

Online supplement

Lumacaftor/ivacaftor changes the lung microbiome and metabolome in cystic fibrosis patients

Anne H. Neerincx ¹, Katrine Whiteson ⁴, Joann L. Phan ⁴, Paul Brinkman ¹, Mahmoud I. Abdel-Aziz¹, Els J.M. Weersink¹, Josje Altenburg¹, Christof J. Majoor ¹, Anke H. Maitland-van der Zee ^{1,2}, Lieuwe D.J. Bos ^{1,3}

On behalf of the Amsterdam mucociliary clearance disease research group*

On behalf of the Amsterdam UMC Breath Research group**.

Content:

- Online methods
- Online tables: E1-E5
- Online figures: E1-E6

Online methods:

Sample collection:

For the oral wash, 20 mL sterile normal saline was used to gargle for 30 seconds followed by expectoration of the fluid into a sterile cup. For the nasal wash, 5 mL sterile normal saline was injected into each nostril and held in the nasal cavity for 10 seconds, after which it was dripped into a sterile cup. All specimens were directly stored at minus 80 degrees Celsius for further analysis. For breath, a single vital capacity volume was exhaled into a Tedlar bag (SKC Inc, Eighty Four, PA, USA), after tidal breathing for 30 seconds through a bacterial filter (3-safety-pack-s, Lemon Medical GmbH, Hammelburg, Germany) connected to a two-way non-rebreathing valve and an organic compound filter (A2, North Safety, Middelburg, Netherlands). The sample was subsequently stored on a Tenax sorbent tube (Tenax GR 60/80, Interscience) at 4 degrees Celsius until analysis. Analysis was performed within 2 weeks.

Sample size

The sample size was calculated for the outcome bacterial diversity, as most preliminary data was available for this outcome. The effect size was chosen based on a minimal clinically important change (25%) in bacterial diversity and the standard deviation based on preliminary data [*]. With an expected effect size of 0.25 and a standard deviation of the outcome of 2.0 with a within-subject correlation of 0.8, the required sample size was 20.

[*] Quinn, R.A., et al., Biogeochemical forces shape the composition and physiology of polymicrobial communities in the cystic fibrosis lung. MBio, 2014. 5(2).

Microbiota analysis:

Total DNA was extracted from sputum, oral wash and nasal wash samples using the PowerFecal DNA Kit (MO BIO Laboratories, USA). Sequencing was performed using the Illumina MiSeq platform (San Diego, CA), using a MiSeq Reagent Kit V2 (500 cycles), according to the manufacturer's instructions. Further processing was performed using the DADA2 pipeline.

Library Preparation 16S rRNA Amplicons:

Amplicon fragments are PCR-amplified from the DNA in duplicate using separate template dilutions (generally 1:1 & 1:10) using the high-fidelity Phusion polymerase. A single round of PCR is done using "fusion primers" (Illumina adaptors + indices + specific regions) targeting 16S V4-V5 (primarily Bacteria; ~410 bp) regions with multiplexing which allows up to 380 samples to be run. PCR products are verified visually by running on a high-throughput Hamilton Nimbus Select robot using Coastal Genomics Analytical Gels. Any samples with failed PCRs (or spurious bands) are re-amplified by optimizing PCR conditions to produce correct bands in order to complete the sample plate before continuing. The PCR reactions from the same samples are pooled in one plate, then cleaned-up and normalized using the high-throughput Charm Biotech Just-a-Plate 96-well Normalization Kit. The (up to) 380 samples are then pooled to make one library which is then quantified fluorometrically before sequencing. Paired-end reads from amplicon sequencing targeting bacterial hypervariable V4-V5 rRNA region were pooled for subsequent processing. The reads were quality checked using FastQC and MultiQC, and then cleaned by removing

primers (515F-Y: 5'-GTGYCAGCMGCCGCGGTAA and 926R: 5'-CCGYCAATTYMTTTRAGTTT) and adapter sequences using Cutadapt.

16S rRNA sequencing analytical pipeline

The 16S rRNA gene V4-V5 sequencing was performed using the CGEB-Integrated Microbiome Resource pipeline. Briefly, the pipeline works by performing quality filtering and trimming, dereplicating sequences, learning dataset-specific error rates, denoising by removing potentially containing errors sequences, merging paired-end reads while removing mismatches to reduce errors, constructing amplicon sequence variants (ASVs), removing chimera by implementing “bimera” method, and running taxonomic classification of ASVs using different publically available databases.

Both negative controls (blanks) and positive controls (mock communities) are part of the quality control of the sequencing process. [*]. For confirmatory purposes, our bioinformatic processing did not retrieve any detectable reads from included negative control sample, which implies that there is no bacterial contamination of the included samples. The EzBioCloud database was selected for down-stream analyses because it is an integrated database which allows for improved bacterial taxonomic identification with reported better accuracy compared to other publically available databases.

[*] Comeau, AM, Douglas GMLangille MG. Microbiome helper: a custom and streamlined workflow for microbiome research. MSystems 2017; 2.

16S rRNA sequencing post processing procedures

After trimming primers with cutadapt, the forward and reverse reads were inspected for quality using fastqc and multiqc. Then, we have followed the dada2 pipeline

developed by Callahan et al. [*]. The complete pipeline can be found in github (<https://benjjneb.github.io/dada2/tutorial.html>). During the trimming and filtering step of the forward and reverse sequencing reads, quartiles of the quality score distribution of nucleotide positions were inspected. Nucleotides with worse quality at the end of the forward and reverse reads were trimmed (particularly, truncation of the forward reads was performed at position 270 and reverse reads at position 220). Parameters for the reads filtering were used as follows; truncQ=2 (Truncate reads at the first instance of a quality score less than or equal to 2), maxEE=7 (reads with higher than 7 expected errors will be discarded), and rm.phix=TRUE (discard reads that match against the phiX genom). Then, learning of the error rates was performed. The estimated error rates were in a good fit to the observed rates, and the error rates dropped with increased quality as expected. Afterwards, The, forward and reverse reads were merged using the default parameters in dada2 (minimum length of overlap=20 nucleotides, and maximum nucleotides mismatches=0). Subsequently, chimeric sequences were removed by implementing the Bimera method (removeBimeraDenovo function in dada2). Only 1 % of the total sequences were identified as chimera and subsequently removed from the final identified amplicon sequence variants (ASVs) reads. Finally, Taxonomy was assigned using the Ribosomal Database Project (RDP) Naive Bayesian Classifier algorithm [**] impleted in dada2 using the default parameters (minimum bootstrap confidence for assigning a taxonomic level=50) against the EzBioCloud database (May 2018) [***].

[*] Callahan, BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJAHolmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* 2016; 13: 581-583.

[**] Wang, Q, Garrity GM, Tiedje JMCole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* 2007; 73: 5261-5267.

[***] Yoon, S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo HChun J. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International journal of systematic and evolutionary microbiology* 2017; 67: 1613.

Sputum metabolomics assays

1: Gas chromatography – time of flight – mass spectrometry (GC-TOF-MS) was used for separation of primary metabolite classes such as amino acids, hydroxyl acids, carbohydrates, sugar acids, sterols, aromatics, nucleosides, amines and miscellaneous compounds. GC-TOF-MS analysis was performed, after metabolite extraction from 20uL of frozen supernatant using the standard plasma extraction protocol in 1mL of 3:3:2 acetonitrile:isopropyl alcohol:water, as described previously. In short, 0.5 uL of sample was injected onto a Rtx-5Sil MS capillary column (30 m, 0.25 mm inner diameter, 0.25 µm film of 95% dimethyl/5%diphenylpolysiloxane; Restek) at 1.0 mL/min column flow. The oven temperature profile started at 50 °C for 1 minute, then ramped at 20 °C/min to 330 °C, and finished with 5 minutes at 330 °C. A Leco Pegasus IV mass spectrometer at -70 eV was used to ionize the molecules, and a quadrupole mass spectrometer (GCMS-GP2010) detected them with a scan range of m/z 80-500 Da.

2: Hydrophilic interaction chromatography quadrupole time of flight mass spectrometry (HILIC-QTOF-MS) was used for analysis of biogenic amines, after extraction in sequential ethyl acetate solvation followed by methanol solvation. Briefly; an Agilent 1290 Infinity LC system (G4220A binary pump, G4226A

autosampler, and G1316C Column Thermostat) was used for separation of polar compounds using an Acquity UPLC BEH Amide Column, 130Å, 1.7 µm, 2.1 mm X 150 mm, maintained at 45°C, at a flow-rate of 0.4 mL/min. The mobile phases consist of: Water, 10 mM Ammonium Formate, 0.125% Formic Acid (A) and Acetonitrile: Water (95/5, v/v), 10 mM Ammonium Formate, 0.125% Formic Acid (B). The gradient started with 0 min 100% (A); followed by 0–2 min 100% (A); 2–7.7 min 30% (A); 7.7–9.5 min 60% (A); 9.5–10.3 min 70% (A); 10.3–12.8 min 0% (A); 12.8–16.8 min 0%. Electrospray ionization (ESI) was used to ionize the molecules, and a SCIEX Triple TOF 6600 mass spectrometer detected them with a full scan range of m/z 50-1200 Da.

Exhaled breath analysis

After sampling, the tubes were transported to the TD unit where the tubes were heated to 250 °C for 15 minutes with a flow of 30 mL/min. A cold trap at 5 °C captured the VOCs, and was subsequently heated to 300 °C to release the molecules through a heated transfer line (split less injection) at 120 °C onto a Inertcap 5MS/Sil GC capillary column (30 m, 0.25 mm inner diameter, 1.0 µm film thickness, and 1,4-bis(dimethylsiloxy)phenylene dimethyl polysiloxane; Restek) at 1.2 mL/min column flow. The oven temperature profile started at 40 °C for 5 minutes, then increased to 270 °C with a 10 °C/min heating rate, and finished with 5 minutes at 270 °C. Electron ionization at 70 eV was used to ionize the molecules, and a quadrupole mass spectrometer (GCMS-GP2010) detected them with a scan range of m/z 37-300 Da. The R package *xcms* was used to align the chromatogram based on known reference peaks in the chromatogram as has been described before. The ion-fragment peaks were grouped based on retention time and co-elution pattern. The

base peak was used for subsequent statistical analyses. Fragment ions of interest were manually checked in the raw chromatograms, and corresponding metabolites were tentatively identified using National Institute of Standards and Technology library (NIST, Gaithersburg, MD).

Supplemental tables

Table E1: Overview analyses performed in samples of each individual patient

Patient number	Visit number	Oral wash	Nasal wash	Sputum	Sputum	Sputum	Sputum	Exhaled breath
		16S rRNA seq	16S rRNA seq	16S rRNA seq	Metagenomics	Metabolomics (1)	Metabolomics (2)	Metabolomics
1	1	X		X	X	X	X	X
	2	X	X					X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	
	5	X	X	X	X			X
2	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
3	1	X	X					X
	2	X	X					X
	3	X	X					X
	4	X	X					X
	5	X	X					X
4	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	X
	5	X	X	X	X	X	X	X

5	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	
	5	X	X	X	X			X
6	1	X	X					X
	2	X	X					X
	3	X	X					X
	4	X	X					
	5	X	X					X
7	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	X
	5	X	X	X	X	X	X	
8	1	X	X	X	X			X
	2	X	X	X	X			X
	3							
	4							
	5							
9	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
10	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3							X
	4	X	X	X	X	X	X	X
	5	X	X	X	X	X	X	X

11	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
12	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X		X	X	X	X	
	4	X		X	X	X	X	
	5	X		X	X	X	X	
13	1	X	X	X	X	X	X	X
	2	X	X					
	3	X	X					X
	4	X	X					X
	5							
14	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X					
	4	X	X					
	5	X	X					X
15	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
16	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X			
	5	X	X	X	X	X	X	

17	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	X
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
18	1	X	X	X	X	X	X	X
	2	X	X	X	X	X	X	X
	3	X	X	X	X	X	X	
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	X
19	1	X	X					X
	2	X	X					X
	3	X	X					
	4	X	X					
	5	X	X					
20	1	X	X					X
	2	X	X					X
	3	X	X					
	4	X	X					
	5	X	X					

Table E2: Metabolites from dataset 1 included sPLS model

Metabolite	Loading var1	Loading var2
Tyrosine	0.000000000	0.312590235
Thymidine	0.000000000	-0.340450698
Ribitol	-0.411808432	0.212304707
Pseudo uridine	0.000000000	0.066330799
Phosphoenolpyruvate	0.000000000	0.366252404
Phenylpyruvate	-0.404887879	0.208736869
Oxoproline	0.000000000	0.023189631
Lysine	0.000000000	0.043476258
Homoserine	-0.118584989	0.061135590
Glyceric acid	-0.082811931	0.042693062

Table E3: Metabolites from dataset 2 included sPLS model

Metabolite	Loading var1	Loading var2
Gamma aminobutyric acid	0.00308421	0.0003547811
2,8-Quinolinediol	0.01695296	0.0019501225
3,5-Dihydroxyphenylglycine	0.00000000	0.4305857875
3-Ureidopropionic.acid	0.00000000	-0.2051775604
4-Hydroxymandelonitrile	0.17286083	0.0198844260
Leucylarginine	0.00000000	0.3817082936
Cortisone	0.00000000	0.4519463198
Cortodoxone	-0.30846008	-0.0354826001
Corydaline	0.00000000	0.2237905727
Diazepam	0.36614394	0.0421180568

Table E4: Breath metabolites included sPLS model

Metabolite	Loading var1	Loading var2
1,3–Butadiene	0.1491300	0.42170985
Co-elution of unknown identity	0.4716426	-0.07654338
4 ethyl benzanoic acid 2-pentyl esther	0.5848016	-0.10027912
	0.5876768	0.07249388
	0.3529467	0.41154142
	0.3742912	0.41443420
1-methylpentryl hydroperoxide	0.1731457	0.59057180
Contaminant	0.2885300	0.52546595
Contaminant	0.5420864	0.11081600
Contaminant	0.5434267	0.13705716

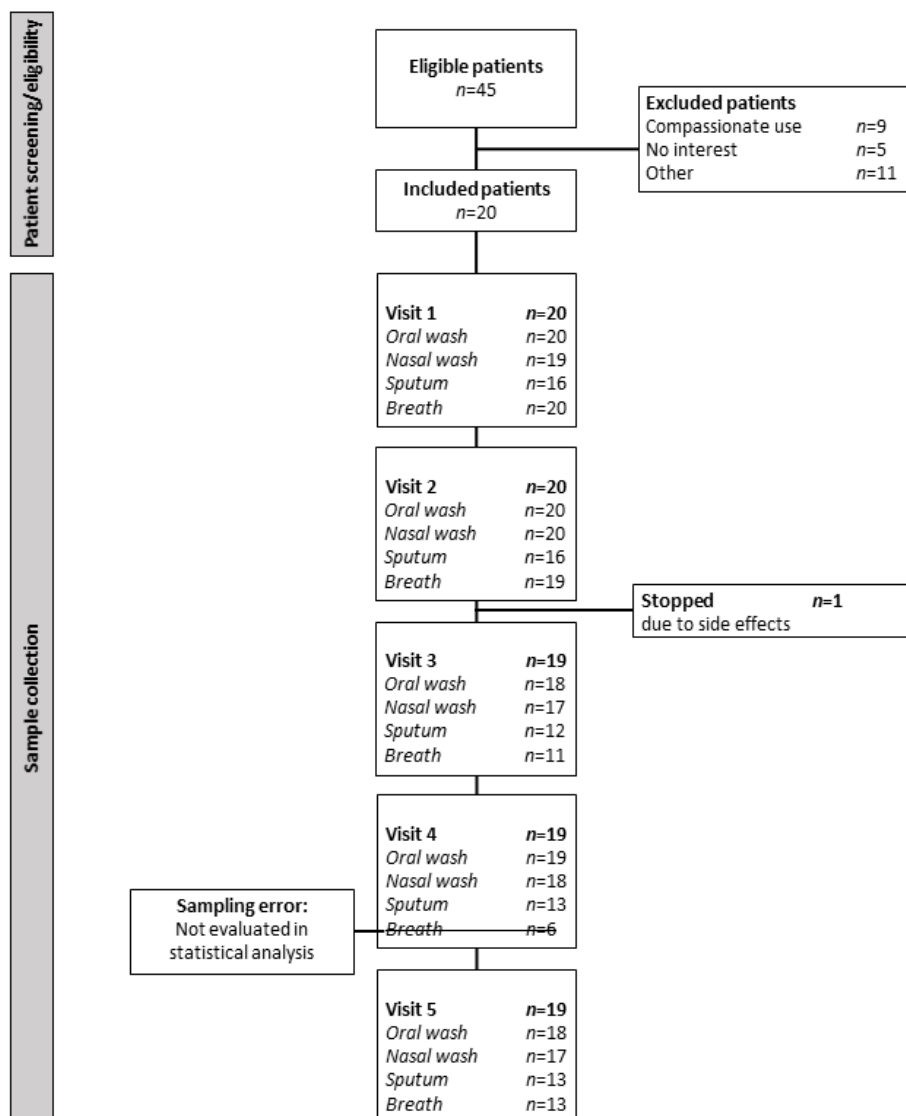
Table E5: Antibiotic use during study visits

PatientID	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
1	Tobramycin	Tobramycin Colistin	None	None	Levofloxacin
2	Azithromycin Aztreonam Colistin Tobramycin	Azithromycin Tobramycin Colistin Aztreonam	Azithromycin Colistin Aztreonam Tobramycin	Azithromycin Colistin Tobramycin	Azithromycin Colistin Tobramycin
3	None	Cotrimoxazole	None	None	None
4	Tobramycin Azithromycin Colistin	Tobramycin Colistin Azithromycin	Tobramycin Colistin Azithromycin	Tobramycin Colistin Azithromycin	Tobramycin Colistin Azithromycin
5	None	None	None	None	None
6	Azithromycin	None	Amoxicillin Clavulanic acid	None	None
7	None	None	None	None	None
8	Colistin	Colistin	-	-	-
9	Doxycycline Tobramycin	Doxycycline Tobramycin	Tobramycin Cotrimoxazole	Tobramycin Cotrimoxazole	Tobramycin Cotrimoxazole
10	None	None	None	None	None
11	Azithromycin Colistin	Colistin Azithromycin	Azithromycin Colistin	Azithromycin Colistin	Erythromycin Colistin Ciprofloxacin
12	None	Meropenem Colistin	Flucloxacillin Piperacillin Tazobactam	Flucloxacillin	Flucloxacillin
13	Tobramycin	Tobramycin	Tobramycin	Tobramycin	-
14	Azithromycin Tobramycin	Azithromycin	Azithromycin	Azithromycin	Azithromycin
15	Colistin	None	Tobramycin	Tobramycin	Tobramycin
16	Azithromycin	Azithromycin	Azithromycin	Azithromycin	Azithromycin

	Colistin	Colistin Flucloxacillin Ceftazidim Tobramycin	Colistin Tobramycin Ceftazidim Ciprofloxacin	Colistin	Colistin
17	Tobramycin	Tobramycin	Tobramycin	Tobramycin	Tobramycin
18	Azithromycin Colistin	Levofloxacin Colistin	Levofloxacin Colistin	levofloxacin Colistin	Levofloxacin
19	None	None	None	None	None
20	Azithromycin Aztreonam Colistin	Aztreonam Colistin	Aztreonam Colistin	None	None

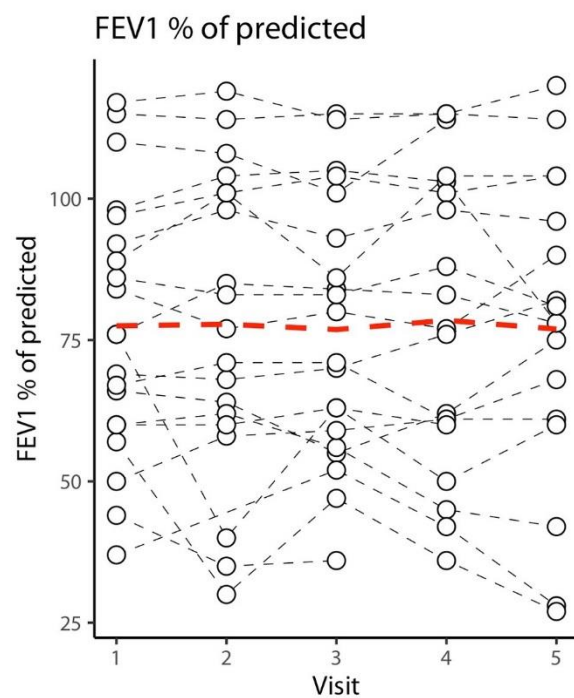
Supplemental figures

Figure E1: Flowchart of patient selection and sample collection during the study.



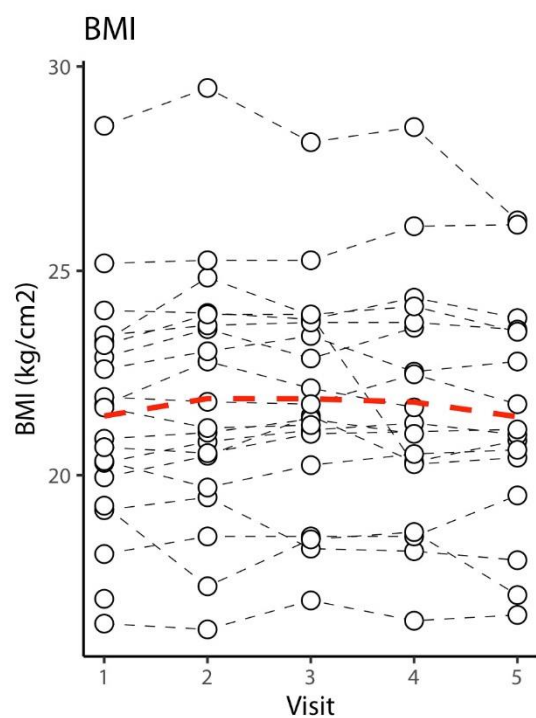
The upper part of this figure shows the number of patients eligible for participation, excluded, and included in the study. The lower part shows a schematic representation of the number of samples that are collected during the study. Each box indicates the number of samples for each of the individual visits, with visit 1 before start of treatment and subsequent visits 3 months apart. One patient stopped due to side effects after 3 months follow-up. Most samples from visit 4 were missing due to sampling error; therefore, this visit was not evaluated in statistical analysis.

Figure E2: Change in predicted FEV1 during 1 year of treatment with lumacaftor/ivacaftor.



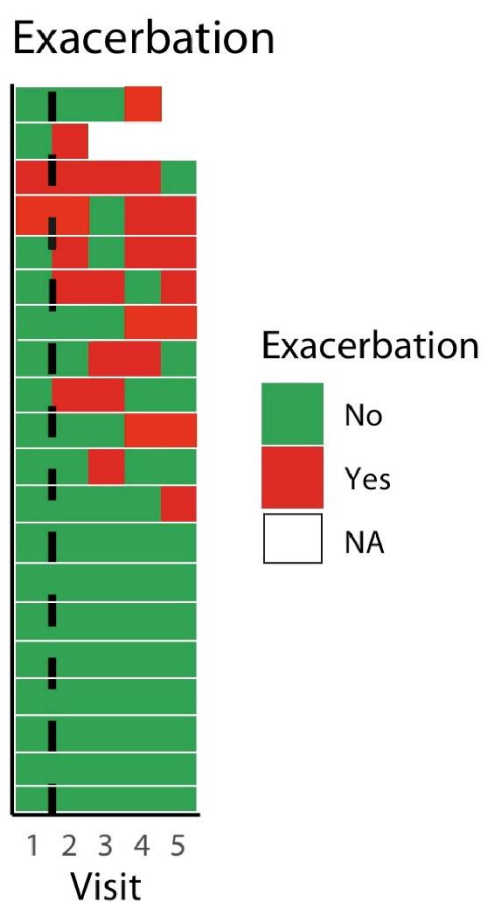
X-axis indicates the visits, with visit 1 before start of treatment and subsequent visits 3 months apart. Y-axis indicates FEV1 % of predicted. The red line indicates the mean value per visit.

Figure E3: Change in BMI during 1 year of treatment with lumacaftor/ivacaftor.



X-axis indicates the visits, with visit 1 before start of treatment and subsequent visits 3 months apart. Y-axis indicates body mass index (BMI) in kilogram/centimeter squared (kg/cm^2). The red line indicates the mean value per visit.

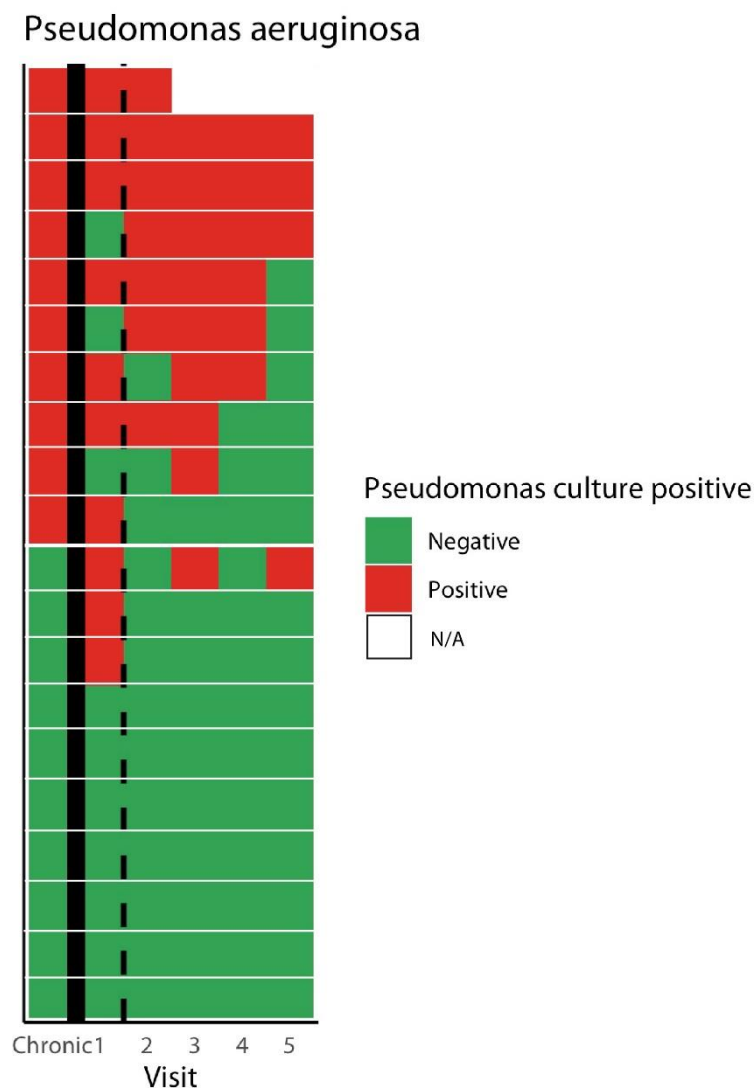
Figure E4: Exacerbations during three months prior to visit



X-axis indicates the visits, with visit 1 before start of treatment and subsequent visits 3 months apart. The red values indicate that a patient had an exacerbation in the three months previous to the visit. The black dashed line indicates the start of treatment. The data is aligned according to exacerbation frequency, with the highest number of exacerbations on top, and the lowest number at the bottom of the figure.

N/A = lost to follow-up.

Figure E5: Change in *Pseudomonas aeruginosa* culture positivity



X-axis indicates the visits, with visit 1 before start of treatment and subsequent visits 3 months apart. The red values indicate that a patient had an culture positive for *Pseudomonas aeruginosa* in the three months previous to the visit. The solid black line indicates the period before the start of the study. The black dashed line indicates the start of treatment. The data is aligned according to culture positivity, with the highest number of positive cultures on top, and the lowest number at the bottom of the figure. N/A = lost to follow-up.

Figure E6: Change in *Staphylococcus aureus* culture positivity



X-axis indicates the visits, with visit 1 before start of treatment and subsequent visits 3 months apart. The red values indicate that a patient had an culture positive for *Staphylococcus aureus* in the three months previous to the visit. The solid black line indicates the period before the start of the study. The black dashed line indicates the start of treatment. The data is aligned according to culture positivity, with the highest number of positive cultures on top, and the lowest number at the bottom of the figure. N/A = lost to follow-up.

