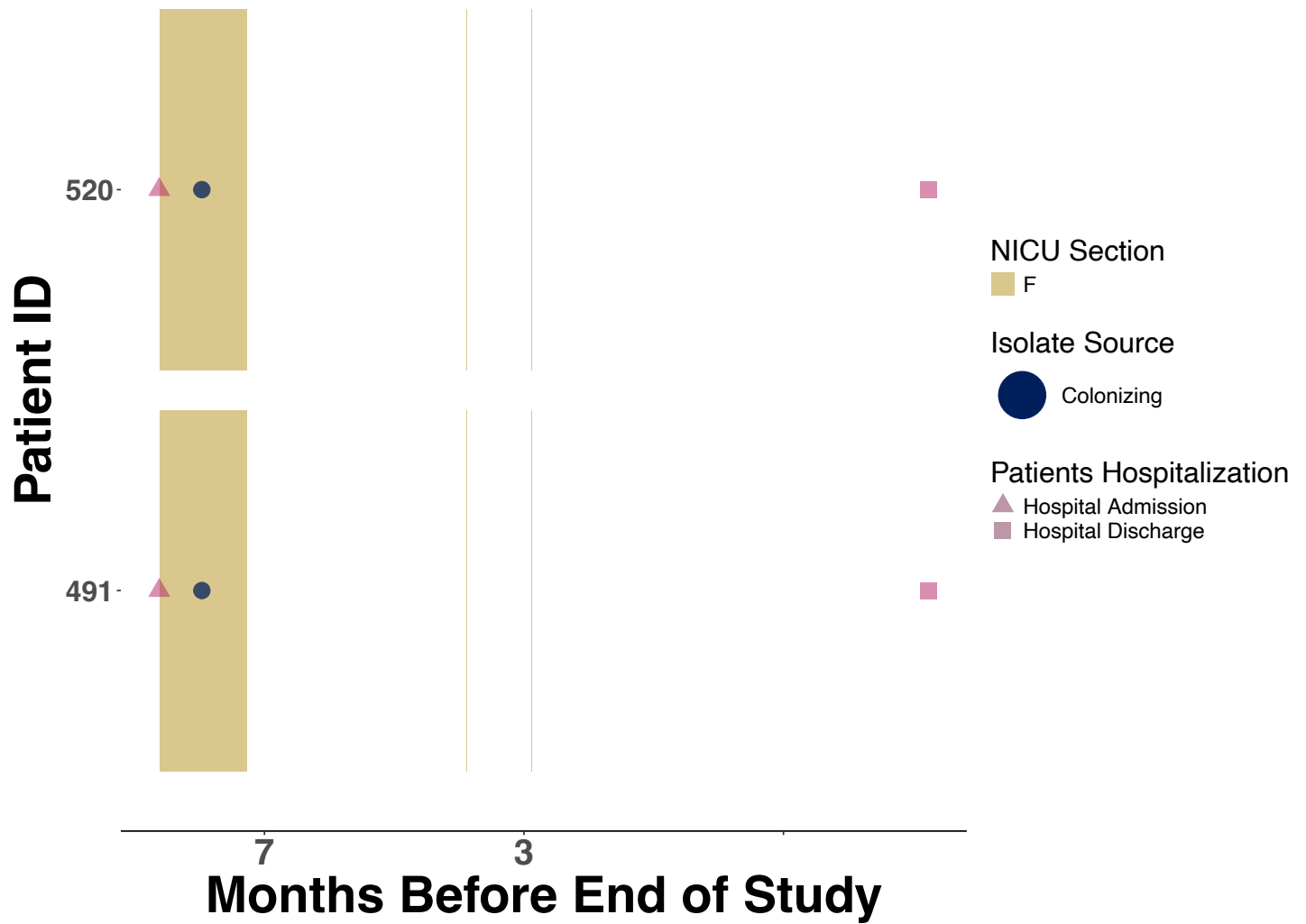
Cluster48



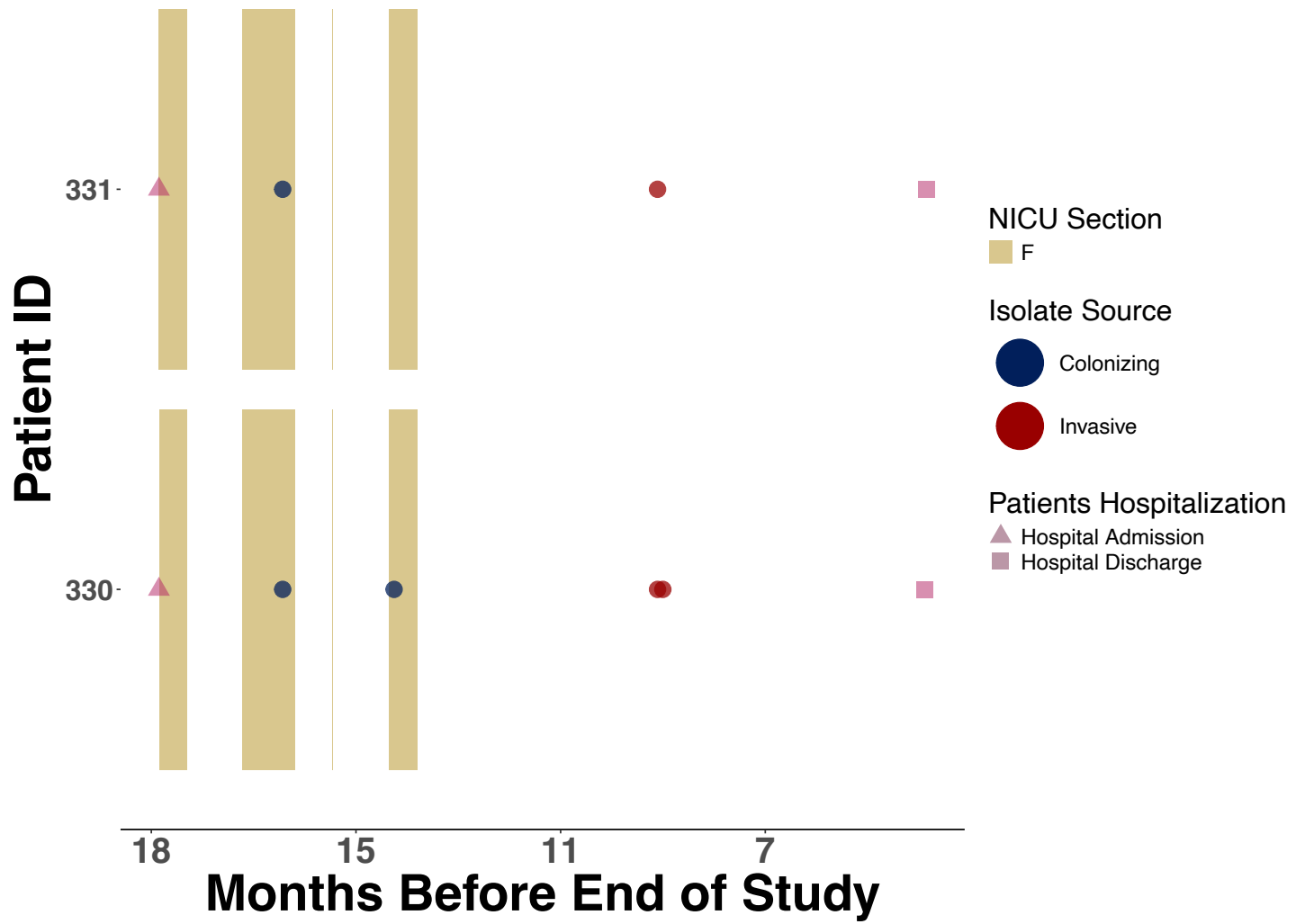
Cluster49



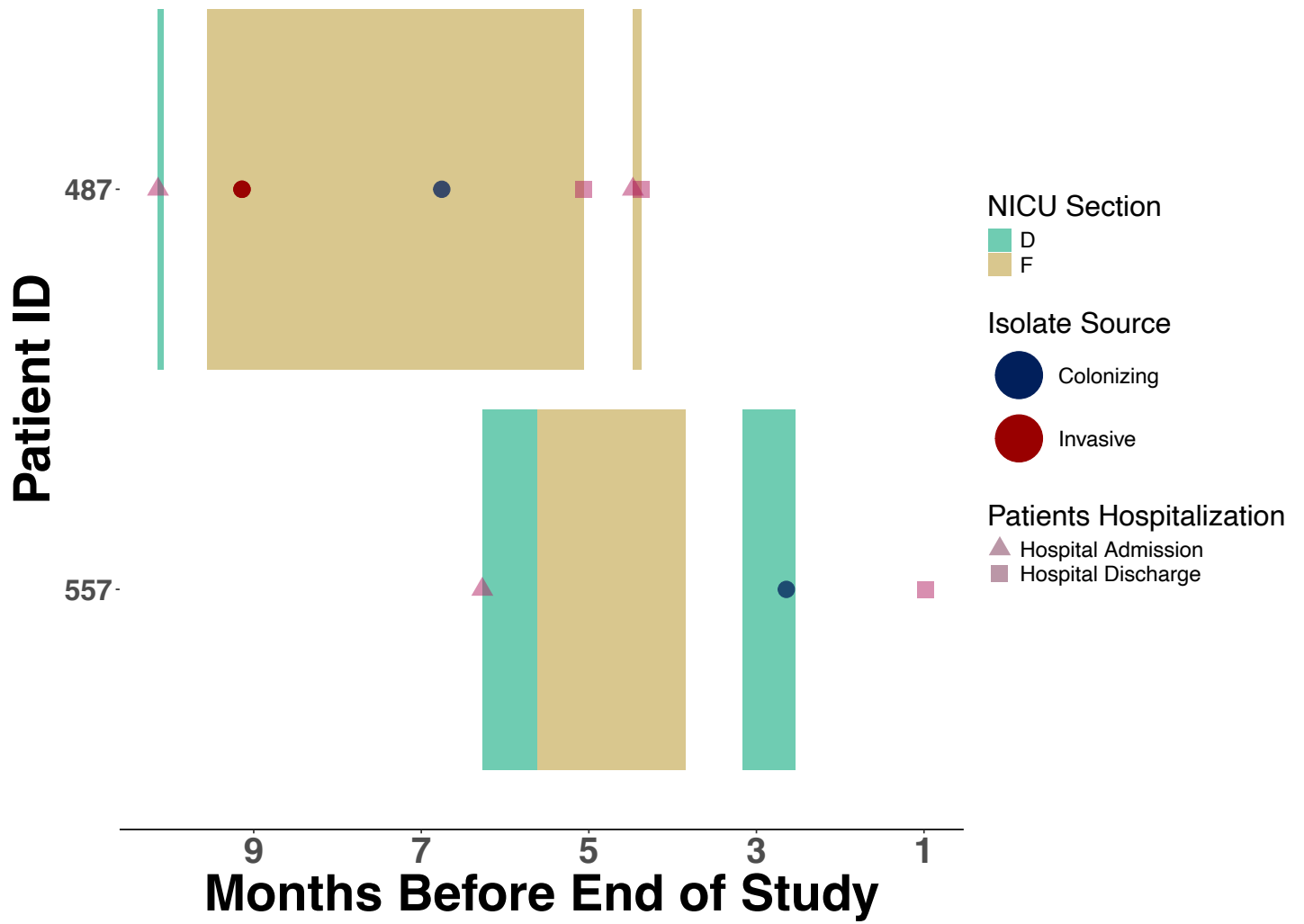
Cluster50



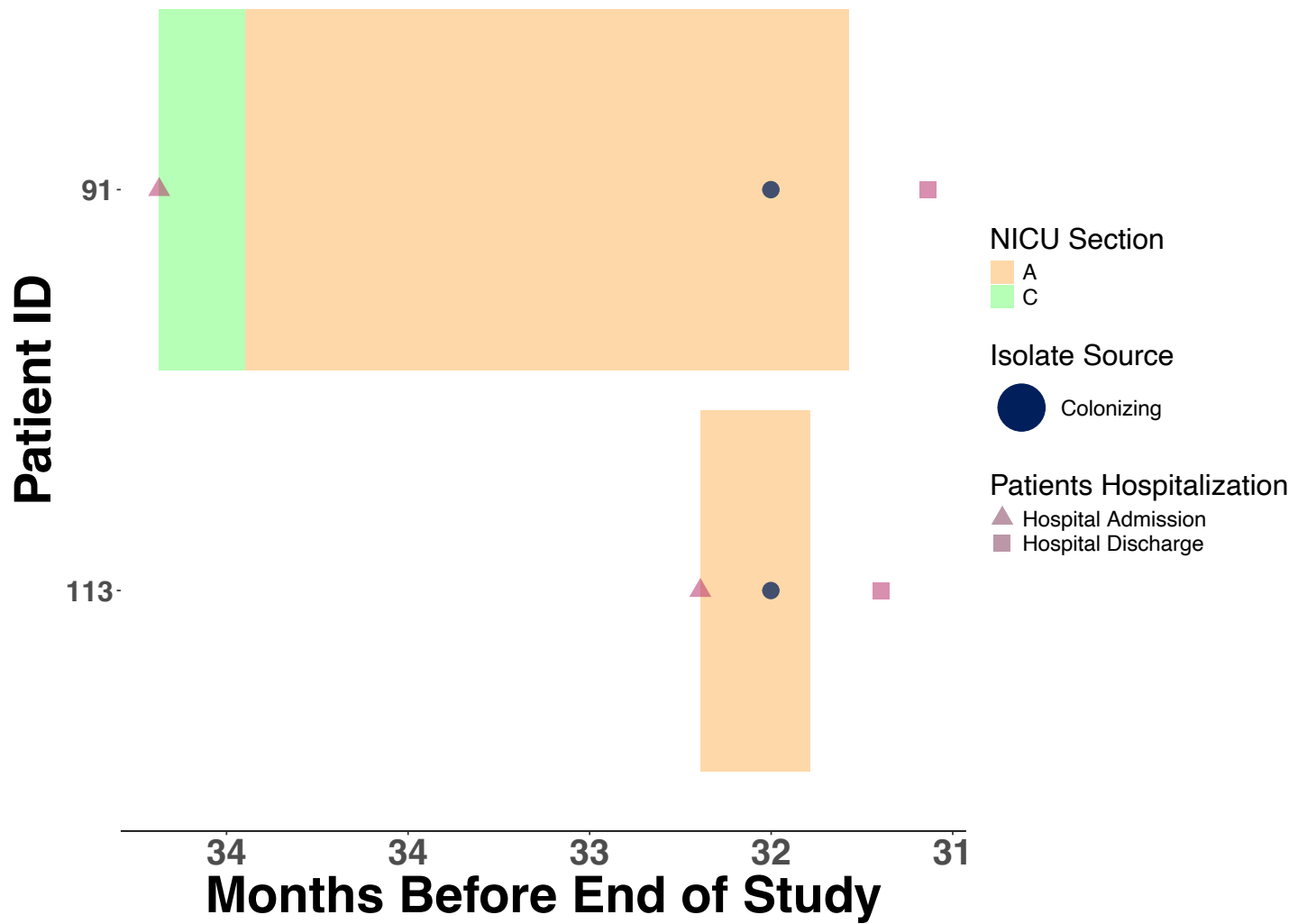
Cluster51



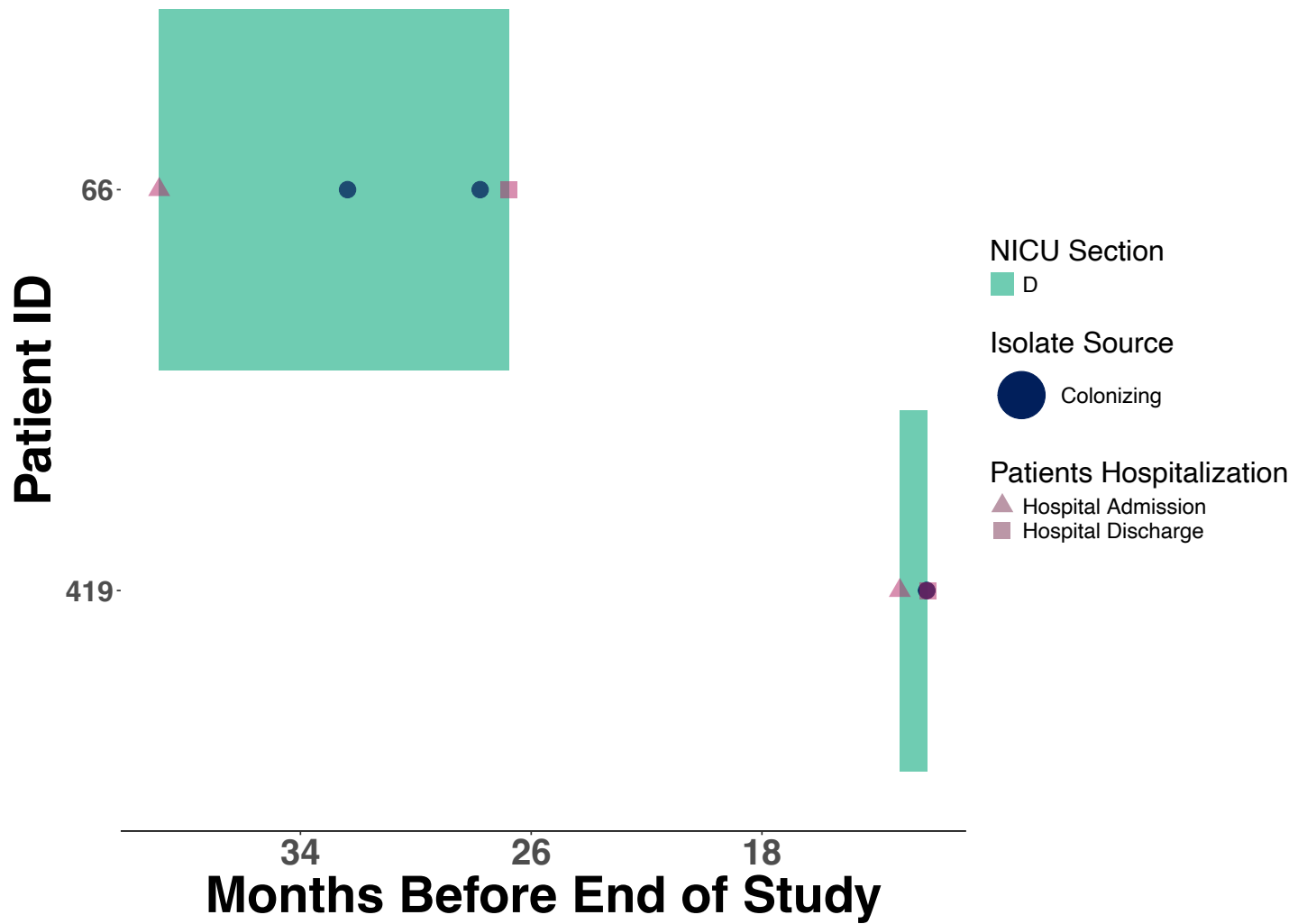
Cluster52



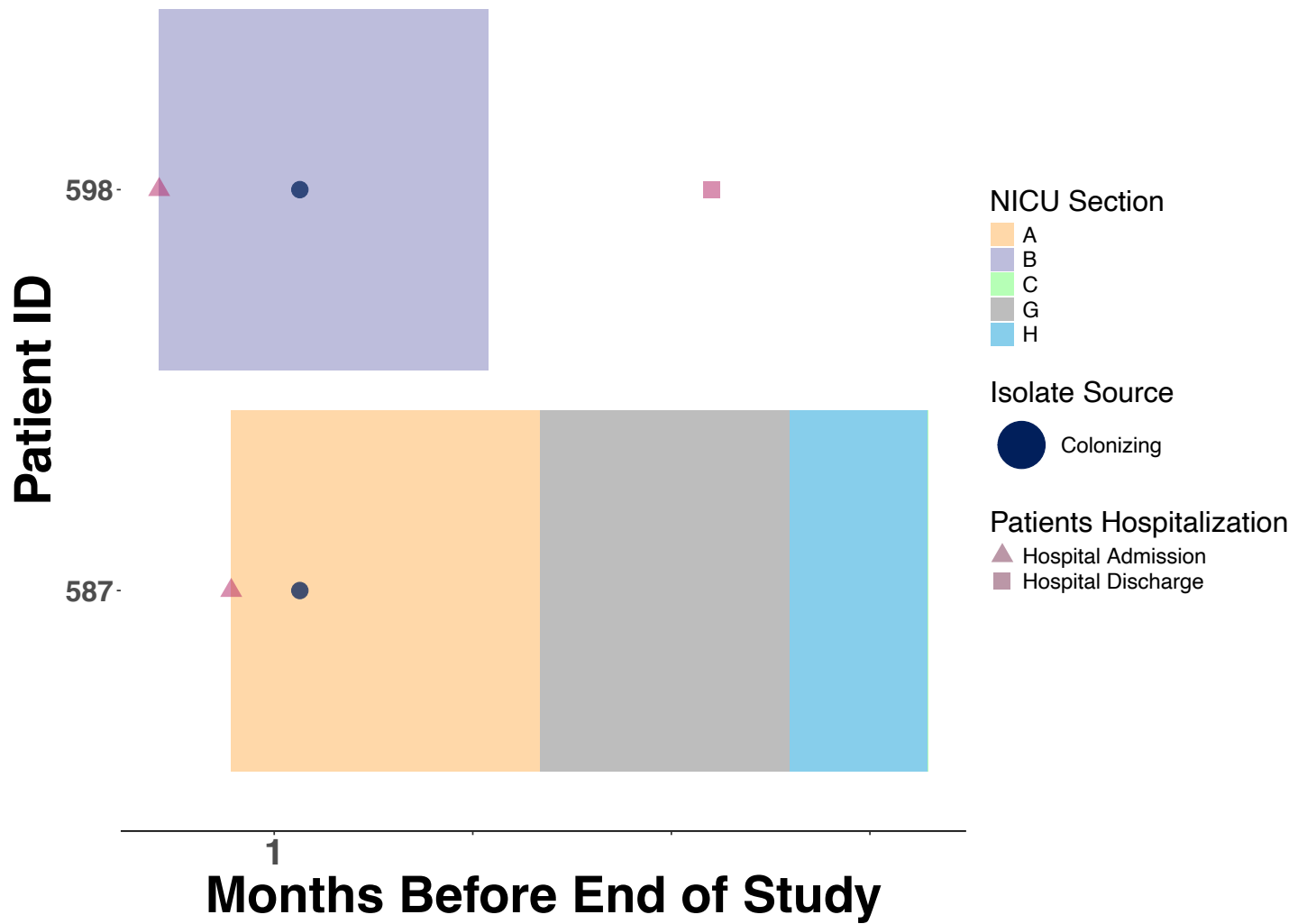
Cluster53



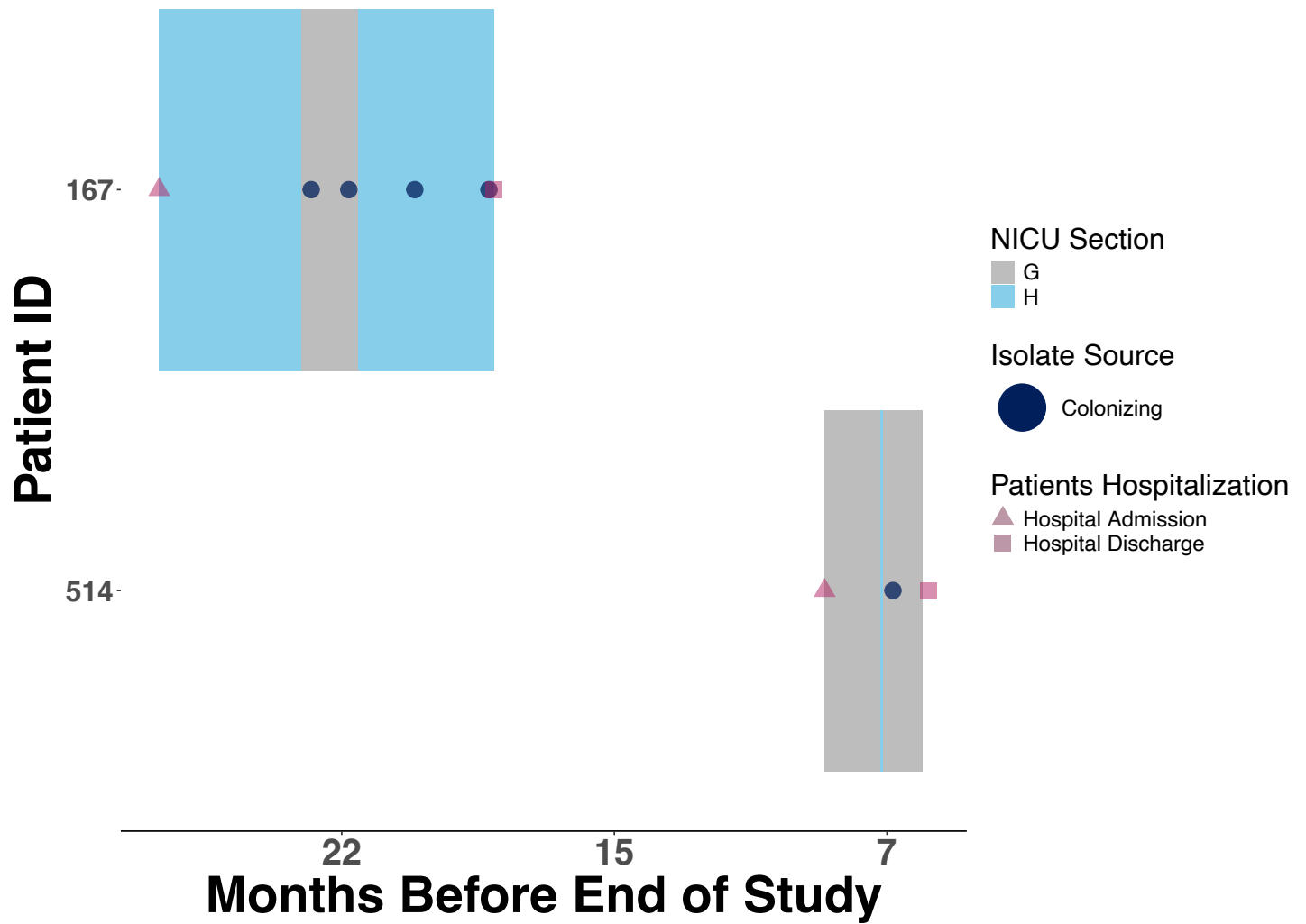
Cluster54



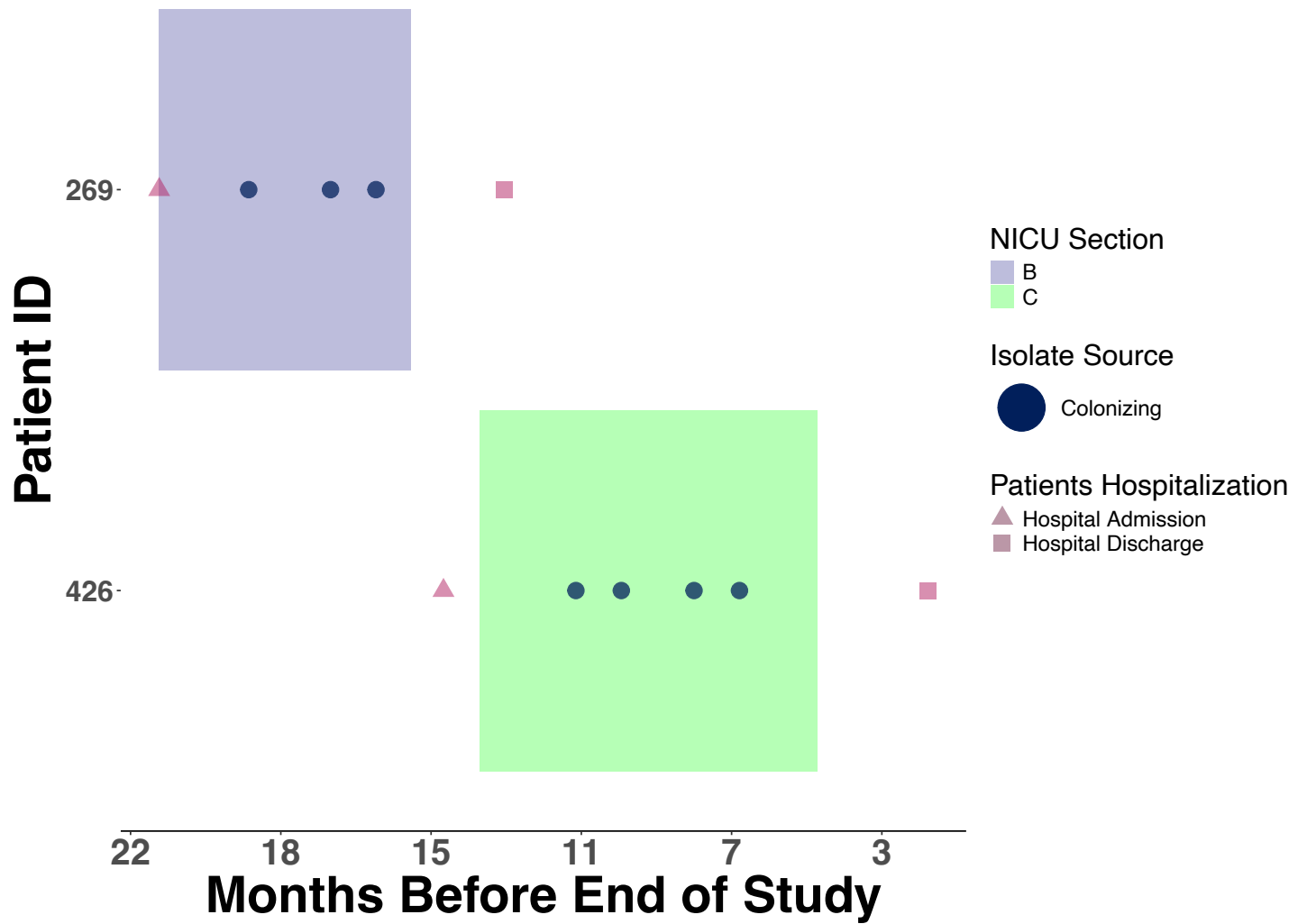
Cluster55



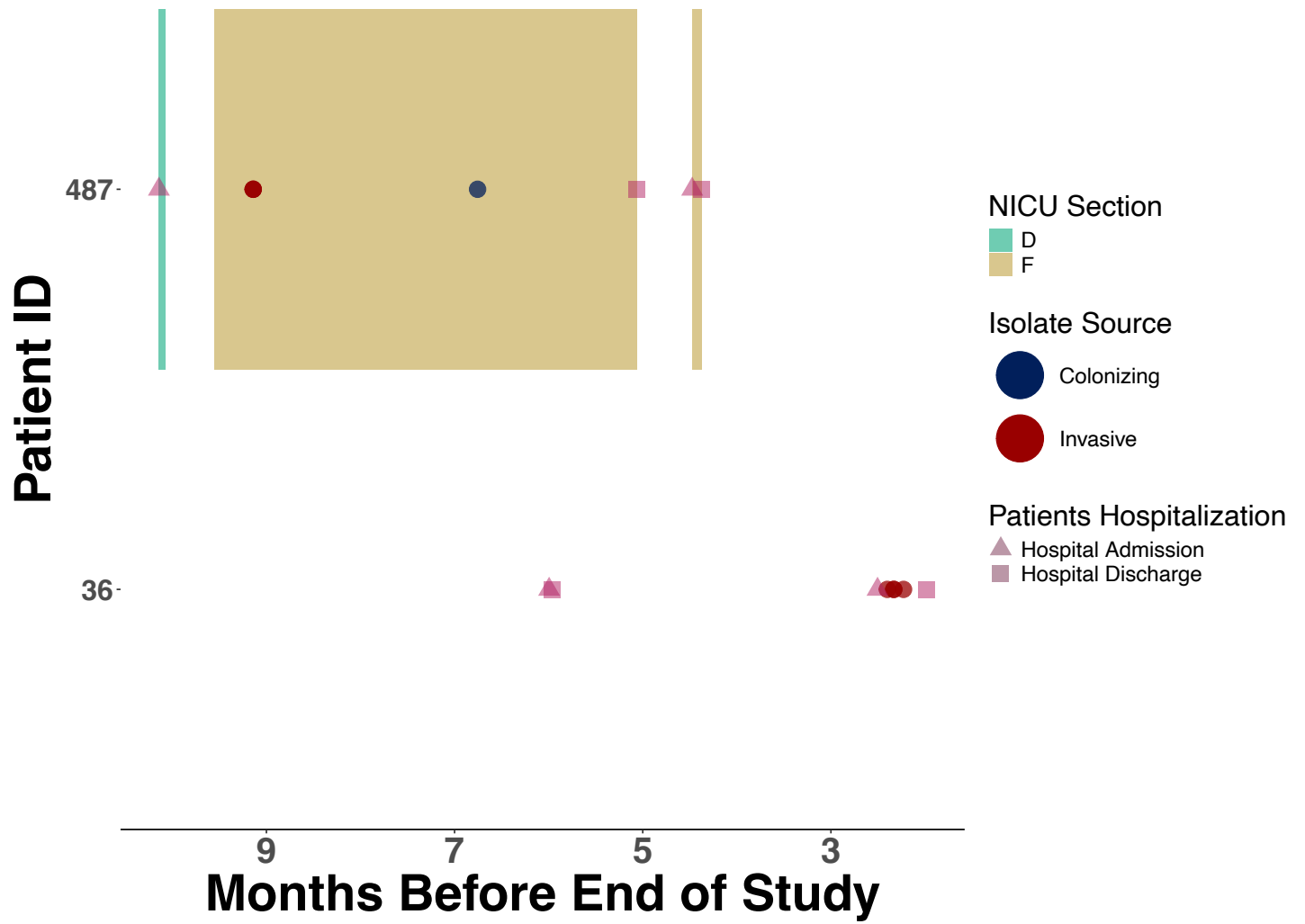
Cluster56



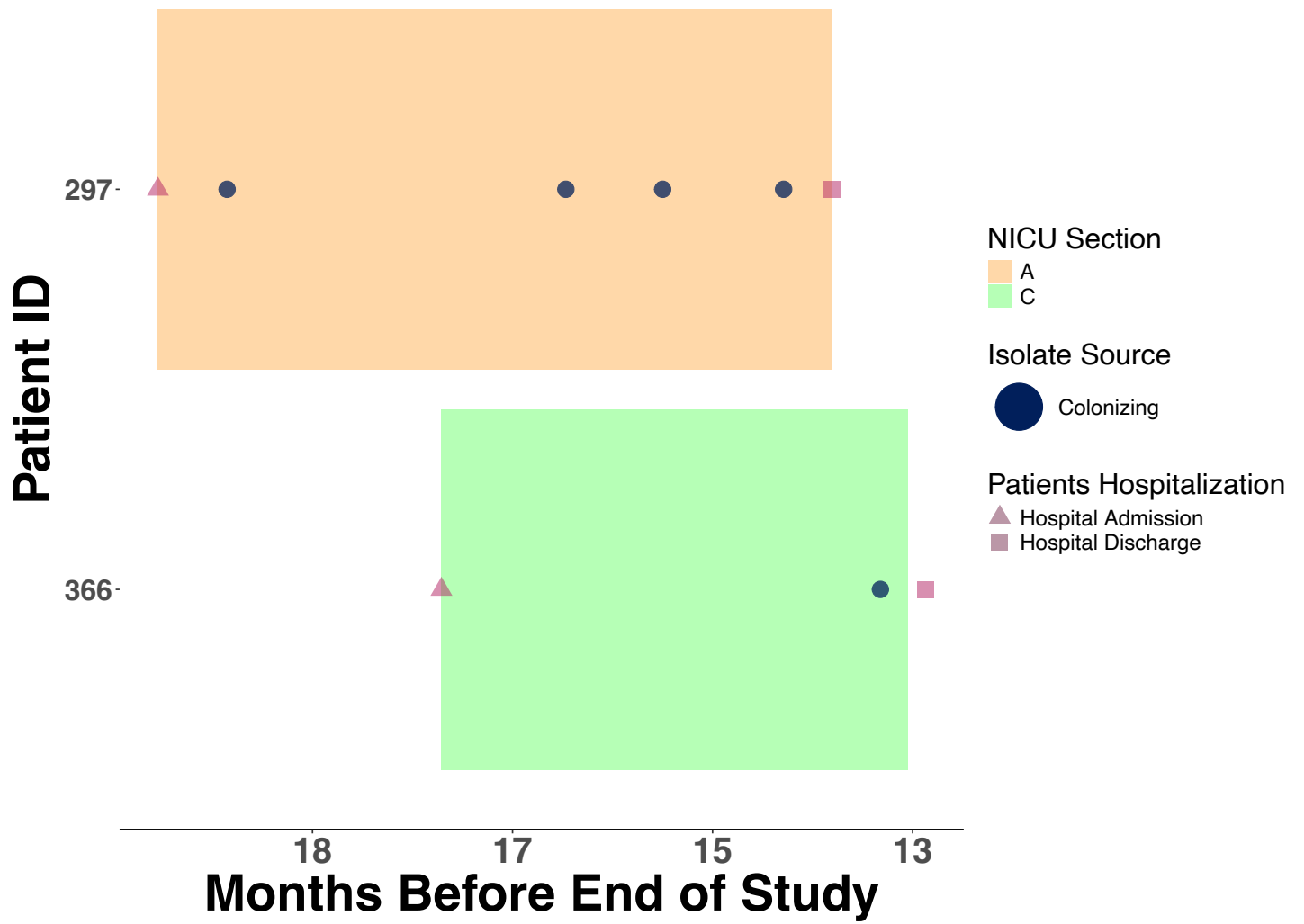
Cluster57



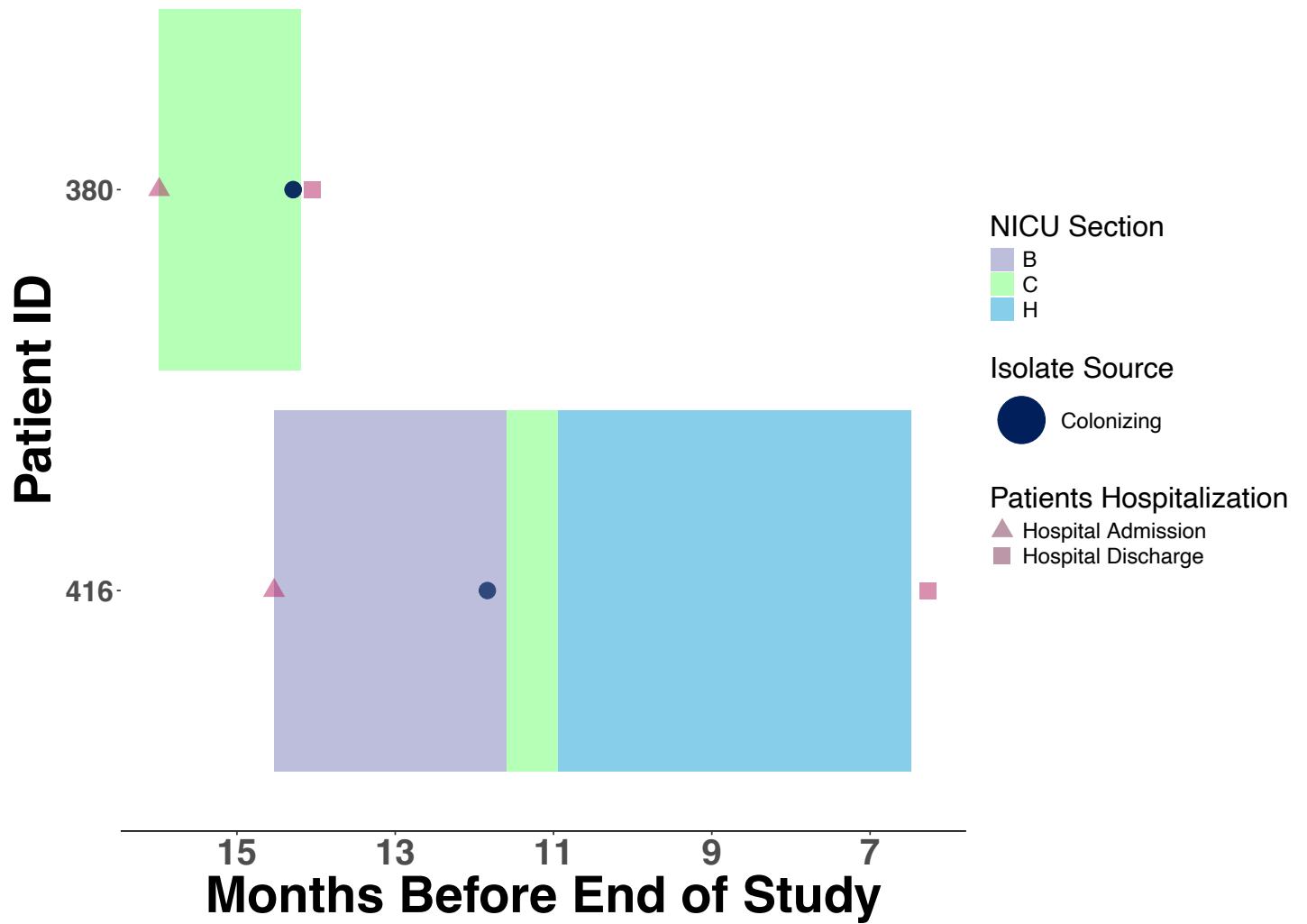
Cluster58



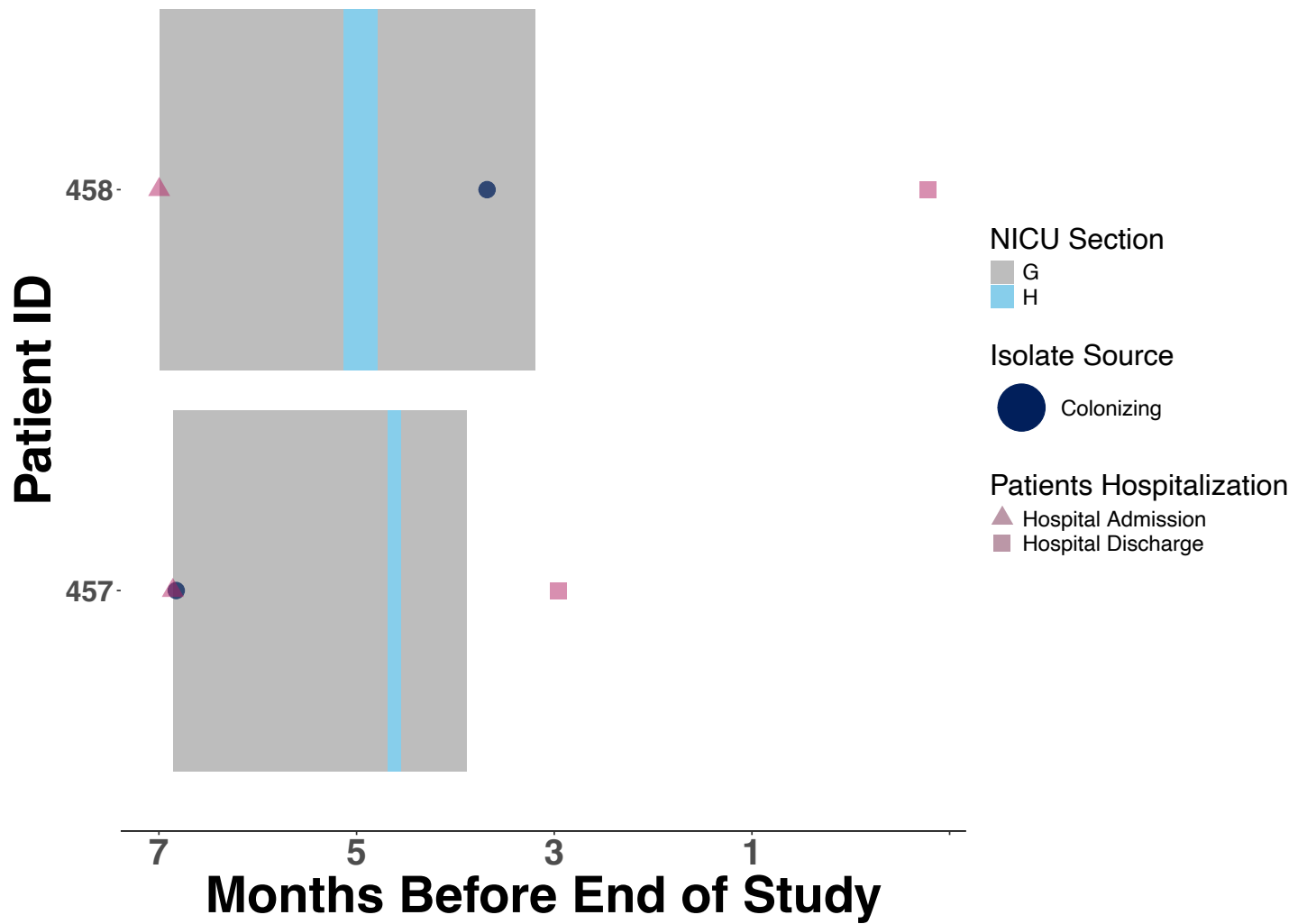
Cluster59



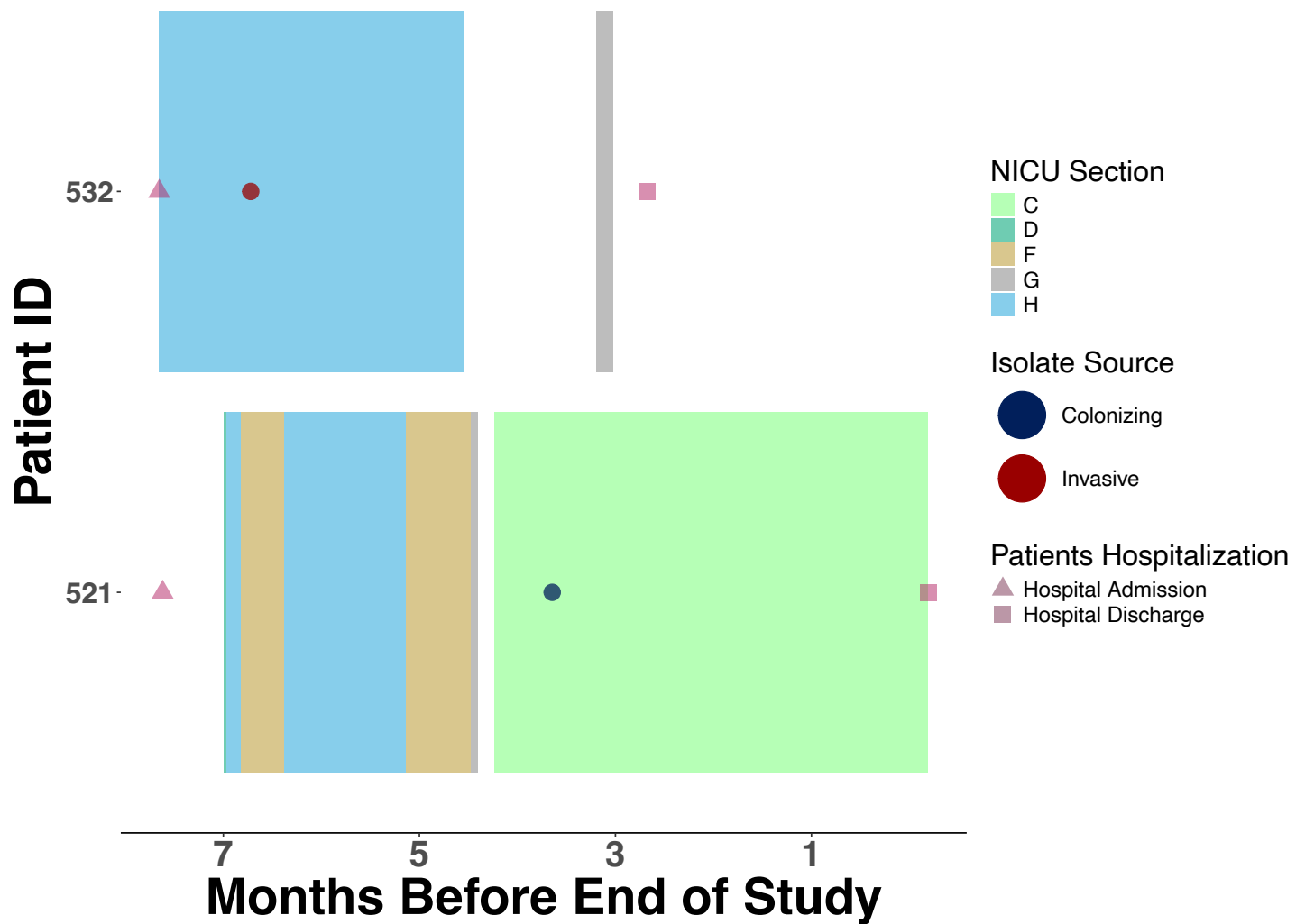
Cluster60



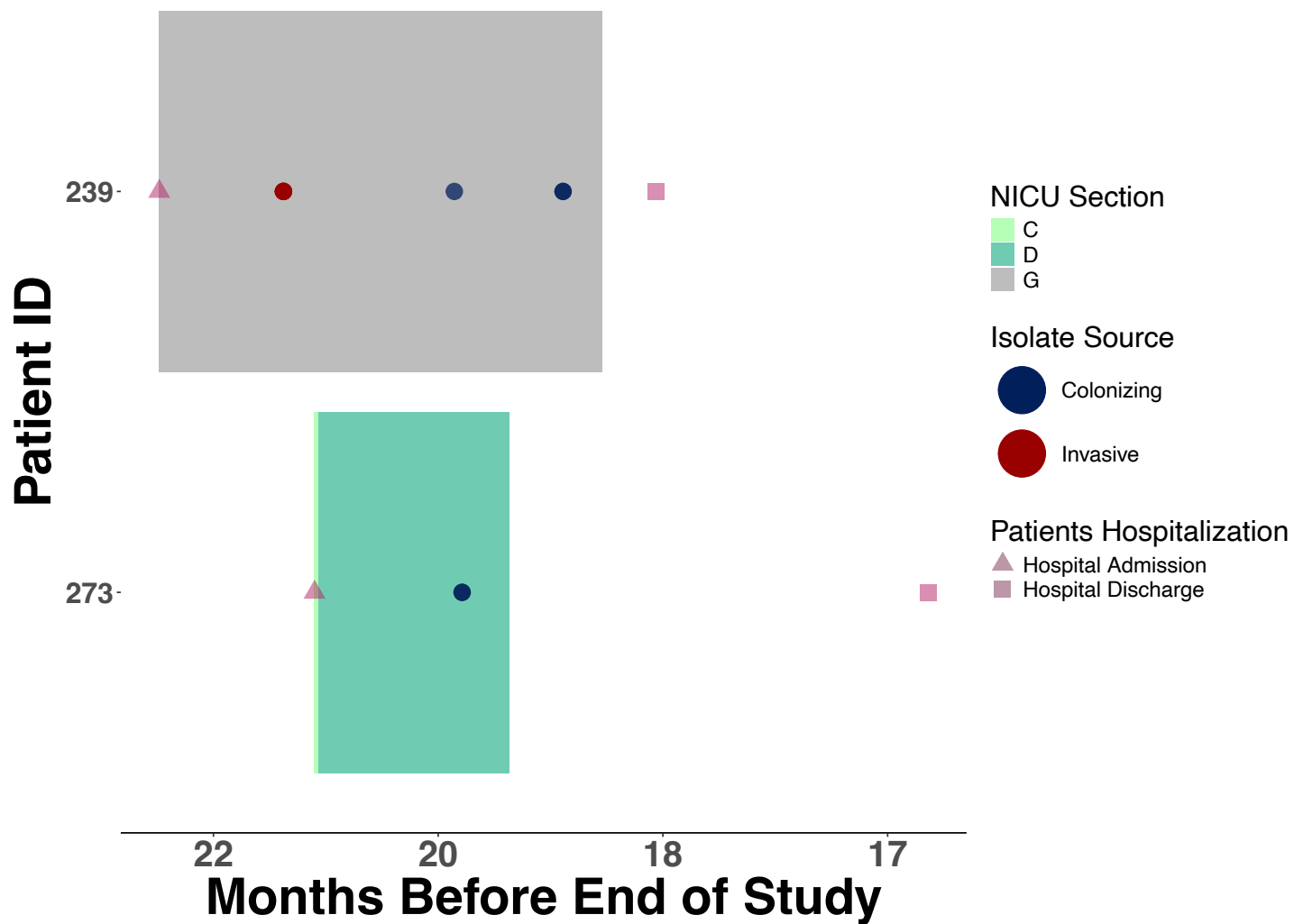
Cluster61



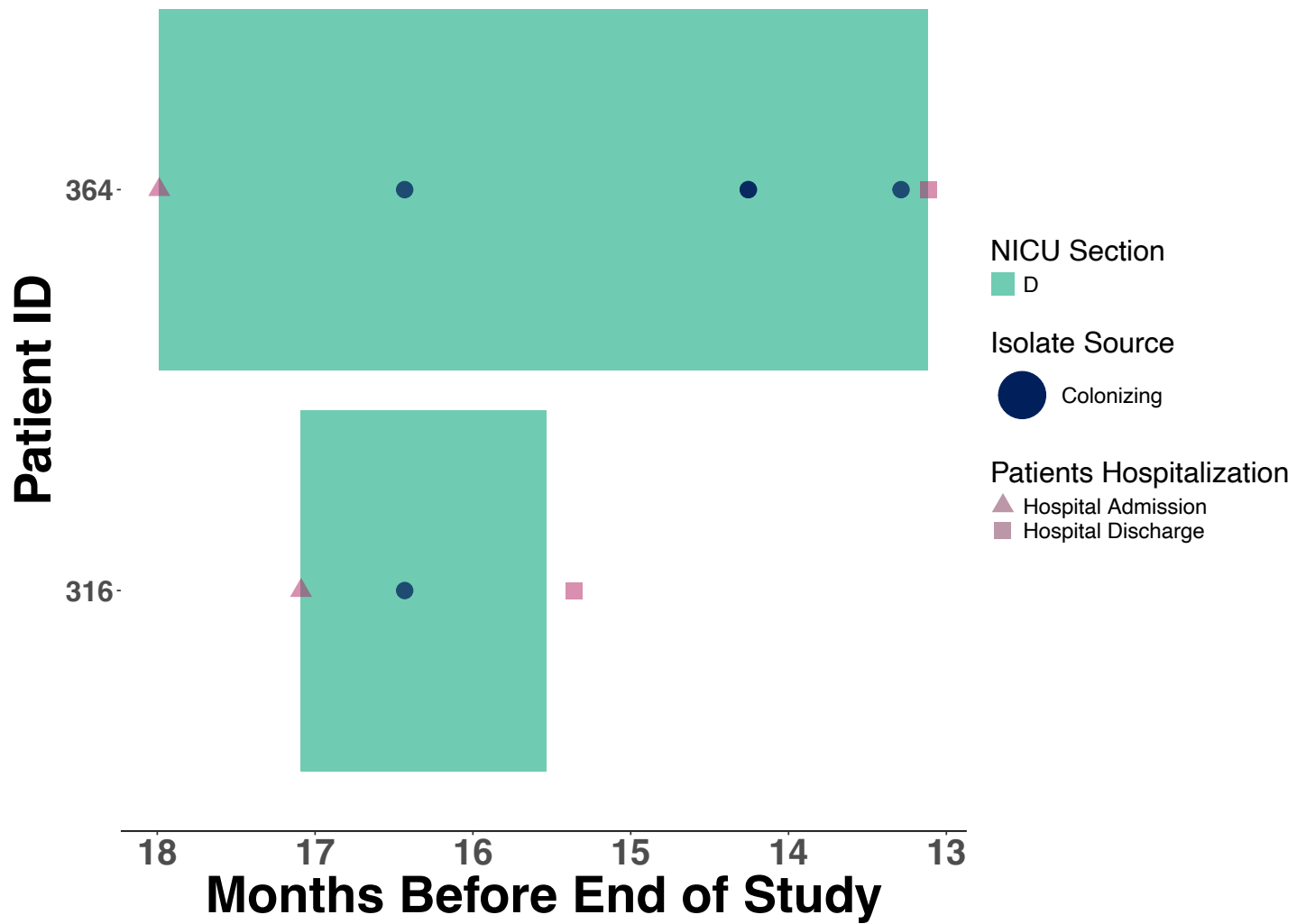
Cluster62



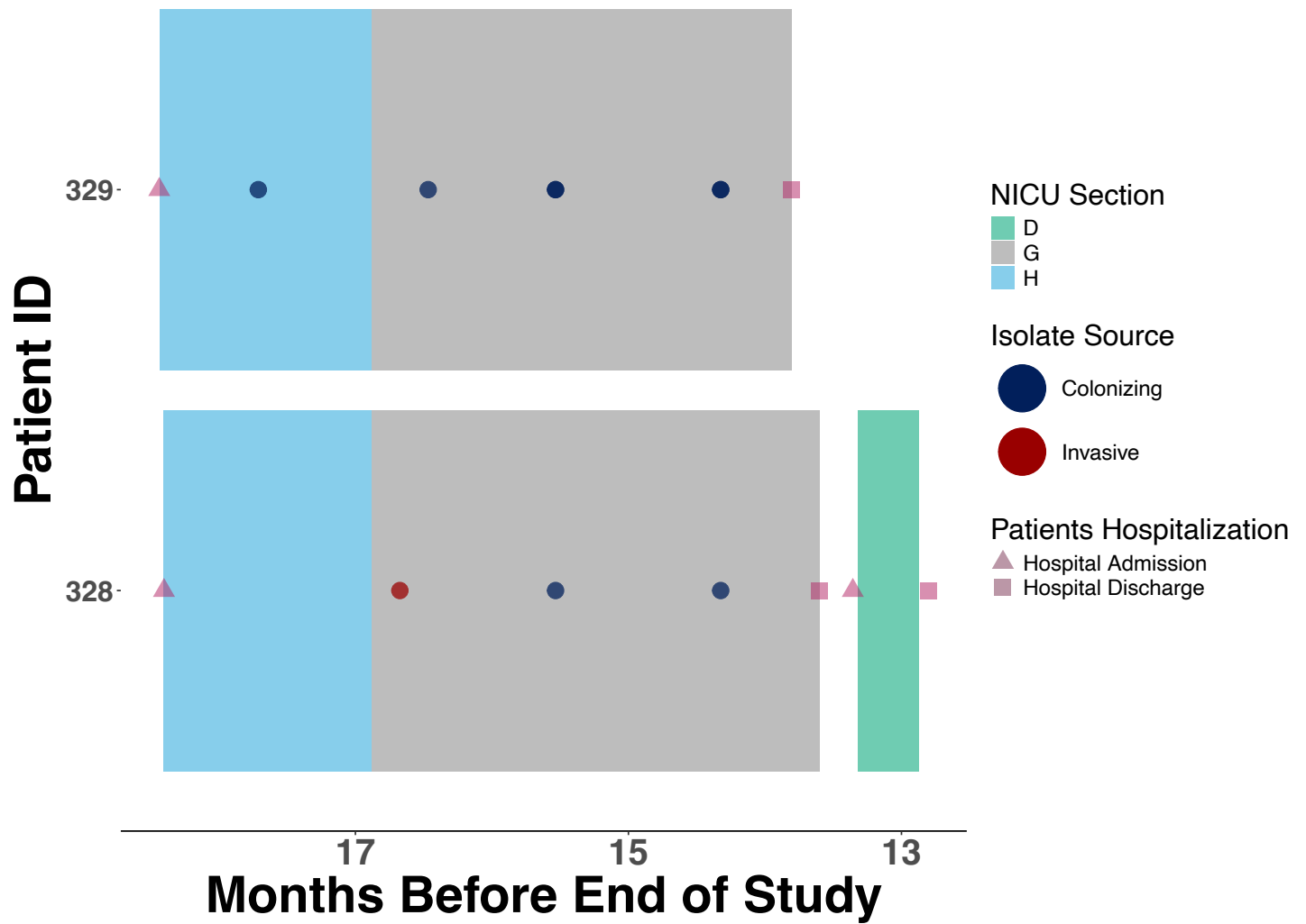
Cluster63



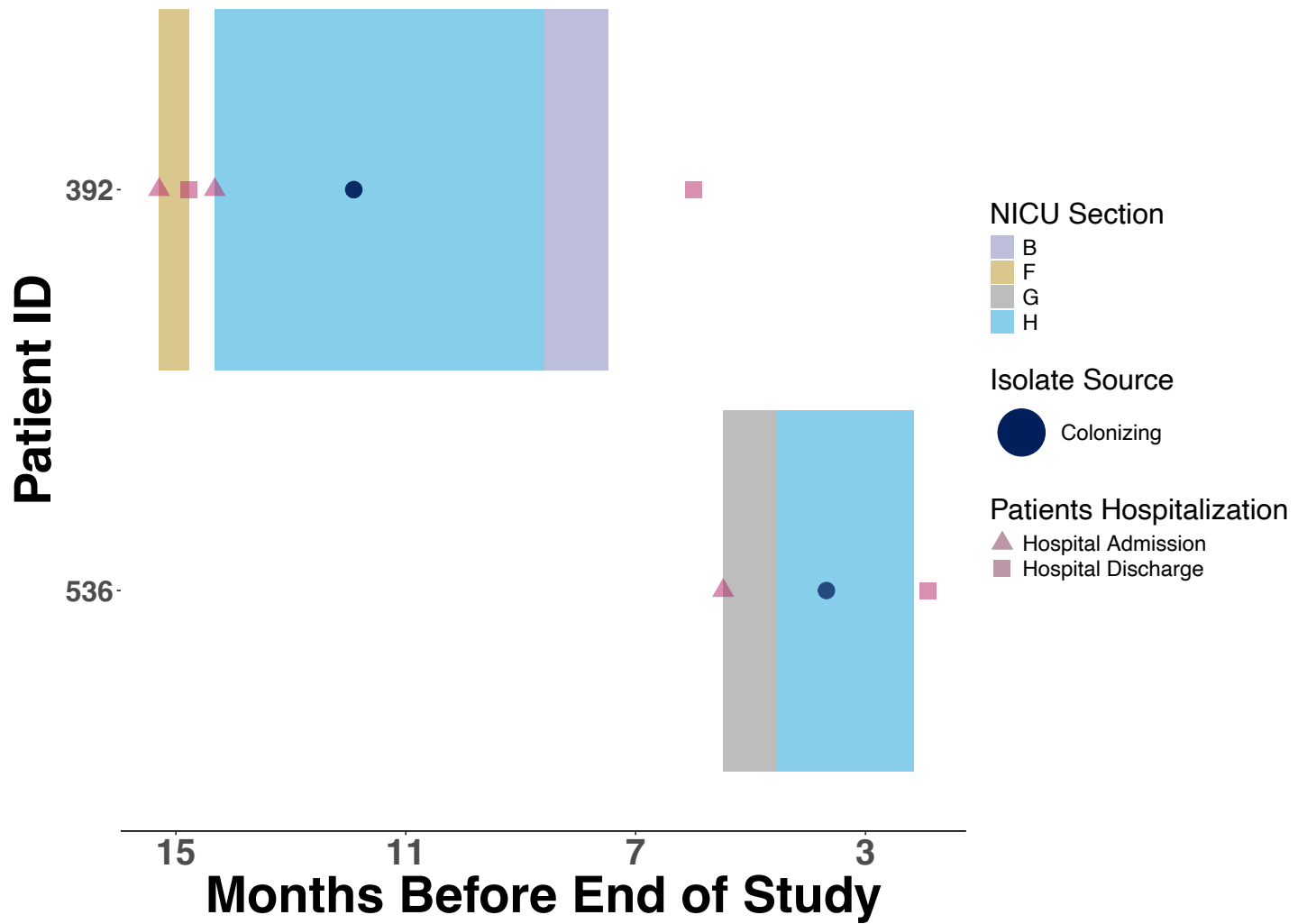
Cluster64



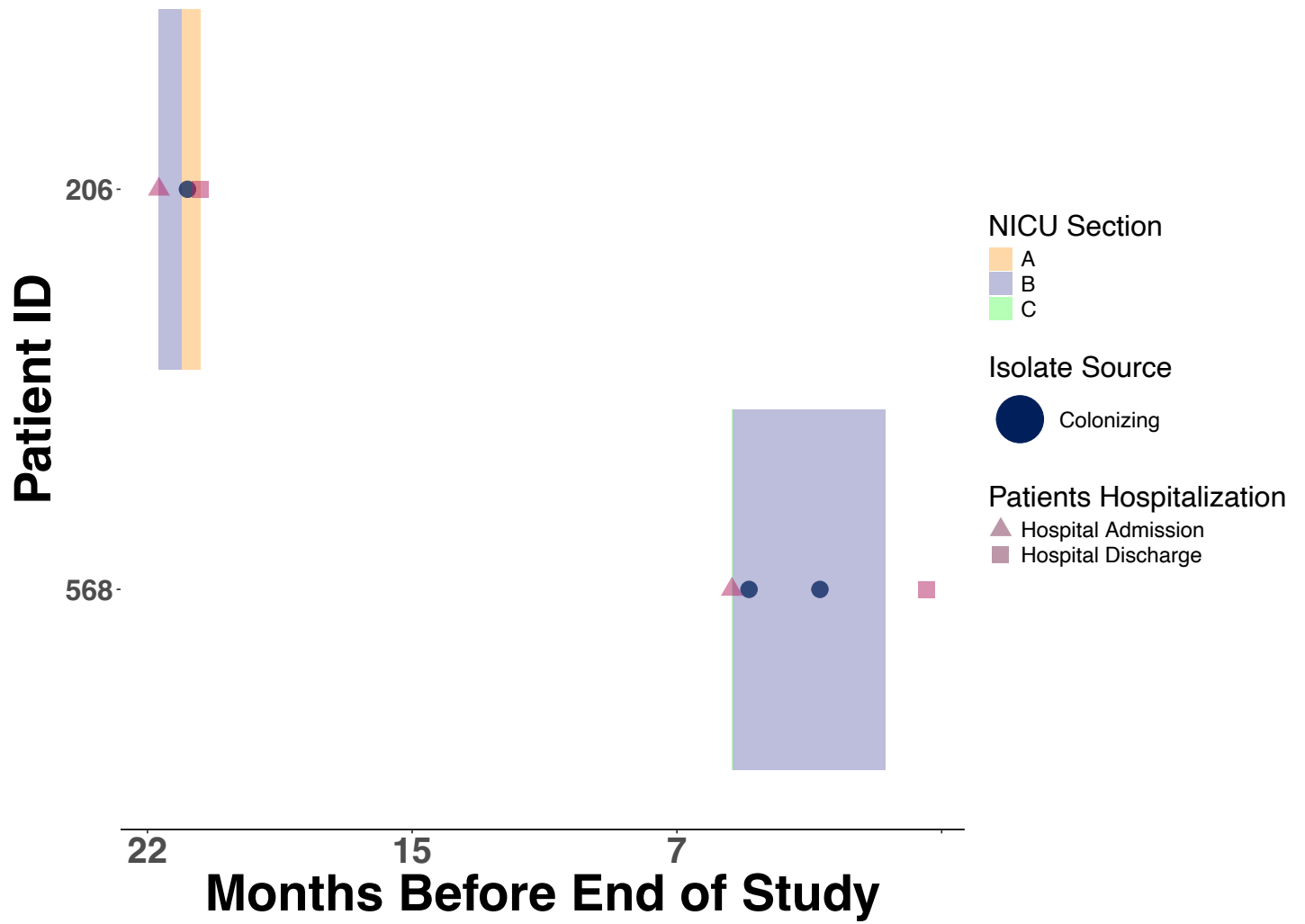
Cluster65



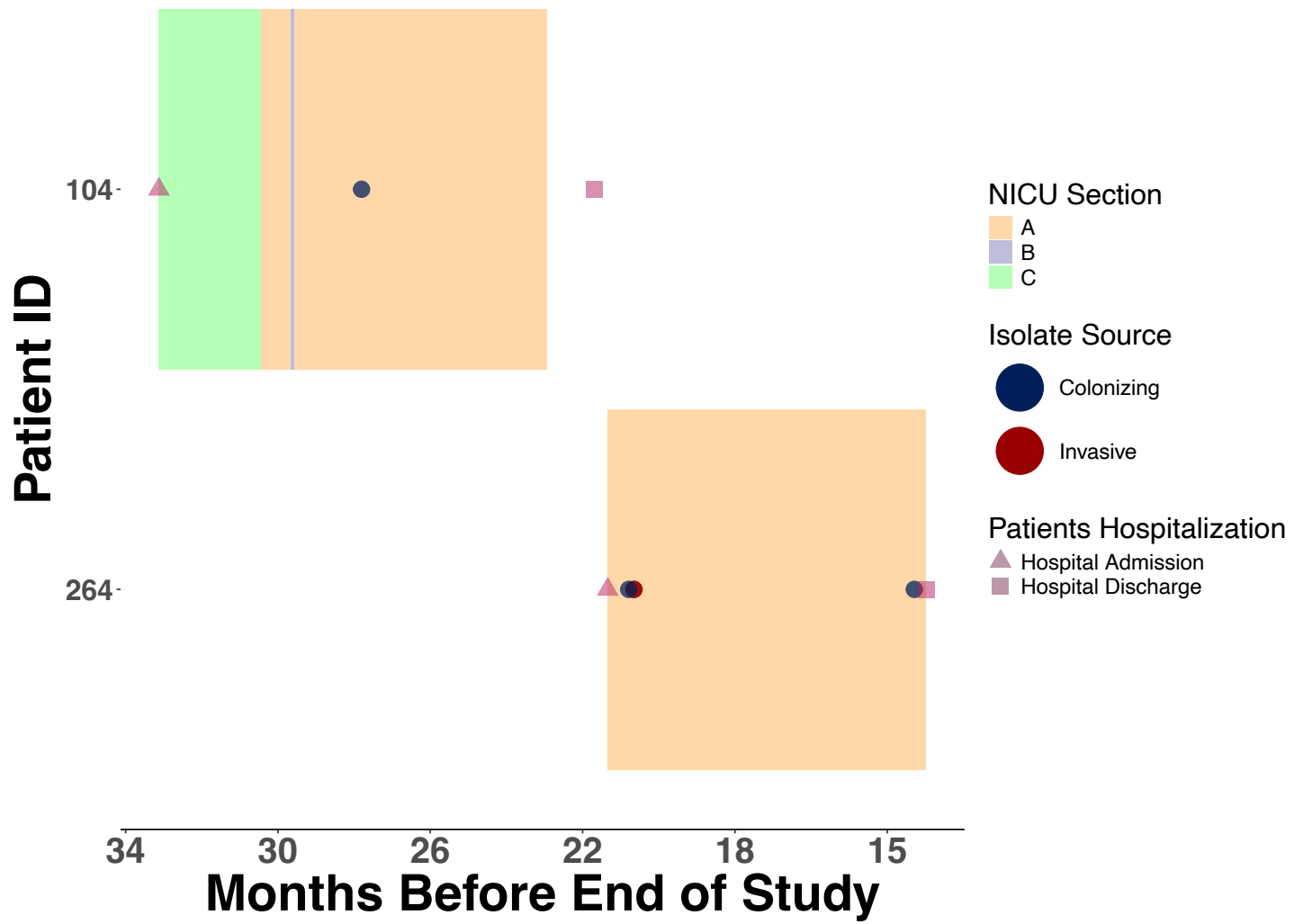
Cluster66



Cluster67



Cluster68



Cluster69

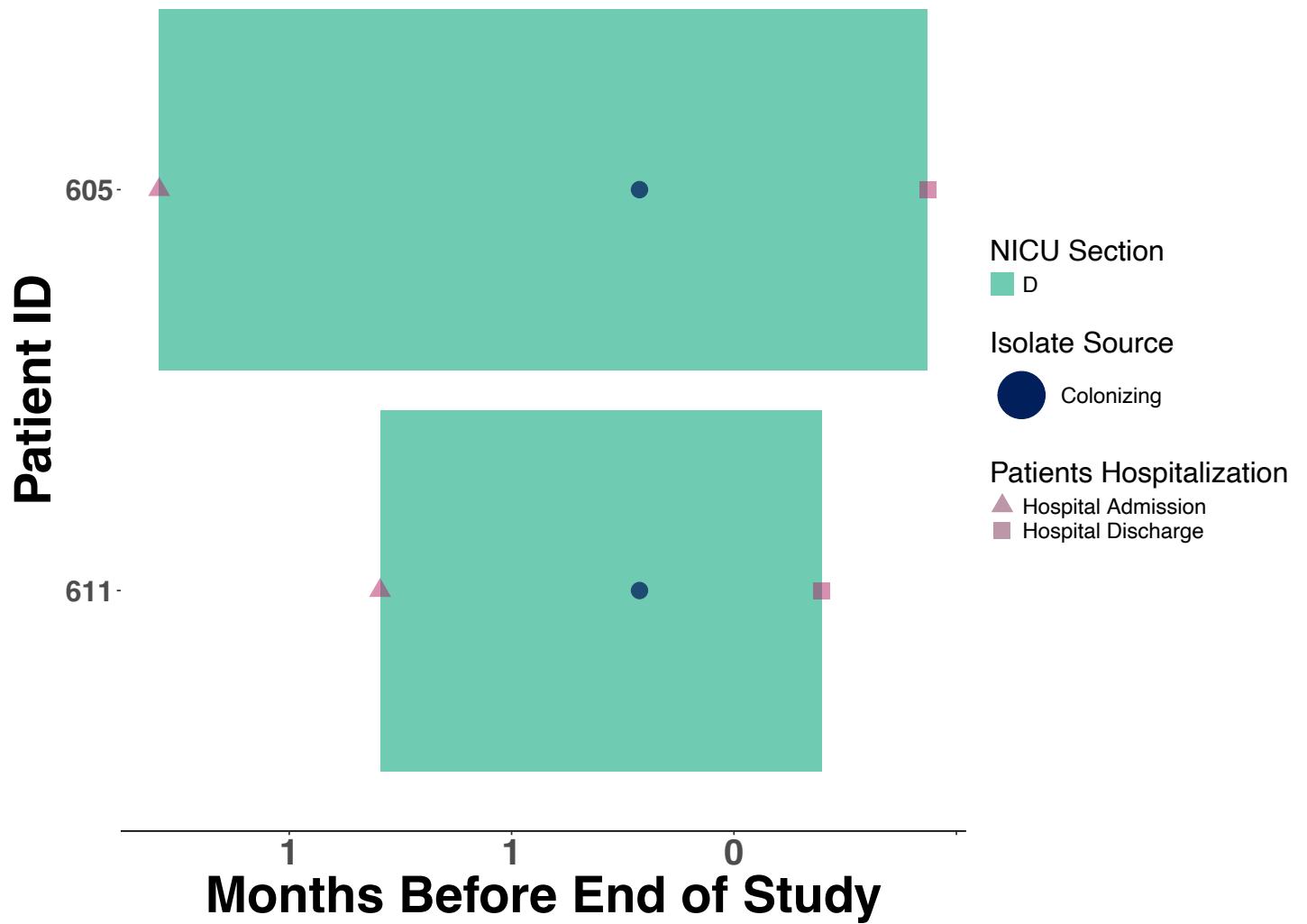
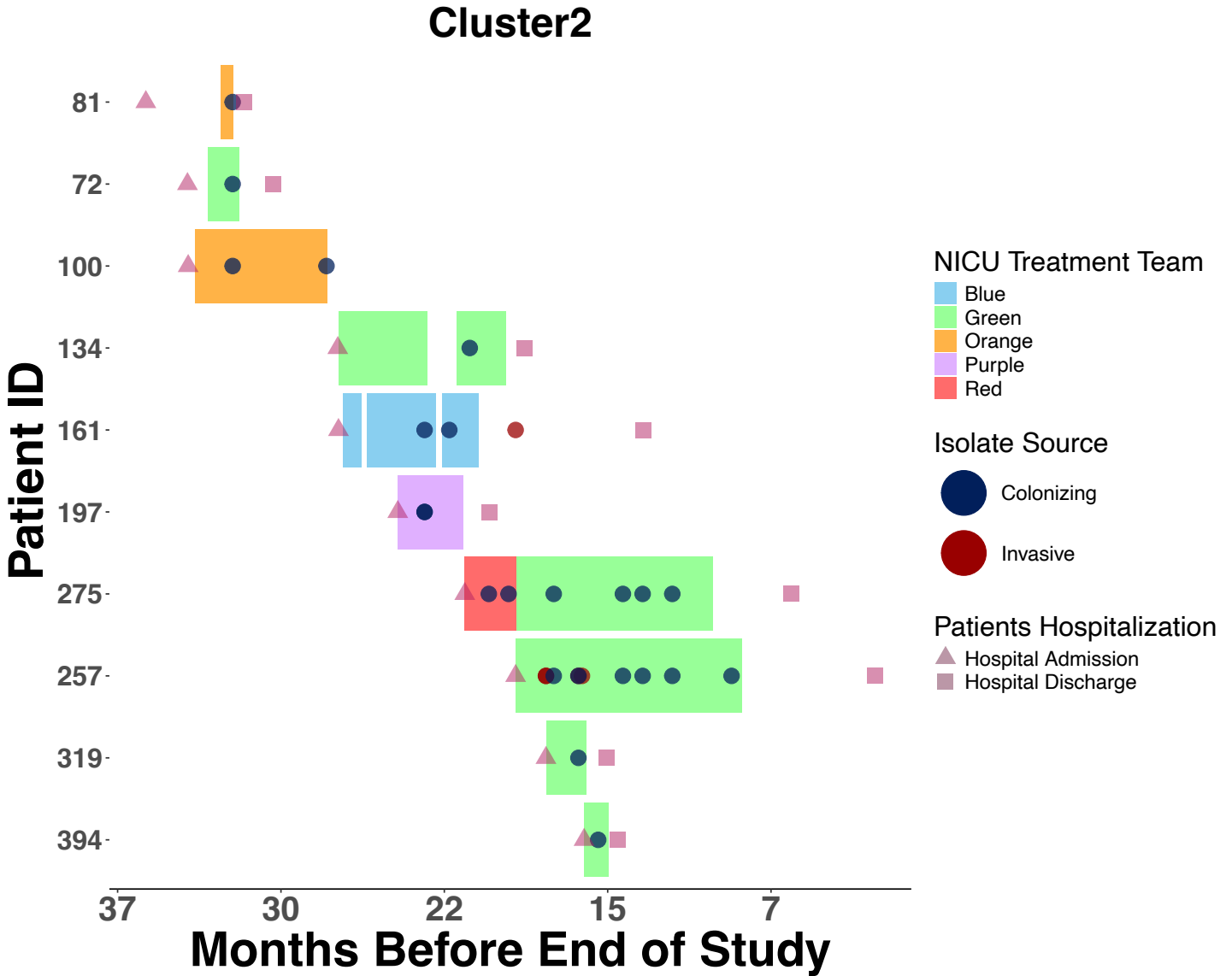
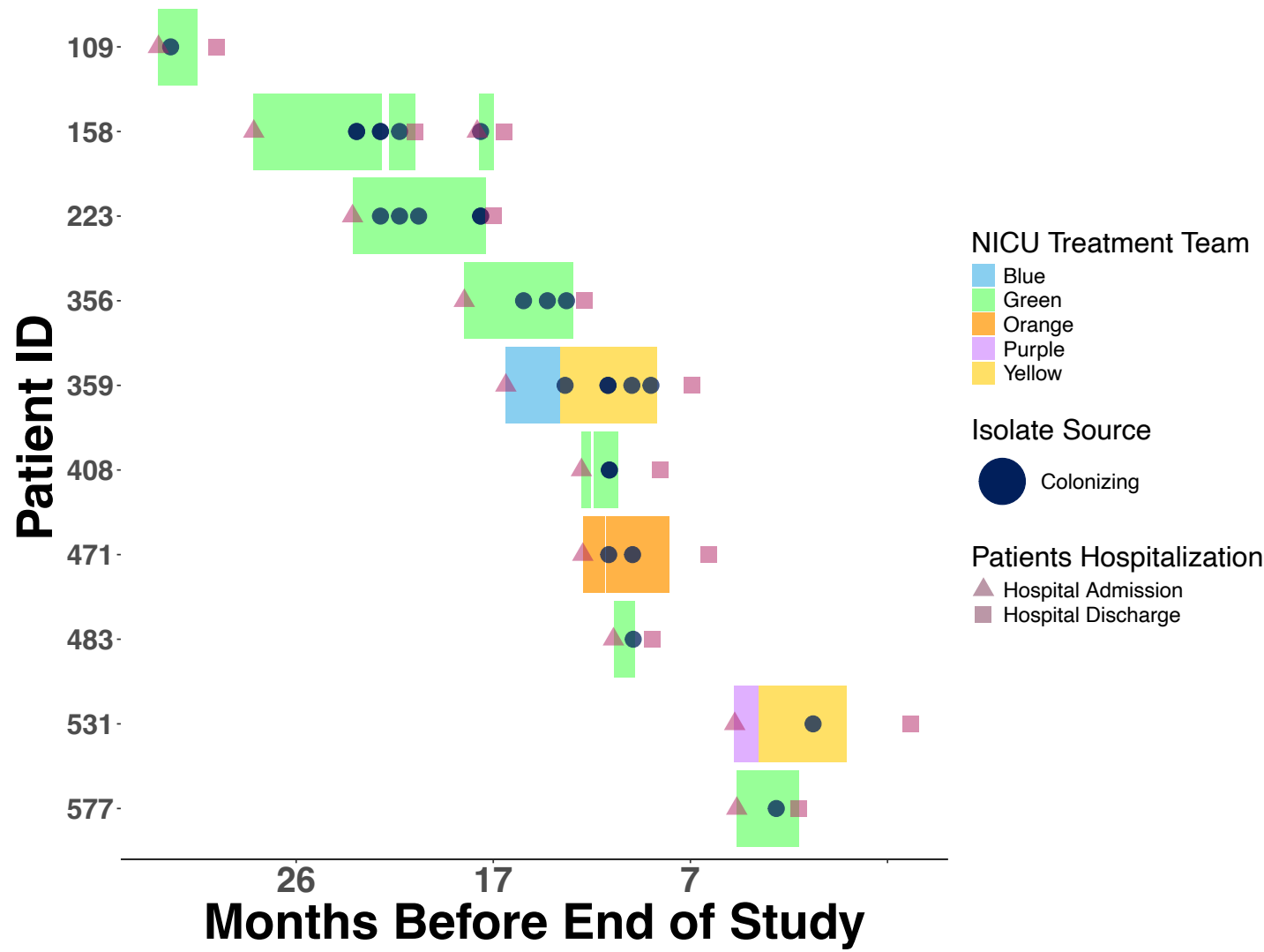


Figure S5: Temporal analysis for each identified transmission cluster by treatment team assignments

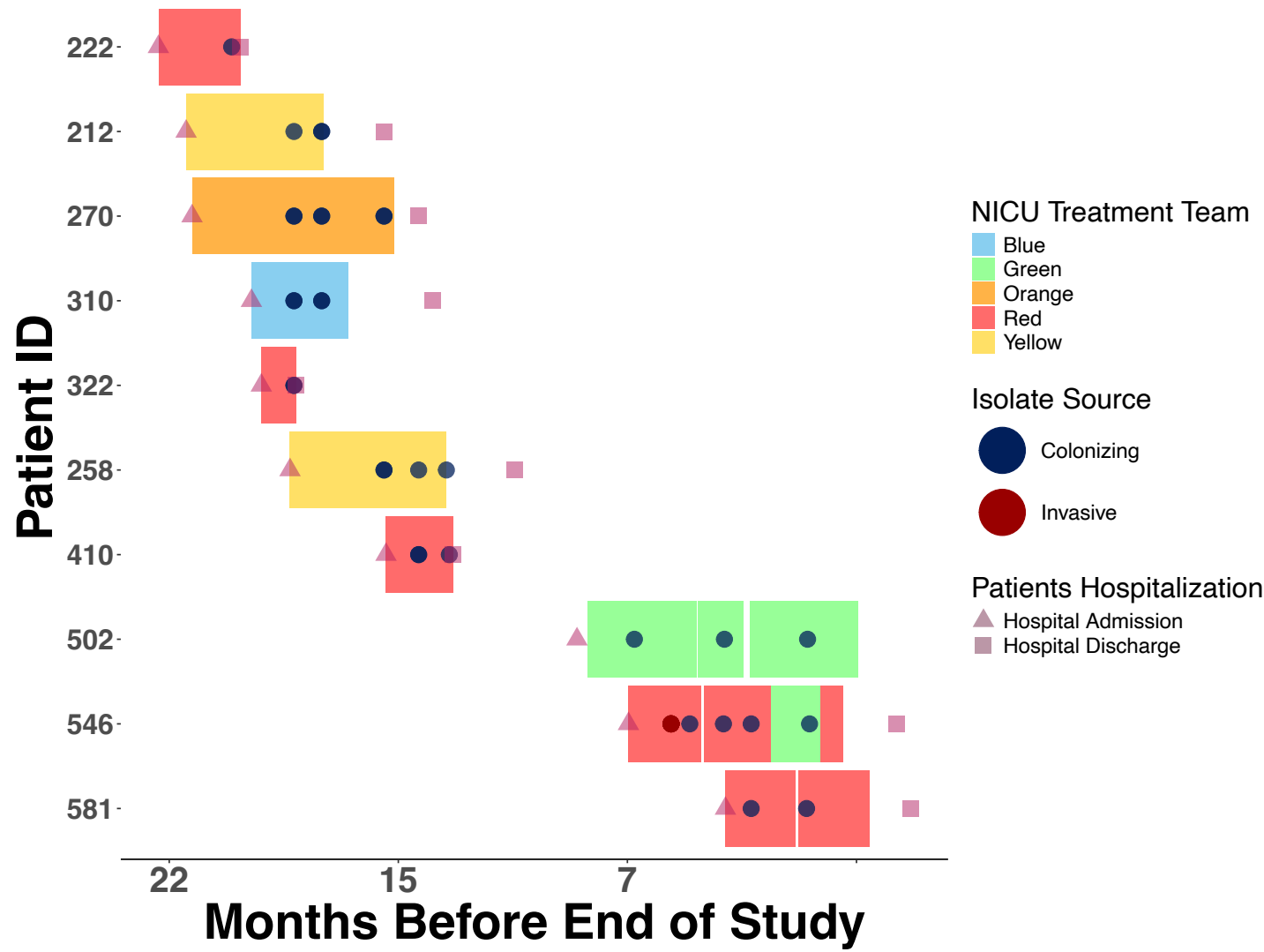
Each horizontal bar represents a single patient's NICU stay, with colored rectangles indicating corresponding treatment team assignments over time. Admission dates are marked with triangles and discharge dates with squares. X-axis (Months Before End of Study) represents the number of months prior to completion of study period.



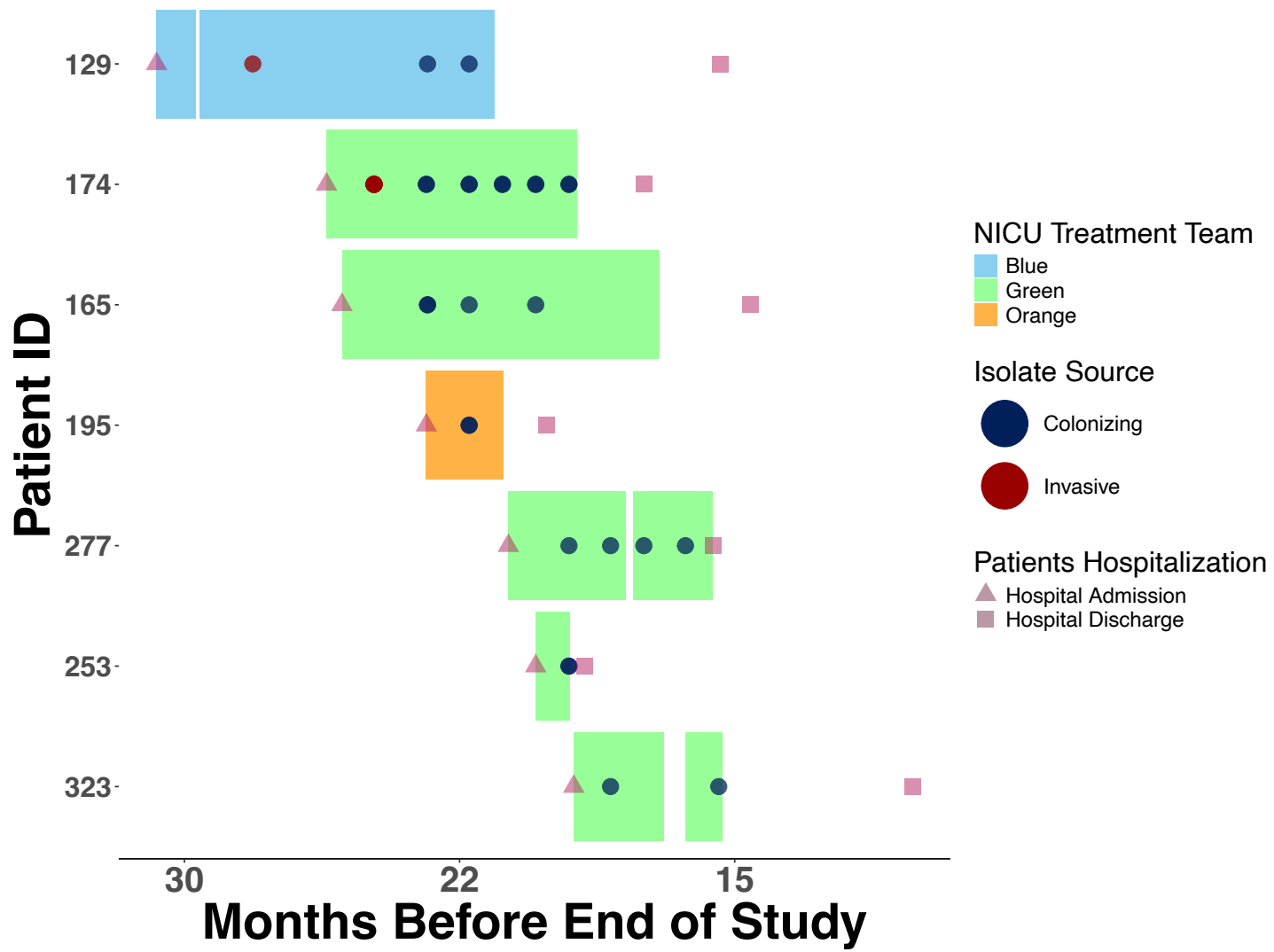
Cluster3



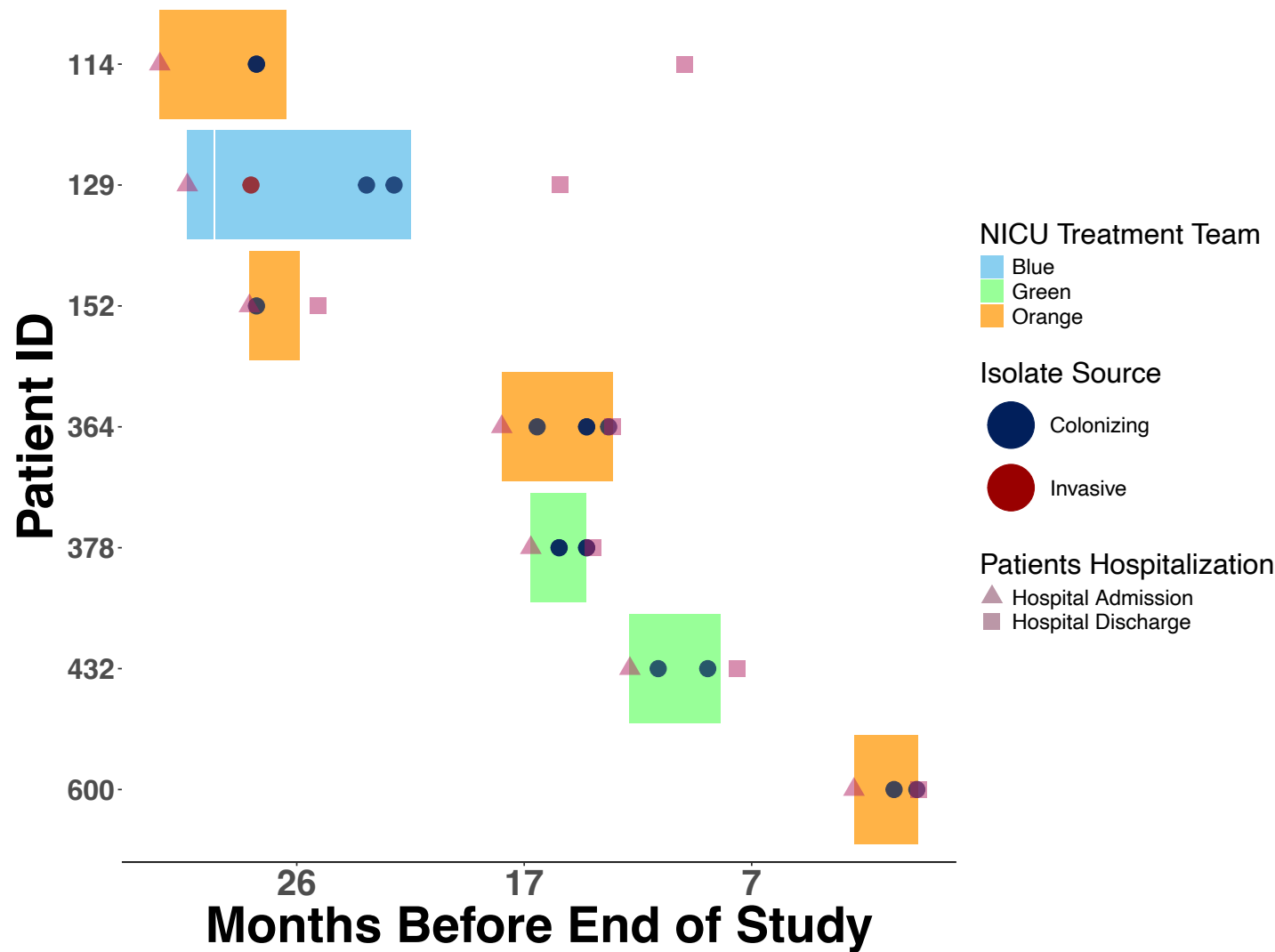
Cluster4



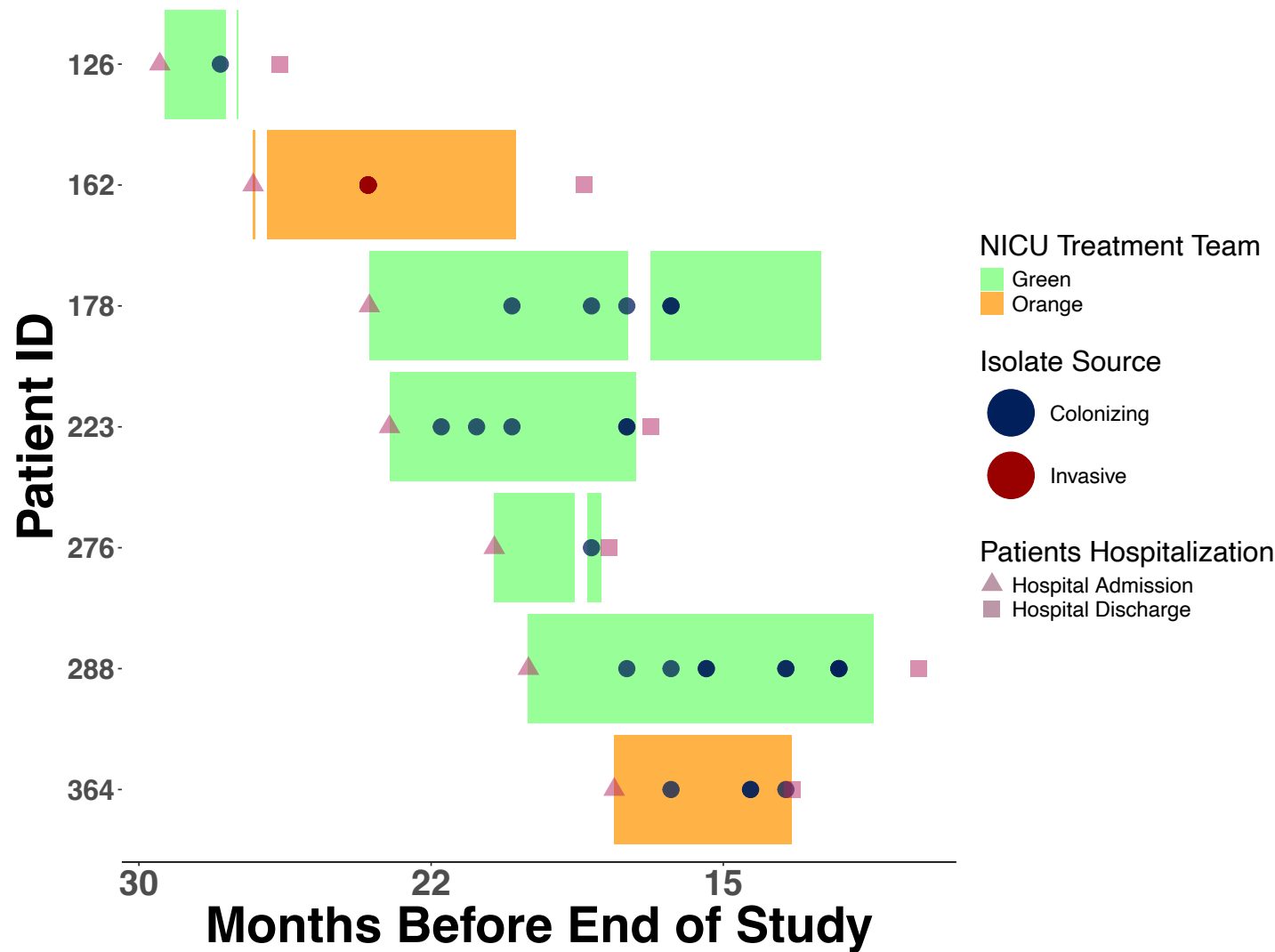
Cluster5



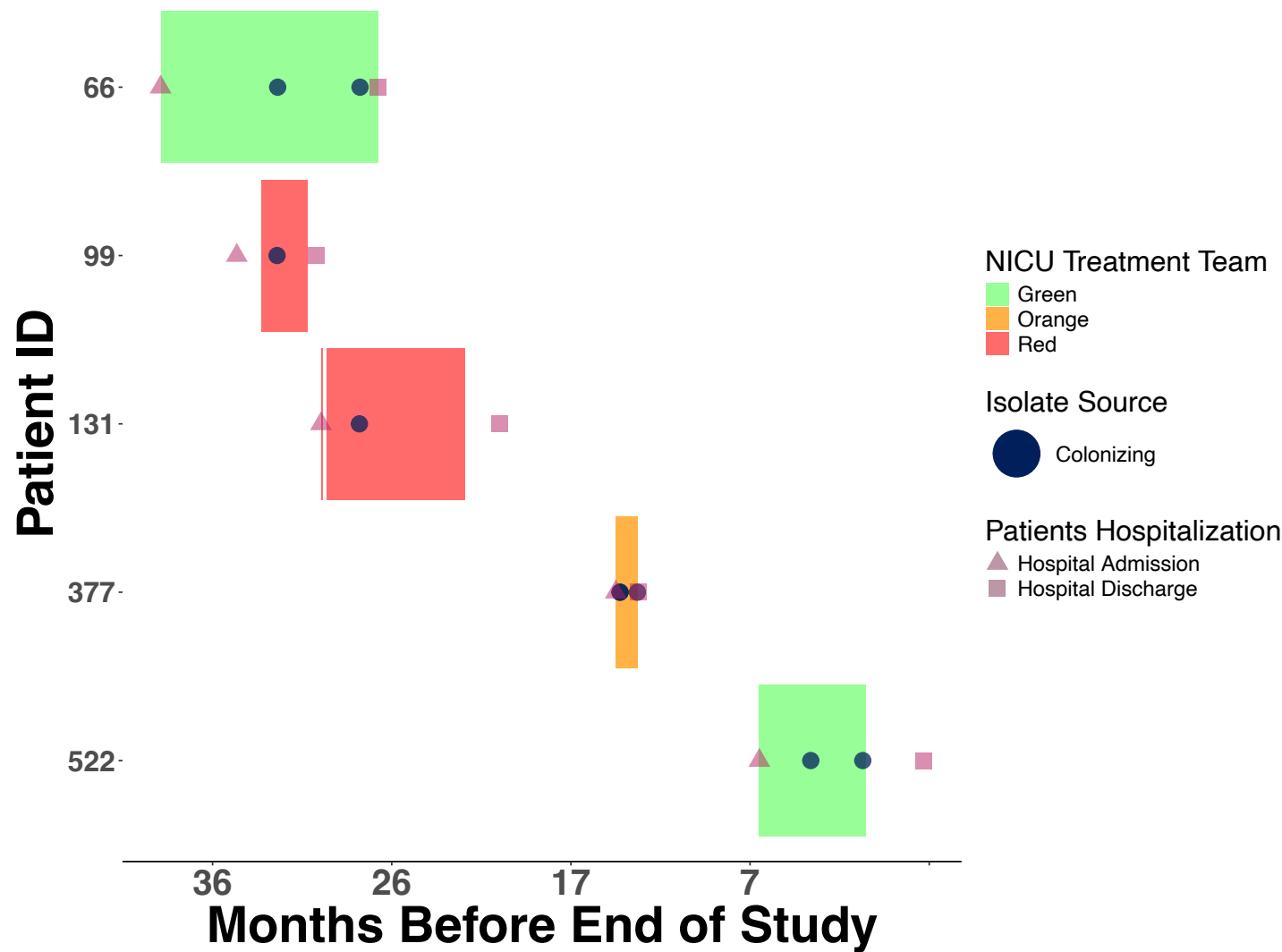
Cluster6



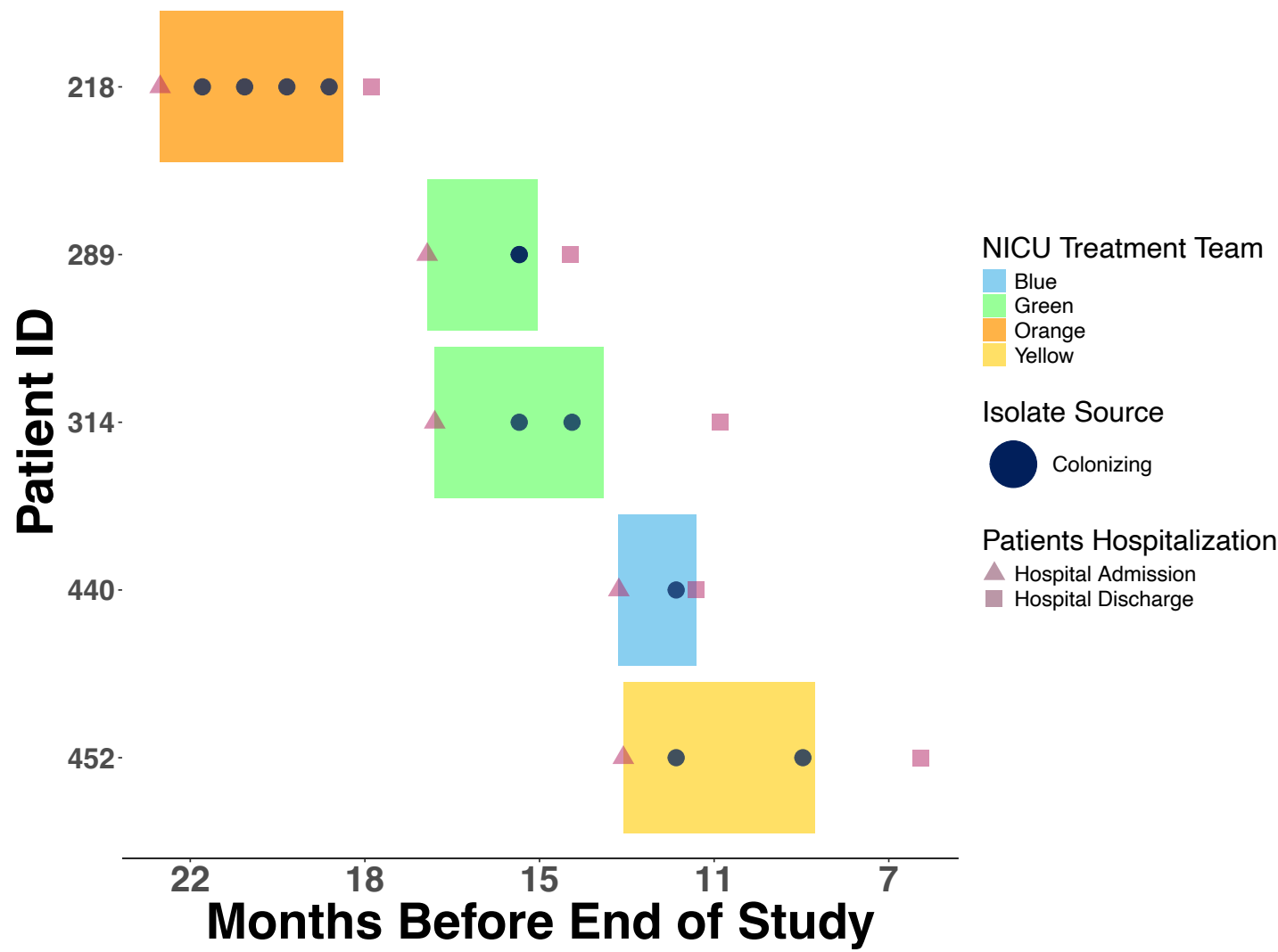
Cluster7



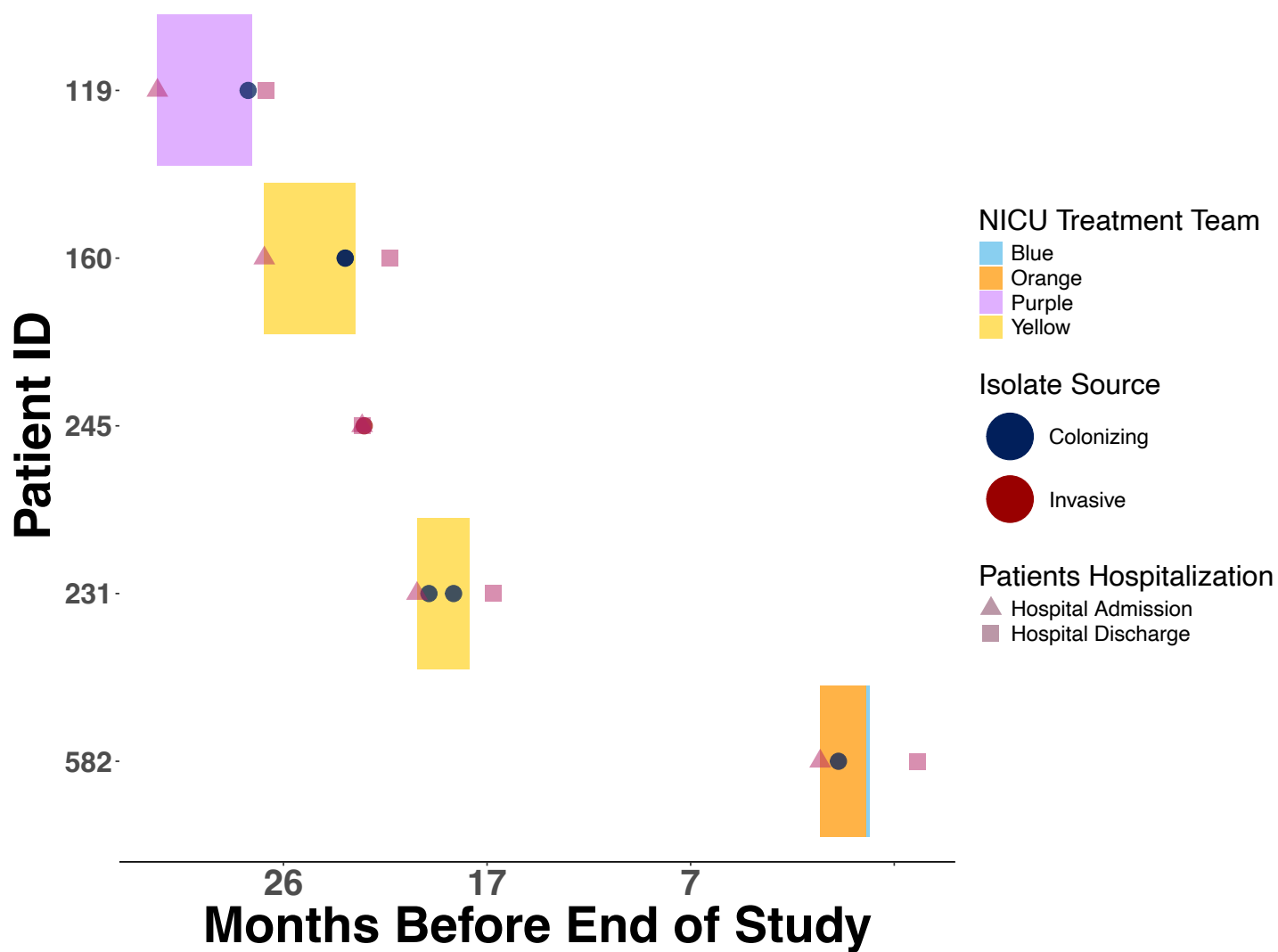
Cluster8



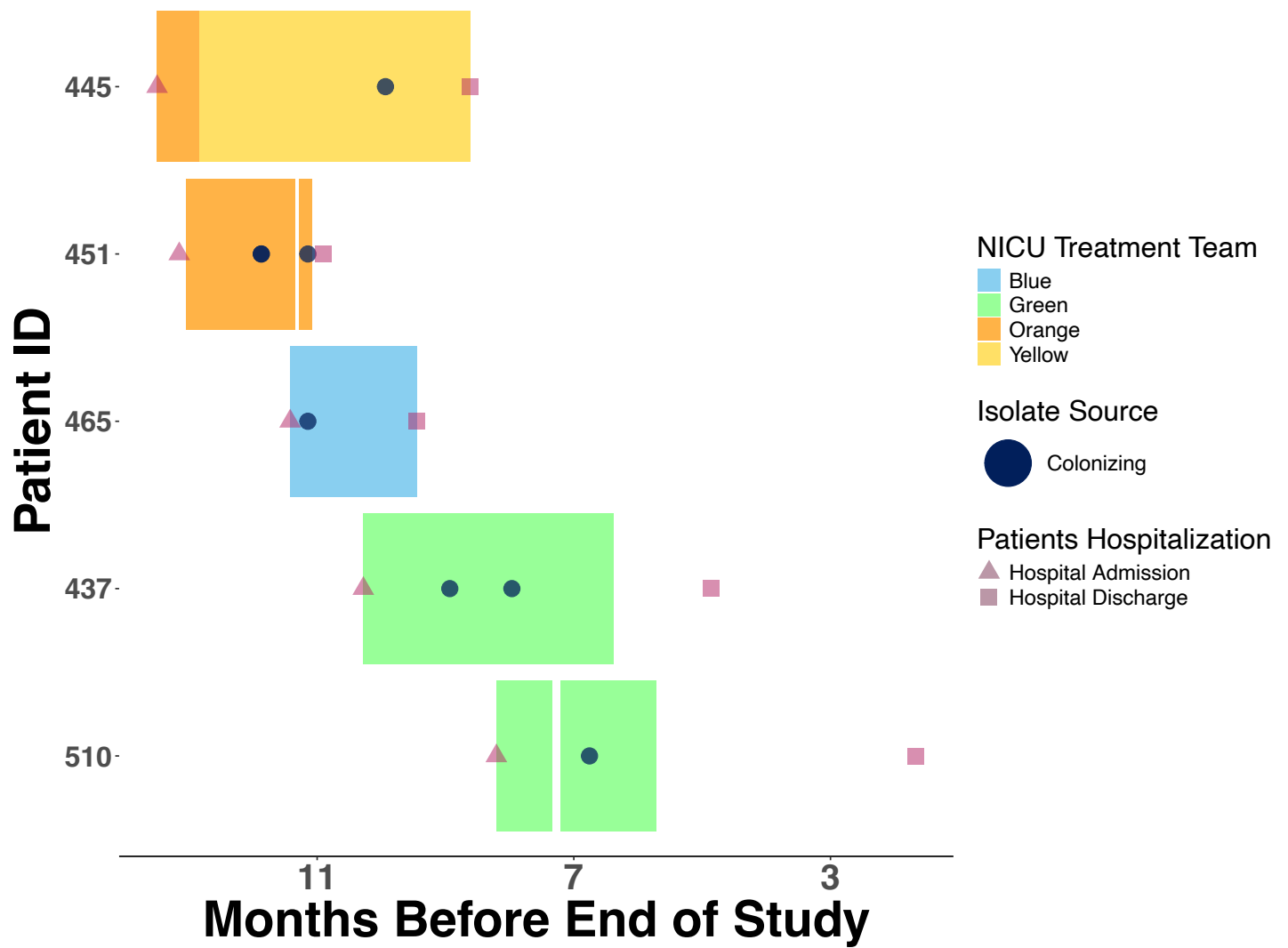
Cluster9



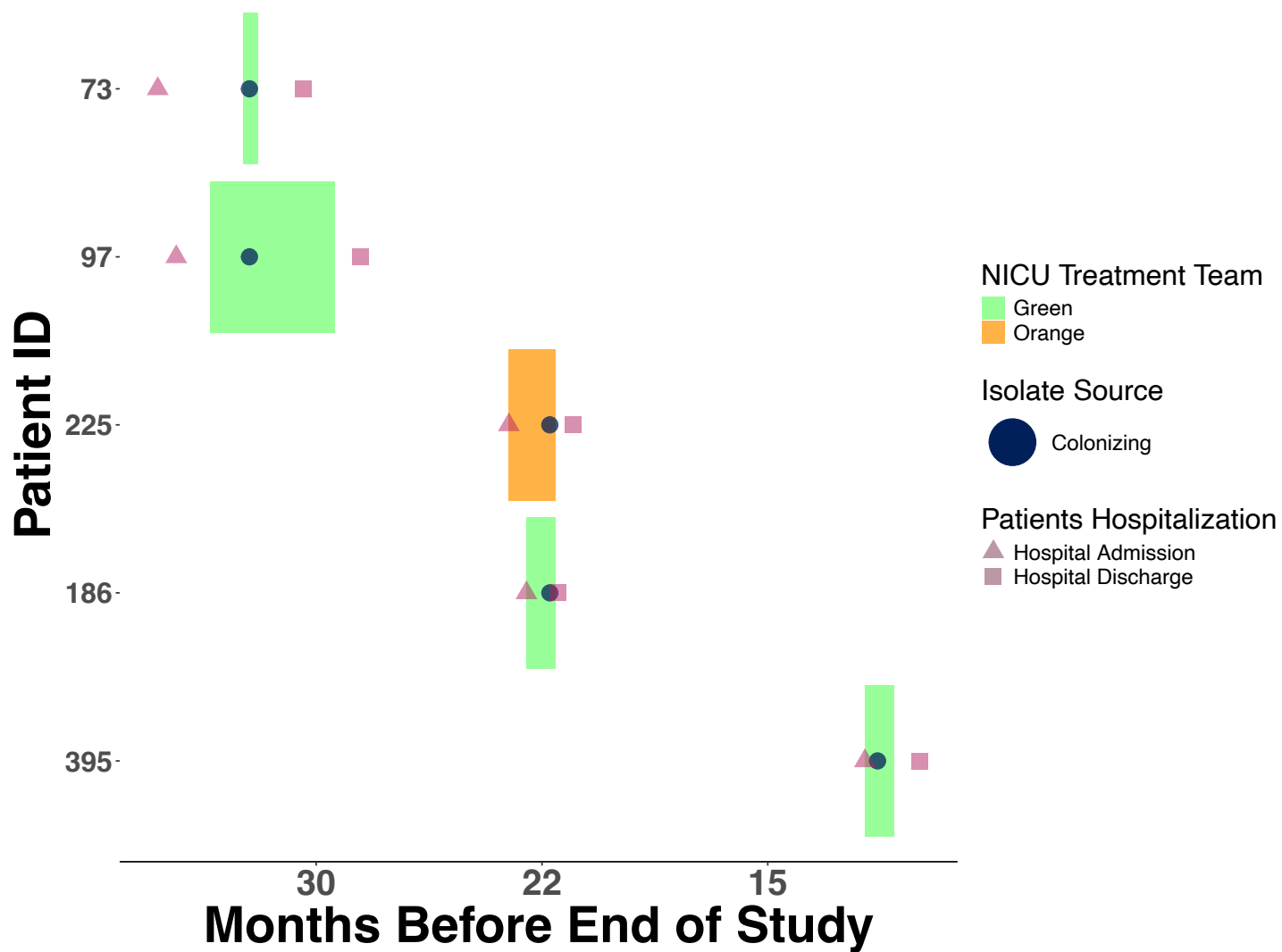
Cluster10



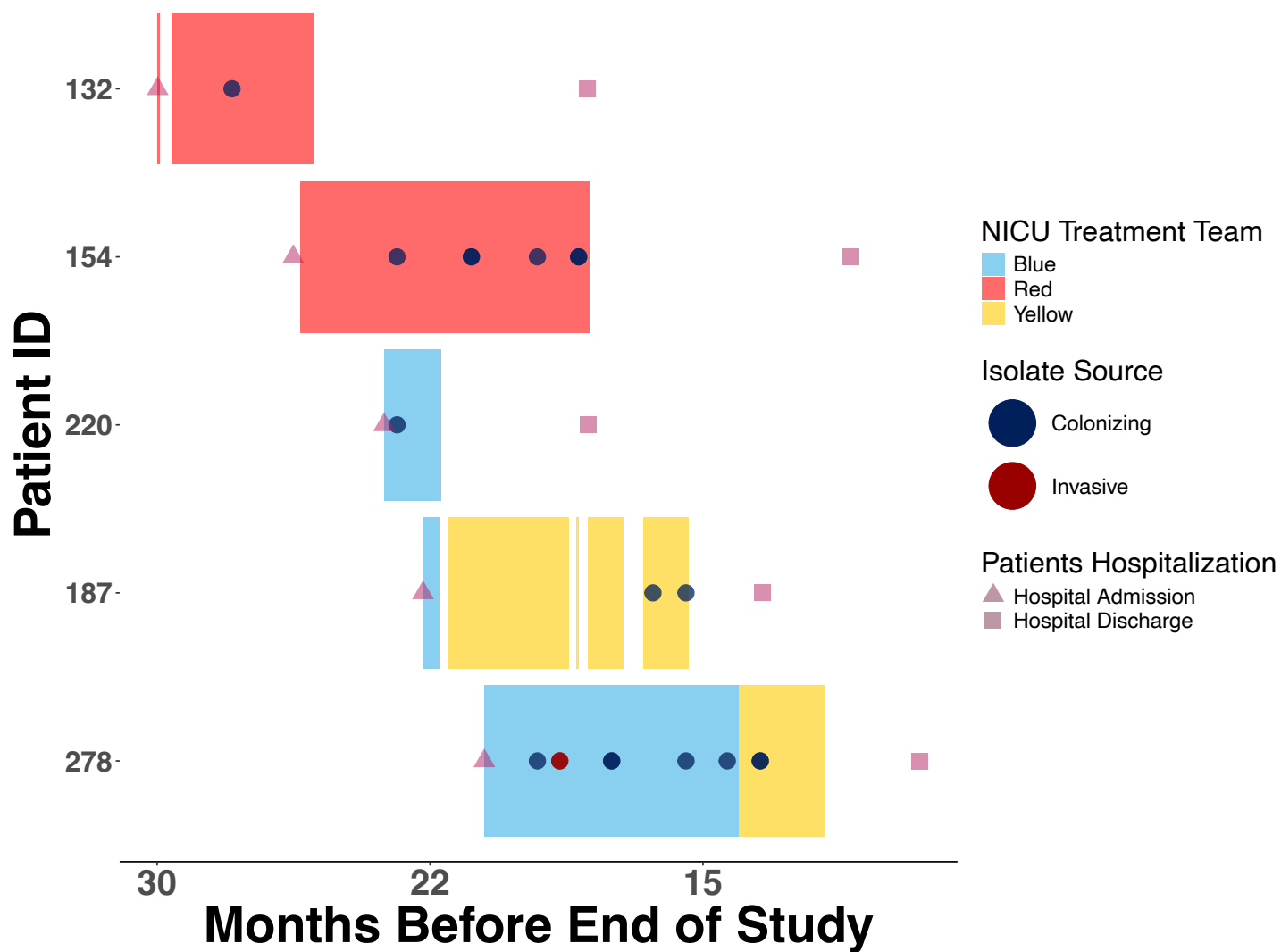
Cluster11



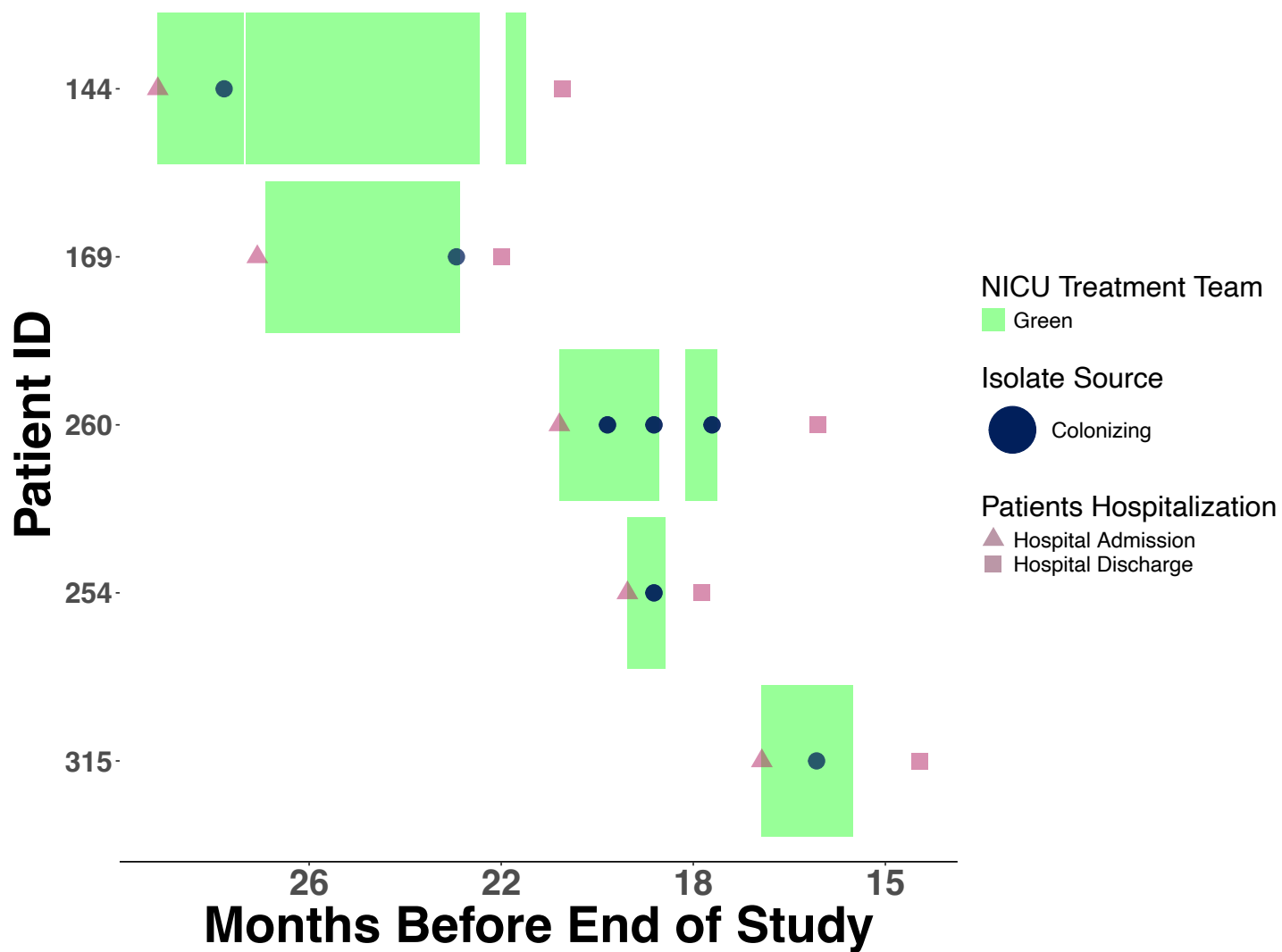
Cluster12



Cluster13



Cluster14



The figure is a dot plot with error bars representing the distribution of time to treatment discontinuation for three groups: Placebo (yellow), Active (green), and Control (red). The x-axis is labeled 'Months Before End of Study' and has major ticks at 36, 26, 17, and 7. The y-axis is labeled 'Number of Patients' and ranges from 0 to 100 in increments of 10. The error bars represent the distribution of time to treatment discontinuation, with a vertical line indicating the median. Individual patient data points are shown as dots, with triangles indicating censored data.

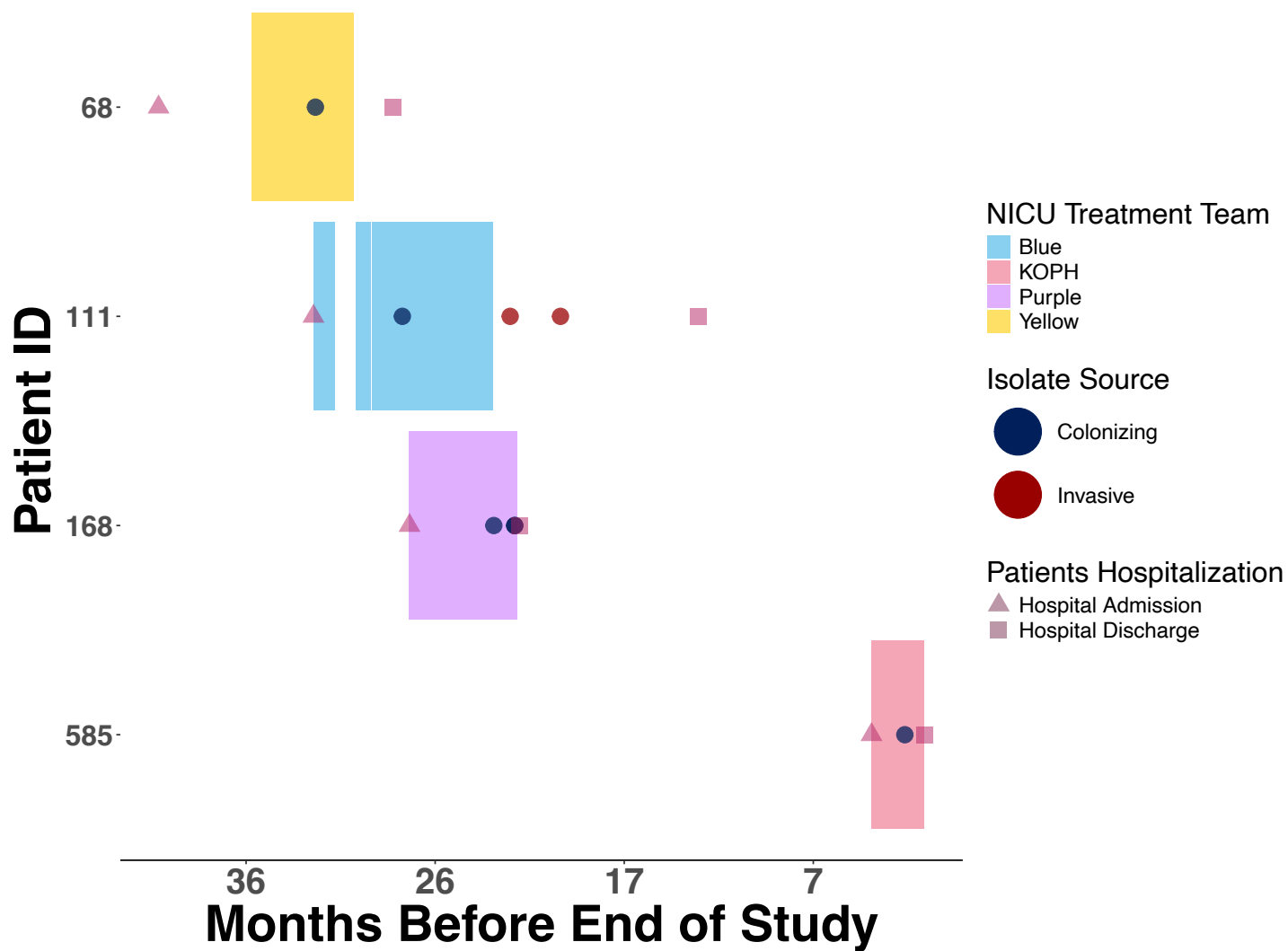
Group	Median (Months)	Q1 (Months)	Q3 (Months)	Min (Months)	Max (Months)
Placebo	~30	~28	~32	~25	~35
Active	~22	~20	~24	~18	~26
Control	~8	~6	~10	~4	~12

Green
Red
Yellow

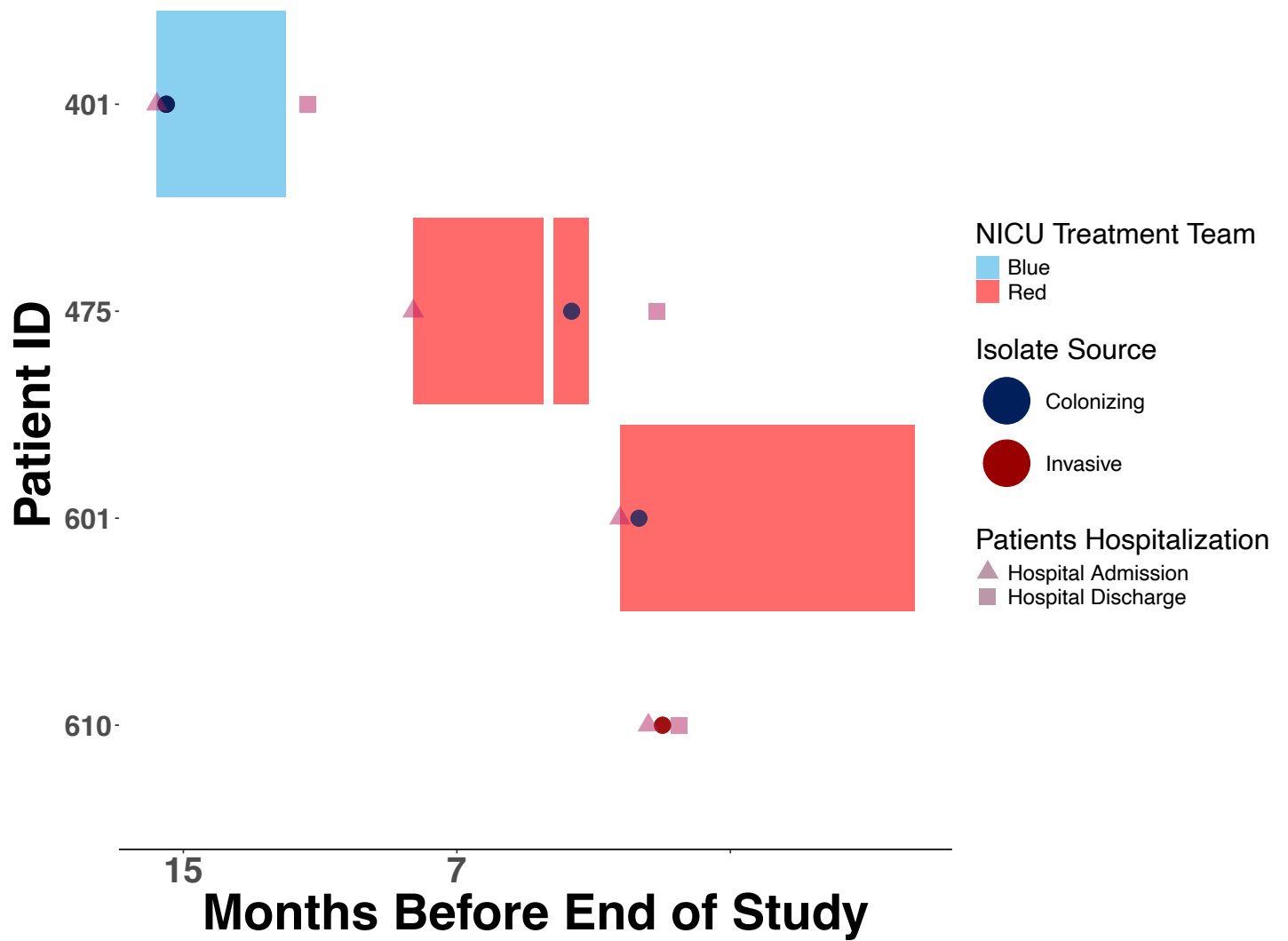
- Colonizing
- Invasive

 Hospital Admission
 Hospital Discharge

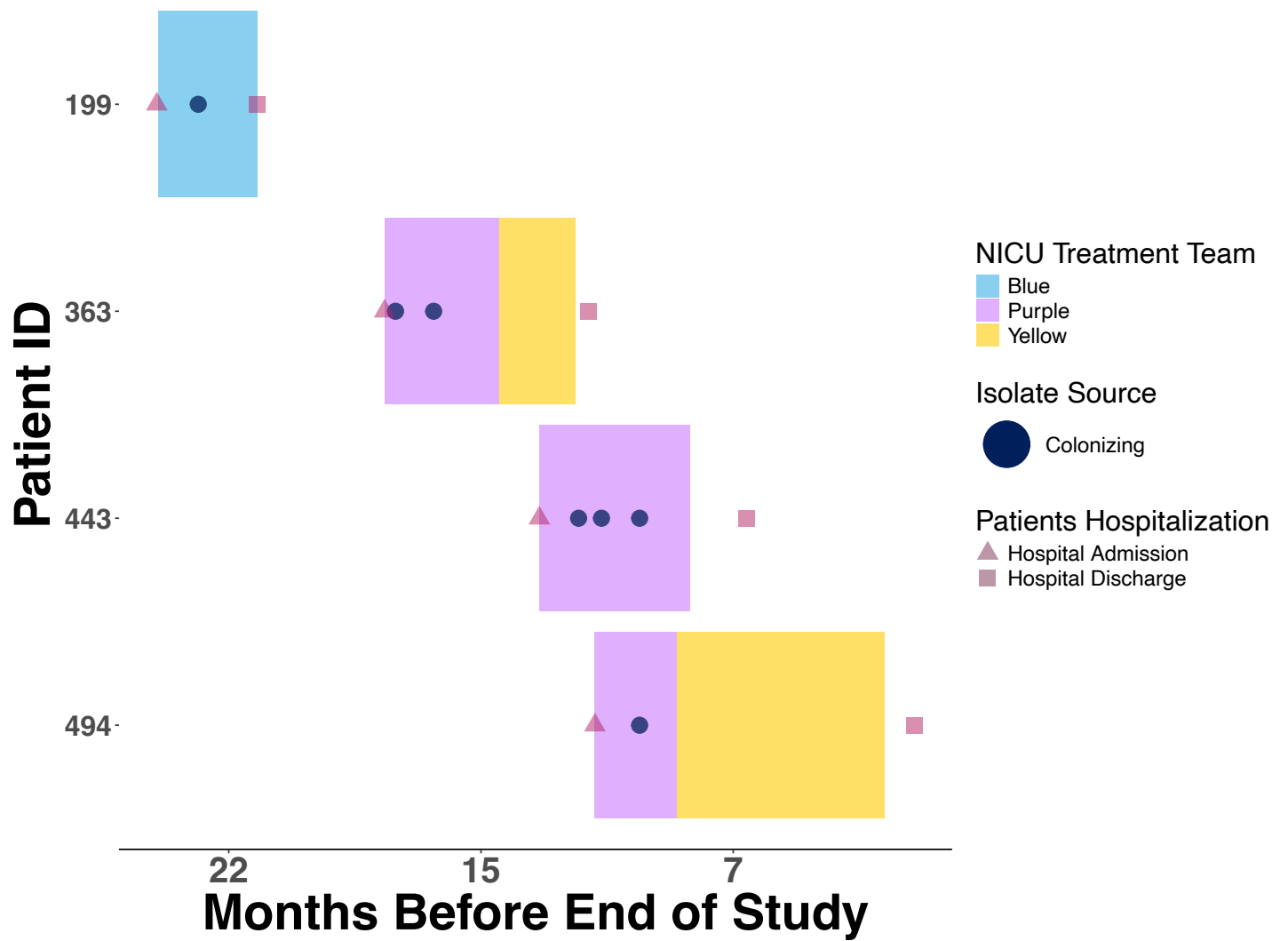
Cluster16



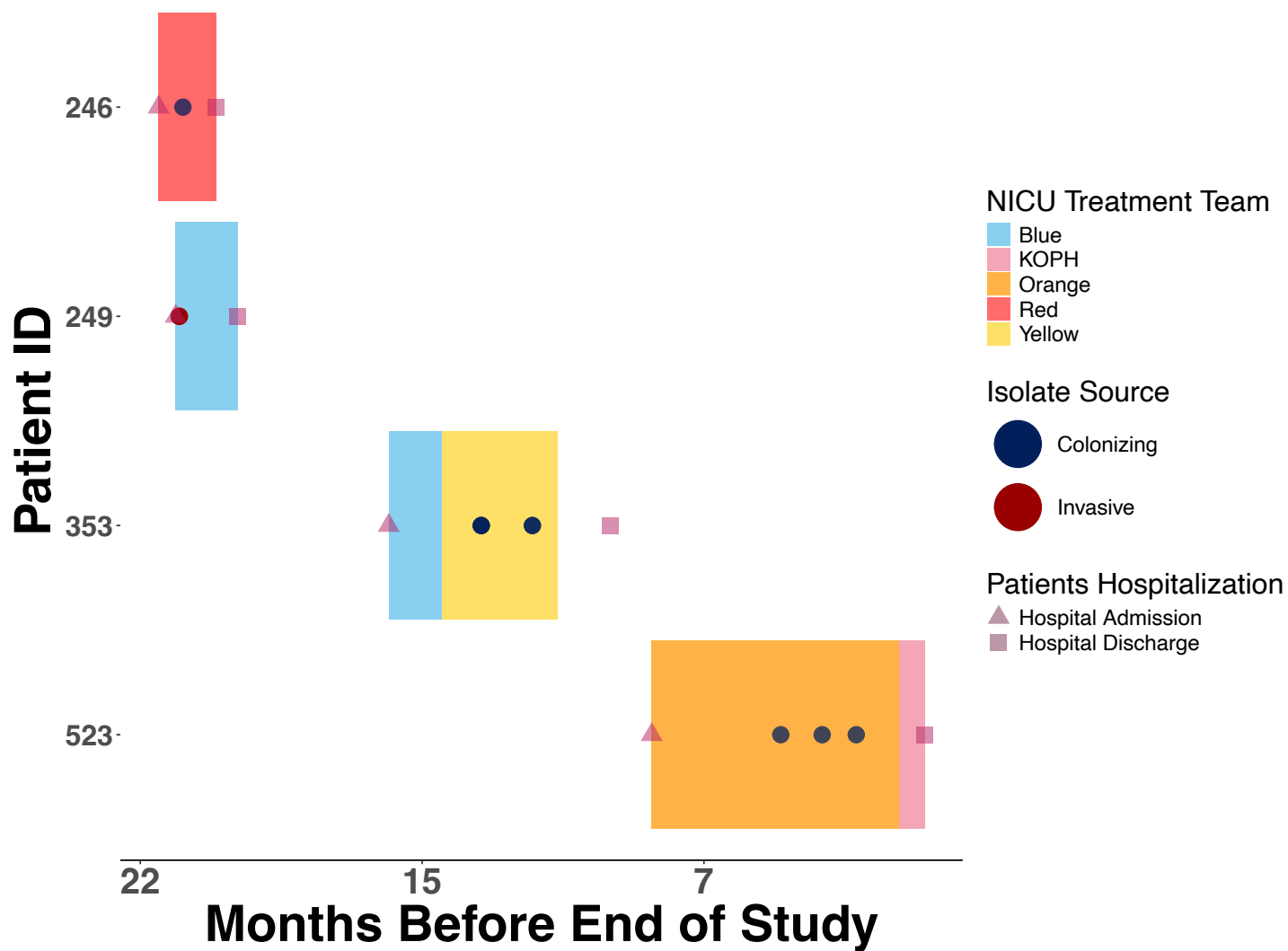
Cluster17



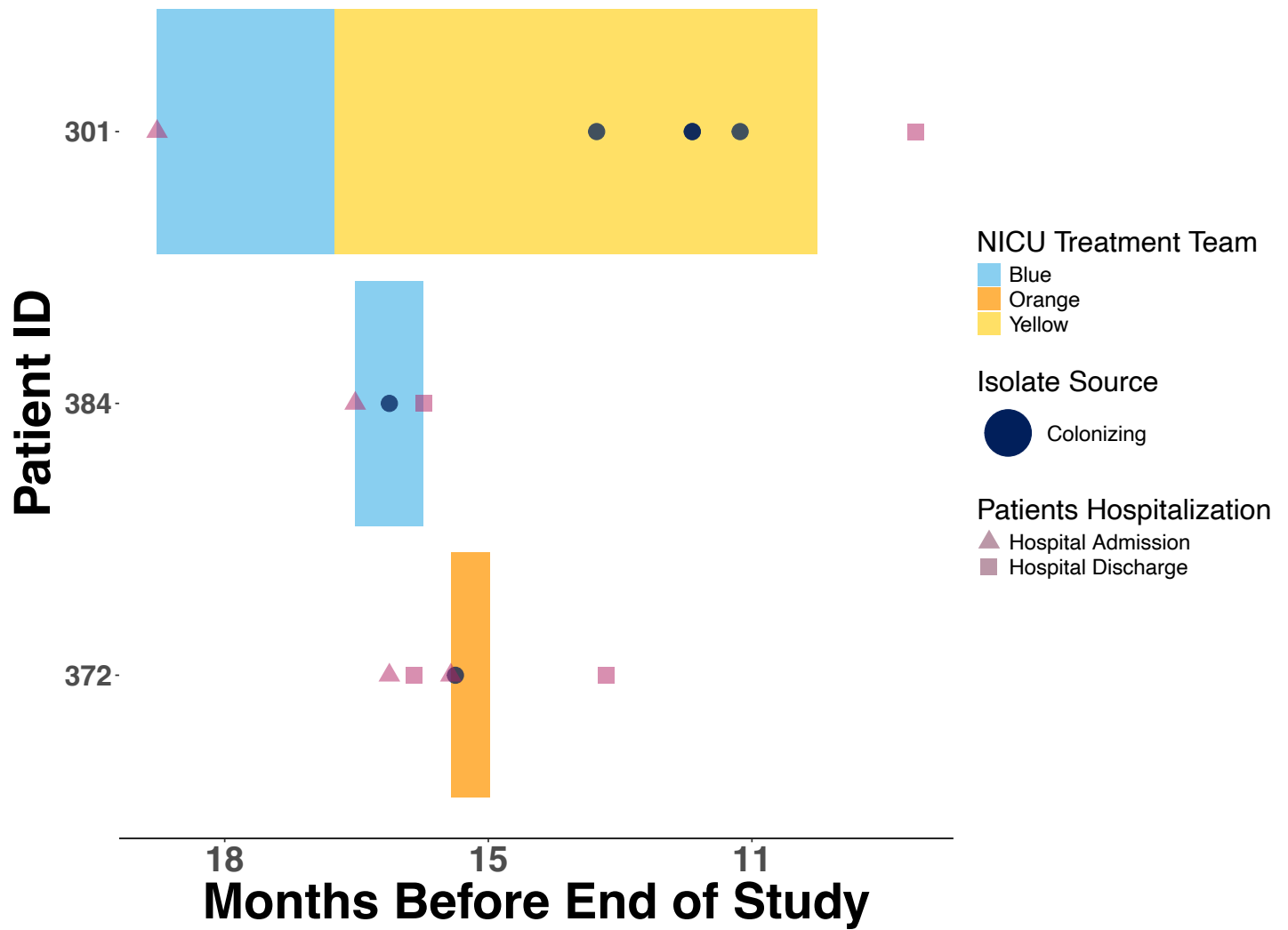
Cluster18



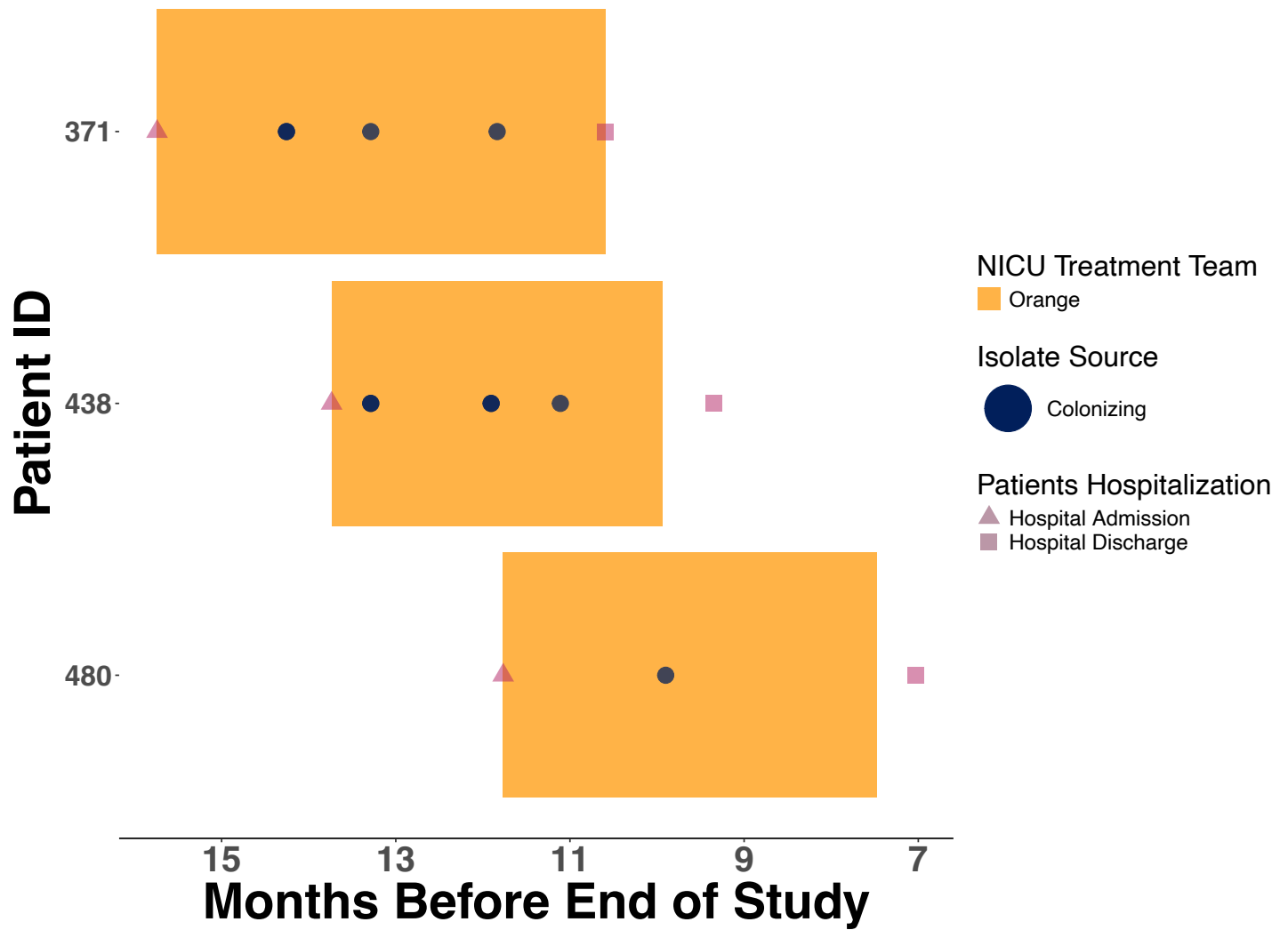
Cluster19



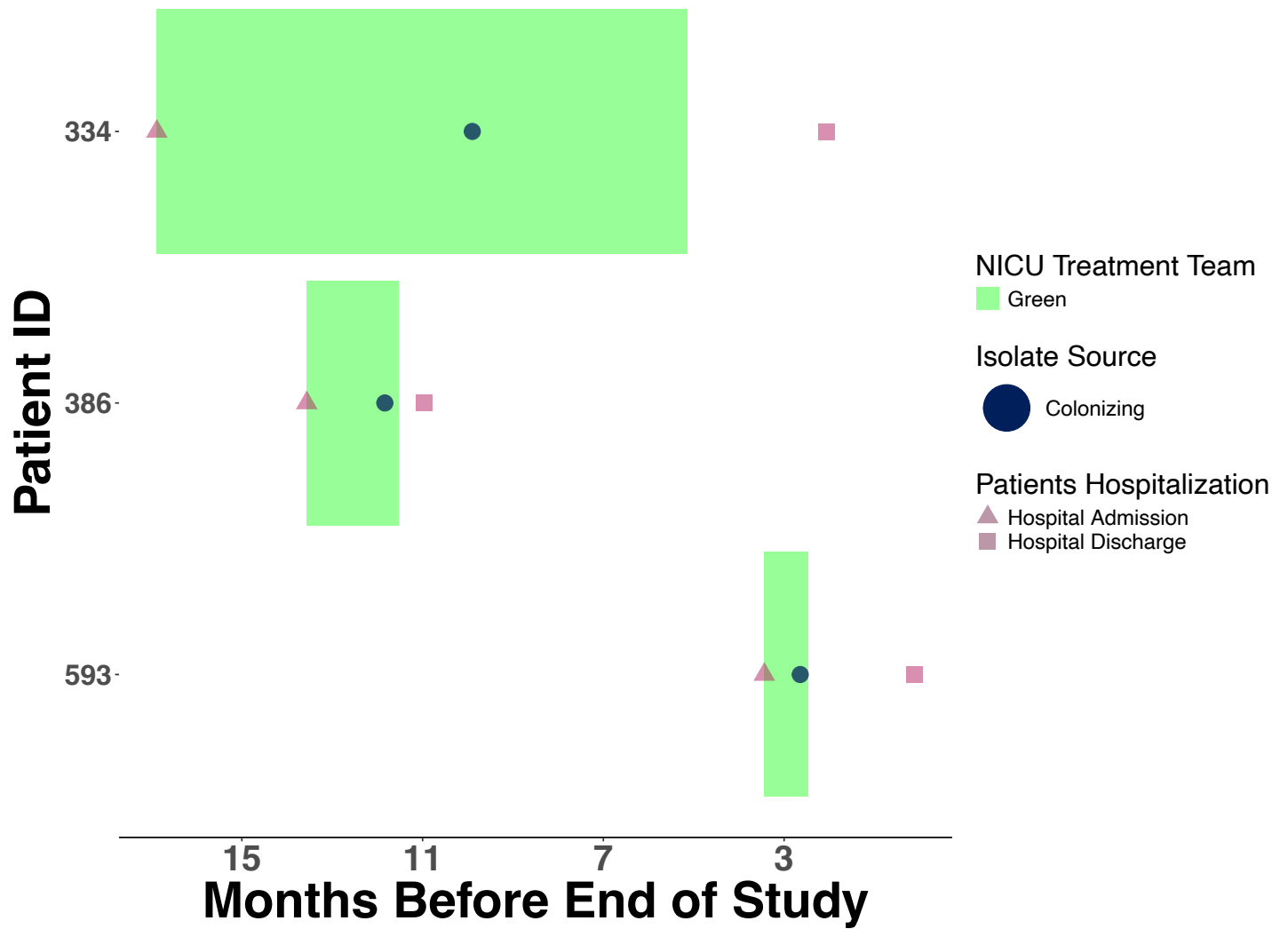
Cluster20



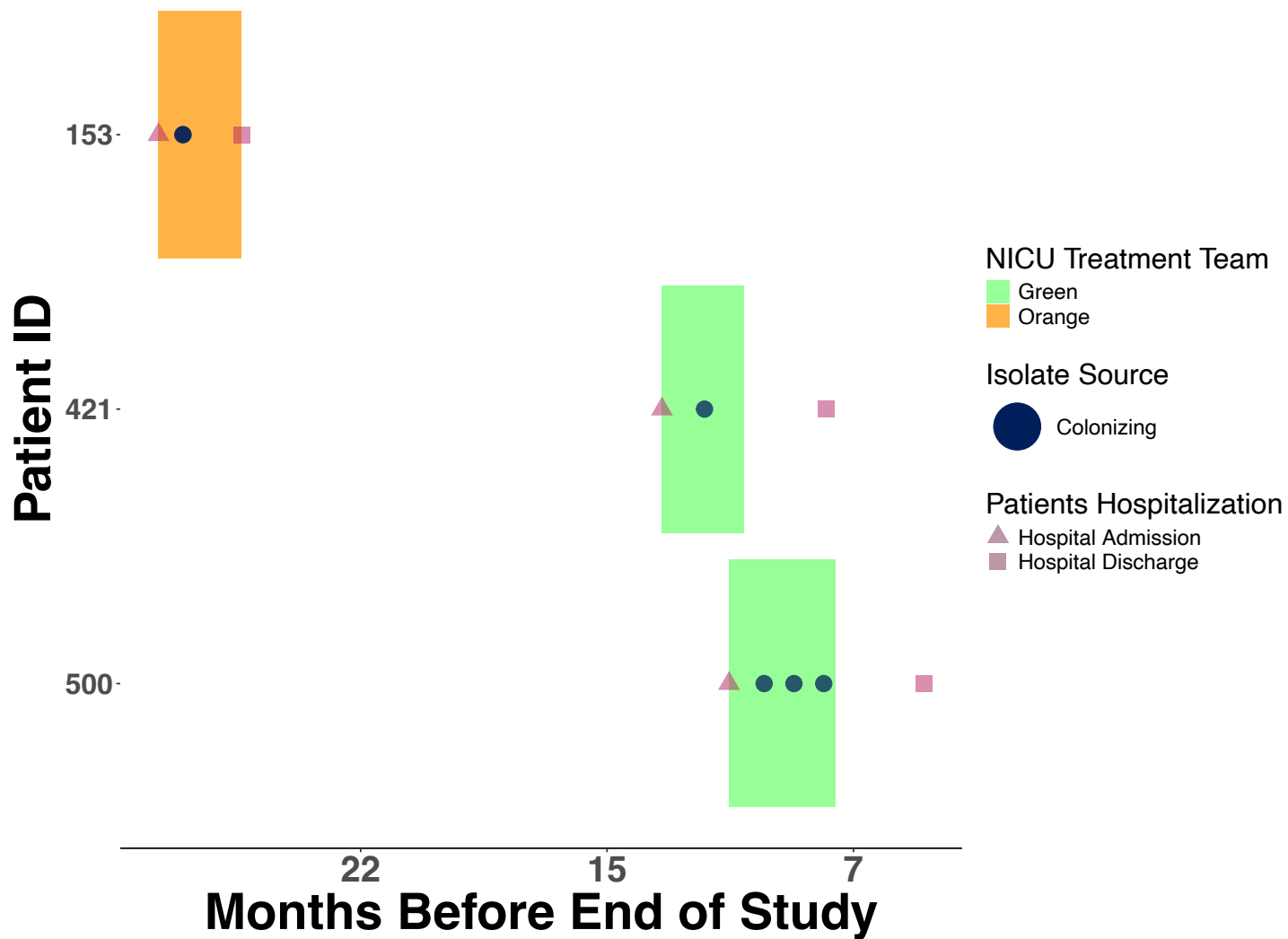
Cluster21



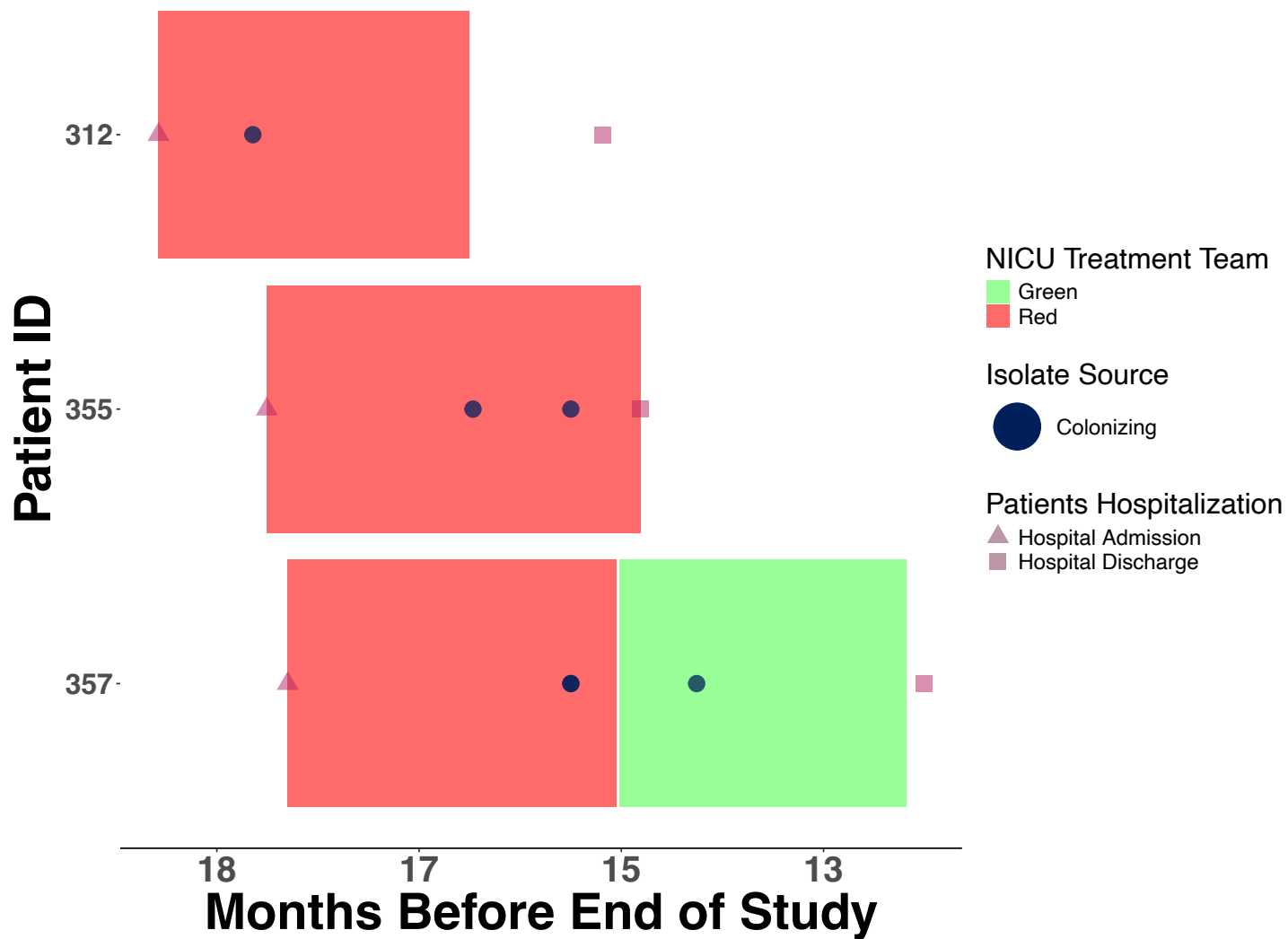
Cluster22



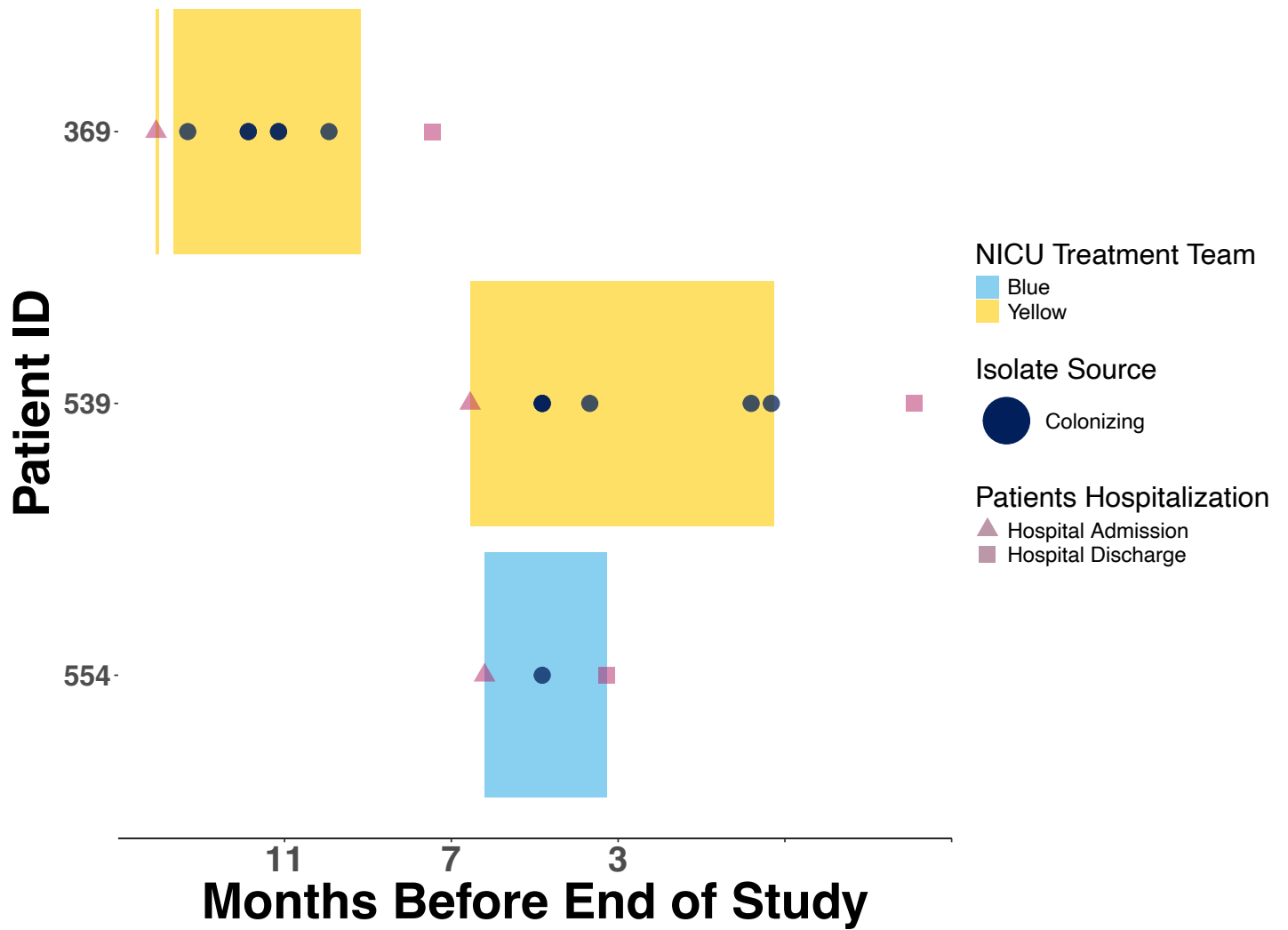
Cluster23



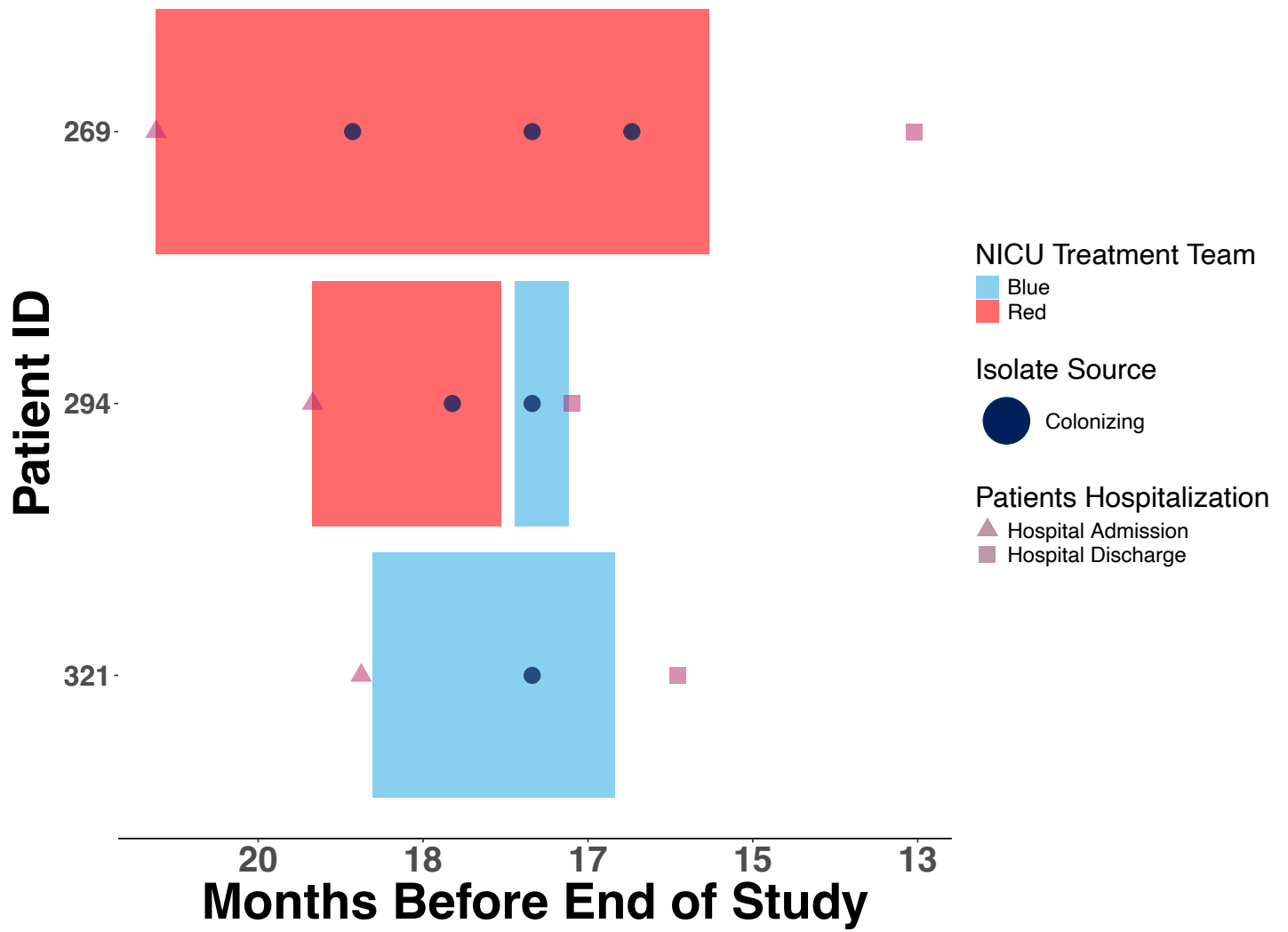
Cluster24



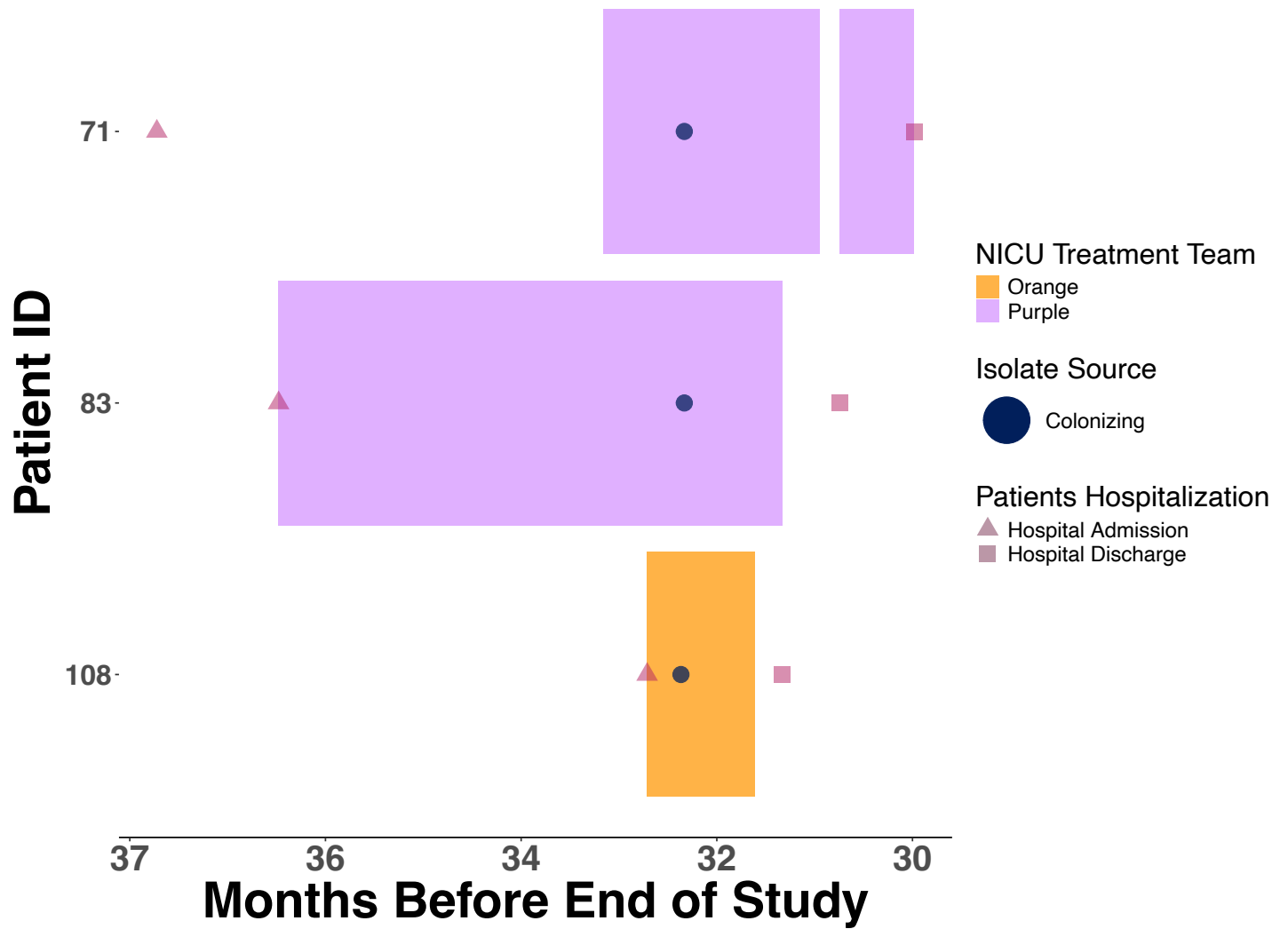
Cluster25



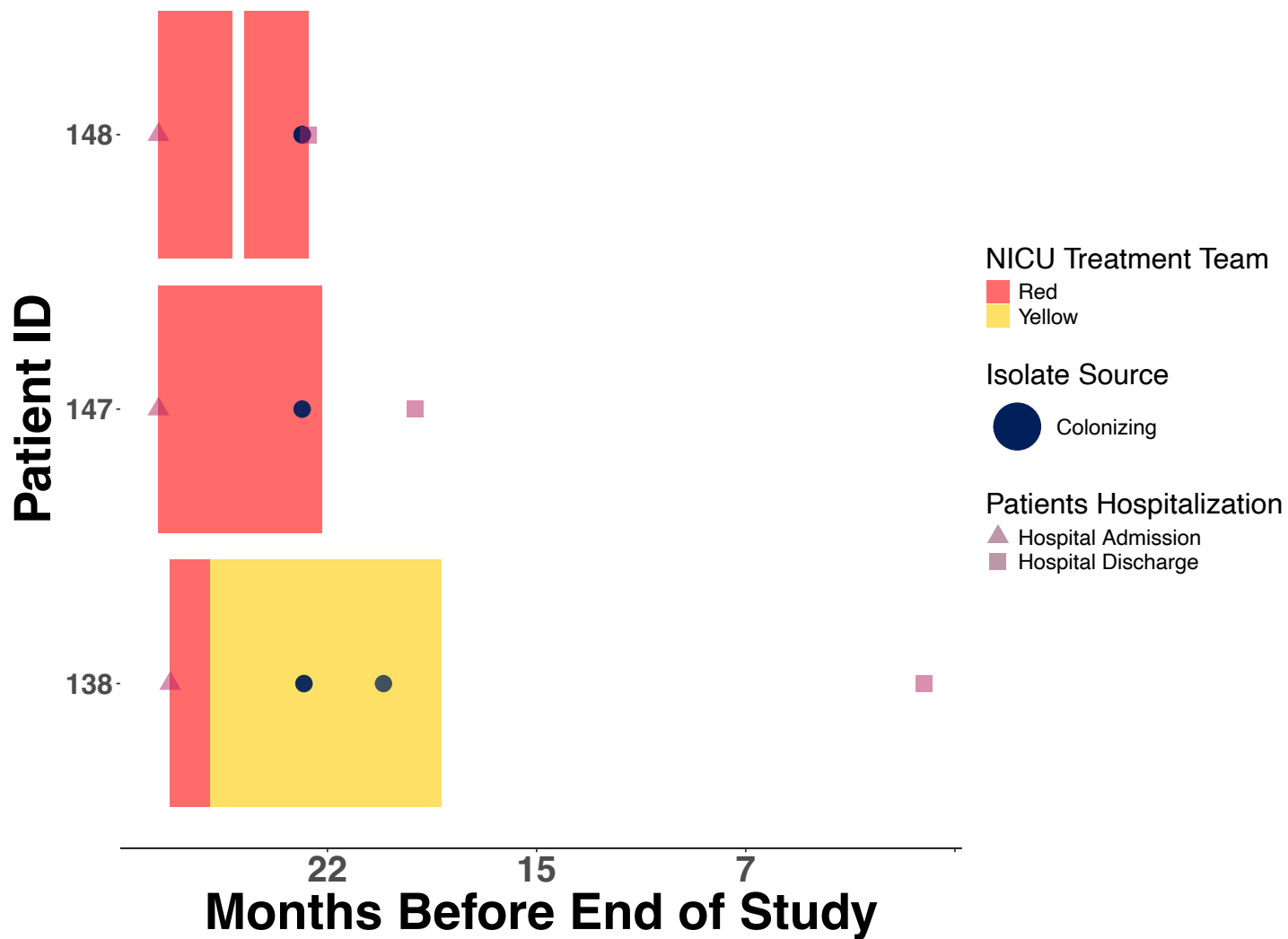
Cluster26



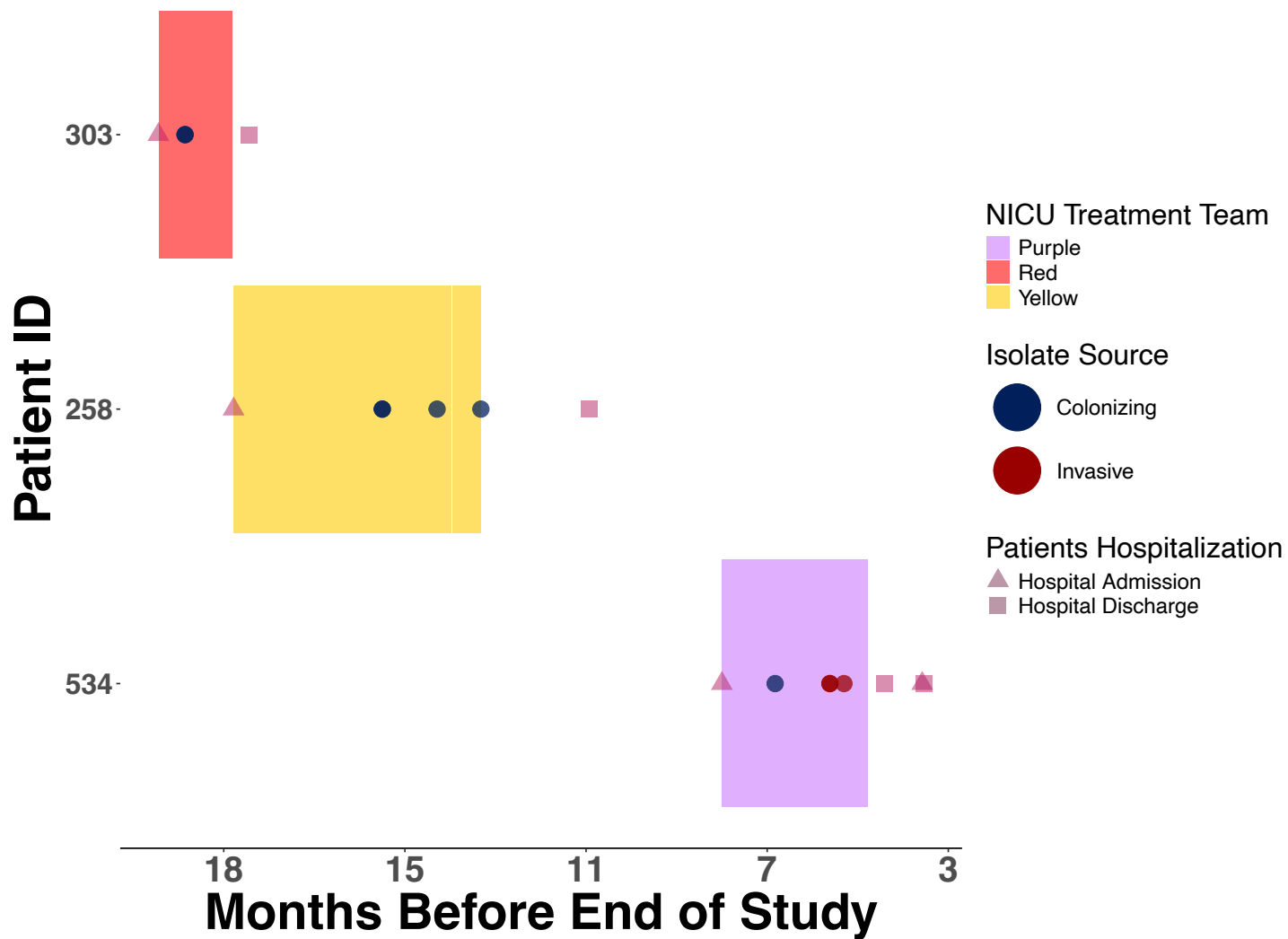
Cluster27



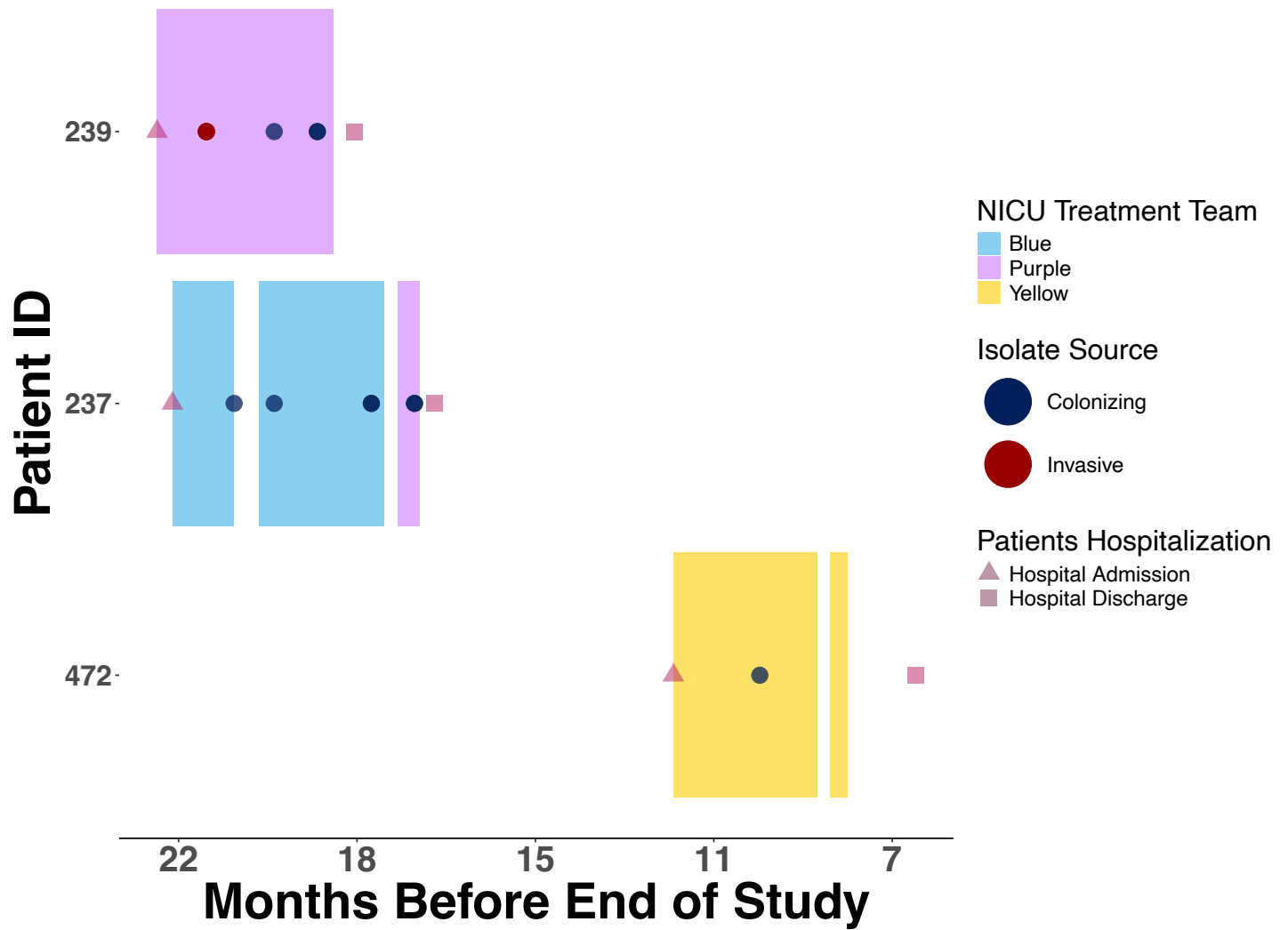
Cluster28



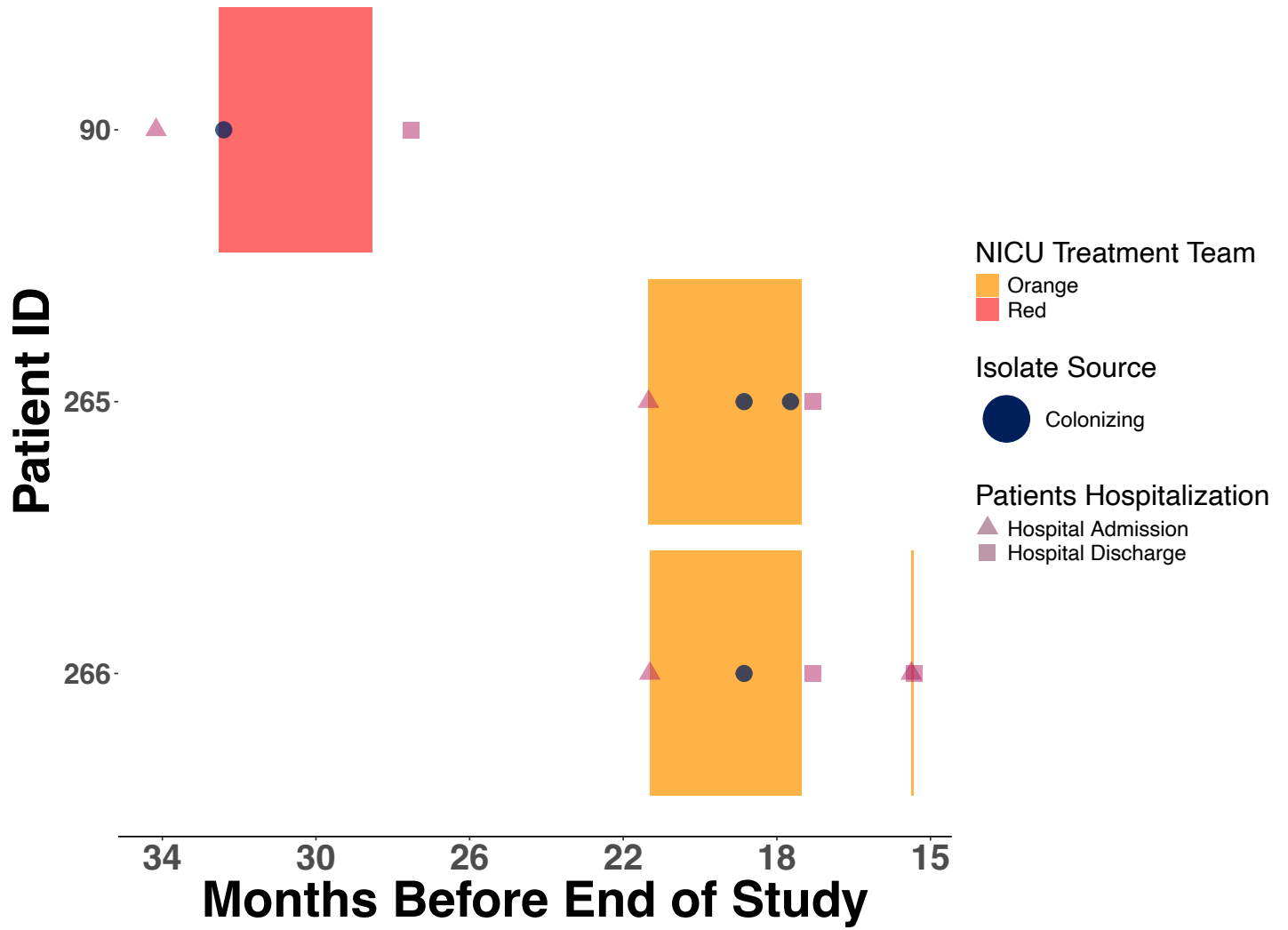
Cluster29



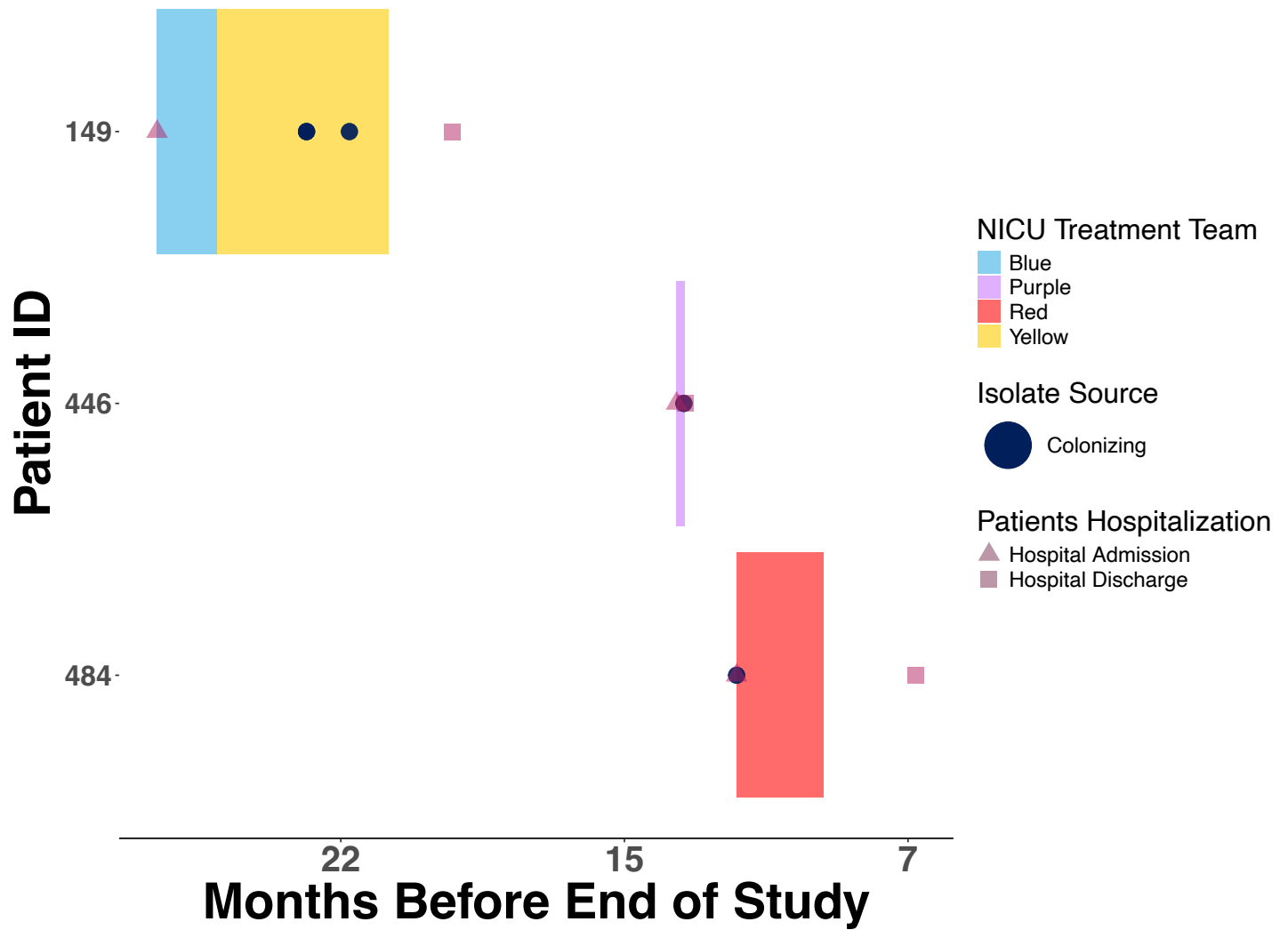
Cluster30



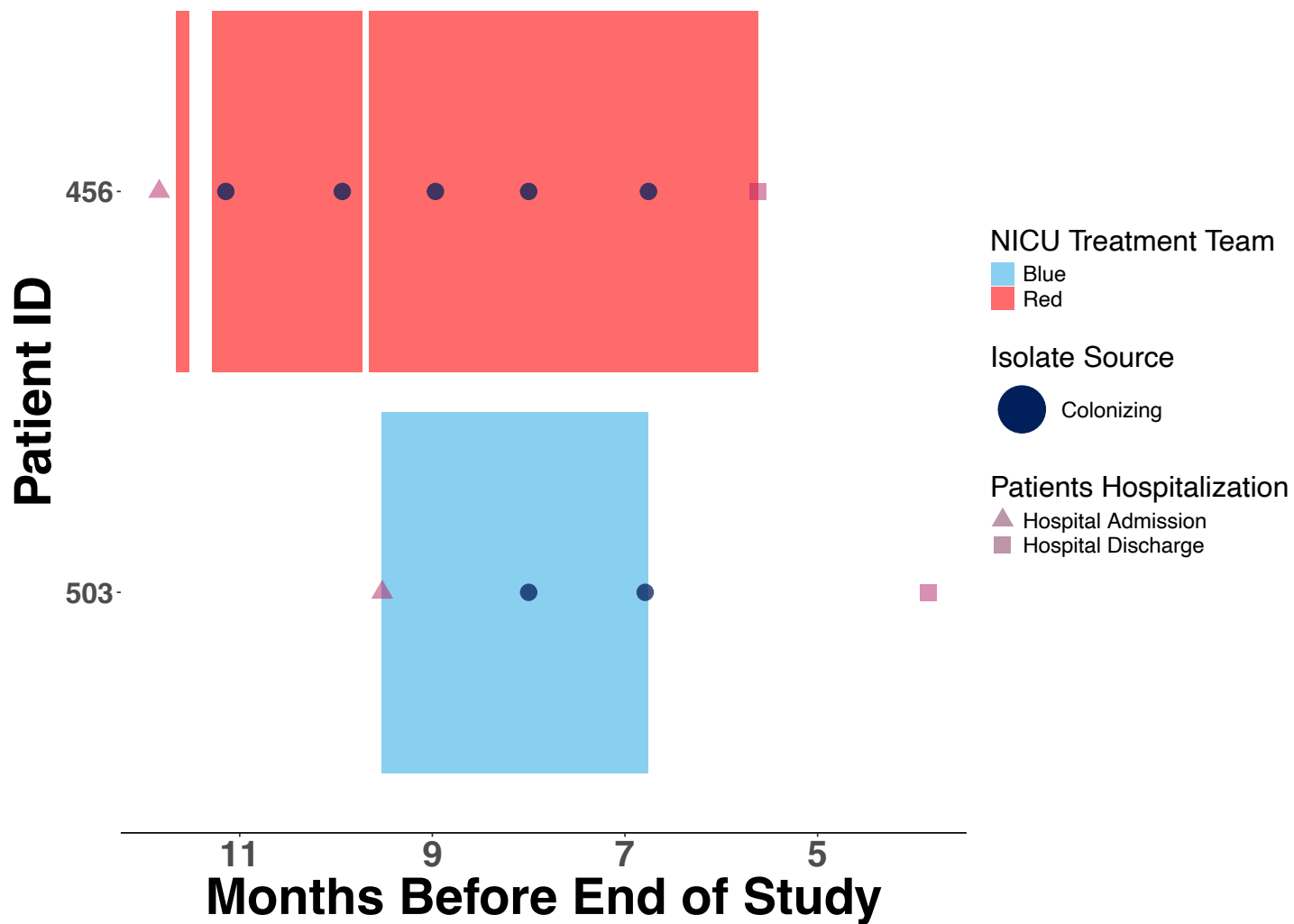
Cluster31



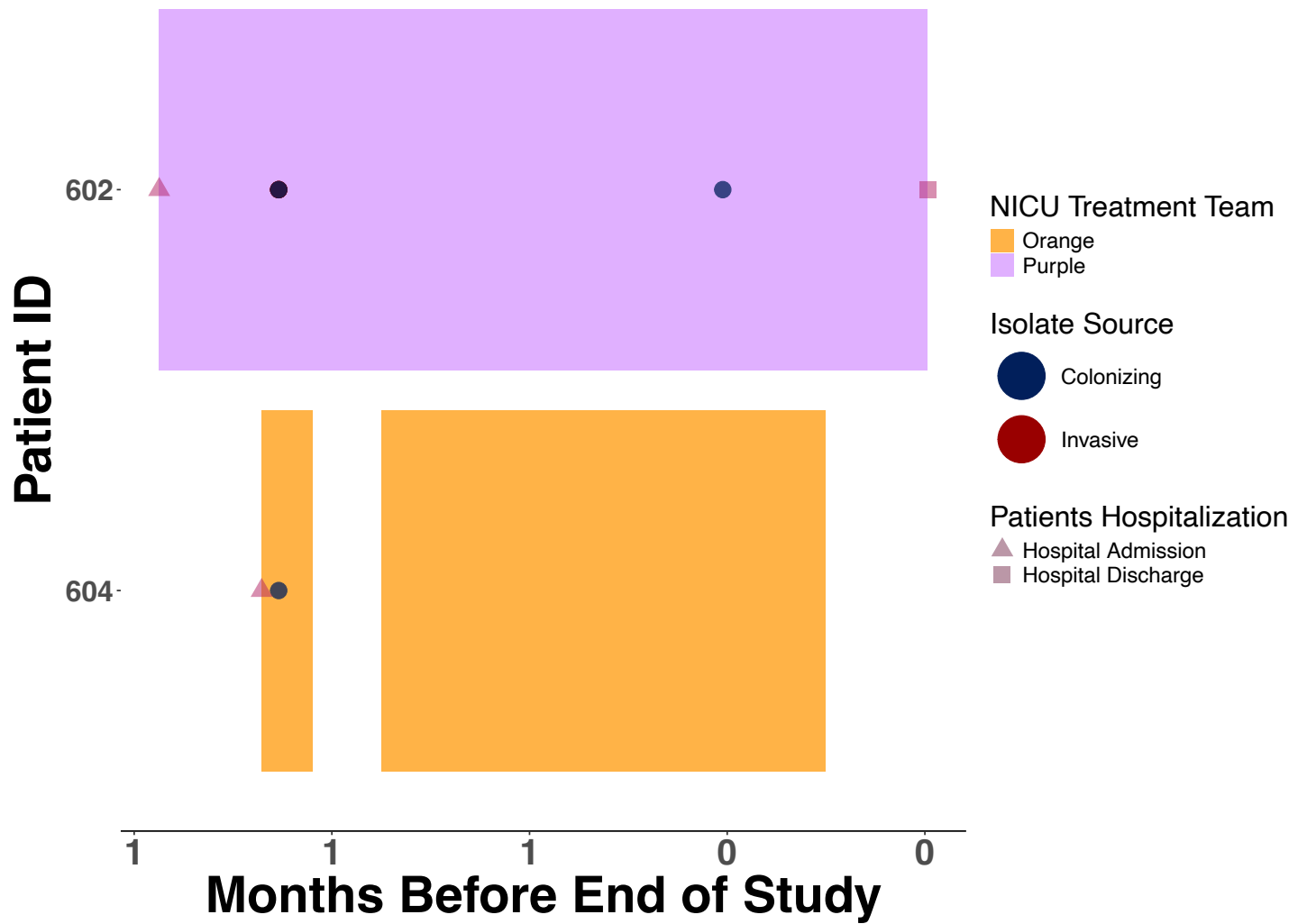
Cluster32



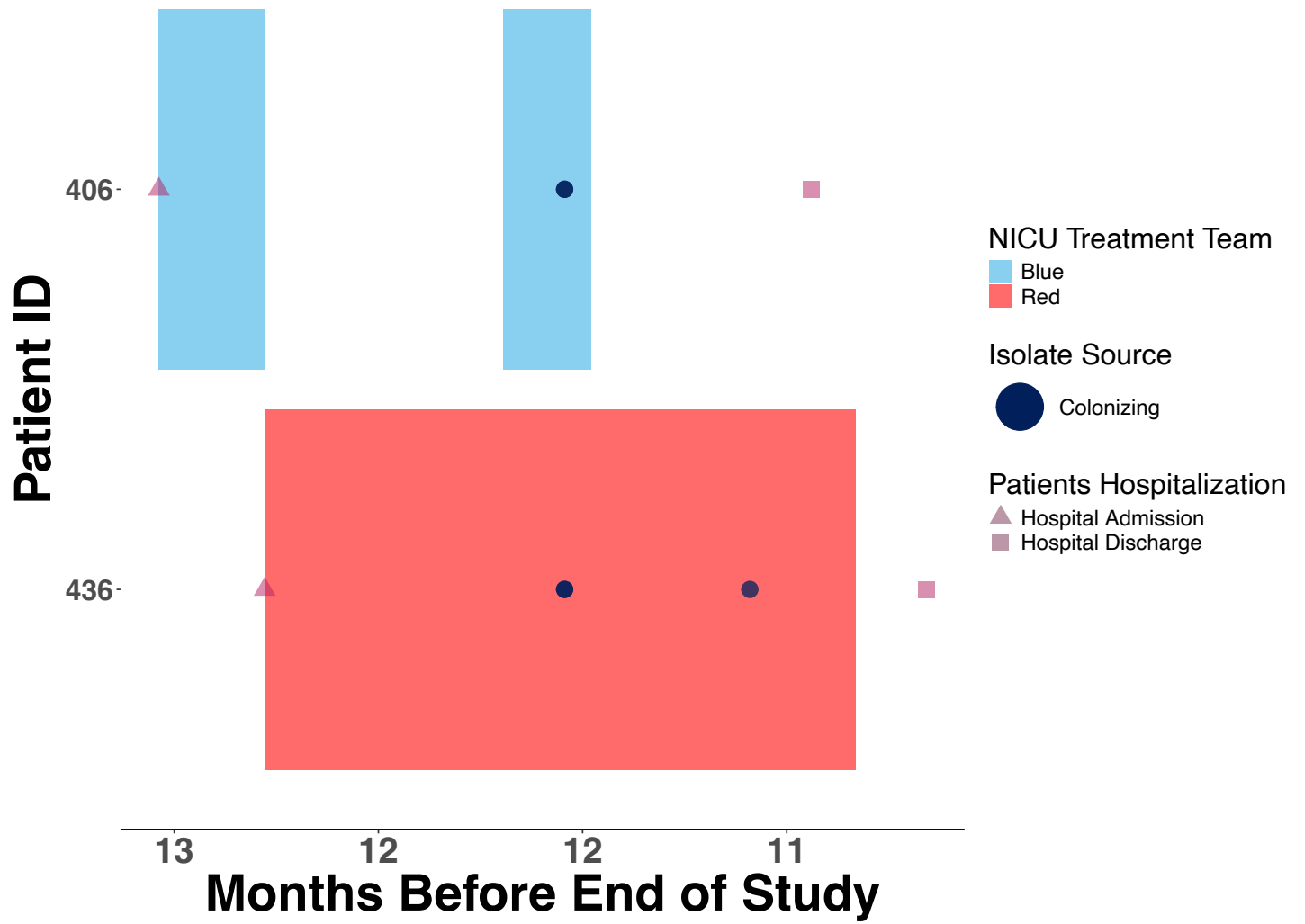
Cluster33



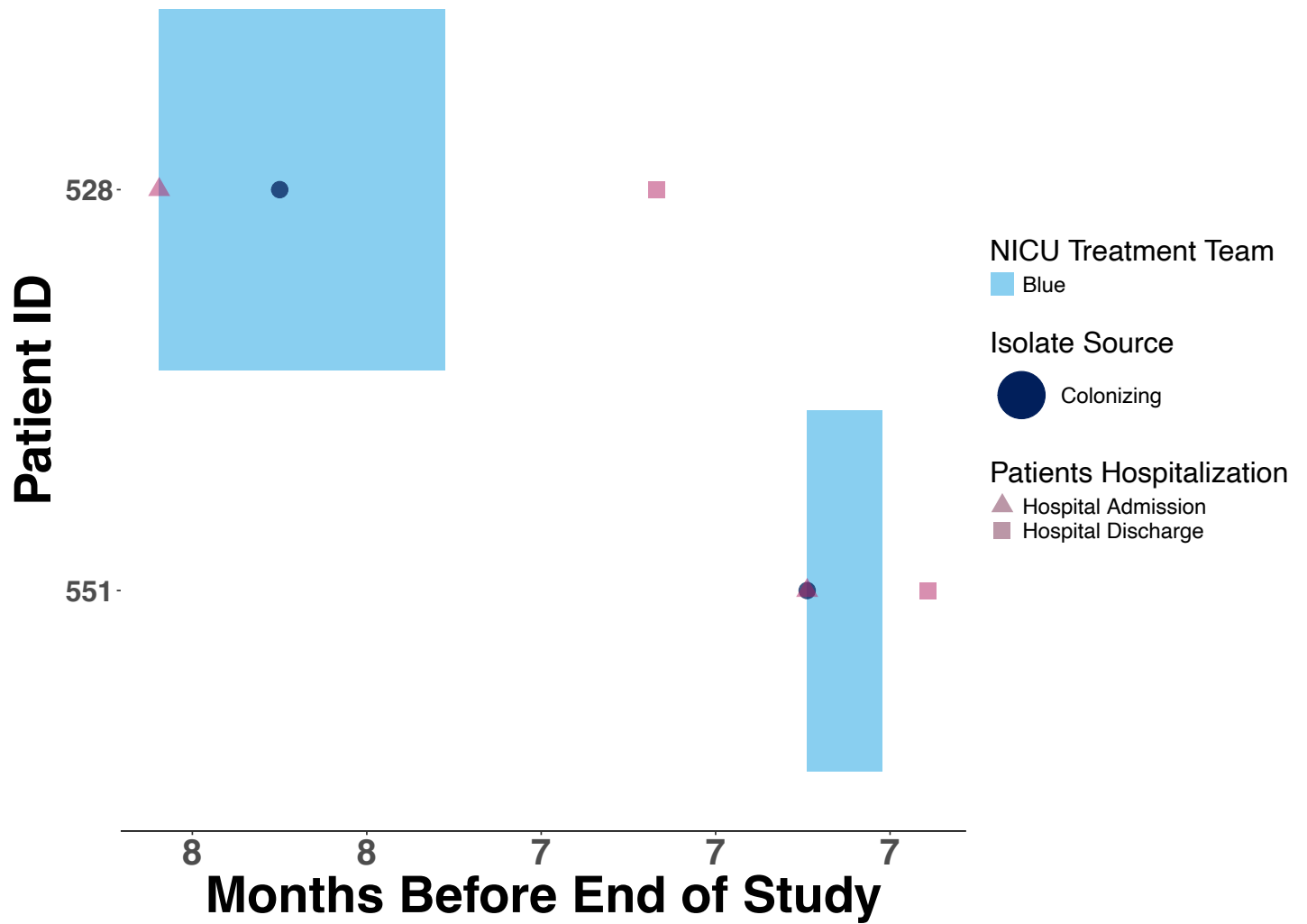
Cluster34



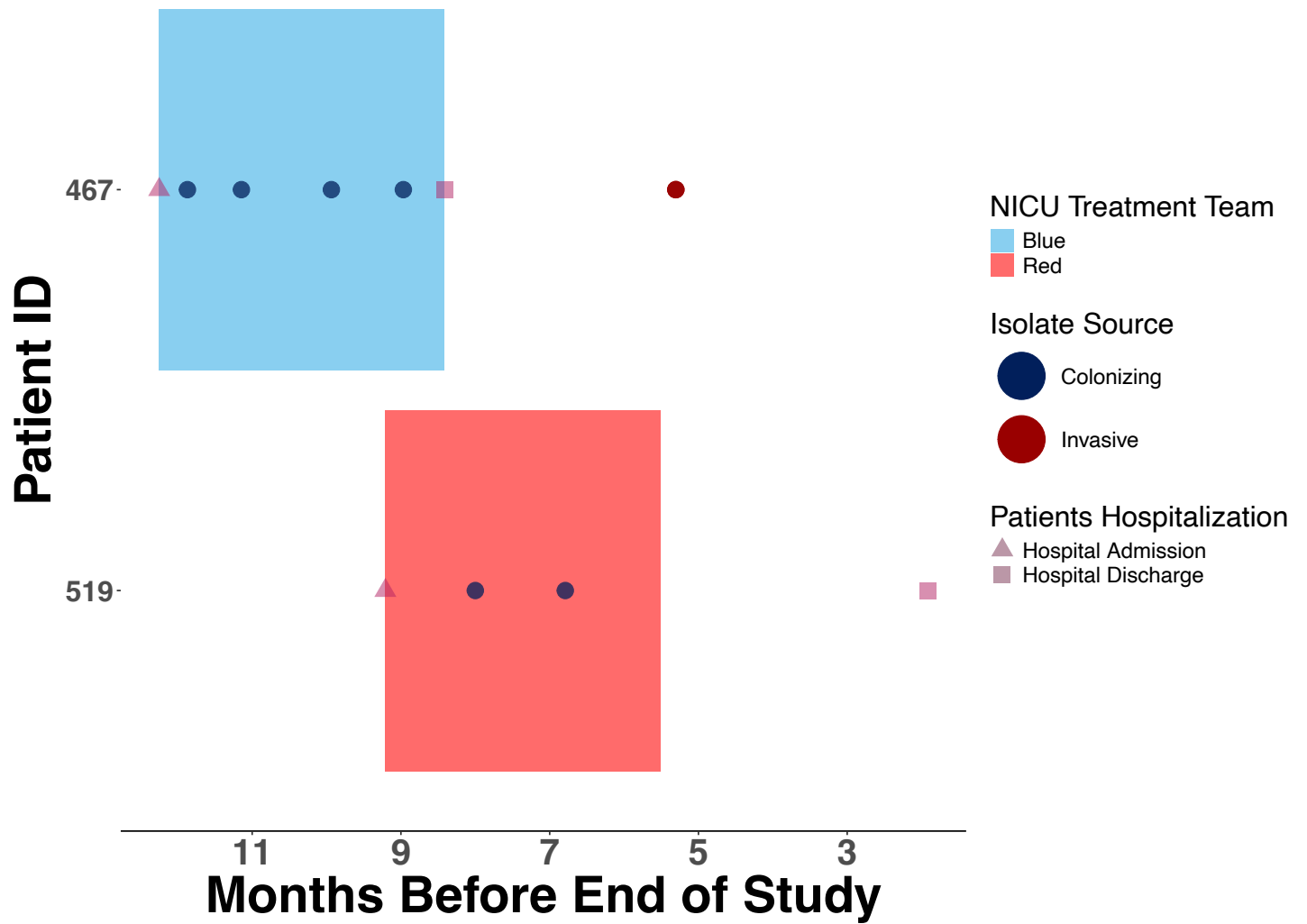
Cluster35



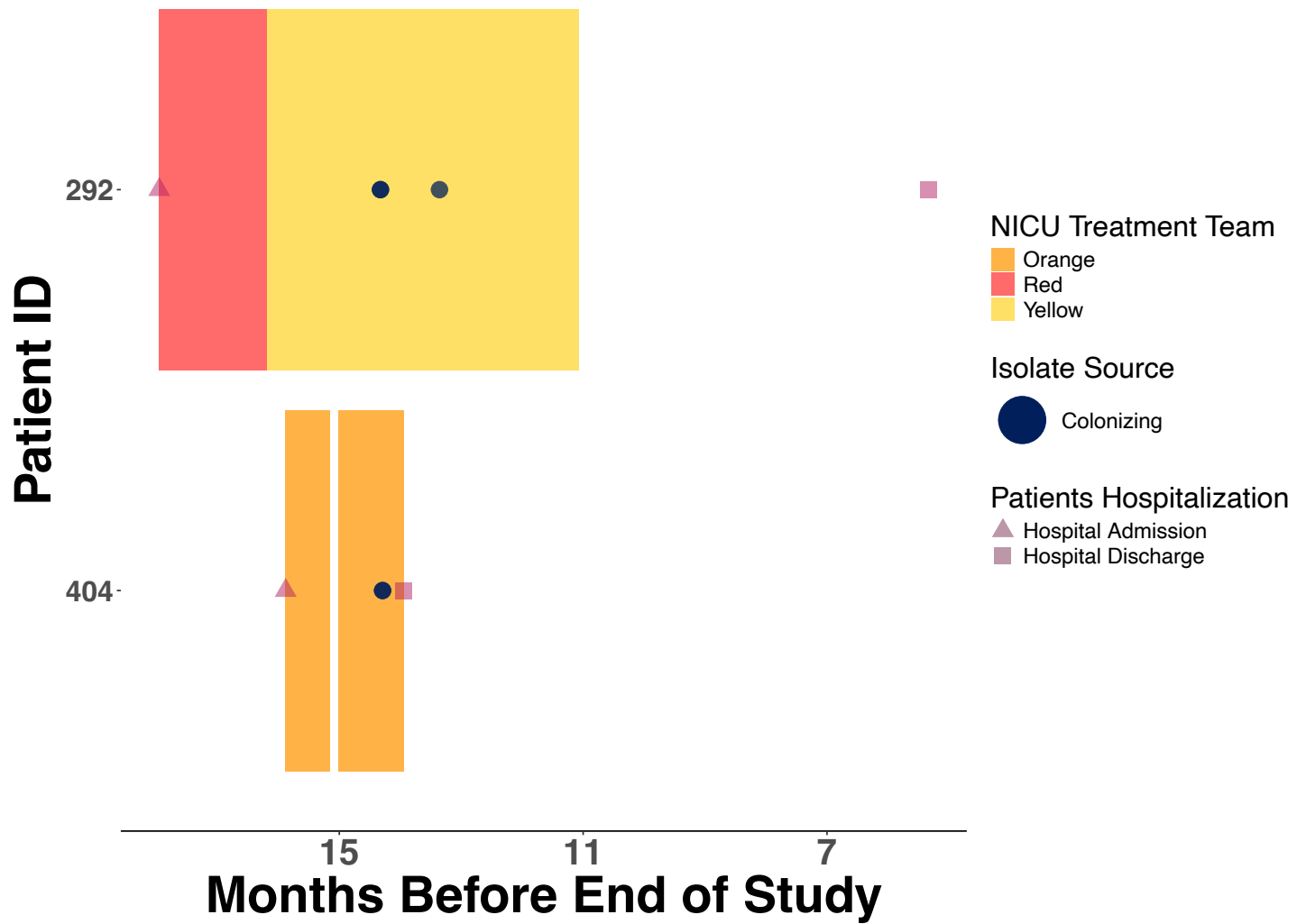
Cluster36



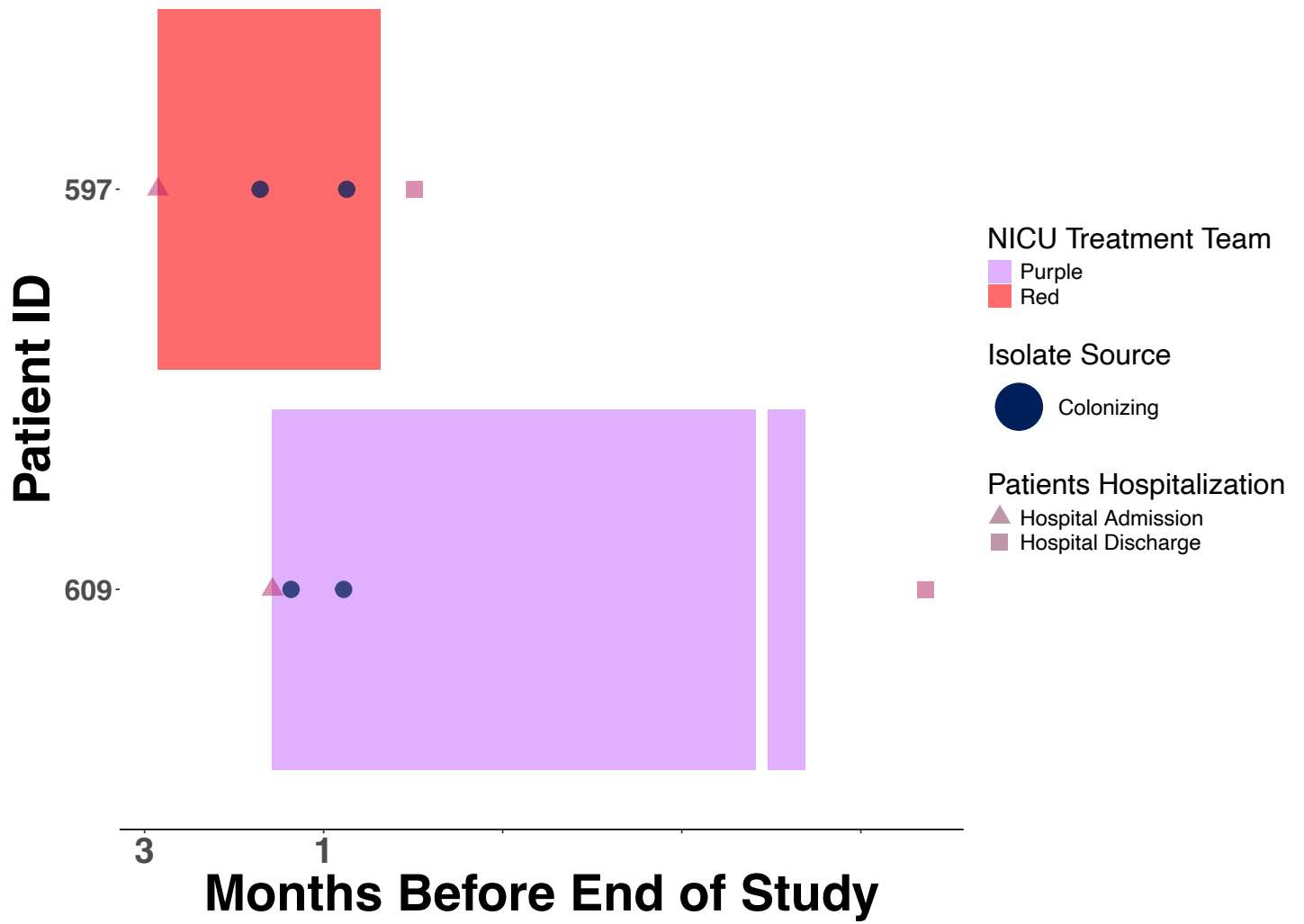
Cluster37



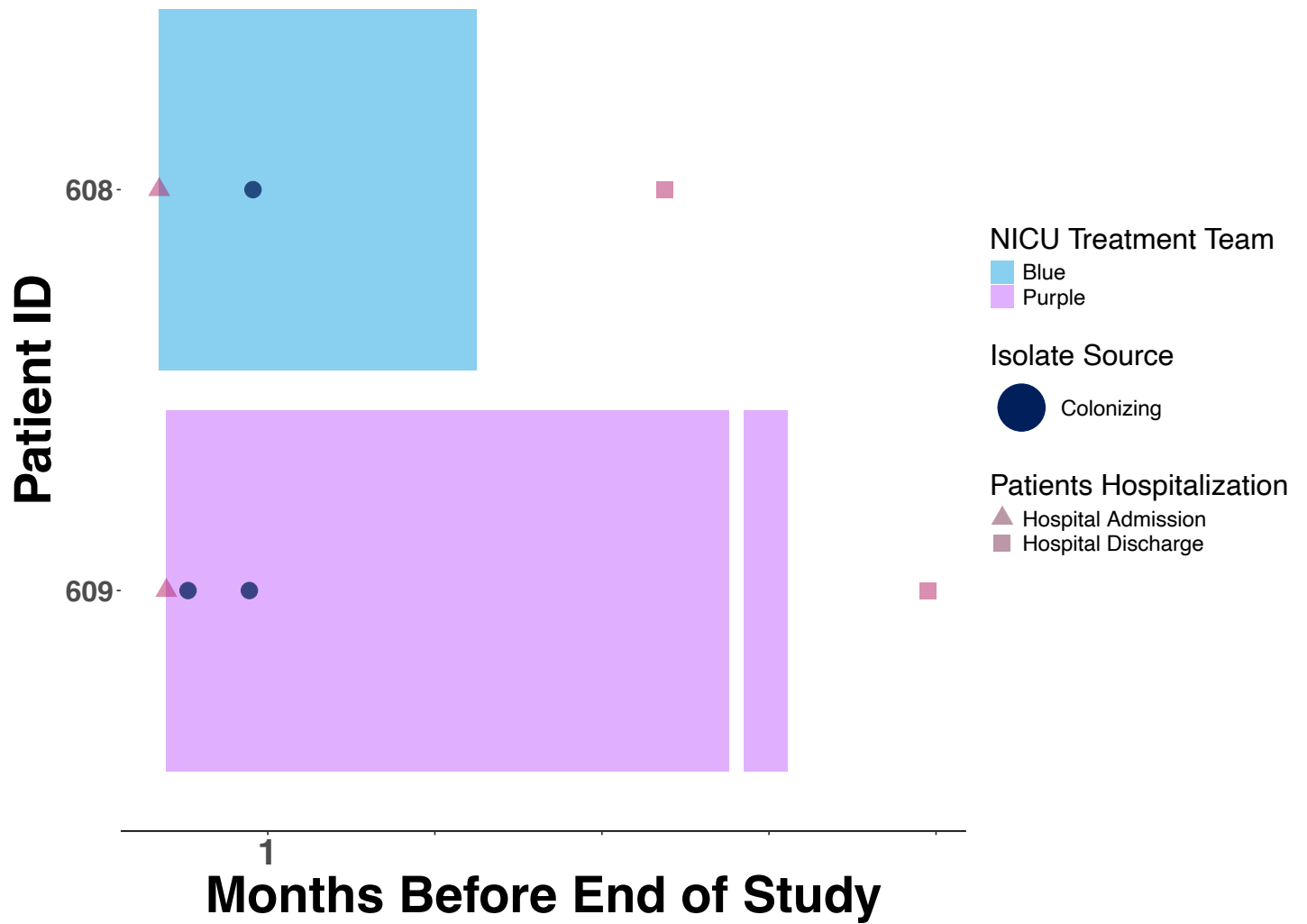
Cluster38



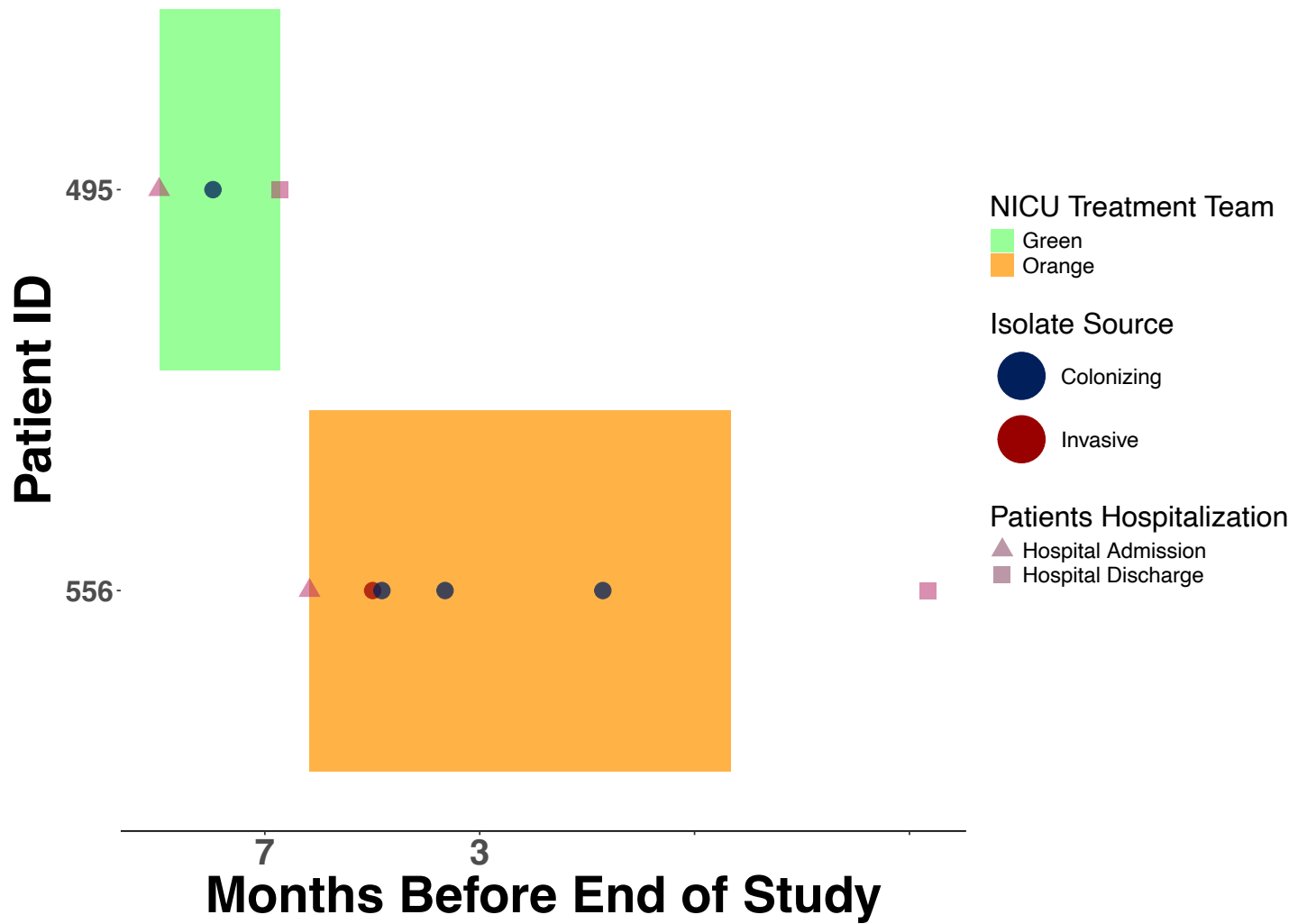
Cluster39



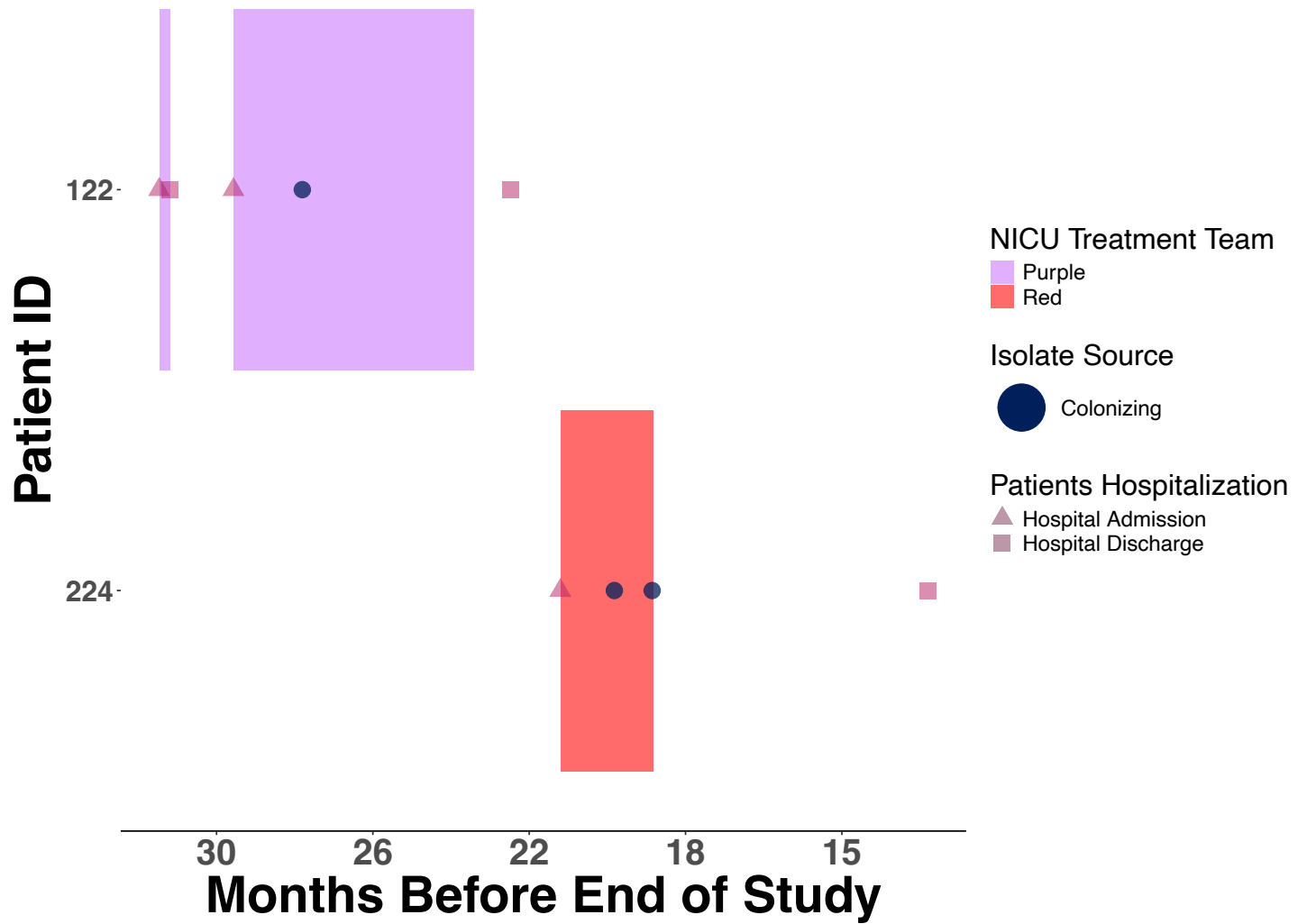
Cluster40



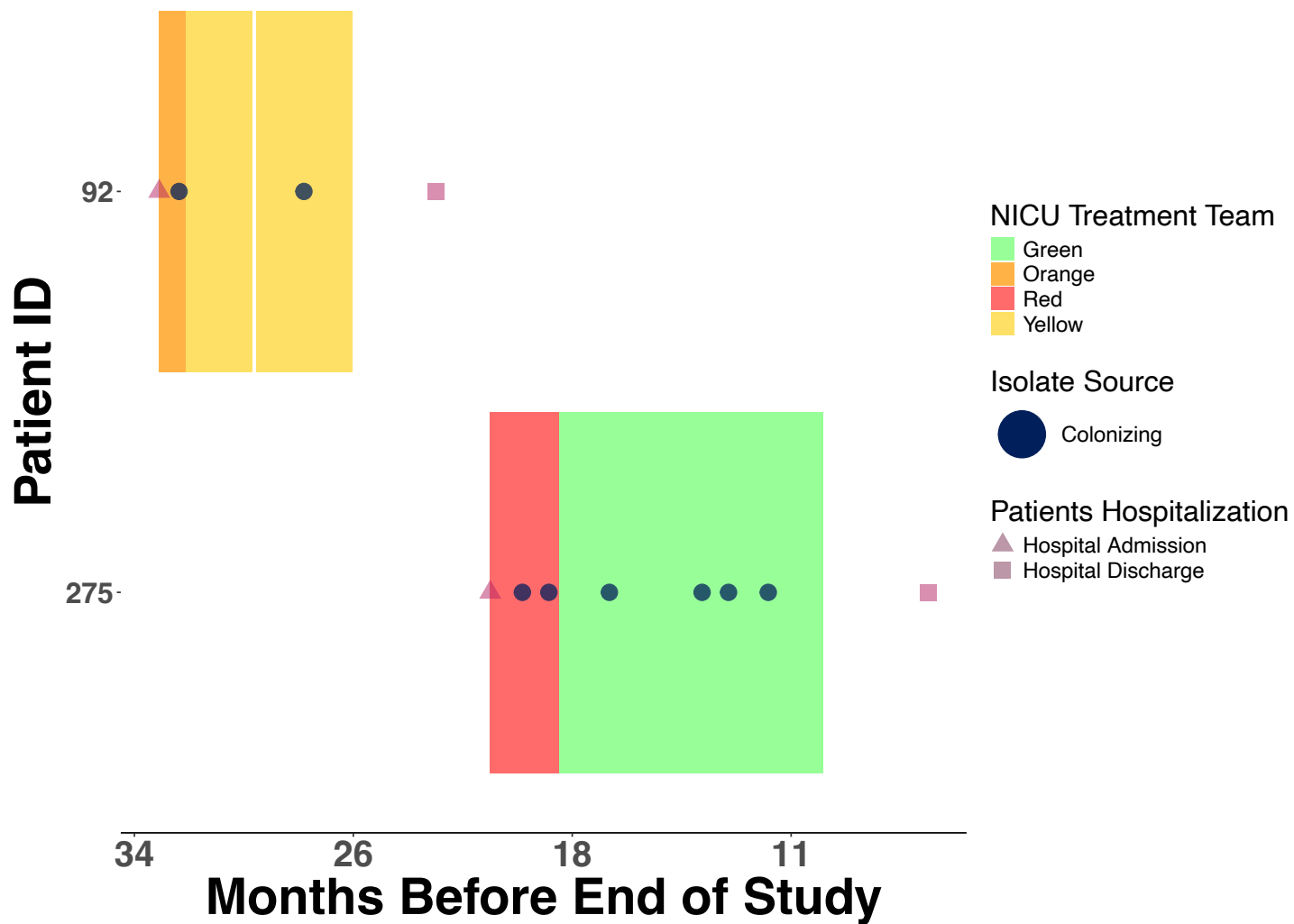
Cluster41



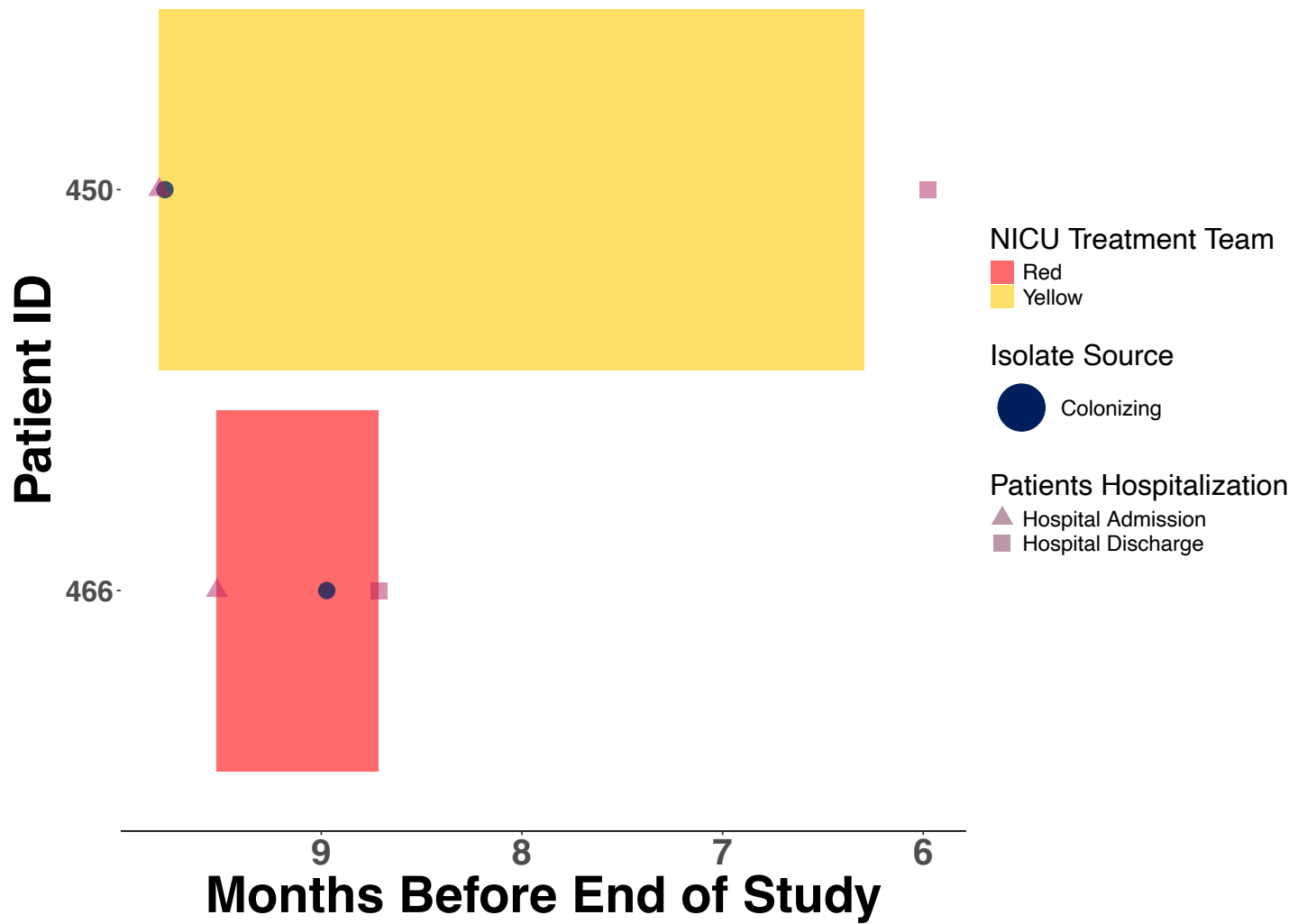
Cluster42



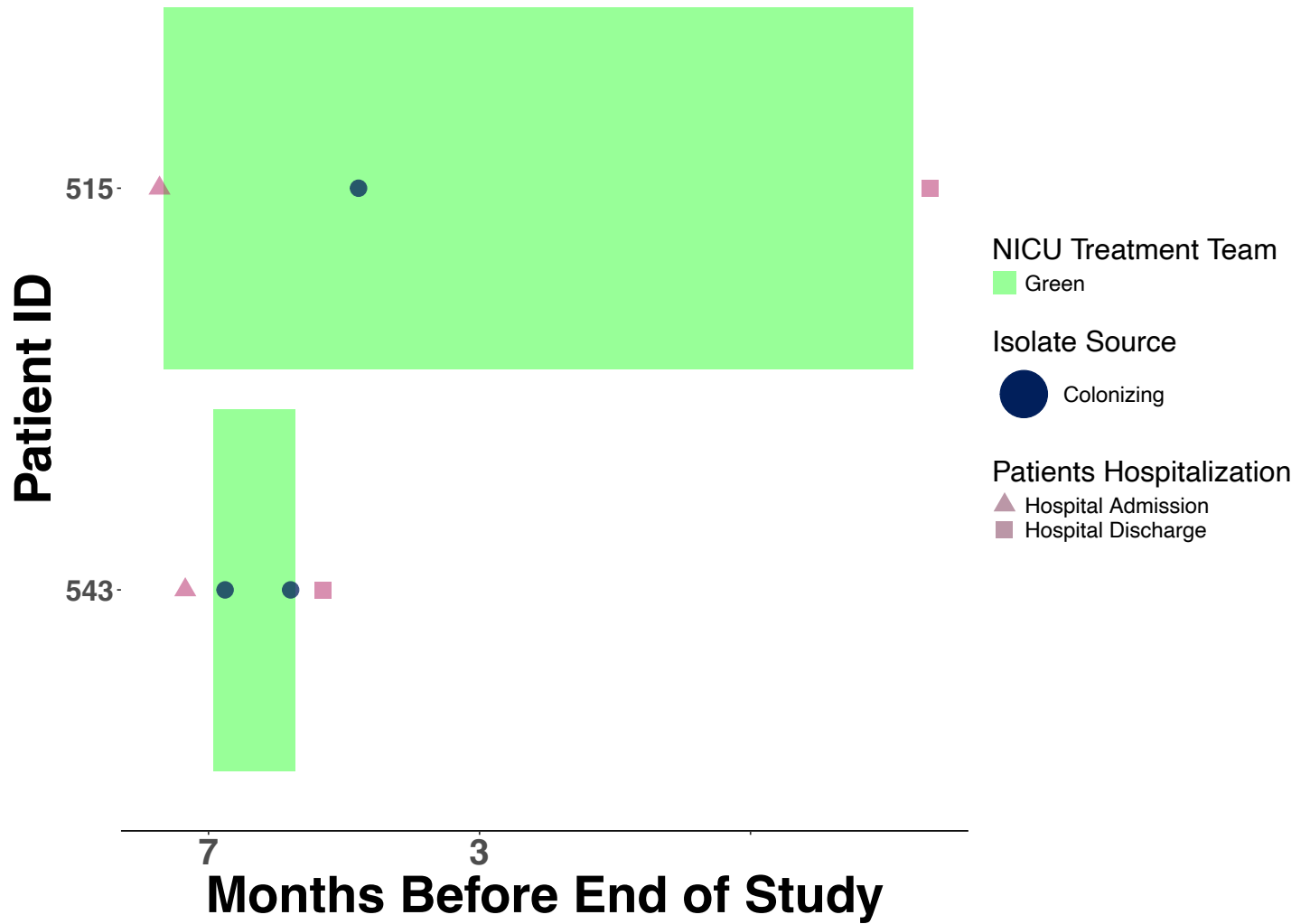
Cluster43



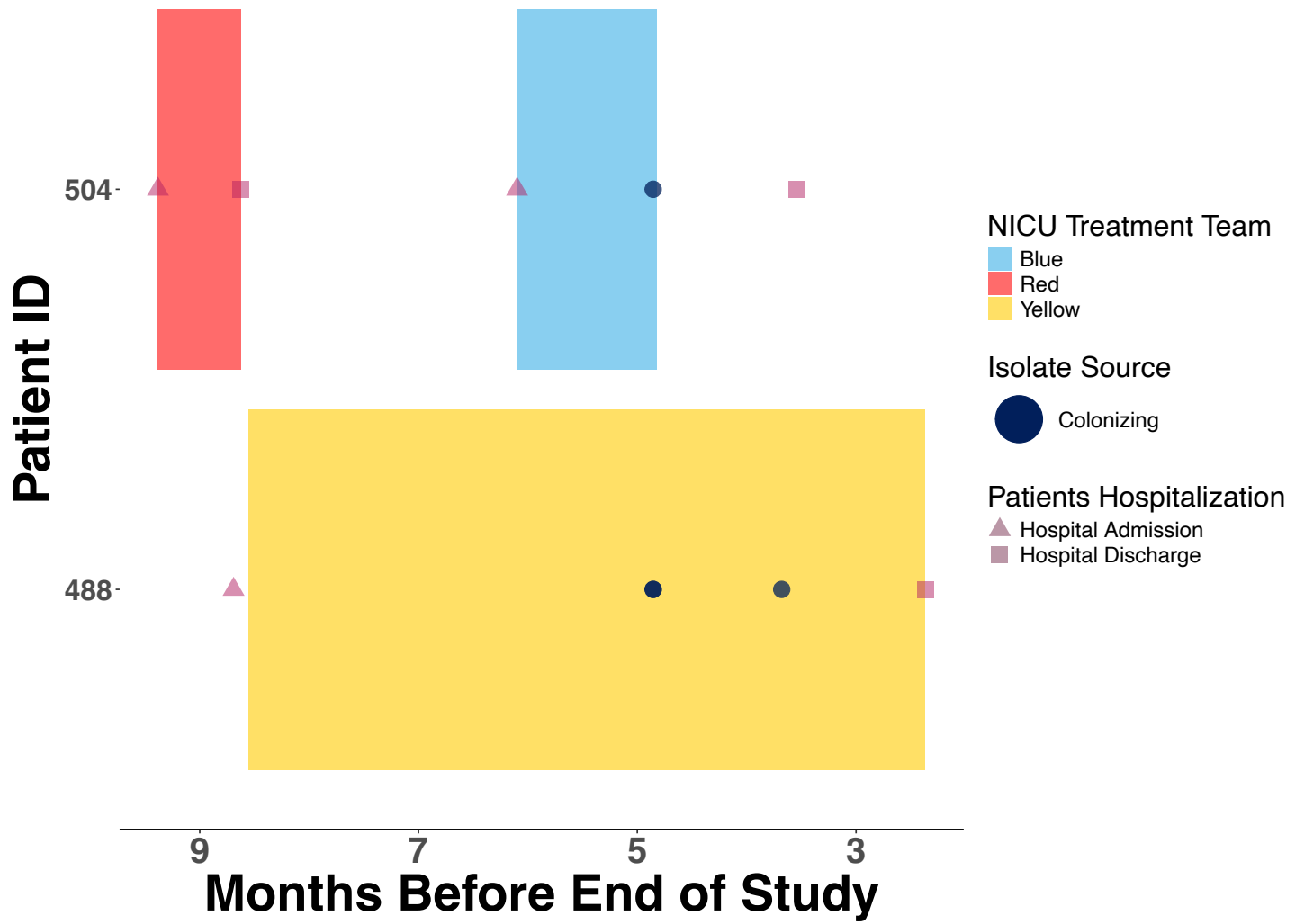
Cluster44



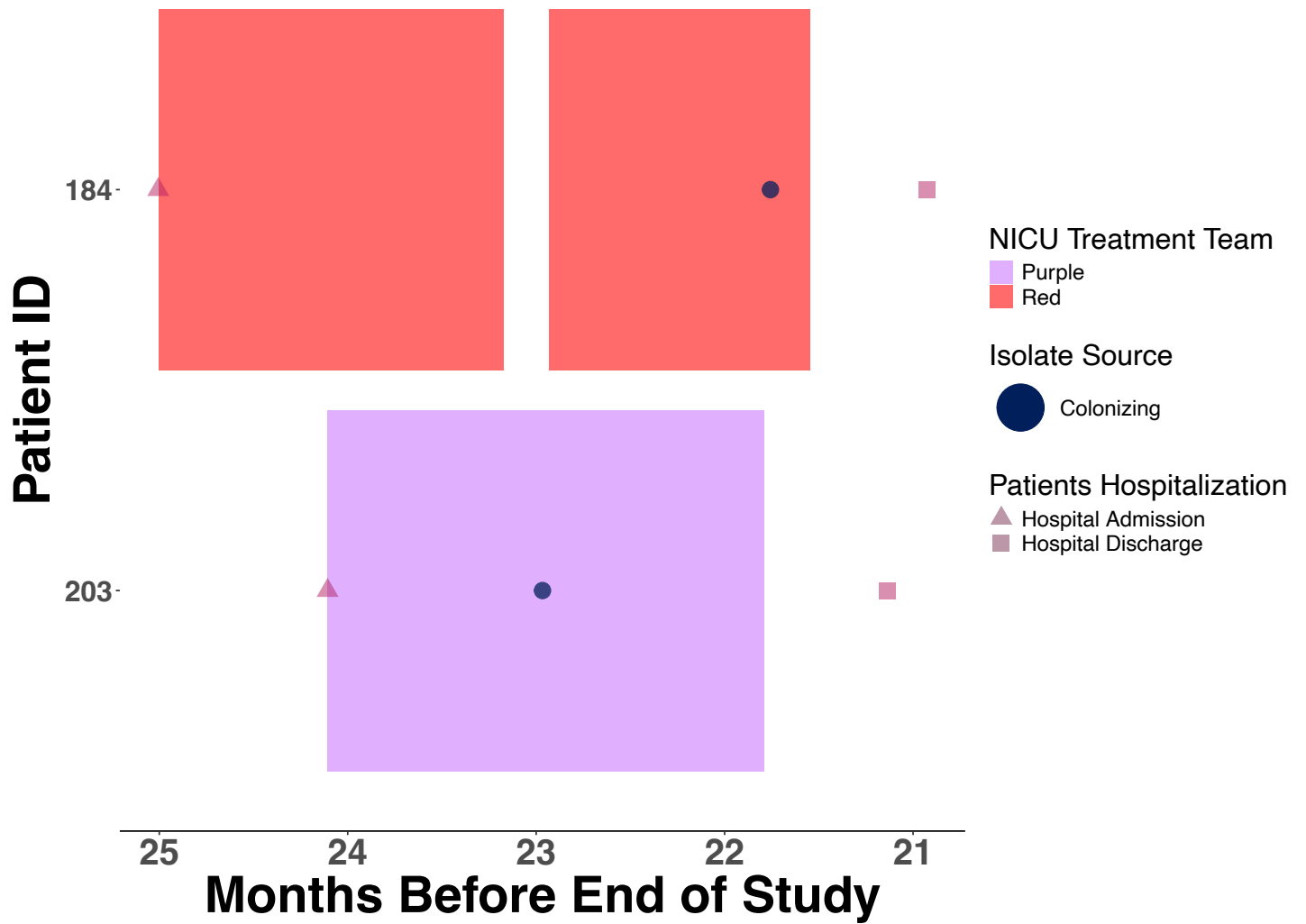
Cluster45



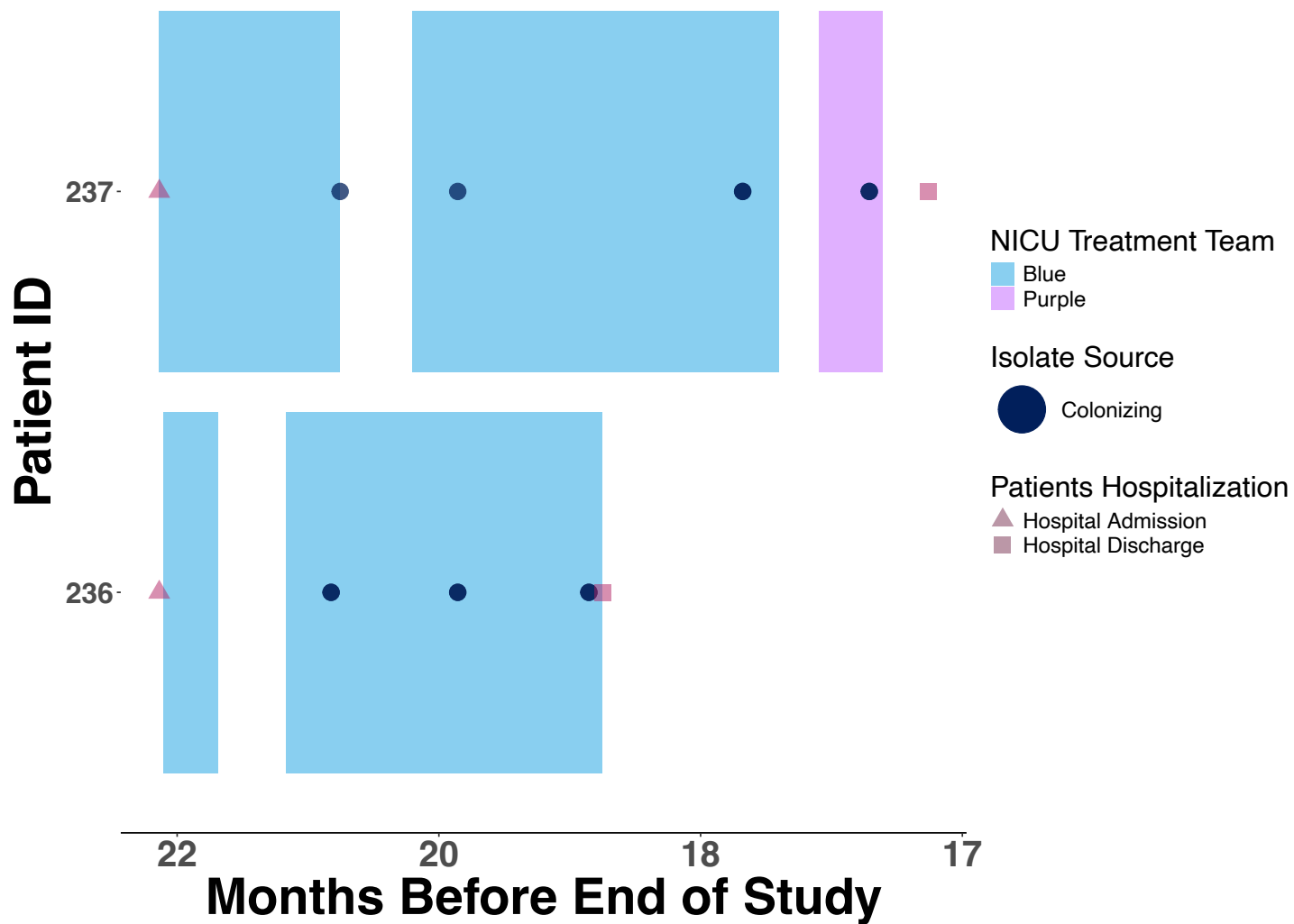
Cluster46



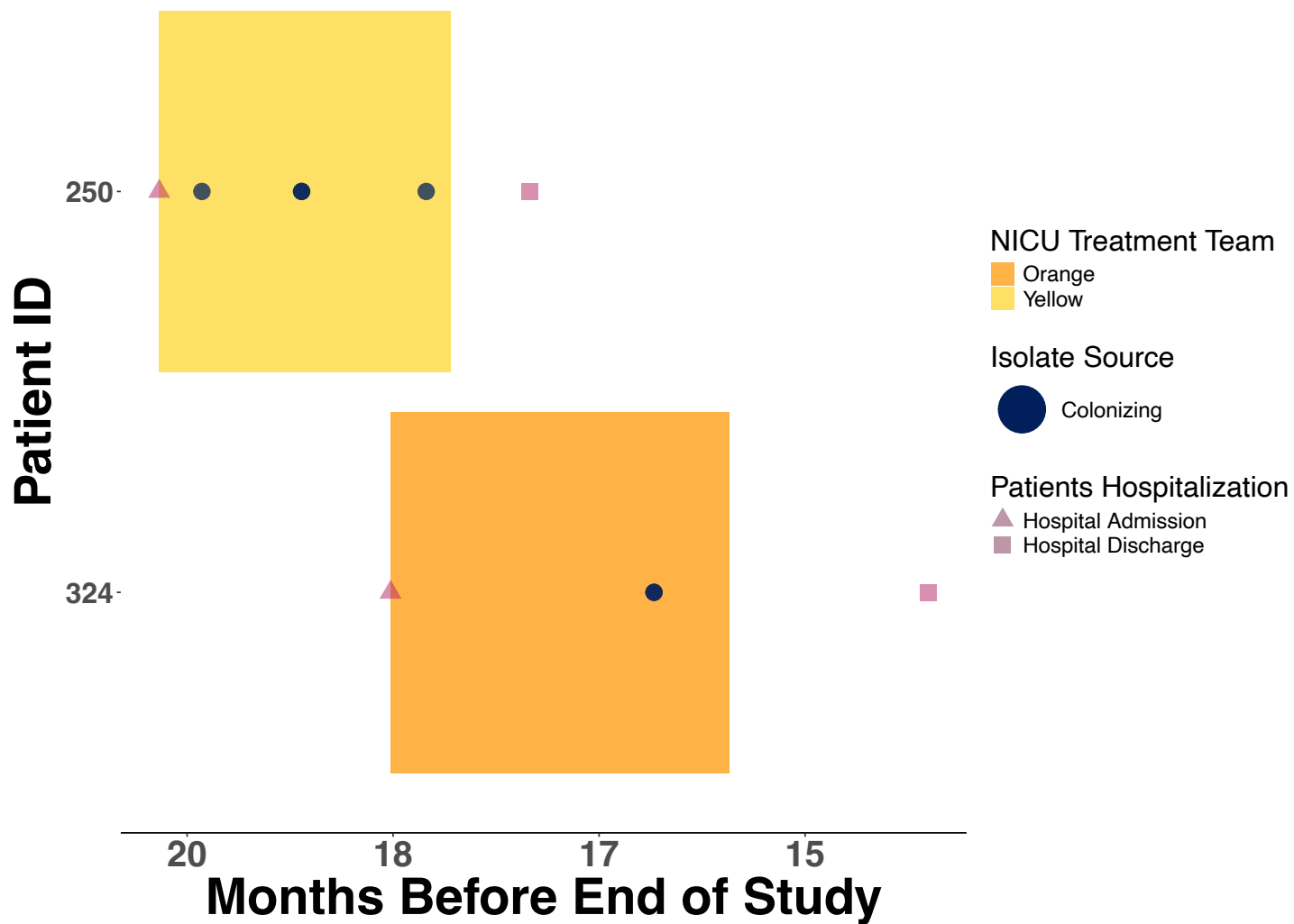
Cluster47



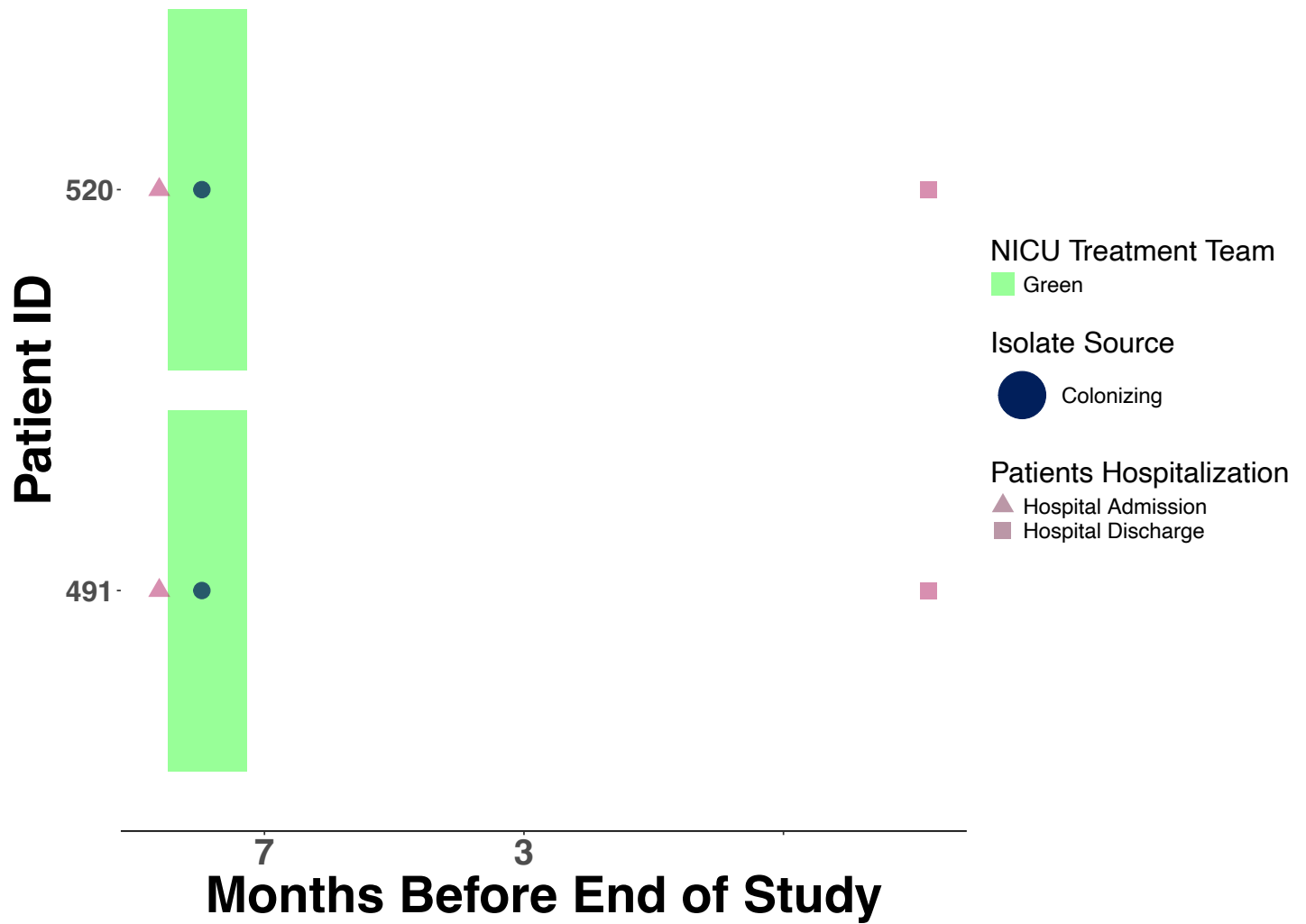
Cluster48



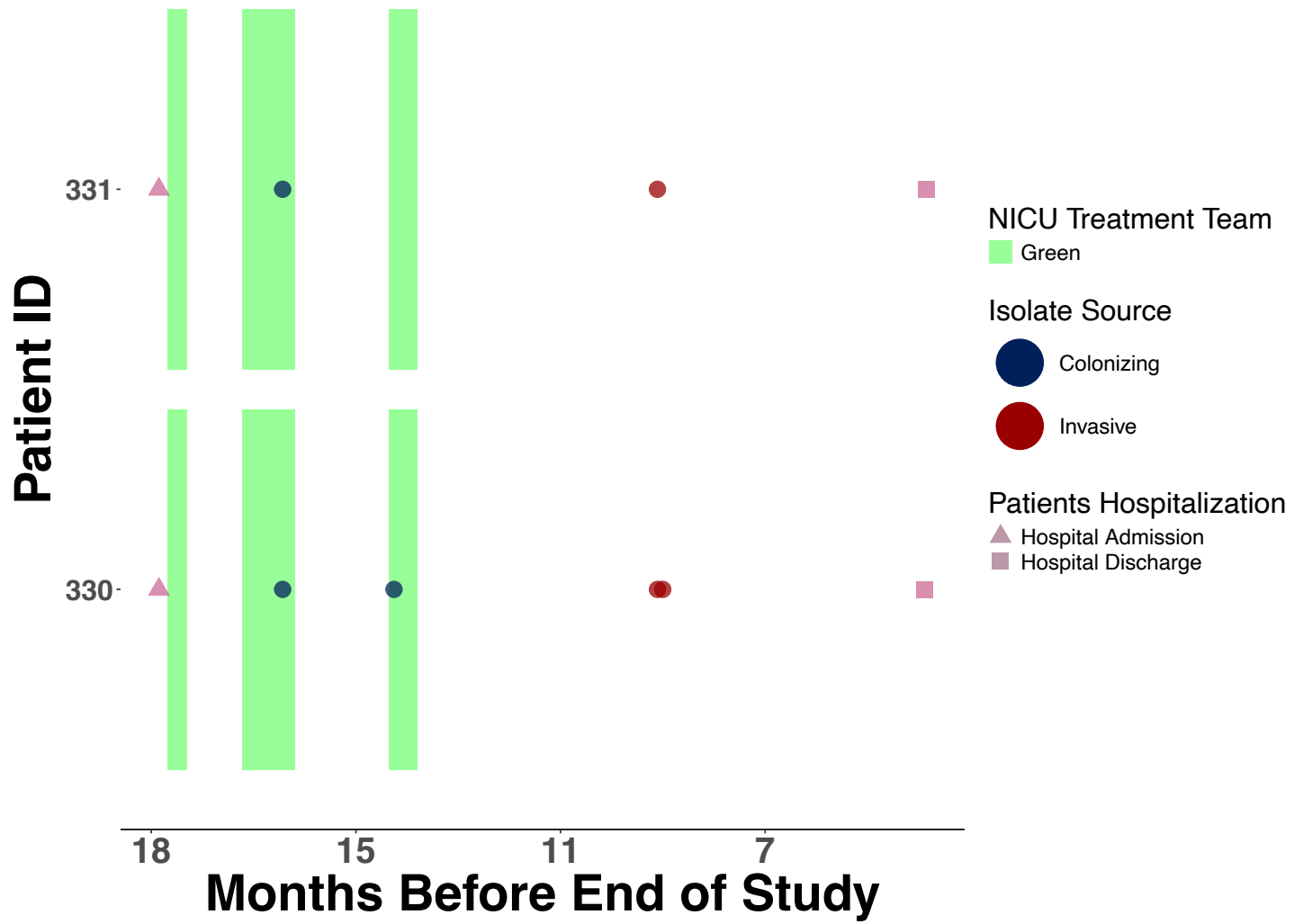
Cluster49



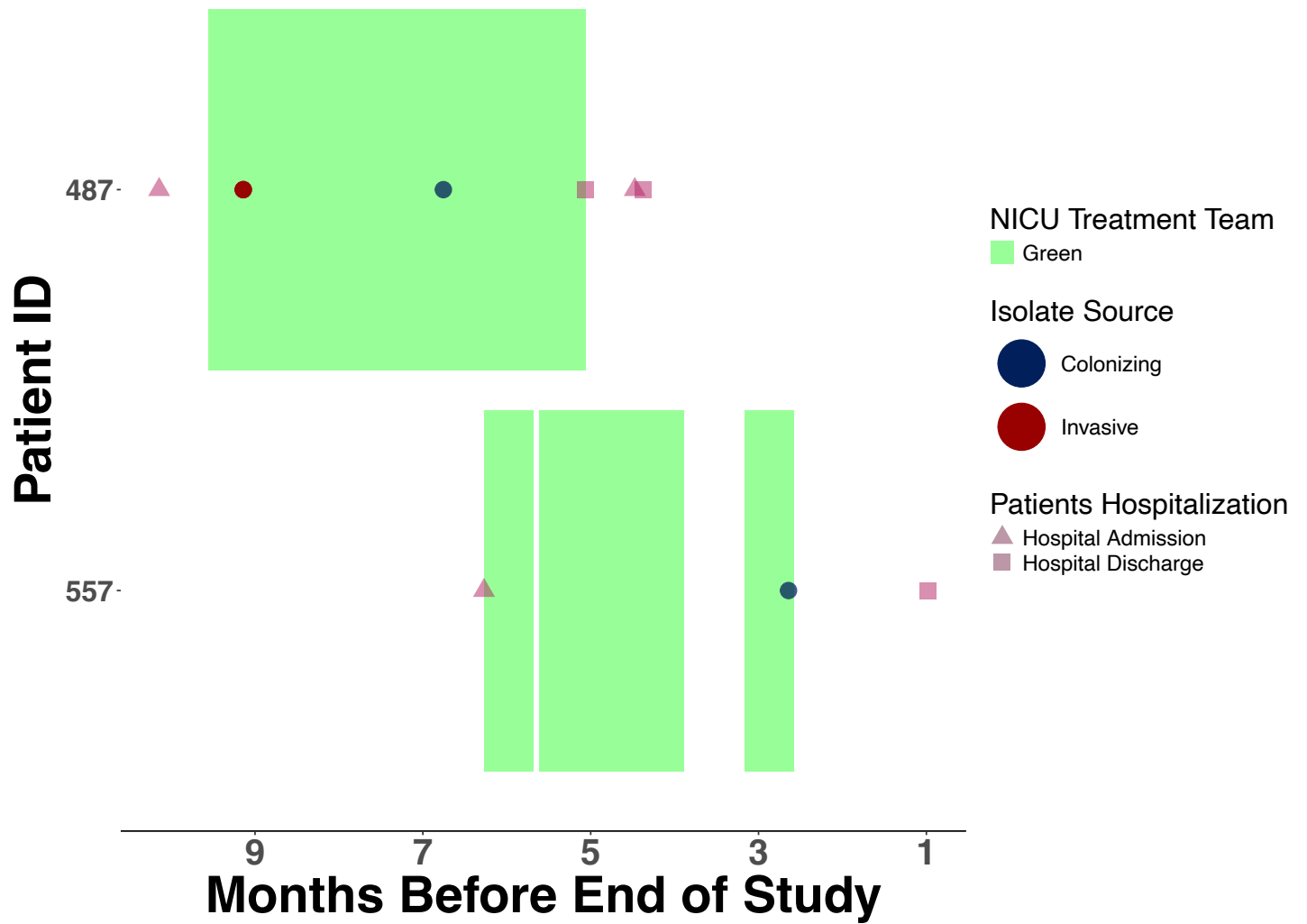
Cluster50



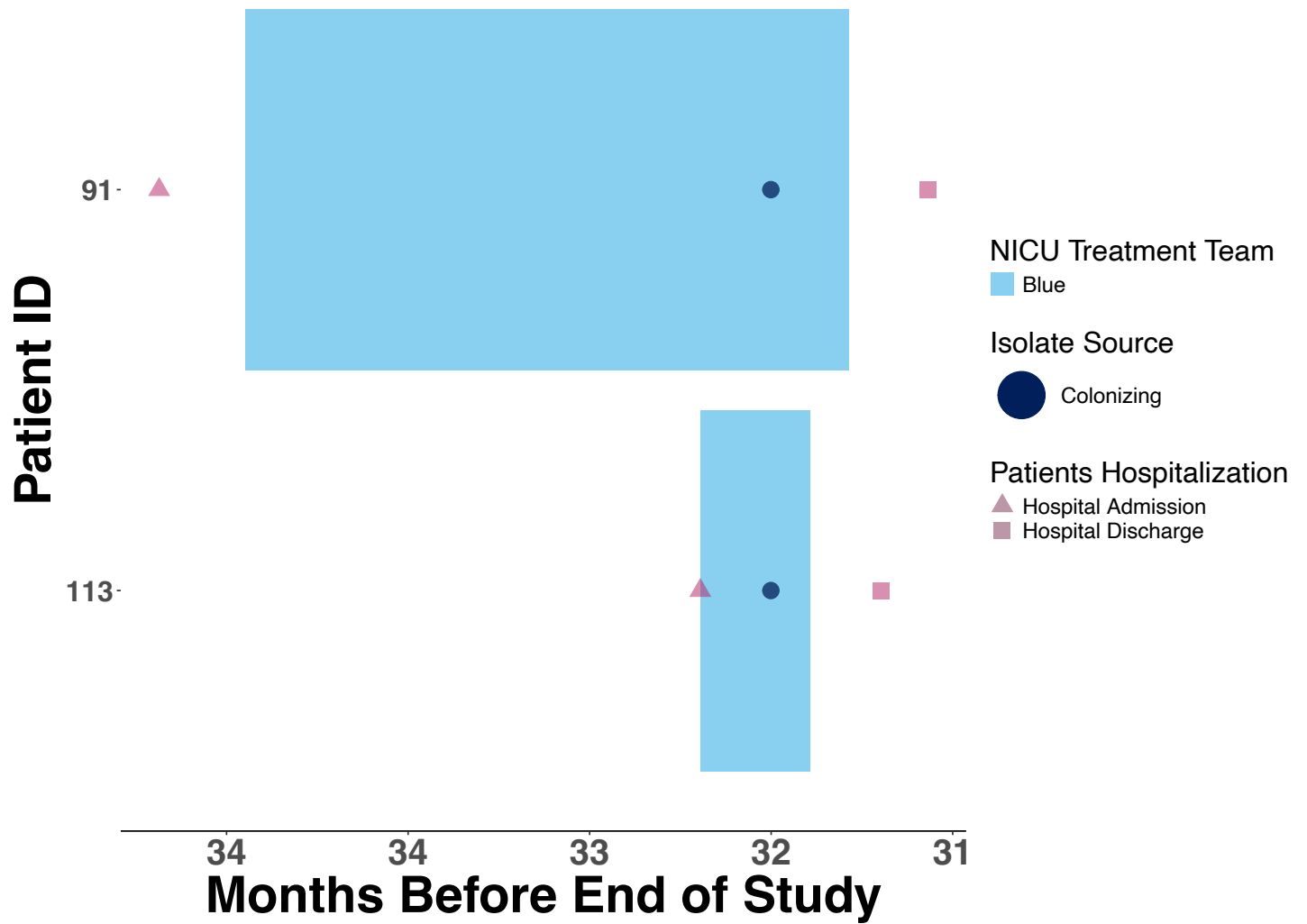
Cluster51



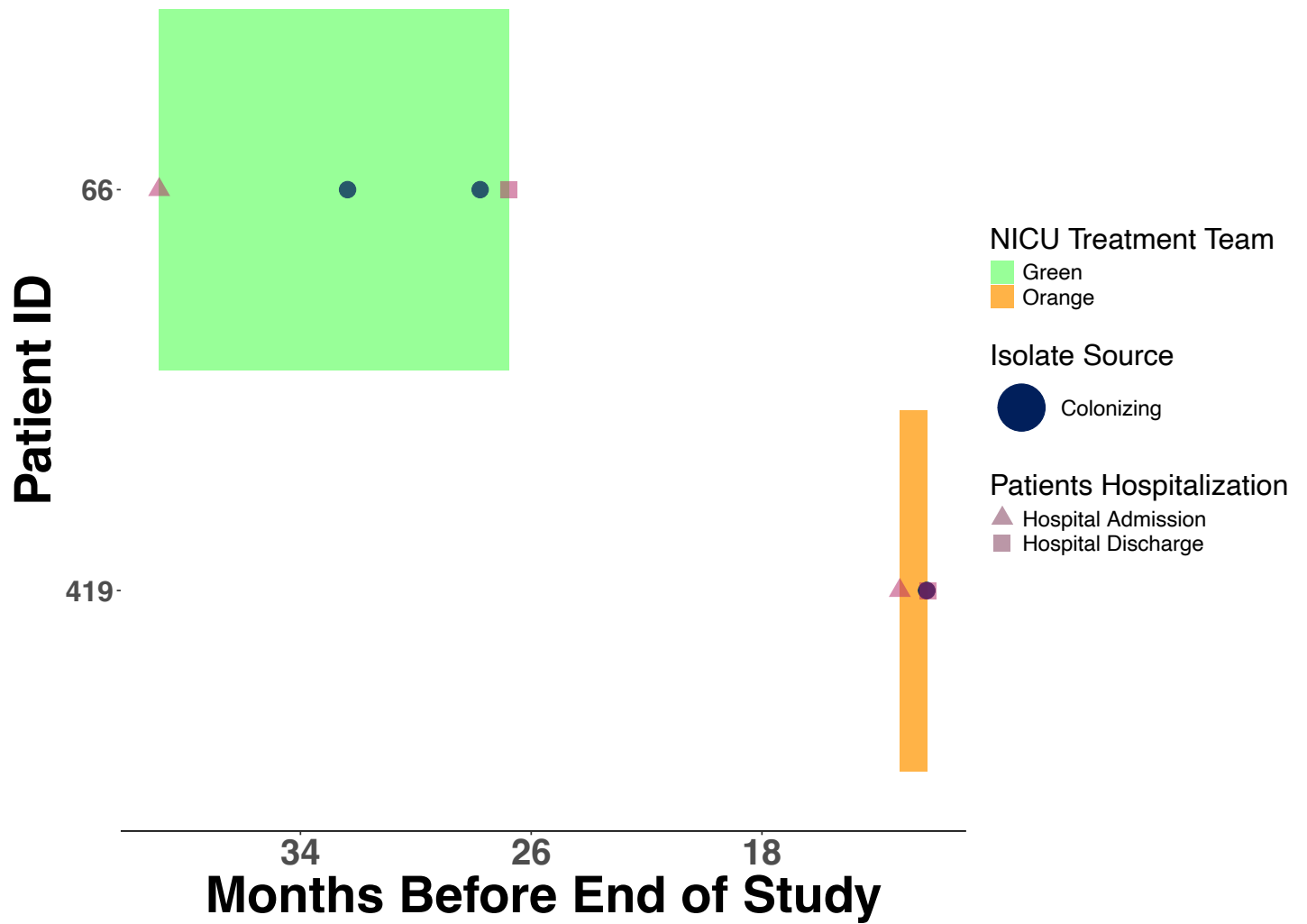
Cluster52



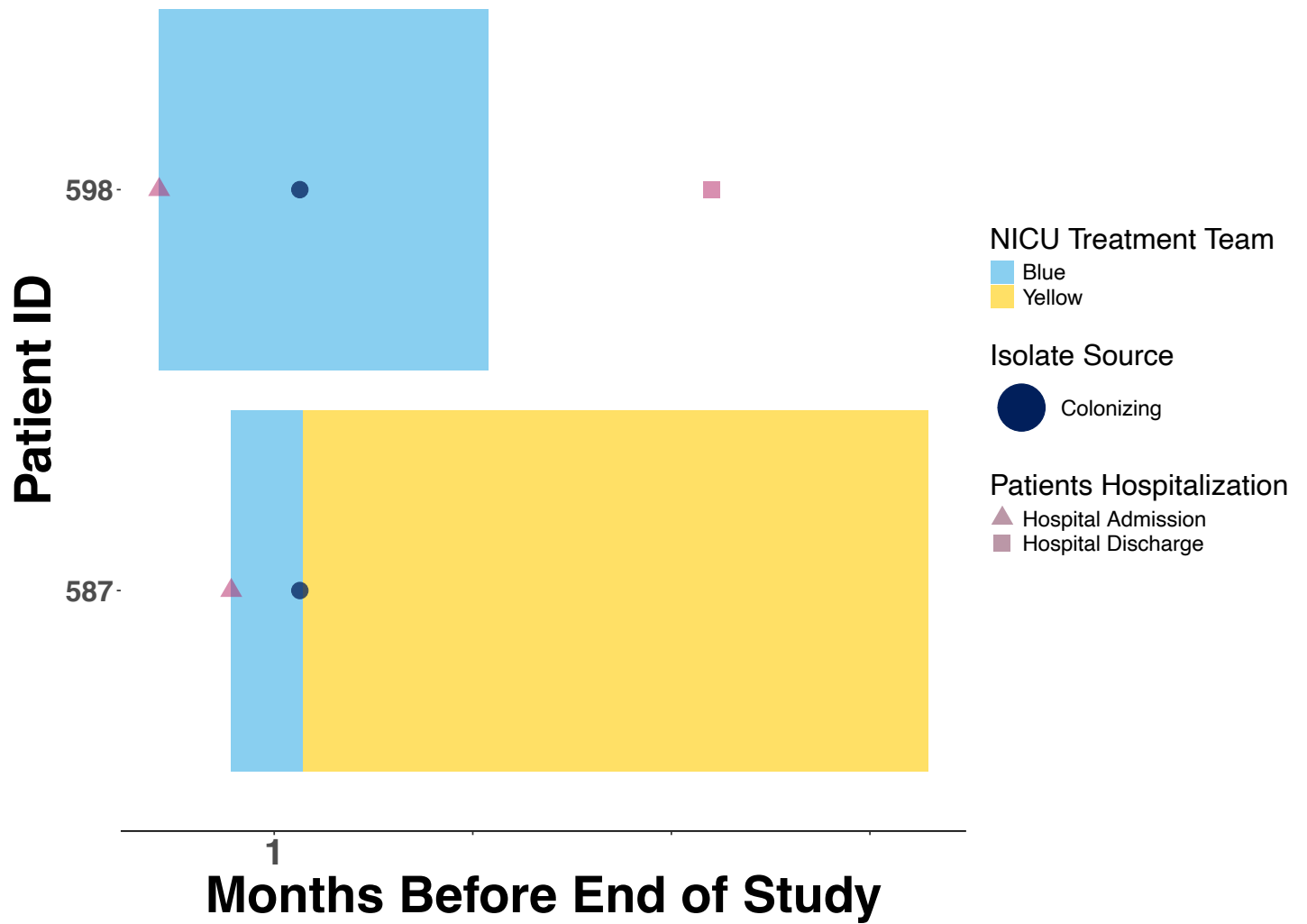
Cluster53



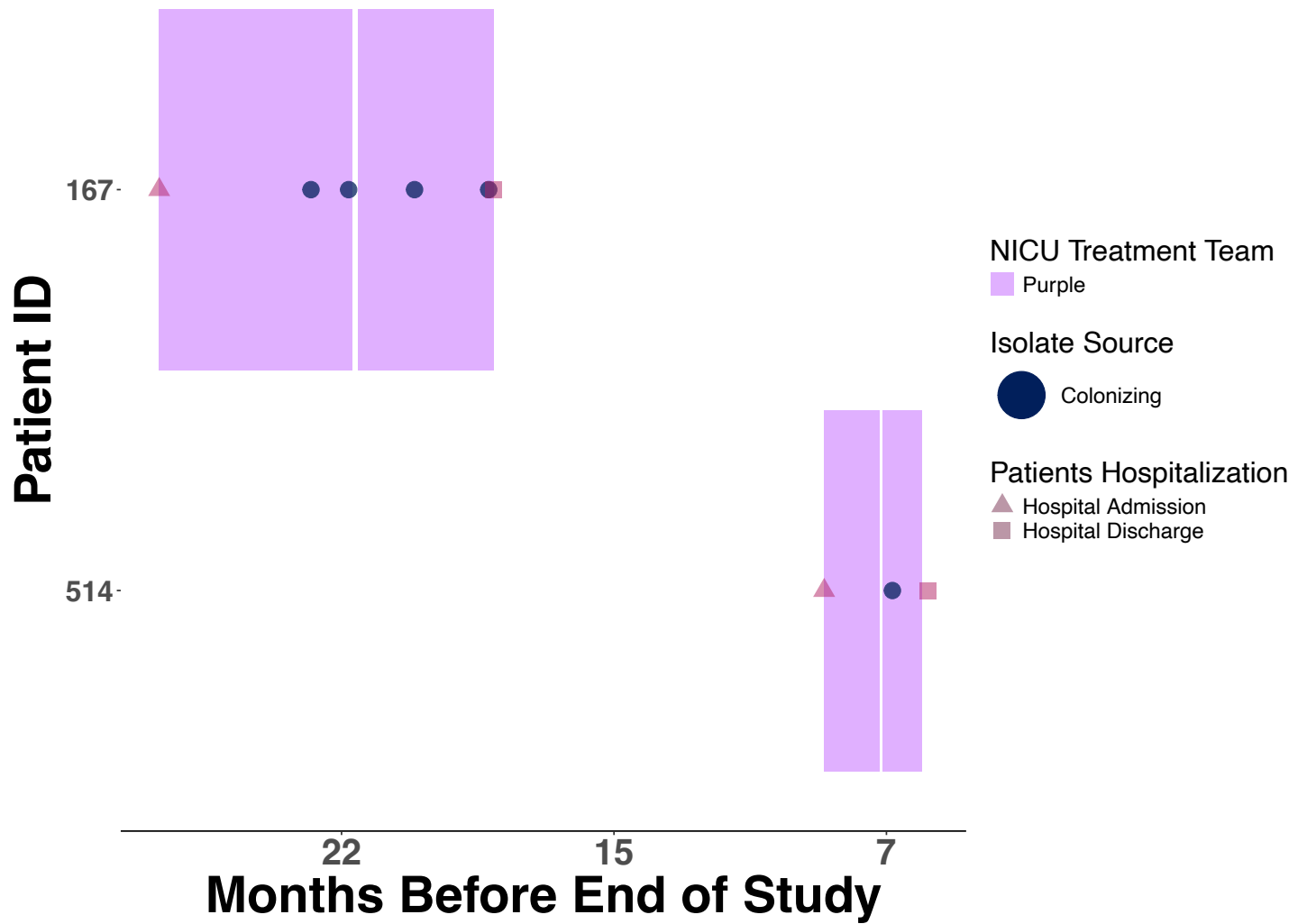
Cluster54



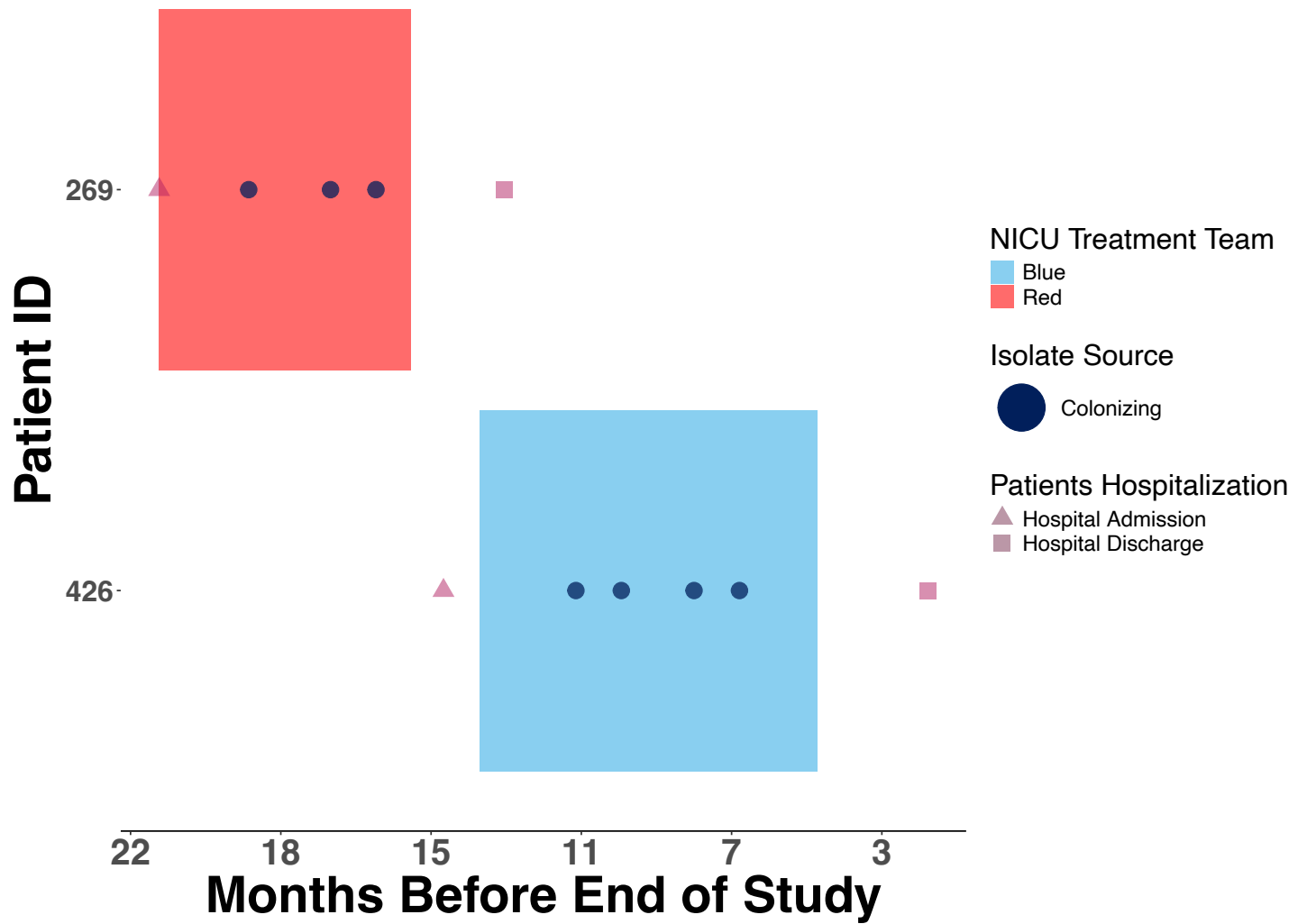
Cluster55



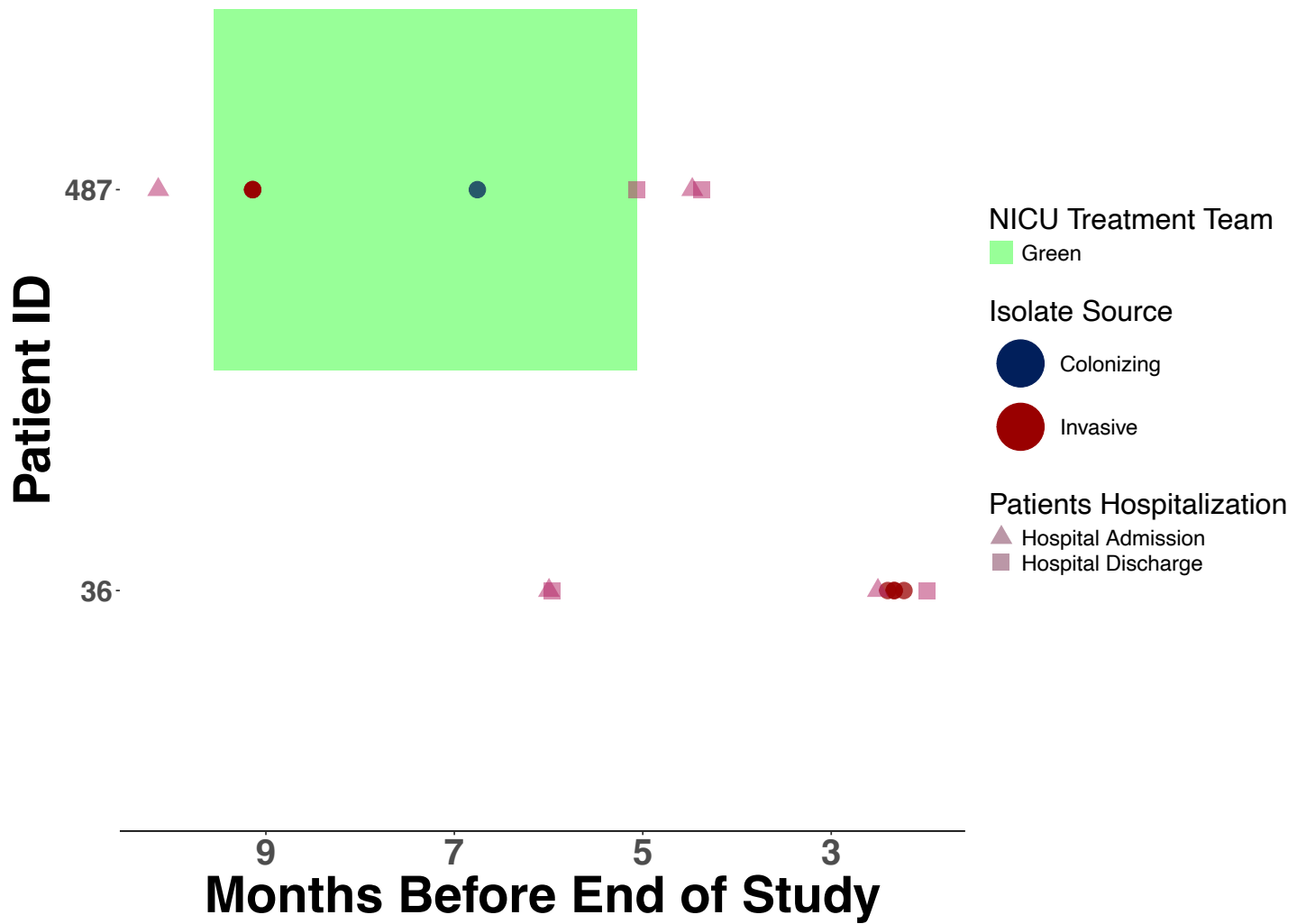
Cluster56



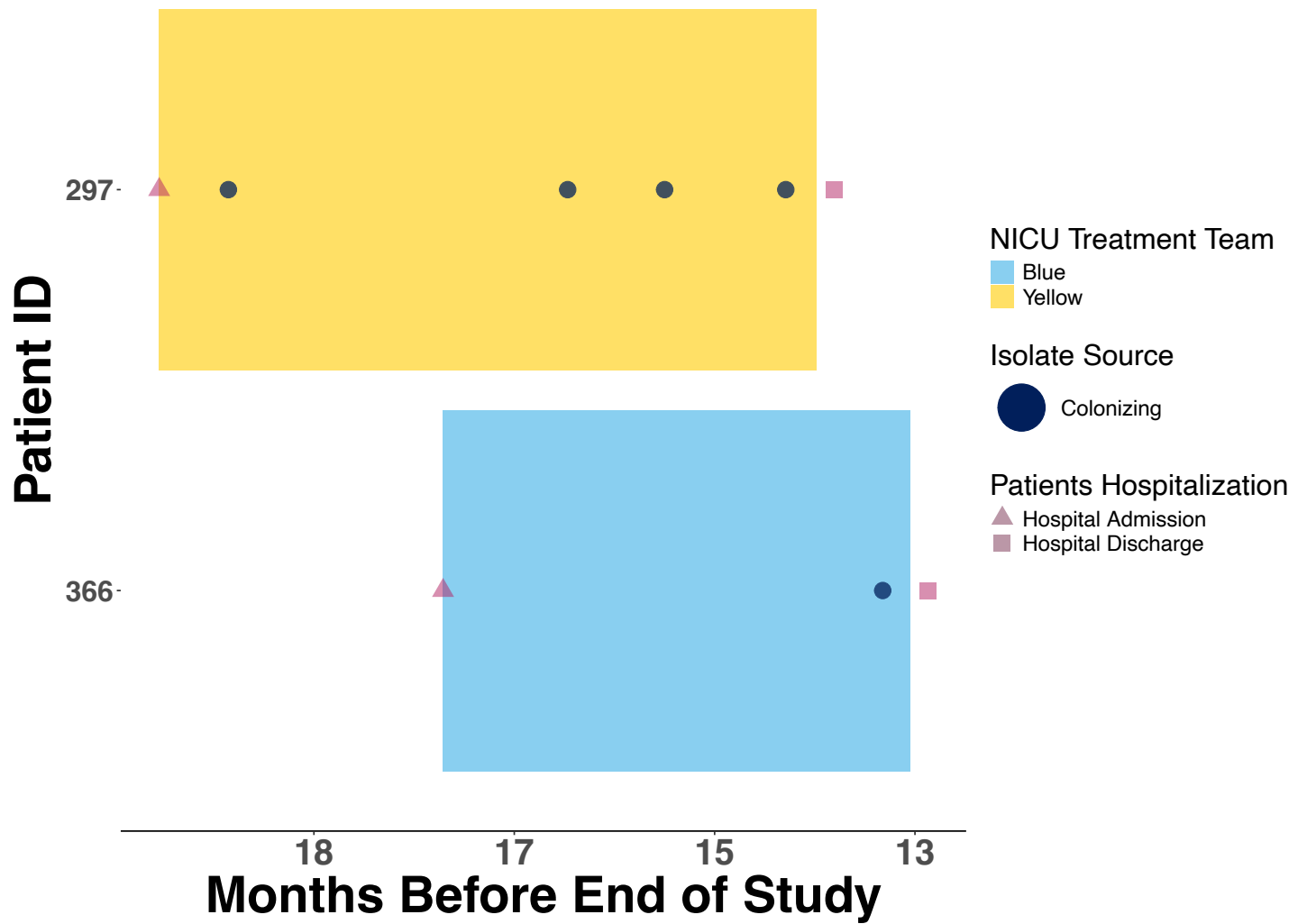
Cluster57



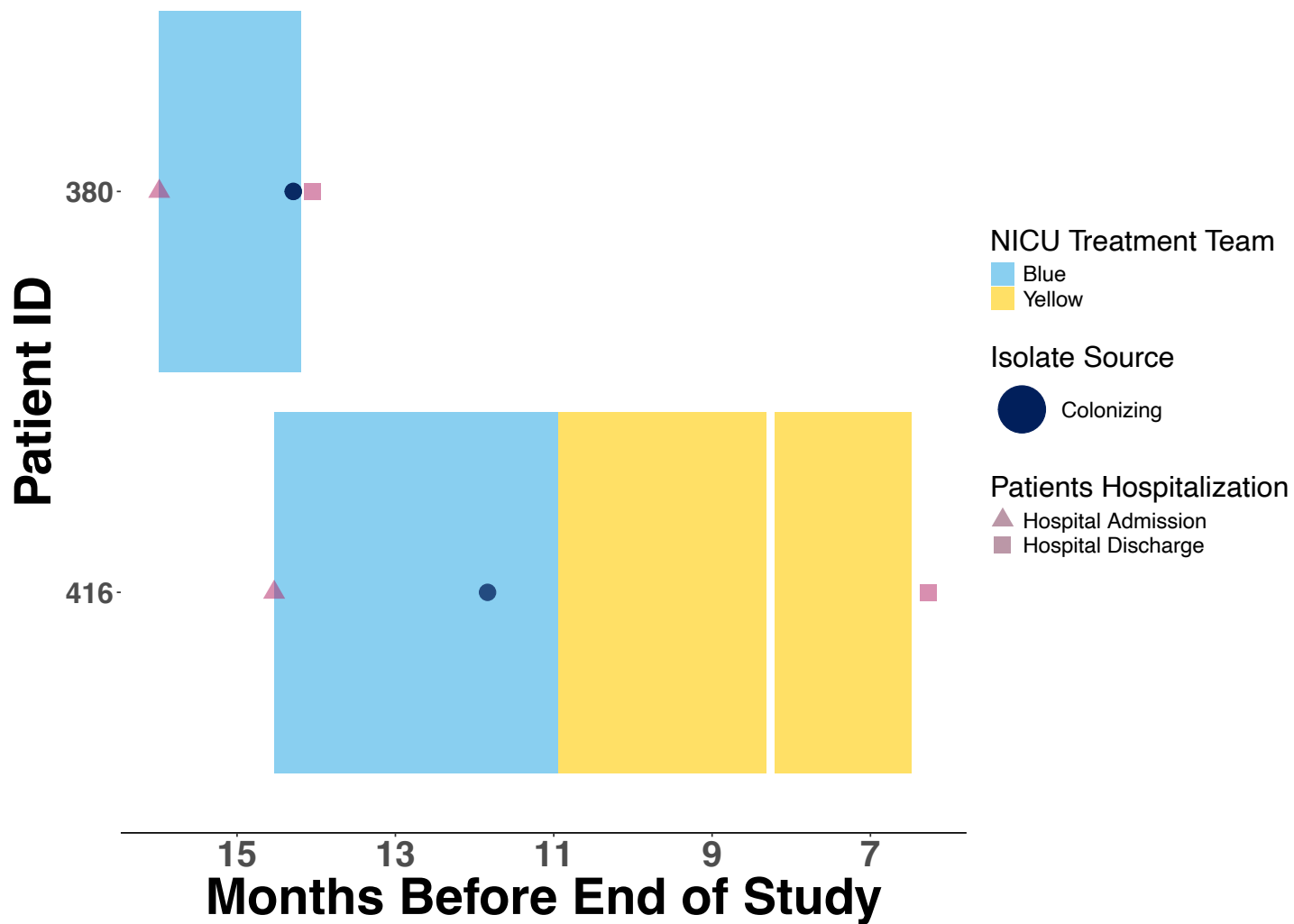
Cluster58



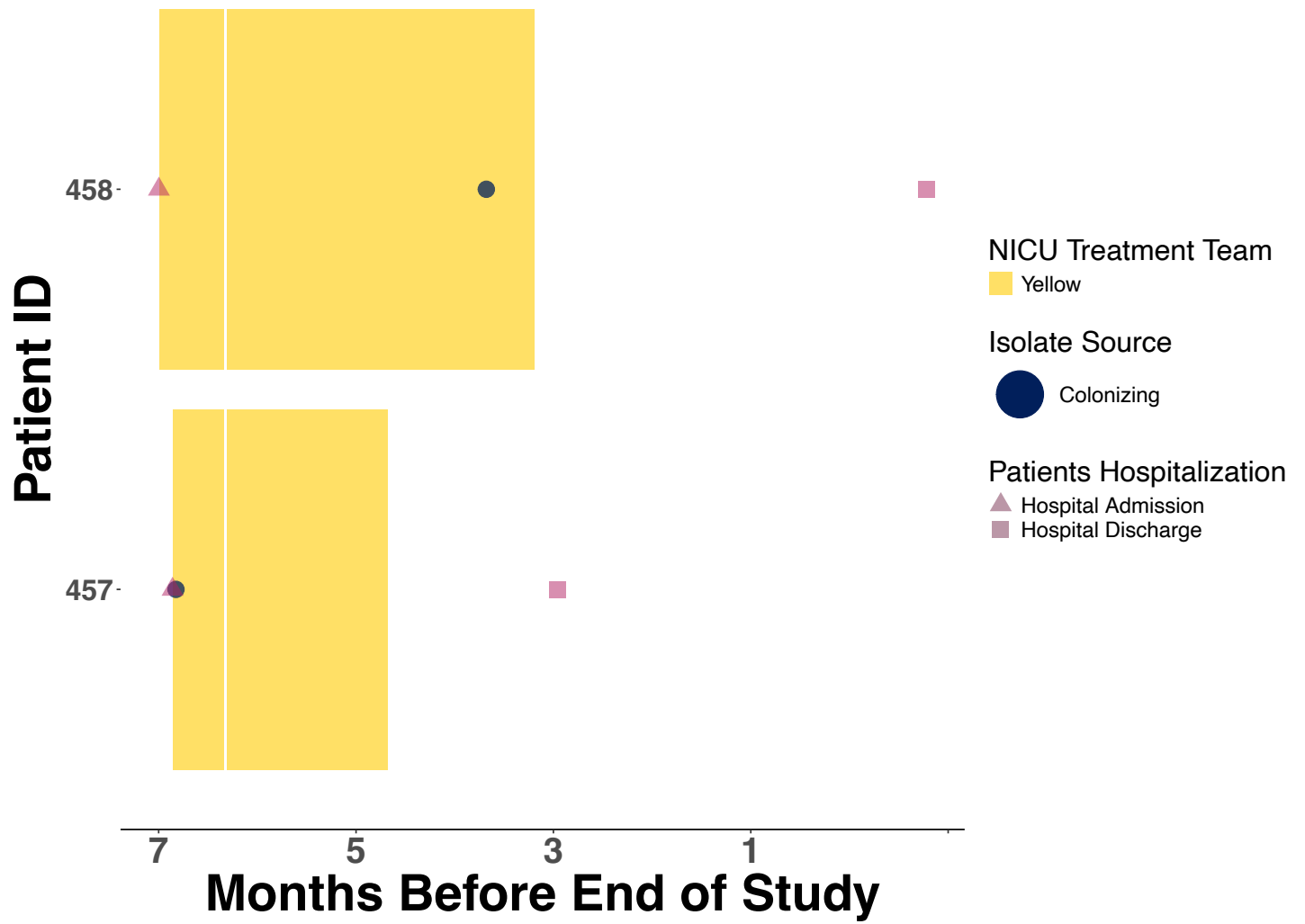
Cluster59



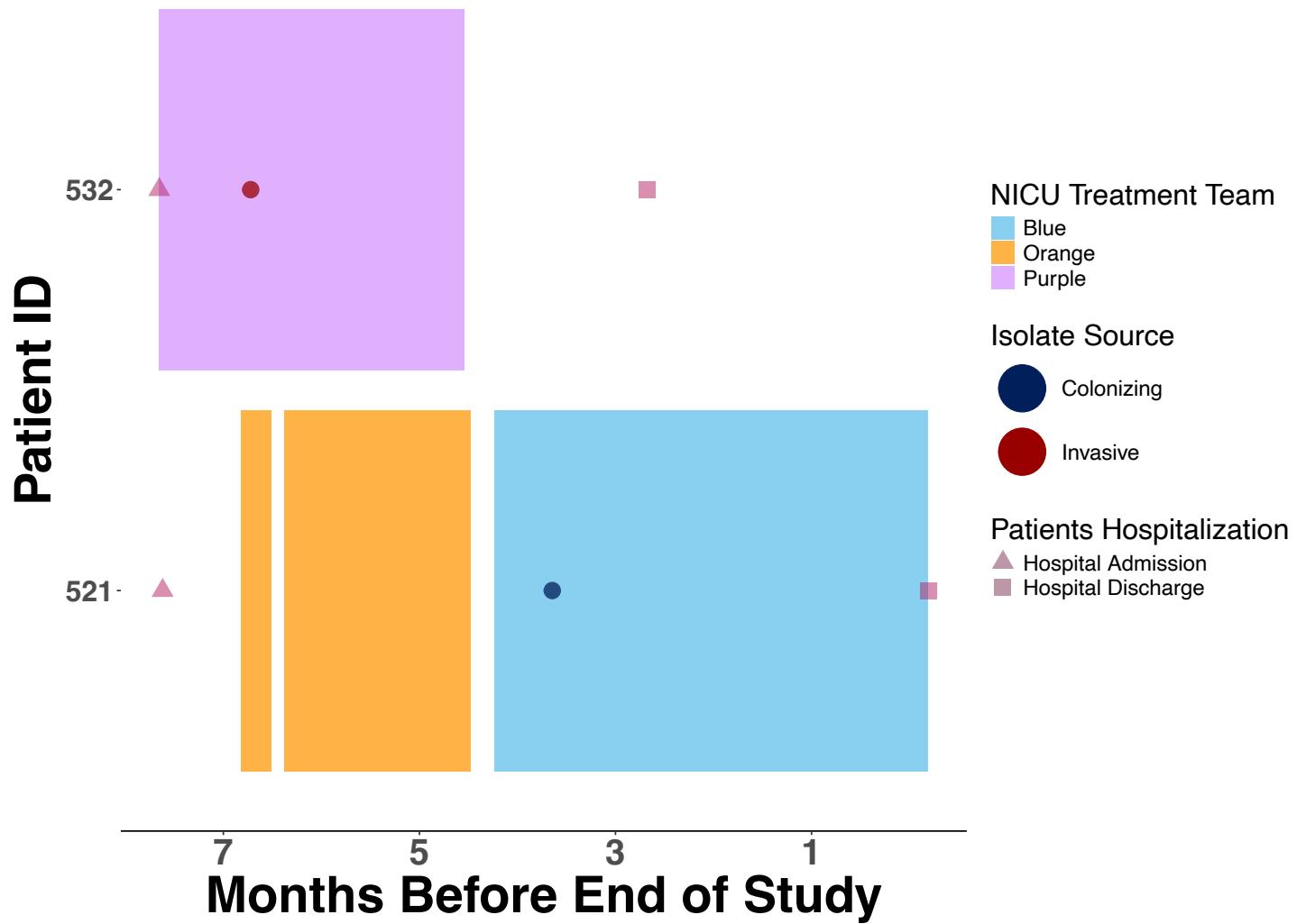
Cluster60



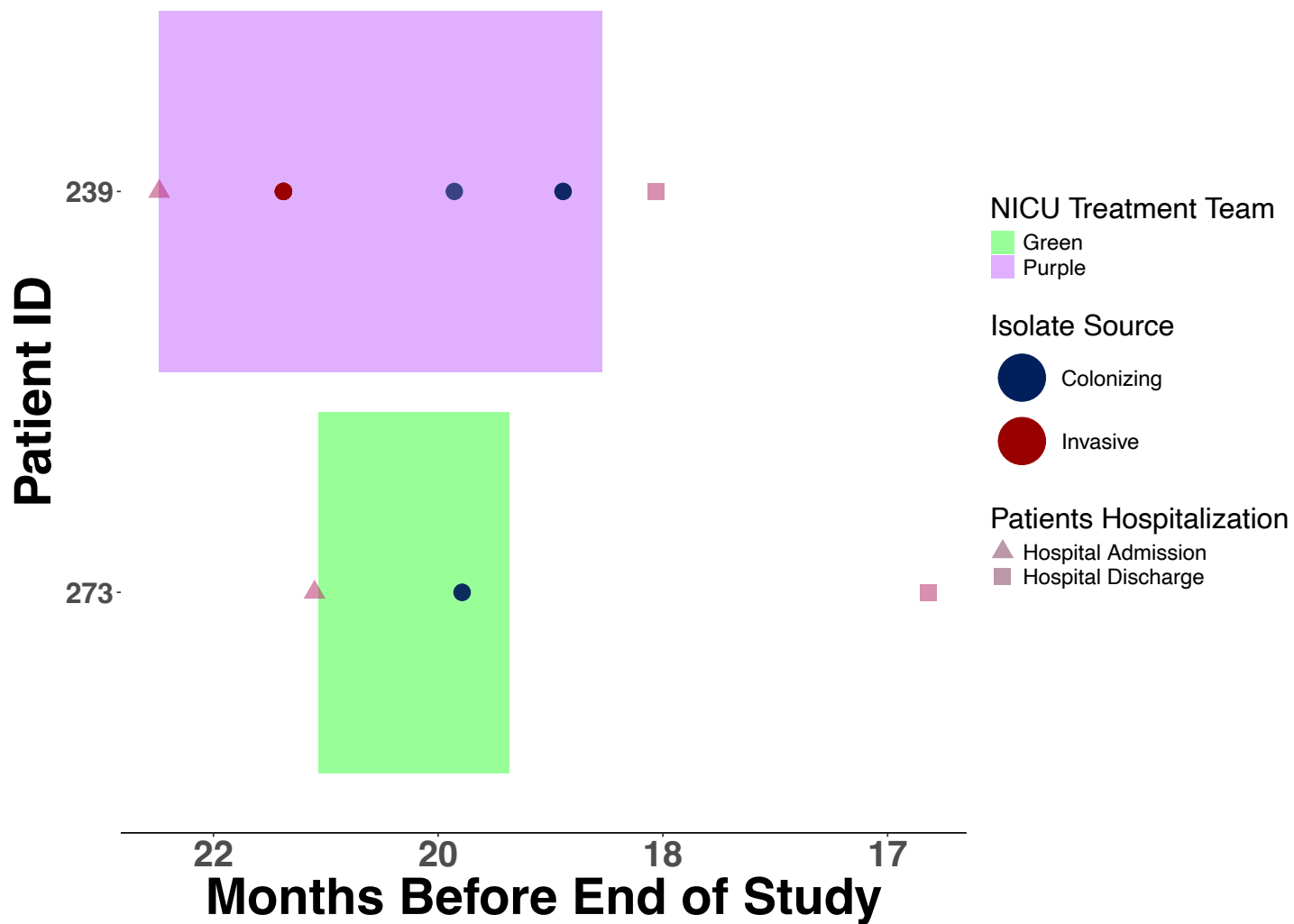
Cluster61



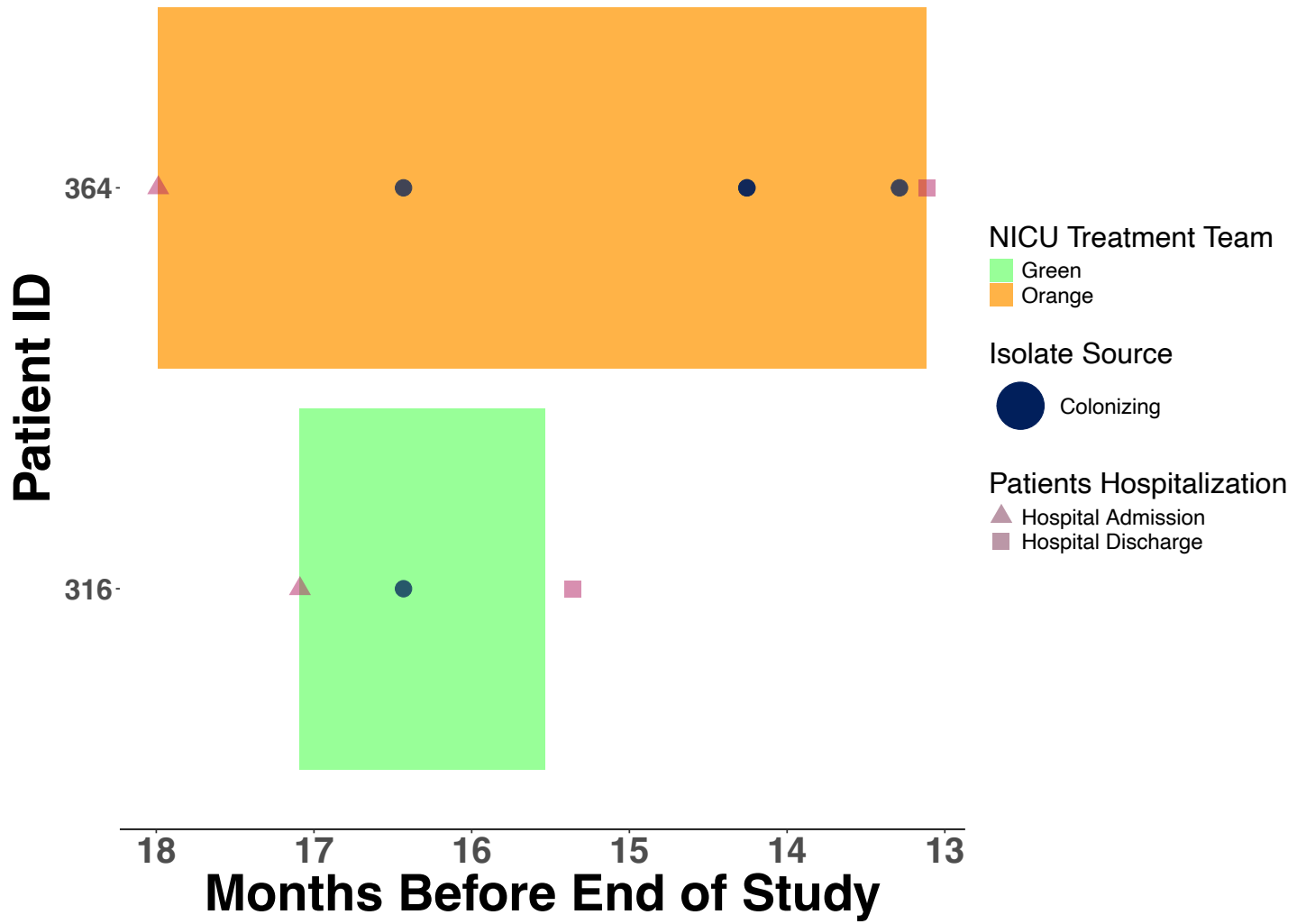
Cluster62



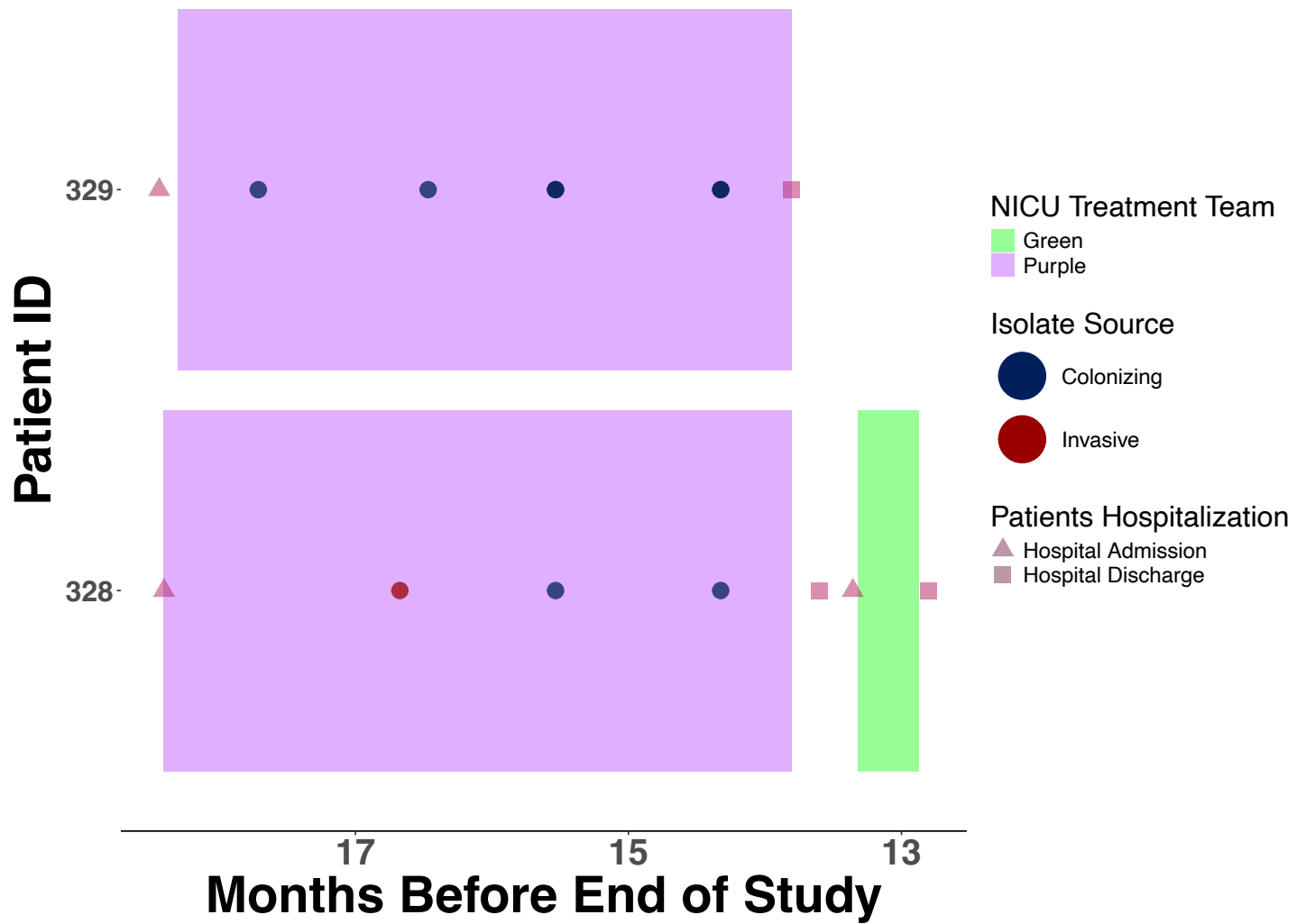
Cluster63



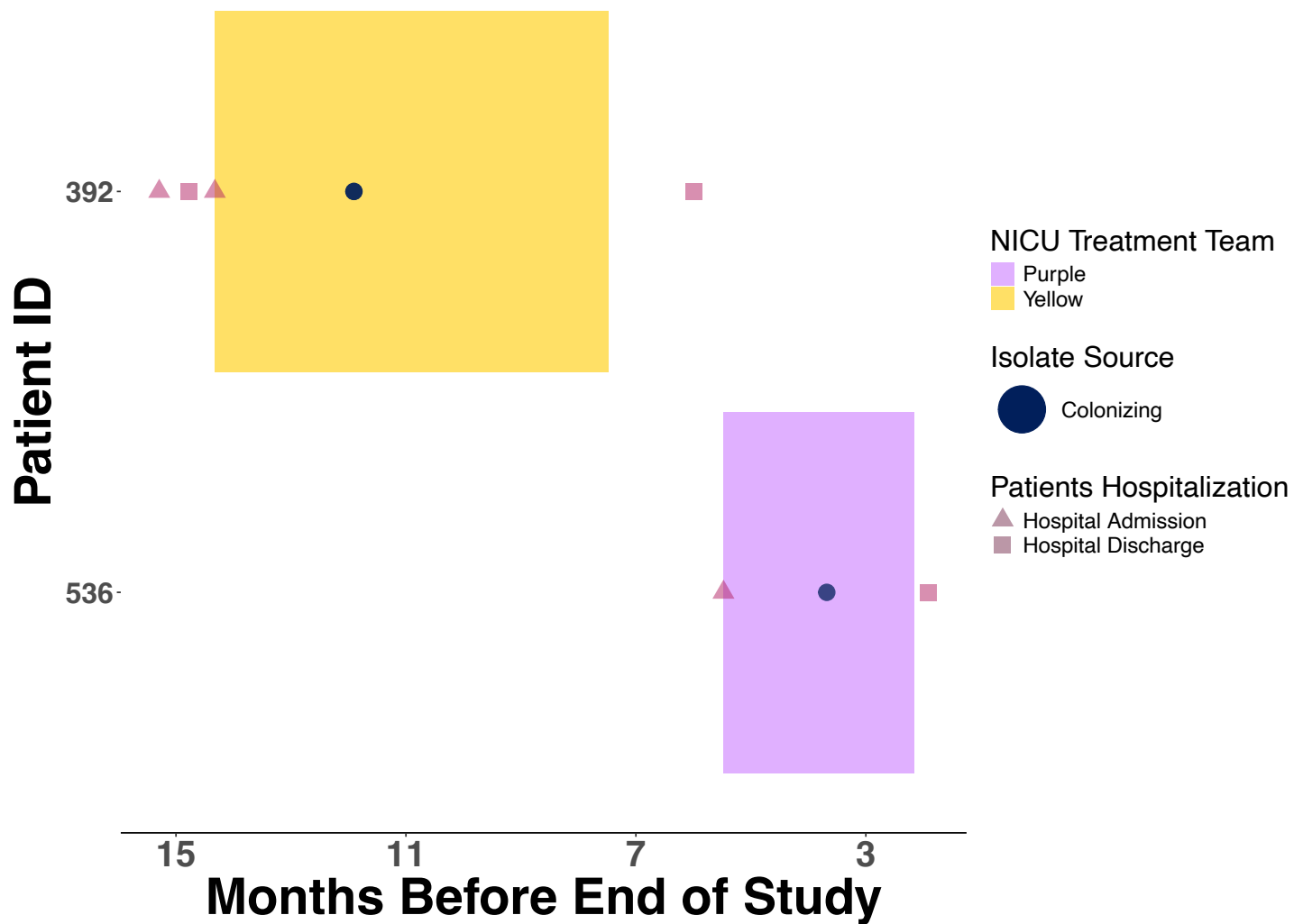
Cluster64



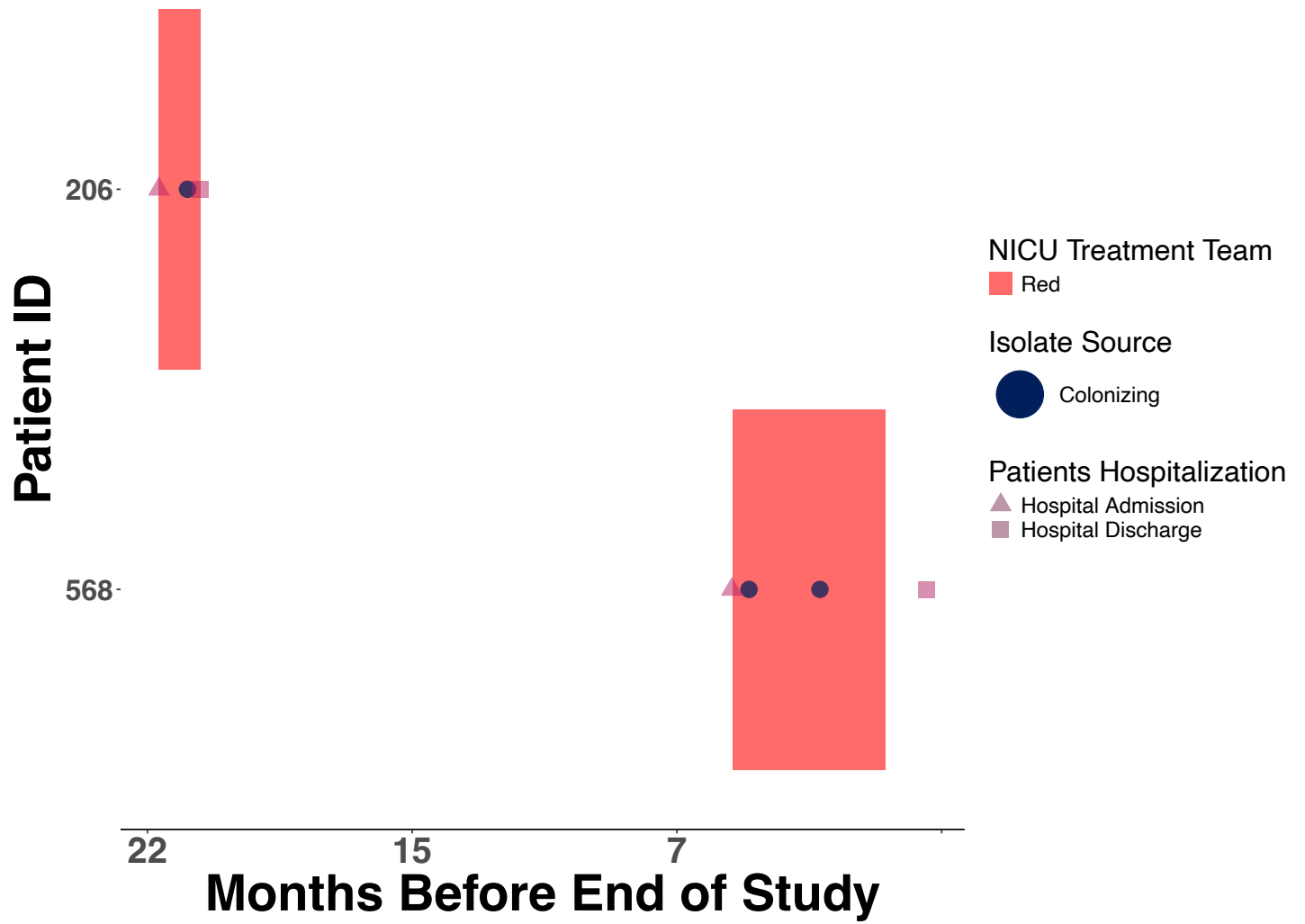
Cluster65



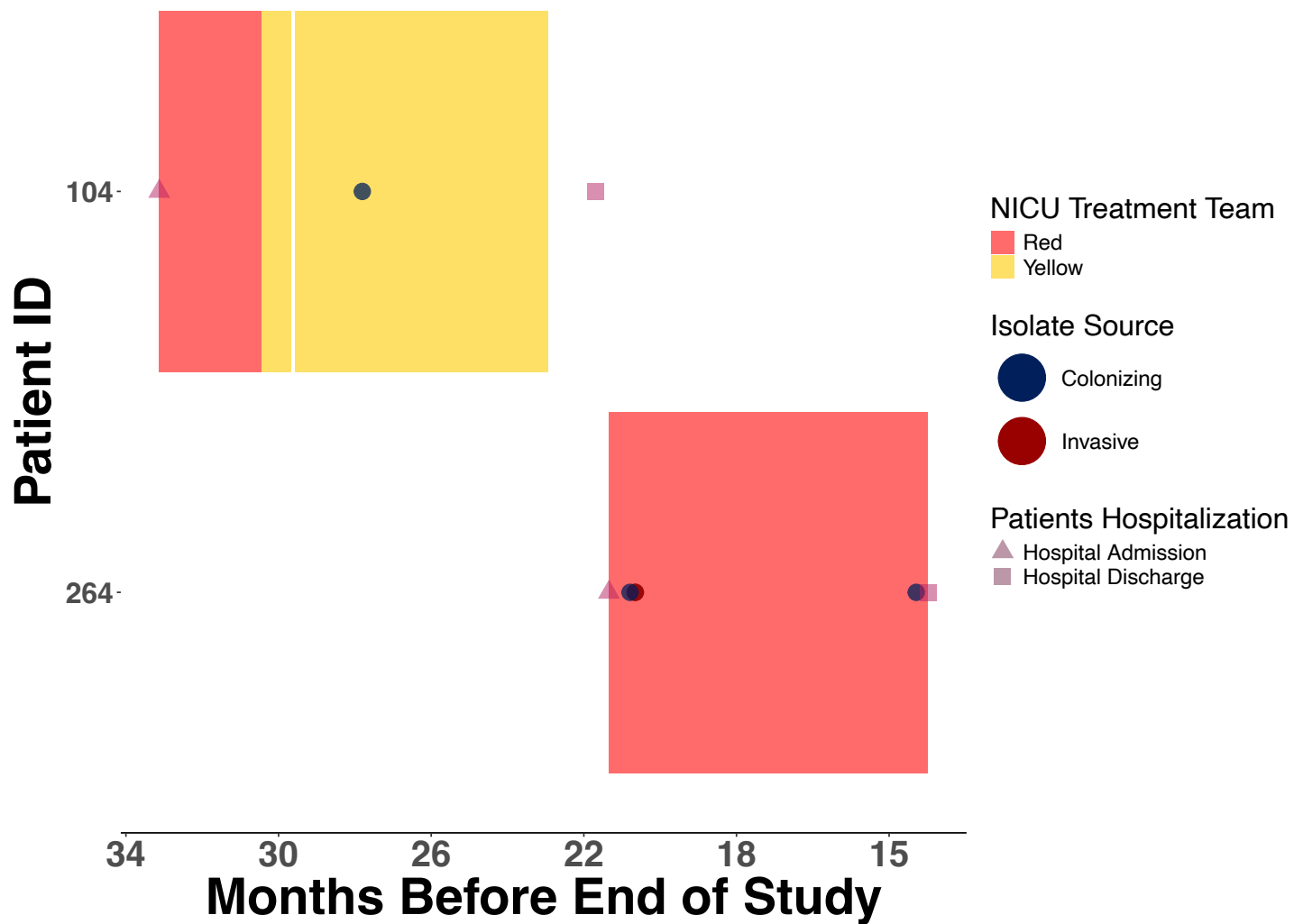
Cluster66



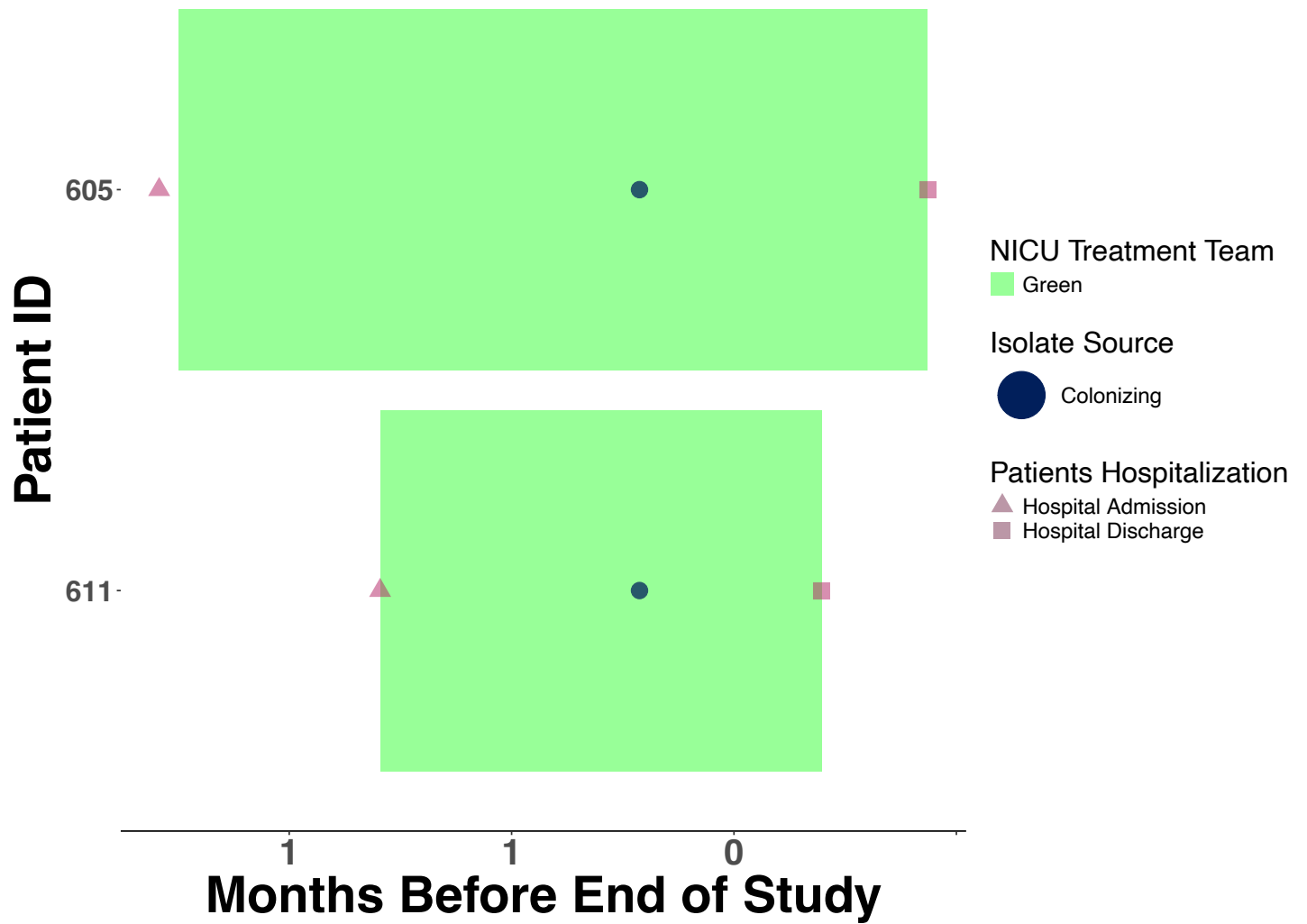
Cluster67



Cluster68



Cluster69



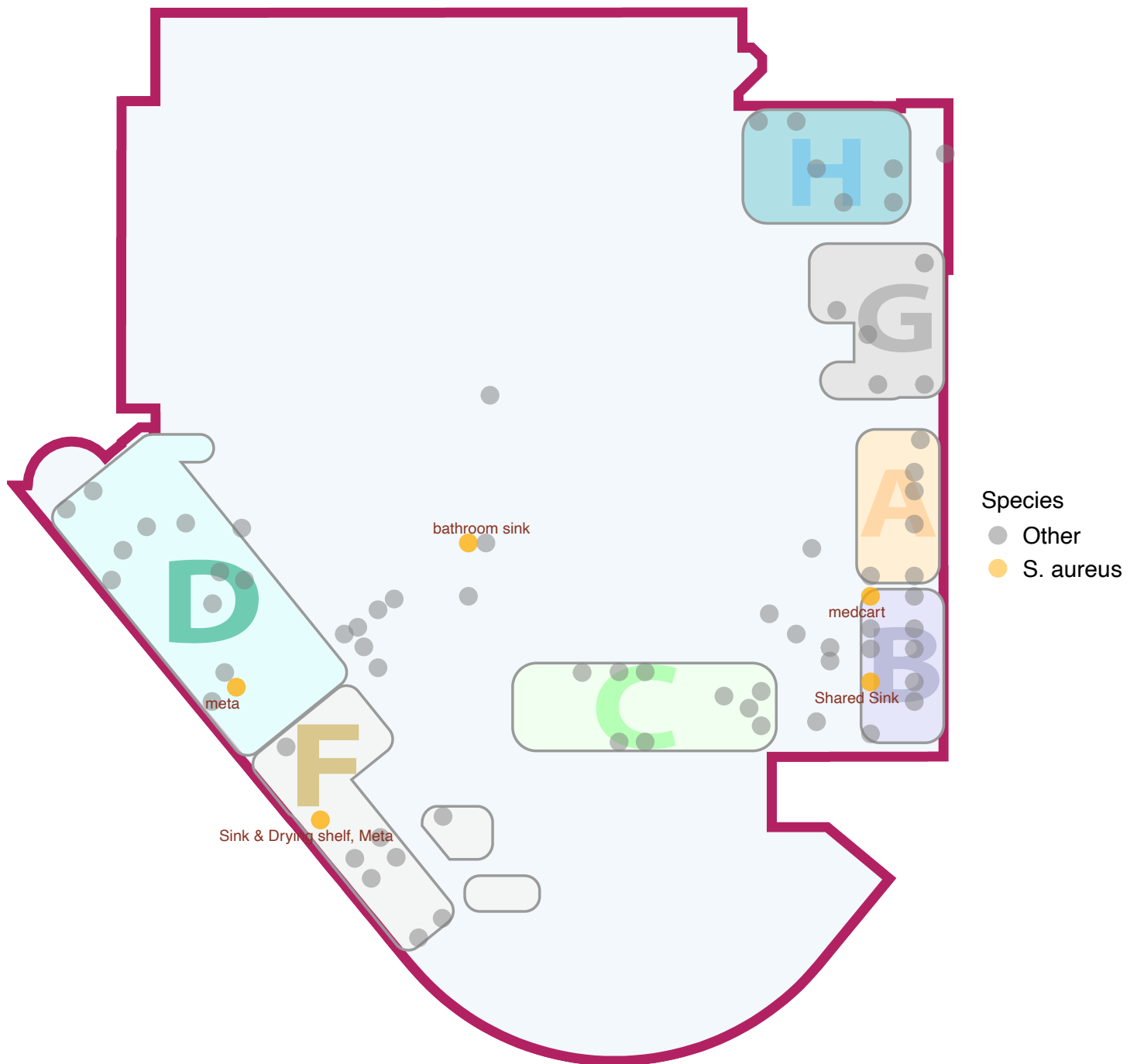


Figure S6: NICU Floorplan Indicating Environmental Sampling Locations

NICU floorplan depicting the 210 environmental sampling locations surveyed across all seven sections. Yellow dots indicate sites with positive *S. aureus* cultures, while gray dots represent locations where other bacterial species were detected.