Supplementary material

Eight novel variants in the *SLC34A2* gene in pulmonary alveolar microlithiasis

Åsa Lina M. Jönsson, Elisabeth Bendstrup, Susie Mogensen, Elizabeth J. Kopras, Francis X. McCormack, Ilaria Campo, Francesca Mariani, Amparo Escribano-Montaner, Are M. Holm, Maria del Mar Martinez-Colls, Guillem Pintos-Morell, Camille Taillé, Bruno Crestani, Ole Hilberg, Jane Hvarregaard Christensen*, and Ulf Simonsen*

*Shared senior authorship

Text

Genetic analyses

Eleven pairs of PCR primers (table E1) were designed to amplify the coding exons and the intron-exon boundaries of *SLC34A2*. Exons 2-3 and exons 11-12 were each amplified as part of a single PCR product. Exon 13 was amplified as two overlapping fragments. The amplifications started with an initial step of denaturation of 2 min at 95°C, followed by 35-40 cycles of 12 seconds denaturation at 95°C, 40 seconds of annealing at 55-66°C (specific temperatures are given in table E1), and 30-40 seconds of extension at 70°C, and a final prolonged extension step of 4 min at 72°C. After amplification, the PCR products were run on Agarose TBE gels to verify the specificity of the amplifications. Subsequently, after purification using Jetquick PCR product purification spin kit (Genomed, Löhne, Germany), the PCR products were subjected to bi-directional direct DNA sequencing (Eurofins Genomics, Ebersberg, Germany) using the same primers as used for PCR amplification. Repeated sequence analysis of an independently generated PCR product containing the variant was performed in all patients not previously genetically tested.

Classification of the variants

Variants were scored on the basis of their predicted effect on NaPi-2b function based on the variant type and localisation within the protein; missense or nonsense/frameshift and the localisation in critical domains or areas (figure 4 in the print version). The domains/areas considered critical were the transmembrane domains (TMDs) 3-4 and 8-9 (green color), which form the substrate coordination site. In addition, important areas for electrogenicity, regulation and targeting are located in the area between TMDs 4-5, 10-11, and at the C-terminal region [E1]. Thus, variants causing premature truncation (frameshift and nonsense variants) or having localisation within a key functional domain or area of the predicted protein scored one point each. Missense variants and

small in frame deletions scored zero points. Maximum score was two. Human Genome Variation Society (HGVS) guidelines were used for sequence variant nomenclature [E2].

Six of 10 of the expected disease-causing allelic variants were present in very low frequency in Genome Aggregation Database (GnomAD) [E3], Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP) [E4], 1000Genomes Project [E5], and Database of single nucleotide polymorphisms (dbSNP) [E6] databases. The variants p.Gly106Arg (rs137853142) with an allele count of four (4:246252), p.Trp413Ter (rs766735856) with an allele count of three (3:246254), p.Leu443Ter (rs775513338) with an allele count of one (1:246270), p.? (c.1333+1G>A) with an allele count of one (1:246268), p.Gly464Arg (rs115151720) with an allele count of six (6:246266), and p.Thr468del (rs765046299) with an allele count of one (1:246264) were all present in these public databases of known genetic variation, however all only reported in heterozygous state. The stated allele counts are based on the largest populations investigated.

Tables

Table E1. Primers and conditions used for PCR amplification of *SLC34A2*

Exon	Primer sequences	Annealing temperature (°C)	Predicted length of amplicons (nucleotides)
2+3	Forward: 5′-GAAATCTTTCCCTTCCTTTTACTGC- 3′ Reverse: 5′-GATGATGACACCGTGAACCAC- 3′	66	553
4	Forward: 5′-TCCCAAACGCAAATCCTTTAGA- 3′ Reverse: 5′-CTCTCGGAACCACATTTCAAGATAA- 3′	56	449
5	Forward: 5′-CTGTTTACTCAGTGCCCACCTAATC- 3′ Reverse: 5′-CACCCTCAGATAGACAGGAGAGTGT- 3′	62	281
6	Forward: 5′-GCATAGGTAACTTTAGCCTGC- 3′ Reverse: 5′-CTGAGTCTAATCAGCTGGTGAGTA- 3′	55	363
7	Forward: 5′-GAGGGTGGCAGATGATACAGGTAAT- 3′ Reverse: 5′-CAGGCCCAGAATAGTGTTGTAGACA- 3′	62	391
8	Forward: 5′-GGCTCACAGTCACATTTATCTCCT- 3′ Reverse: 5′-TTCAGGGACTCCCAATATTTCTTT- 3′	55	331
9	Forward: 5′-GTGGTGTCTGCGCCTGTTCA- 3′ Reverse: 5′-TGGCAAAGGAAATGGACTTCAAGTT- 3′	56	458
10	Forward: 5′-AAGGAGGCTGAGAAGGGCTGTC- 3′ Reverse: 5′-ATGCCAAGAACATTTCCAGGTGAAT- 3′	66	491
11+12	Forward: 5′-GAGTGGTTACTTGGGTGGGTTCAC- 3′ Reverse: 5′-TGGCTCTTGTTAGAGACCAGTTTGC- 3′	62	692
13A	Forward: 5′-TGTCAAGAAATAATGGTTGCCACTC- 3′ Reverse: 5′-CAGCCGGTGAACTTGGAGAC- 3′	63	647
13B	Forward: 5′-GCCGTCTTCTACCTGATCATC- 3′ Reverse: 5′-AGCCAAAGGGAATCGAGTTAG- 3′	66	644

Table E2. Clinical disease severity parameters

Parameters	Points
1. FVC (% pred.)	
>70	0
50-70	1
<50	2
2. DLCO (% pred.)	
> 60	0
35-60	1
< 35	2
3. Pulmonary hypertension	
No	0
Yes	1
4. Dyspnoea	
MRC 1	0
MRC 2-3	1
MRC 4-5	2
5. Chest pain	
No	0
Yes	1
6. Fatigue	
No	0
Yes	1
7. Clinical progression	
No	0
Yes	1
8. Limitations of daily activities	
No	0
Yes	1
9. Incapability to work	
No	0
Yes	1

FVC: forced vital capacity; DLCO: carbon monoxide diffusing capacity; % pred.: % of predicted. Clinical progression was defined as progression in symptoms, pulmonary function tests, and/or radiographical appearance. Clinical disease severity score was calculated for each patient based on these nine parameters. The value for each parameter was summed and divided by the total number of parameters.

Table E3. Patients demographics and clinical characteristics

Id.	Country	Consanguinity (Relatives with PAM)	Smoking	Extrapulmonary calcifications (Co-morbidity)	Clubbing	Auscultation (Oxygen saturation)	Blood levels calcium and phosphate
1[*]	USA	Yes, NS (ND)	No	No (MM)	No	Normal (Normal)	Ca: normal Pho: ND
2[*]	SPA	Yes, parents first-degree cousins (No)	No	No (DM type 1)	No	Normal (Normal)	Normal
3[*, ‡]	ITA	ND (Yes, NS)	No	Gallbladder stones (PH)	ND	Crackles (Abnormal)	Ca: normal Pho: ND
4[†]	FRA[§]	No (2 siblings)	No	No (No)	No	Crackles (Normal)	Ca: normal Pho: slightly low
5[*]	ITA	No (ND)	No	No (No)	No	Crackles (Normal)	Ca: slightly low Pho: ND
6[*]	USA	Yes, NS (ND)	ND	ND (ND)	ND	ND (ND)	ND
7[*, ‖]	NOR	No (Sister)	ES (20 PY)	No (PH)	Yes	Crackles (Abnormal)	Normal
8[*]	NOR	No (Sister)	ES (30 PY)	No (No)	Yes	Crackles (Abnormal)	Normal
9[*]	USA	ND (No)	ES (NS)	No (AS, PH, IHD, AA)	No	ND (Normal)	ND
10[*]	SPA	No (Sister)	No	No (No)	No	Normal (Normal)	Normal
11[*]	SPA	No (Brother)	No	No (No)	No	Normal (Normal)	Normal
12[*]	USA	ND (Yes[**])	ES (6 PY)	Kidney stone (PH, MVP, AH)	Yes	Crackles (Abnormal)	Normal
13[*]	DEN	ND (No)	Yes (16 PY)	Kidney- and gallbladder stones, gastric wall calcifications (AS)	No	Crackles (Normal)	Normal
14[*]	DEN	No (No)	Yes (37 PY)	No (IHD)	Yes	Crackles (Normal)	Normal

AA: aortic aneurism; AH: arterial hypertension; AS: aortic valve stenosis; Ca: calcium; DEN: Denmark; DM: diabetes mellitus; ES: ex-smoker; FRA: France; Id.: patient identification; IHD: ischemic heart disease; ITA: Italy; MM: mitochondrial myopathy; MVP: mitral valve prolapse; ND: no data; NOR: Norway; NS: not specified; PH: pulmonary hypertension (based on echocardiography and/or right heart catheterisation); Pho: phosphate; PY: pack years; SPA: Spain. [*]: Caucasian. [†]: Arabic. [‡]: Lost to follow-up. [§]: Originated from Morocco. [ll]: Deceased. [**]: Patient is a part of a cluster of PAM patients in Castellana, Italy.

Table E4. Clinical disease severity score

Patient	Clinical disease severity score	Clinical disease severity*
1	0.11	Mild
2	0.00	Mild
3	1.11	Severe
4	0.44	Mild
5	0.22	Mild
6	NA†	NA†
7	1.11	Severe
8	1.11	Severe
9	0.57	Moderate
10	0.00	Mild
11	0.11	Mild
12	1.22	Severe
13	1.00	Severe
14	0.67	Moderate

NA: not applicable. *: Based on the clinical disease severity score, which was based on a composite measure of the patient´s symptoms, lung function, limitation of daily life due to PAM, signs of advanced disease, and disease course (nine parameters in total). Each parameter was scored 0-1 point or 0-2 points. Max average score 12/9 = 1.33. Mild: 0-0.44, Moderate: 0.45-0.88, Severe: 0.89-1.33. †: Clinical disease severity score was not calculated due to missing data.

Table E5. Correlation between clinical disease severity and smoking in adults and children

Clinical disease severity*	Smoking status†		
	No	**Yes**	**Total**
Mild	6	0	6
Moderate	0	2	2
Severe	1	4	5
Total	7	6	13

*: Clinical disease severity in each patient was categorised as mild, moderate or severe. The grade of severity was stated as a clinical disease severity score, which was based on a composite measure of the patient´s symptoms, lung function, limitation of daily life due to PAM, signs of advanced disease, and disease course (nine parameters in total). †: Smoking status was categorised into two groups; No (never smoker) and Yes (former or current smoker). *Children were included in the calculations.*

H0 = Phenotype and smoking status are independent. Spearman rank correlation coefficient = 0.762, p = 0.003. STATA (version 11.2, StataCorp 2009, College Station, TX, USA).

Table E6. Correlation between clinical disease severity and smoking in adults

Clinical disease severity[*]	Smoking status[†]		Total
	No	Yes	
Mild	3	0	3
Moderate	0	2	2
Severe	1	4	5
Total	4	6	10

[*]: Clinical disease severity in each patient was categorised as mild, moderate or severe. The grade of severity was stated as a clinical disease severity score, which was based on a composite measure of the patient´s symptoms, lung function, limitation of daily life due to PAM, signs of advanced disease, and disease course (nine parameters in total). [†]: Smoking status was categorised into two groups; No (never smoker) and Yes (former or current smoker). *Only adults (>18 years) were included in the calculations.*

H0 = Phenotype and smoking status are independent. Spearman rank correlation coefficient = 0.617, p = 0.057. STATA (version 11.2, StataCorp 2009, College Station, TX, USA).

Table E7. Classification of the variants based on type and localisation

Id.	Nucleotide change	Protein change	Type and localisation scoring	Variant severity*
1	c.316G>A c.1238G>A	p.Gly106Arg p.Trp413Ter	Type: Missense: 0, Localisation: 0 Type: Nonsense: 1, Localisation: 0	Moderate
2	c.560G>A	p.Gly187Glu	Type: Missense: 0, Localisation: 1	Moderate
3	c.646G>T	p.Gly216Ter	Type: Nonsense: 1, Localisation: 1	Severe
4	c.906G>A	p.Trp302Ter	Type: Nonsense: 1, Localisation: 0	Moderate
5	c.1136G>A	p.Cys379Tyr	Type: Missense: 0, Localisation: 0	Mild
6	c.1238G>A	p.Trp413Ter	Type: Nonsense: 1, Localisation: 0	Moderate
7	c.1327delC	p.Leu443Ter	Type: Nonsense: 1, Localisation: 1	Severe
8	c.1327delC	p.Leu443Ter	Type: Nonsense: 1, Localisation: 1	Severe
9	c.1333+1G>A	p.?	Type: *Considered* nonsense: 1, Localisation: 1	Severe
10	c.1390G>C	p.Gly464Arg	Type: Missense: 0, Localisation: 1	Moderate
11	c.1390G>C	p.Gly464Arg	Type: Missense: 0, Localisation: 1	Moderate
12	c.1402_1404delACC	p.Thr468del	Type: *Considered* missense: 0, Localisation: 1	Moderate
13	c.1402_1404delACC	p.Thr468del	Type: *Considered* missense: 0, Localisation: 1	Moderate
14	c.1402_1404delACC	p.Thr468del	Type: *Considered* missense: 0, Localisation: 1	Moderate

Id.: patient identification. *: Variant severity on the basis of the predicted effect on NaPi-2b function based on the variant type and localisation within the protein. Variants causing premature truncation (frameshift/nonsense) or had a presumed crucial localisation within the protein (transmembrane domains (TMDs) 3-4, 8-9 and the area between TMDs 4-5, 10-11) scored one point each. Missense variants and small deletions scored zero point. Mild (0 point), moderate (1 point), and severe (2 points). *SLC34A2* DNA reference sequence: Ensembl Transcript ID ENST00000382051.7 (GRCh38.p12 assembly).

Table E8. Variant properties and *in silico* predictions of pathogenicity

Id.	Nucleotide change	Protein change	Exon	Localisation in protein*	Results *in silico* tools†
1	c.316G>A	p.Gly106Arg	4	TMD 1	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: Probably damaging PROVEAN: Deleterious
2	c.560G>A	p.Gly187Glu	6	1st pore, TMD 4	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: Probably damaging PROVEAN: Deleterious
3	c.646G>T	p.Gly216Ter	7	Critical area for electrogenicity, small IC-Loop between TMDs 4 and 5	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: NA, PROVEAN: NA
4	c.906G>A	p.Trp302Ter	8	Long EC-Loop	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: NA, PROVEAN: NA
5	c.1136G>A	p.Cys379Tyr	10	TMD 6	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: Probably damaging PROVEAN: Deleterious
1, 6	c.1238G>A	p.Trp413Ter	11	TMD 7	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: NA, PROVEAN: NA
7-8	c.1327delC	p.Leu443Ter	11	2nd pore, TMD 8	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: NA, PROVEAN: NA
9	c.1333+1G>A	p.?	Int. 11	2nd pore, TMD 8, donor-splice site immediately after the 3' end of exon 11	Human Splicing Finder: Alteration of the donor site, most probably affecting splicing. MutationTaster: Disease causing
10-11	c.1390G>C	p.Gly464Arg	12	2nd pore, TMD 9	MutationTaster: Disease causing PANTHER: Probably damaging Polyphen-2: Probably damaging PROVEAN: Deleterious

12-14	c.1402_1404delACC	p.Thr468del	12	2nd pore, TMD 9	MutationTaster: Disease causing PANTHER: NA, Polyphen-2: NA PROVEAN: Deleterious

Id.: patient identification; Int.: intron; TMD: transmembrane domain; NA: not applicable. *: A model of NaPi-2b, made by superimposing human NaPi-2b on rat NaPi-2a predicted topology and modified from Forster et al. 2013 (E1) and Virkki et al. 2007 (E7), was used to predict the locations of the variants in the protein (figure 4 in the print version). The protein sequences used for alignment in Clustal Omega version 1.2.4 (E8): Ensembl Transcript ID ENST00000382051.7, release 92 (Human (GRCh38.p12) assembly) and Ensembl Transcript ID ENSRNOT00000033749.5 (Rat (Rnor_6.0) assembly)). †: Human Splicing Finder (E9), Mutation Taster (E10), PANTHER (E11), Polyphen-2 (E12), and PROVEAN (E13). *SLC34A2* DNA reference sequence: Ensembl Transcript ID ENST00000382051.7 (GRCh38.p12 assembly).

Table E9. Correlation between clinical disease severity and variant severity score

Clinical disease severity*	Variant severity score†			
	0	**1**	**2**	**Total**
Mild	1	5	0	6
Moderate	0	1	1	2
Severe	0	2	3	5
Total	1	8	4	13

*: Clinical disease severity in each patient was categorised as mild, moderate or severe. The grade of severity was stated as a clinical disease severity score, which was based on a composite measure of the patient´s symptoms, lung function, limitation of daily life due to PAM, signs of advanced disease, and disease course (nine parameters in total). †: Variant severity score: variants were scored on the basis of their predicted effect on NaPi-2b function based on variant type and localisation within the protein.

H0 = Phenotype and localisation and variant type are independent. Spearman rank correlation coefficient = 0.629, p = 0.021, STATA (version 11.2, StataCorp 2009, College Station, TX, USA).

References

E1. Forster IC, Hernando N, Biber J, Murer H. Phosphate transporters of the SLC20 and SLC34 families. *Mol Aspects Med* 2013: 34(2-3): 386-395.

E2. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PE. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* 2016: 37(6): 564-569.

E3. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016: 536(7616): 285-291.

E4. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [date (03, 2019) accessed].

E5. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015: 526(7571): 68-74.

E6. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001: 29(1): 308-311.

E7. Virkki LV, Biber J, Murer H, Forster IC. Phosphate transporters: a tale of two solute carrier families. *Am J Physiol Renal Physiol* 2007: 293(3): F643-654.

E8. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011: 7: 539.

E9. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009: 37(9): e67.

E10. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014: 11(4): 361-362.

E11. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 2017: 45(D1): D183-D189.

E12. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010: 7(4): 248-249.

E13. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015: 31(16): 2745-2747.