Cystic Fibrosis: A Simple and Customized Strategy for Genetic Screening Able to Detect Over 90% of Identified Mutated Alleles in Brazilian Newborns

Article *in* Molecular Diagnosis & Therapy · March 2020

DOI: 10.1007/s40291-020-00456-9

CITATIONS
0

READS
80

13 authors, including:

Thaiane Rispoli Serrano
Universidade Federal do Rio Grande do Sul
7 PUBLICATIONS **59** CITATIONS
SEE PROFILE

Simone Castro
Universidade Federal do Rio Grande do Sul
48 PUBLICATIONS **523** CITATIONS
SEE PROFILE

Mayara Jorgens Prado
Universidade Federal do Rio Grande do Sul
7 PUBLICATIONS **43** CITATIONS
SEE PROFILE

Gilberto Bueno Fischer
Universidade Federal de Ciências da Saúde de Porto Alegre
103 PUBLICATIONS **1,811** CITATIONS
SEE PROFILE

Molecular Diagnosis & Therapy
https://doi.org/10.1007/s40291-020-00456-9

ORIGINAL RESEARCH ARTICLE

Cystic Fibrosis: A Simple and Customized Strategy for Genetic Screening Able to Detect Over 90% of Identified Mutated Alleles in Brazilian Newborns

Thaiane Rispoli[1,2] · Grazielle M. Rodrigues[2,3] · Simone M. de Castro[3,4] · Mayara J. Prado[1,2] · Cláudia M. D. da Silva[2] · Tarciana Grandi[2] · Gilberto B. Fischer[5] · Leonardo A. Pinto[6] · Paulo J. C. Maróstica[7] · Laís C. R. Scortegagna[8] · Helena T. Mocelin[4] · José Eduardo Vargas[9] · Maria L. R. Rossetti[1,10]

Abstract

Introduction The incorporation of molecular genetic testing into cystic fibrosis (CF) screening programs increases the specificity of the diagnostic strategy and has the potential to decrease the rate of false- positive results. In this sense, our objective was to develop a genotyping assay that could detect 25 pathogenic variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene with high sensitivity and that could be incorporated into the routine of newborn screening, complementing the current existing protocol used in our public health institution.

Methods A mini-sequencing assay was standardized using single-base extension in a previously genotyped control sample. This strategy was validated in a Brazilian cohort of CF patients by Sanger sequencing.

Results The inclusion of the 25 variants in the current newborn screening program increased the identification rates of two alleles from 33 to 52.43% in CF patients. This new approach was able to detect a total of 37 variants, which represents 93.01% of all mutated alleles described in the last CF Brazilian Register.

Conclusions Mini-sequencing for the simultaneous detection of 25 *CFTR* gene variants improves the screening of Brazilian newborns and decreases the number of inconclusive cases. This method uses minimal hands-on time and is suited for rapid screening, which reduces sample processing costs.

Key Points

A customized panel for the most frequent Brazilian cystic fibrosis (CF) variants increased sensitivity of detection in patients via molecular diagnosis in CF newborn screening.

The panel achieved 93% coverage for CF mutated alleles in the Brazilian population.

This is a low-cost approach for the detection of CF mutated alleles, suitable for developing countries.

1 Introduction

Cystic fibrosis (CF) [OMIM: 219700] is one of the most common life-threatening autosomal recessive diseases described in humans and is caused by variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene [1–3]. Currently, more than 2000 variants have been described in the *CFTR* gene, of which 336 are annotated as pathogenic variants (http://www.genet.sickkids.on.ca/ and https://www.cftr2.org/). CF is characterized by pancreatic insufficiency and chronic obstructive airway disease, which promotes bacterial infection; salty sweat is also detected [4–6], and meconium ileus manifests in 20% of children with CF [7].

Several newborn screening (NBS) approaches are used to identify early CF, improving the treatment of diagnosed patients. All Brazilian NBS protocols include the immunoreactive trypsinogen (IRT) assay in the first 5 days of life as the initial step and the second assessment within 30 days

José Eduardo Vargas and Maria L. R. Rossetti contributed equally to this work.

✉ Thaiane Rispoli
thaianerispoli@gmail.com

✉ José Eduardo Vargas
josevargas@upf.br

Extended author information available on the last page of the article

Published online: 17 March 2020

of newborn life [8–10]; finally, the chloride concentration in sweat, considered the gold standard of CF diagnosis, is performed. Sweat chloride concentrations above 60 mmol/L, as measured by quantitative methods, are used to confirm the diagnosis [10–13]. Genetic analysis is also recommended in NBS to corroborate abnormal sweat test results, focusing on two mutated alleles, which can be identified using different approaches. One of these is next-generation sequencing (NGS) of the *CFTR* gene. However, NGS is not a simple solution, because it generates a large amount of data that requires computational infrastructure and expertise in data interpretation. From a cost perspective, this technology is a practically prohibitive strategy for public health system services, especially in developing countries.

Newborn screening for CF variants was implemented in 2012 in the southernmost state of Brazil, Rio Grande do Sul (RS), by the public health system (Sistema Único de Saúde [SUS]). Only the c.1521_1523delCTT (legacy: F508del) variant was screened at that time, due to technical limitations and cost. The need for a more accurate test for CF diagnosis led Brazilian institutions to sign an agreement to increase detection of CF and improve diagnosis through a complementary test to be implemented through SUS.

In this context, in a previous work, our group developed a molecular panel based on the 11 most frequent pathogenic variants in the state (c.1624G > T [legacy: G542X], c.1652G > A [legacy: G551D], c.350G > A [legacy: R117H], c.1000C > T [legacy: R334W], c.1657C > T [legacy: R553X], c.3484C > T [legacy: R1162X], c.3846G > A [legacy: W1282X], c.254G > A [legacy: G85E], c.1585-1G > A [legacy: 1717-1G > A], c.2657 + 5G > A [legacy: 2789 + 5G > A], and c.2988 + 1G > A [legacy: 3120 + 1G > A]) [14]. However, the spectrum of pathogenic CF variants continues to expand, and the frequency of the CF-causing variant is different among countries, a consequence of wide ethnic heterogeneity [5, 15]. Panels for molecular screening of CF, when designed and customized for specific populations, can increase sensitivity and specificity levels and have proven to be cost effective and to improve diagnosis through NBS [14].

In this paper, we designed an additional low-cost molecular panel capable of complementing the current molecular protocol for CF diagnosis in southern Brazil used in the public health system, with potential application in all Brazilian states.

2 Materials and Methods

2.1 Samples and Data Collection

Genomic DNA from a cohort composed of 103 individuals with abnormal sweat tests (values above 60 mmol/L) was extracted from peripheral blood using a salting out protocol [16]. These samples were selected from the Referral Service in Newborn Screening of RS state located in Hospital Materno Infantil Presidente Vargas and from CF reference centers (Hospital São Lucas and Hospital da Criança Santo Antônio/Santa Casa de Misericórdia de Porto Alegre) between June 2016 and June 2018. In addition, whole-*CFTR* sequencing data were collected from medical records, when available, including data from pediatric CF patients of the Hospital de Clínicas de Porto Alegre, RS, from which no blood samples were collected. To validate the genotype concordance and accuracy, 42 control samples were tested. These samples were previously genotyped and kindly provided by the Centers for Disease Control and Prevention (CDC) (USA) and the Molecular Genetics Laboratory of the University Institute of Clinical Research (Montpellier, France).

2.2 Mini-sequencing Assay

2.2.1 Variants Selection, Primer Design, and Multiplex PCR

According to the latest CF Brazilian registry [17] and the data obtained from the review of medical records from RS (Table 1), the 25 most frequent variants in CF patients for the RS state and the country of Brazil were selected to compose this new panel: c.3454G > C (legacy: D1152H), c.1645A > C or c.1647T > G (legacy: S549R [rs121908757 and rs121909005]), c.1519_1521delATC (legacy: I507del), c.3140-26A > G (legacy: 3272-26A > G), c.1682C > A (legacy: A561E), c.948delT (legacy: 1078delT), c.613C > T (legacy: P205S), c.1040G > A (legacy: R347H), c.1040G > C (legacy: R347P), c.3196C > T (legacy: R1066C), c.1766 + 1G > A (legacy: 1898 + 1G > A), c.11C > A (legacy: S4X), c.3276C > A or c.3276C > G (legacy: Y1092X), c.579 + 5G > A (legacy: 711 + 5G > A), c.579 + 1G > T (legacy: 711 + 1G > T), c.617T > G (legacy: L206W), c.2991G > C (legacy: L997F), c.1675G > A (legacy: A559T), c.3909C > G (legacy: N1303K), c.3659delC (legacy: 3791delC), c.2052delA (legacy: 2184delA), c.1397C > A or c.1397C > G (legacy: S466X), c.3528delC (legacy: 3659delC), and c.1680-1G > A (legacy: 1812-1G > A). It is important to note that the 11 variants currently applied in screening routine [14] were excluded. However, all patients used to test this new panel were previously analyzed for the 11 variants and c.1521_1523delCTT (legacy: F508del), according to [14].

The Primer3 v.0.4.0 software (Whitehead Institute, Cambridge, UK; http://bioinfo.ut.ee/primer3-0.4.0/) was used to construct the primers used in the multiplex (Table 2) and mini-sequencing (Table 3) reactions. The IDT OligoAnalyzer (Integrated DNA Technologies, Coralville, Iowa, USA; https://www.idtdna.com/pages) and Multiple Primer

Table 1 Patients variants identified in the medical records review in RS state

Variant cDNA (legacy name)	Reference SNP(rs)[#]	Alleles number	Allele frequency (%)	Variant cDNA name (legacy name)	Reference SNP(rs)[#]	Alleles number	Allele frequency (%)
c.1521_1523delCTT (F508del)[a]	113993960	182	49.19	c.3G > A (M1I)	397508657	1	0.27
c.3484C > T (R1162X)[a]	74767530	21	5.68	Exons deletion 20 and 21	Not found	1	0.27
c.1624G > T (G542X)[a]	113993959	20	5.41	c.274G > T (E92X)	121908751	1	0.27
c.3909C > G (N1303K)[b]	80034486	14	3.78	c.1519_1521delATC (I507del)[b]	121908745	1	0.27
c.2988 + 1G > A (3120 + 1G > A)[a]	75096551	10	2.70	c.3528delC (3659delC)[b]	121908747	1	0.27
c.254G > A (G85E)[a]	75961395	10	2.70	c.1397C > A or c.1397C > G (S466X)[b]	121908805	1	0.27
c.579 + 5G > A (711 + 5G > A)[b]	78440224	6	1.62	Deletion exon 10	Not found	1	0.27
c.579 + 1G > T (711 + 1G > T)[b]	77188391	5	1.35	c.1399C > T (L467F)	1800089	1	0.27
c.3196C > T (R1066C)[b]	78194216	5	1.35	c.349C > T (R117C)	77834169	1	0.27
c.2052delA (2184delA)[b]	121908746	4	1.08	c.3454G > C (D1152H)[b]	75541969	1	0.27
c.2657 + 5G > A (2789 + 5G > A)[a]	80224560	4	1.08	c.3424_3425insAGTA	397508559	1	0.27
c.3302T > A (M1101K)	36210737	3	0.81	c.2012delT (2143delT)	121908812	1	0.27
c.1680-1G > A (1812-1G > A)[b]	121908794	3	0.81	I506del	Not found	1	0.27
c.3276C > A or c.3276C > G (Y1092X)[b]	121908761	3	0.81	Exon deletion 2	Not found	1	0.27
c.1000C > T (R334W)[a]	121909011	3	0.81	c.1585-1G > A (1717-1G > A)[a]	76713772	1	0.27
c.1652G > A (G551D)[a]	75527207	2	0.54	c.1727G > C (G576A)	1800098	1	0.27
c.1657C > T (R553X)[a]	74597325	2	0.54	c.3846G > A (W1282X)[a]	77010898	1	0.27
c.613C > T (P205S)[b]	121908803	2	0.54	c.3659delC (3791delC)[b]	121908811	1	0.27
c.11C > A (S4X)[b]	397508173	2	0.54	c.3197G > A (R1066H)	121909019	1	0.27
c.38C > T (S13F)	397508395	2	0.54	2185insA	Not found	1	0.27
c.1052C > G (T351S)	1800086	2	0.54	c.601G > A (V201M)	138338446	1	0.27
c.2002C > T (R668C)	1800100	2	0.54	c.617T > G (L206W)[b]	121908752	1	0.27
c.2089C > G (p.Arg697Glyfs*25)	759952845	2	0.54	c.2052dupA (2184insA)	121908786	1	0.27
c.3002_3003delTG (3132delTG)	397508477	2	0.54	c.948delT (1078delT)[b]	121908744	1	0.27
c.3140-26A > G (3272-26A > G)[b]	76151804	2	0.54	CFTRdel2,3	Not found	1	0.27
c.3039delC (3171delC)	121908781	2	0.54	Exons deletion 25_27	Not found	1	0.27
c.3889dupT (4016insT)	121908808	1	0.27	c.1841A > G (D614G)	201124247	1	0.27
c.1682C > A (A561E)[b]	121909047	1	0.27	c.3160C > G (H1054D)	397508510	1	0.27
c.3169 A > G (T1057A)	397508511	1	0.27	c.1040G > A (R347H)[b]	77932196	1	0.27

cDNA complementary DNA, *RS* Rio Grande do Sul, *SNP* single-nucleotide polymorphism

[#]Available at http://cftr2.org and at https://www.ncbi.nlm.nih.gov/snp/

[a]Variants included in current RS state molecular screening in public health system. The c.350G > A (legacy: R117H) variant was not identified in any patient in RS until the time of this analysis

[b]Variants included in this study

Analyzer (Thermo Fisher Scientific, Walthman, Massachusetts, USA; https://www.thermofisher.com) were used to analyze possible hairpin and dimer formations between the sequences of the chosen primers. Sequence alignment was tested using the BLAST online tool (US National Library of Medicine, https://www.ncbi.nlm.nih.gov/BLAST/) and CLUSTALW (Wellcome Genome Campus in Hinxton, Cambridge, UK; https://www.ebi.ac.uk/Tools/msa/clustalw2/).

The multiplex polymerase chain reaction (PCR) was first performed in a final volume of 25 µL containing 10–20 ng

of DNA, 2.5 µL of 1× Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 96 nM of each primer (forward and reverse directions) (Table 2), and ultrapure water to make up the final volume. The cycling conditions were initial denaturation at 95 °C for 15 min; 35 cycles of 94 °C for 30 s, 58 °C for 1 min and 30 s, and 72 °C for 1 min; and a final extension at 72 °C for 15 min. The amplicons obtained were separated in a 3% polyacrylamide gel (0.75-mm thick, 8-cm high) (Bio-Rad Laboratories, Hercules, California, USA) at a constant voltage (110 V) for 4 h in 1× Tris–borate-ethylenediaminetetraacetic acid (EDTA) buffer (1×TBE). The gel was stained with ethidium bromide (0.5 µg/mL) in 100 mL of 1× TBE. PCR products were purified with 2 µL ExoSAP-IT (Affymetrix, Santa Clara, California, USA) to 5 µL PCR product and incubated for 1 h at 37 °C followed by 20 min at 80 °C to inactivate the enzyme.

2.2.2 Single-Base Extension

Detection of target nucleotides was performed by the mini-sequencing technique using single-base extension (SBE) with the commercial ABI PRISM® SNaPshot™ Multiplex kit (Life Technologies, Carlsbad, California, USA). This reaction was carried out using 5 µL of the SNaPshot Multiplex Ready reaction mix (Life Technologies, Carlsbad, California, USA), 3 µL of the purified multiplex PCR product, 0.15 µL of the mixture of extension primers (0.05 µM of each primer) (Table 3), and ultrapure water to make up the final volume to 12 µL. Thermal cycling consisted of an initial denaturation at 95 °C for 5 min, followed by 35 cycles at 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s. Sequentially, the SBE reaction products were treated with 1 U of shrimp alkaline phosphatase (SAP) enzyme (Affymetrix, Santa Clara, California, USA) at 37 °C for 60 min and 80 °C for 20 min.

2.2.3 Capillary Electrophoresis

Samples were prepared for capillary electrophoresis by adding 9 µL of a mix containing 8.2 µL Formamide Hi-Di (Life technologies, Carlsbad, California, USA) and 0.8 µL of GeneScan 120 LIZ size standard (Life Technologies, Carlsbad, California, USA) to 1 µL of purified SBE reaction product. The reaction was denatured at 95 °C for 5 min, immediately cooled to 0 °C for 3 min, and then placed in a 3130XL gene analyzer (Applied Biosystems, Foster City, California, USA) using POP-4 polymers and 36-cm long capillaries. Data were analyzed using GeneMapper v4.0 software (Applied Biosystems, Foster City, California, USA) as previously described.

2.3 Sanger Sequencing Validation

Sanger sequencing of PCR products was used to validate the variants found by mini-sequencing in the control samples. Validation was performed experimentally by PCR amplification of each primer pair comprising the region where the variant is located, using 10–20 ng of DNA extracted from peripheral blood. The PCR was prepared by addition of 266 nM of each primer (forward and reverse) (Table 2), 2 mM $MgCl_2$, 3 µL 1×PCR Buffer without Mg, and 2 U Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, California, USA) to ultrapure water in a final reaction volume of 30 µL. The cycling conditions were initial denaturation at 95 °C for 5 min; 35 cycles at 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 45 s, and then a final extension at 72 °C for 7 min. PCR products were purified with ExoSAP-IT Product (Affymetrix, Santa Clara, California, USA) as described above in Sect. 2.2.1. The sequencing reaction was performed in the forward and reverse directions by using the Big Dye Terminator Cycle Sequencing Kit v.3.1 (Applied Biosystems, Foster City, California, USA), according to the manufacturer's instructions, followed by a precipitation protocol with ethanol/EDTA, recommended by the Applied Biosystems chemical guide. The sequencing system used was the ABI 3130xL genetic analyzer (Applied Biosystems, Foster City, California, USA). Sequences obtained were aligned with the gene region where each variant was located and examined using SeqMan v.7.0.0.0 software (DNAstar, Madison, Wisconsin, USA).

2.4 Statistical Analysis

The Cohen Kappa test was used with IBM SPSS statistical software v.2.1 (Westlands Road, Quarry Bay, Hong Kong) to compare the results obtained with mini-sequencing and Sanger sequencing of control samples. The positive predictive values (PPVs) and negative predictive values (NPVs) with 95% confidence interval (CI) were also analyzed considering an approximate sum of 2000 variants. The threshold for statistical significance was $p < 0.05$.

3 Results and Discussion

A limited number of Brazilian states (São Paulo, Minas Gerais, and Parana e Santa Catarina) perform complementary molecular analysis, and this is usually restricted to the c.1521_1523delCTT (legacy: F508del) variant [18, 19]. The remaining Brazilian states have limited financial resources to perform molecular tests for CF. The costs associated with detecting a broader spectrum of pathogenic variants continue

Table 2 Primers used for multiplex PCR amplification

Variant cDNA name (legacy name)	Class	Reference SNP(rs)[#]	Primer name*	Sequence (5′→3′ direction)	Amplicon sizes (bp)
c.3276C > A or c.3276C > G (Y1092X)	I	121908761	Y1092X/3272/R1066C_F	AACTCCCAGTGGTAGCCAAG	490
c.3140-26A > G (3272-26A > G)	V	76151804	Y1092X/3272/R1066C_R	TGGAAATGAAGGTAACAGCAA	
c.3196C > T (R1066C)	I	78194216	.		
c.1645A > C or c.1647T > G (S549R)	II	121908757	S549R/A559T_F	TGGAGATGCAATGTTCAAAA	449
		121909005	S549R/A559T_R	GGCACAGATTCTGAGTAACCA	
c.1675G > A (A559T)	II	75549581	.		
c.579 + 1G > T (711 + 1G > T)	I	77188391	711 + 1G > T/711 + 5G > A_F	TTAAAGCTGTCAAGCCGTGT	414
c.579 + 5G > A (711 + 5G > A)	I	78440224	711 + 1G > T/711 + 5G > A_R	CCCAGAACAAGAATTGCTCA	
c.1682C > A (A561E)	II	121909047	A561E/1898 + 1G > A/1812-1G > A_F	TTTCAGTGAATCGATGTGGTG	374
c.1766 + 1G > A (1898 + 1G > A)	I	121908748	A561E/1898 + 1G > A/1812-1G > A_R	CCCAGTAGGGCAGATCAGATT	
c.1680-1G > A (1812-1G > A)	I	121908794	.		
c.3909C > G (N1303K)	II	80034486	N1303K_F	TGTTCACAAGGGACTCCAAA	328
			N1303K_R	AGCAGCCTTACCTCATCTGC	
c.1519_1521delATC (I507del)	II	121908745	I507del/S466X_F	TGAATCCTGAGCGTGATTTG	302
			I507del/S466X_R	TGGGTAGTGTGAAGGGTTCAT	
c.1397C > A or c.1397C > G (S466X)	I	121908805	.		
c.1040G > A (R347H)	IV	77932196	R347H/R347P/1078delT_F	TCCATTCCAAGATCCCTGAT	281
c.1040G > C (R347P)			R347H/R347P/1078delT_R	GTTTGTACAGCCCAGGGAAA	
c.948delT (1078delT)	I	121908744	.		
c.3659delC (3791delC)	I	121908811	3791delC/3659delC_F	TTCAGATGCGATCTGTGAGC	256
c.3528delC (3659delC)		121908747	3791delC/3659delC_R	ACCCTCTGGCCAGGACTTAT	
c.617T > G (L206W)	?	121908752	L206W/P205S_F	TGCTCAGAACCACGAAGTGT	217
c.613C > T (P205S)	?	121908803	L206W/P205S_R	AGGCAGACGCCTGTAACAAC	
c.2991G > C (L997F)	?	1800111	L997F_F	GACACACTTTGTCCACTTTGC	199
			L997F_R	GAGTTGCTGTGAGGTTTGGA	
c.3454G > C (D1152H)	IV	75541969	D1152H_F	ATTGCAGTGGGCTGTAAACT	189
			D1152H_R	CAGATACACAGTGACCCTCAA	
c.2052delA (2184delA)	I	121908746	2184delA_F	GGAGATGCTCCTGTCTCCTG	161
			2184delA_R	TCTTCGATGCCATTCATTTG	
c.11C > A (S4X)	I	397508173	S4X_F	CAGGTCAGAGAAAAAGGGTTG	146
			S4X_R	GAGACAACGCTGGCCTTT	

bp base pairs, *cDNA* complementary DNA, *PCR* polymerase chain reaction, *SNP* single-nucleotide polymorphism, *?* not yet classified

*Primer sets used in PCR multiplex and sequence analysis

Table 3 Primers used for *CFTR* mini-sequencing assay

Variant cDNA name (legacy name)	Sequence (5′→3′ direction)	Orientation	Primer length (bp)	Allele (wt/v)	Peak size-range set in GeneMapper
c.3528delC (3659delC)	CAACAGAAGGTAAACCTAC	Forward	19	C/A	21.74–24.78
c.1647T > G (S549R)	GGAGGTCAACGAGCAAGAATTTC	Reverse	23	A/C,G	27.00–29.25
c.1397C > A or c.1397C > G (S466X)	ATAATGATGGGTTTTATTTCCAGA CTT	Forward	27	C/A,G,T	29.99–33.38
c.2052delA (2184delA)	ATGCTCCTGTCTCCTGGACAGAAA CAAAAAA	Forward	31	A/C	33.13–35.45
c.3659delC (3791delC)	CCAAATGACTGTCAAAGATCTCAC AGCAAAATACA	Forward	35	C/A	35.38–37.73
c.3909C > G (N1303K)	CCATATTTCTTGATCACTCCACTG TTCATAGGGATCCAA	Forward	39	G/C	40.24–42.82
c.1675G > A (A559T)	TACAGCAAATGCTTGCTAGACCAA TAATTAGTTATTCACCTTG	Reverse	43	C/T	45.08–47.29
c.2991G > C (L997F)	ATGTGAAAATGTTTACTCACCAAC ATGTTTTCTTTGATCTTACAGTT	Forward	47	G/C	48.00–50.37
c.617T > G (L206W)	GCAGACGCCTGTAACAACTCCCAG ATTAGCCCCATGAGGAGTGCCACT TGC	Reverse	51	A/C	51.79–53.90
c.579 + 1G > T (711 + 1G > T)	AAGTATTGGACAACTTGTTAGTCT CCTTTCCAACAACCTGAACAAATT TGATGAA	Forward	55	G/T	55.34–58.43
c.579 + 5G > A (711 + 5G > A)	CCGCCTTTCCAGTTGTATAATTTA TAACAATAGTGCCTAAAAGATTAA ATCAATAGGTA	Reverse	59	C/T	60.01–62.36
c.3276C > A or c.3276C > G (Y1092X)	GCCTTACTTTGAAACTCTGTTCCA CAAAGCTCTGAATTTACATACTGC CAACTGGTTCTTGTA	Forward	63	C/A,G	63.85–66.16
c.11C > A (S4X)	GGCATTAGGAGCTTGAGCCCAGAC GGCCCTAGCAGGGACCCCAGC GCCCGAGAGACCATGCAGAGGT	Forward	67	C/A,T	68.15–70.32
c.1682C > A (A561E)	CTTTTTCTGTTAAAACATCTAGGT ATCCAAAAGGAGAGTCTAATAAAT ACAAATCAGCATCTTTGTATACT	Reverse	71	G/T	69.74–72.30
c.3196C > T (R1066C)	CTATGGAAATATTTCACAGGCAGG AGTCCAATTTTCACTCATCTTGTT ACAAGCTTAAAAGGACTATGG ACACTT	Forward	75	C/A,T	74.43–77.81
c.1040G > A (R347H)/c.1040G > C (R347P)	TCTGTGCTTCCCTATGCACTAATC AAAGGAATCATCCTCCGGAAAATA TTCACCACCATCTCATTCTGCATT GTTCTGC	Forward	79	G/A,C	78.88–81.95
c.613C > T (P205S)	GGGGTGGAAGATACAATGACACCT GTTTTTGCTGTGCTTTTATTTTCC AGGGACTTGCATTGGCACATTTCG TGTGGATCGCT	Forward	83	C/T	85.42–88.14
c.948delT (1078delT)	TTTTTATAGAACAGAACTGAAACT GACTCGGAAGGCAGCCTATGT GAGATACTTCAATAGCTCAGCCTT CTTCTTCTCAGGGTTCTT	Forward	87	T/G	87.68–90.08
c.1645A > C (S549R)	TGACTCTCTAATTTTCTATTTTTG GTAATAGGACATCTCCAAGTTTGC AGAGAAAGACAATATAGTTCTTGG AGAAGGTGGAATCACACTG	Forward	91	A/C	92.09–93.86

Table 3 (continued)

Variant cDNA name (legacy name)	Sequence (5′→3′ direction)	Orientation	Primer length (bp)	Allele (wt/v)	Peak size-range set in GeneMapper
c.3140-26A > G (3272-26A > G)	TTTAACCAATGACATTTGTGATAT GATTATTCTAATTTAGTCTTTTTC AGGTACAAGATATTATGAAATTAC ATTTTGTGTTTATGTTATTTGCA	Forward	95	A/G	95.44–97.43
c.1519_1521delATC (I507del)	GAACTGGAGCCTTCAGAGGGTAAA ATTAAGCACAGTGGAAGAATTTCA TTCTGTTCTCAGTTTTCCTGGATT ATGCCTGGCACCATTAAAGAAAAT ATC	Forward	99	A/T	99.56–101.81
c.1680-1G > A (1812-1G > A)	GACCATATTGTAATGCATGTAGTG AACTGTTTAAGGCAAATCATCTAC ACTAGATGACCAGGAAATAGA GAGGAAATGTAATTTAATTTCCAT TTTCTTTTTA	Forward	103	G/A	103.04–105.41
c.1766 + 1G > A (1898 + 1G > A)	CAATGGTGAACATATTTCTCAAGA GGTAAAATGCAATCTATGATGGGA CAGTCTGTCTTTCTTTTATTTTAG CATGAGCATTATAAGTAAGGTATT CAAAGAACATA	Reverse	107	C/T,G,A	106.30–108.78
c.3454G > C (D1152H)	TCAATATACGGTATATAGTTCTTC CTCATGCTATTACTCATACTTTGT TACTTGTCTGAATTTTTTTCATAA AAGTTAAAAAGATGATAAGACTTA CCAAGCTATCCACAT	Reverse	111	C/G	109.46–111.68

bp base pairs, *cDNA* complementary DNA, *CFTR* cystic fibrosis transmembrane conductance regulator, *wt* wild type, *v* variant

to be a barrier for CF diagnosis in Brazil. The current algorithm in RS state for CF screening of newborns provided by the public health system detects the most frequent pathogenic variant c.1521_1523delCTT (legacy: F508del) and 11 variants on a molecular panel previously employed by our group [14]. In this new approach, we selected 25 additional variants frequently identified in CF patients from Brazil [17] with a high potential for inclusion in the Brazilian public health system (Table 3). Mini-sequencing validation was performed on 42 previously genotyped control samples and was highly accurate ($\kappa = 1.00$; $p < 0.001$), without any non-specific peaks or ambiguous results. The strategy can identify 25 variants in the same reaction and in the same capillary electrophoresis, with high discrimination power. A representative electropherogram of a control DNA sample is shown in Fig. 1a. Similarly, electropherograms of samples with wild-type and mutated alleles and the Sanger sequencing validation are given in Fig. 1b–d, respectively. However, all electropherograms are available on request from the corresponding author.

To reflect the real impact of our approach on diagnosis, we performed a comparative analysis to define coverage thresholds for detection of CF-related variants. In this analysis we compared the two approaches used in routine NBS (classical c.1521_1523delCTT [legacy: F508del] variant and the 11-variant panel [14]) with the complementary 25 variants included in this new panel. A cohort composed of 103 patients with CF was used for sequential validation of these screening tests (Fig. 2a). All these patients were positive to sweat test; of these, the proportion of newborns that were IRT positive is 75.5% (34/45). From the cohort, 14.6% were identified as homozygous for the c.1521_1523delCTT (legacy: F508del) variant and 37.9% as heterozygous with only one allele detected, while 47.5% corresponded to different variants. In the next step, based on the previous results, we used the current panel available from diagnostic services, which can detect 11 variants, aiming to identify unknown alleles. This molecular screening registered 33% of individuals with both mutated alleles identified for one variant, 29.1% heterozygous individuals, and the remaining 37.9% with no alleles identified. Finally, we tested the complementary 25 variants suggested to be included in the previous molecular protocol used in the Brazilian medical institutions from the public health service. Of our tested cohort, 52.43% had variants in both alleles identified, 18.5% were heterozygous, and the remaining 29.1% had unknown variants.

In previous work, the NGS technique has been performed to identify variants throughout the whole *CFTR* gene [20, 21]. In another project from the Brazilian CF Study Group (http://portalgbefc.org.br/), use of patients from the analyzed

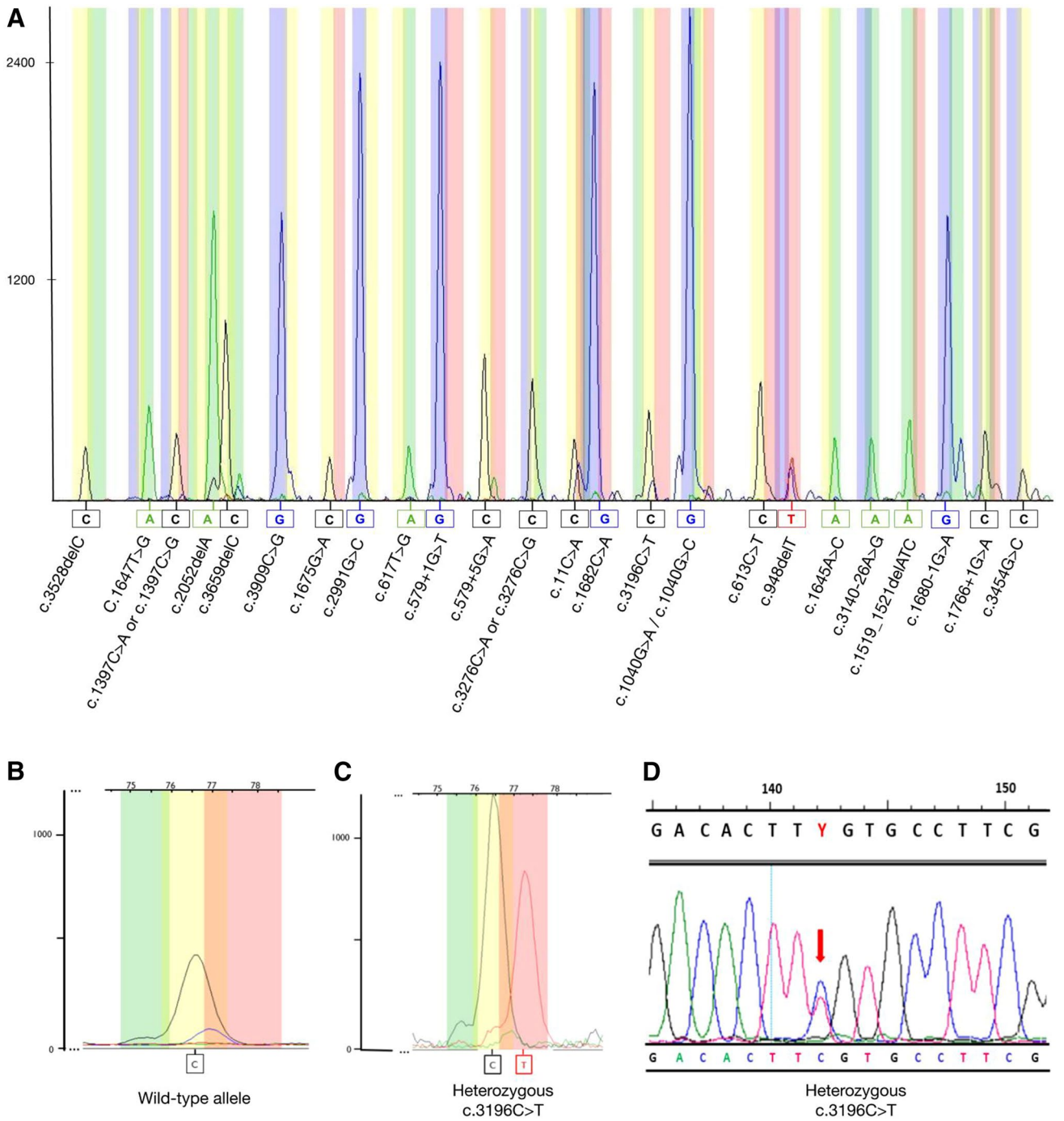

Fig. 1 Electropherogram of a mini-sequencing assay. **a** Electropherograms of SNaPshot for detection of 25 CFTR variants in one sample along with standard size GeneScan 120 LIZ. **b** Electropherograms of SNaPshot shows the wild-type allele C (dark peak), and the example **c** shows mutated allele T (red peak) for c.3196C > T (legacy: R1066C) variants. **d** DNA sequencing of the example **c**: heterozygosity pattern (red arrow), C allele (blue peak), and T allele (red peak). The first sequence is the standard sequence. Y: C or T. CFTR cystic fibrosis transmembrane conductance regulator

cohort is ongoing. In this aforementioned work, a high-throughput approach to the *CFTR* gene was performed by Mendelics® (http://www.mendelics.com/), and the genotyping results from this *CFTR* sequencing were available for our analysis. These subjects were included in our cohort. In this research, 60% of all mutated alleles were identified, and 10.77% showed only one variant. Considering a total of 37 variants screened, it was possible to identify both alleles

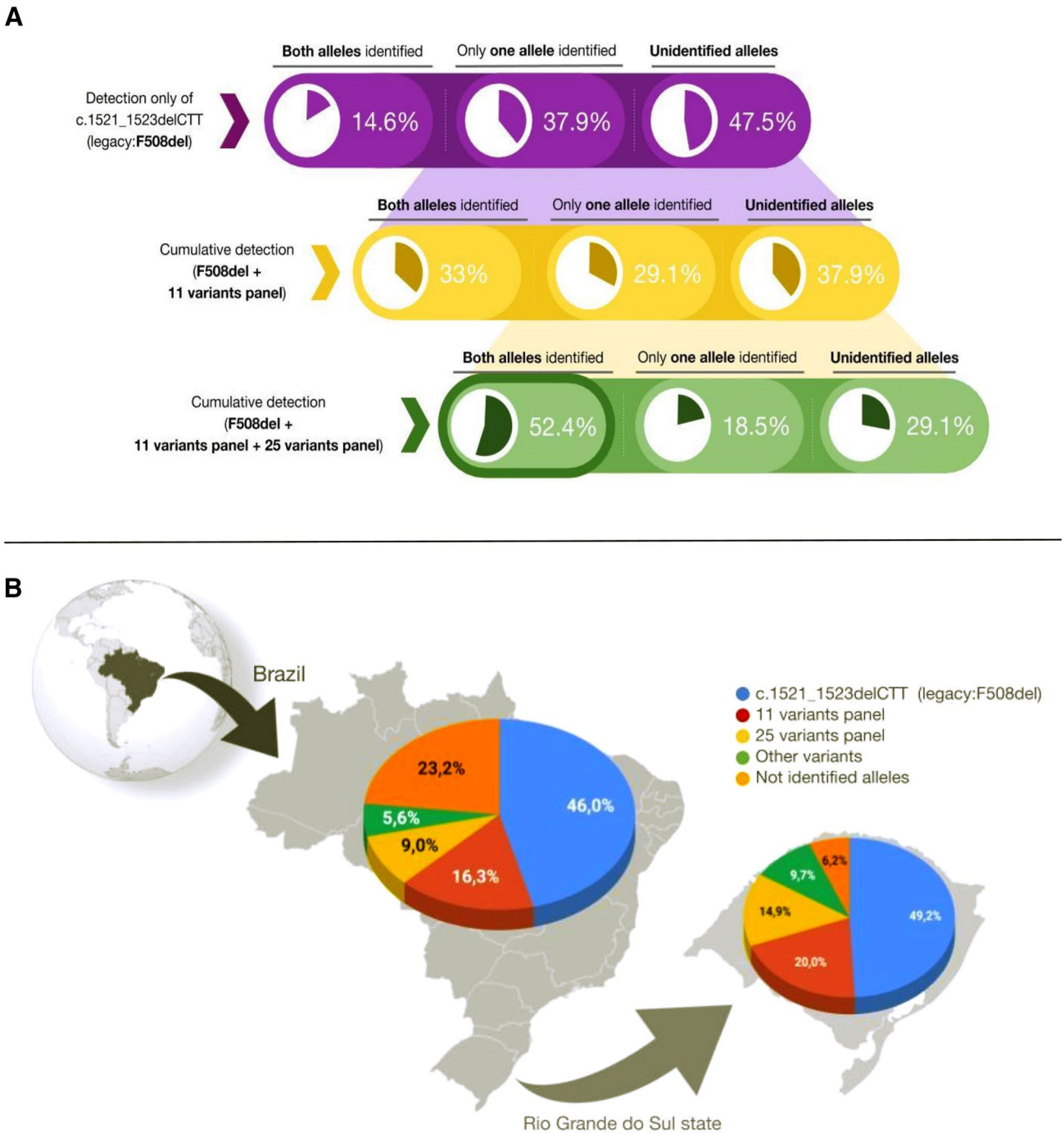

Fig. 2 Comparative analysis of CF-related variant detection and overview of the detectable allelic frequencies present in each approach in RS state and Brazil. **a** CF diagnosis is confirmed when two pathogenic variants are identified. In a cohort of 103 individuals, both alleles were identified in 14.6% when only the c.1521_1523delCTT (legacy: F508del) variant was researched (in purple). Patients with only one identified allele or two unidentified alleles proceeded to the 11-variant research panel, in addition to c.1521_1523delCTT (legacy: F508del) (in yellow). The identification rate of both alleles increased to 33%. Finally, for the remaining patients with no identified alleles, the 25-variant panel was performed. The circled green box shows the increase in the detection rate (52%) of two identified alleles. We suggest that employing the 11-variant panel with this new panel plus c.1521_1523delCTT (legacy: F508del) (totaling 37 variants) will increase the number of variants detected. **b** Summary of Brazilian *CFTR* variants recorded in the Brazilian Registry of Cystic Fibrosis 2016 database, as reported on November 2018 (Brazil) and by medical record review of patients in follow-up in CF reference centers of RS state. The frequencies of mutated alleles were grouped according to the panel in which they are included. This molecular screen of 37 proposed variants comprises 90% and 93% of alleles, with variants identified in the state RS and Brazil, respectively, demonstrating high coverage in the entire country. This percentage was calculated only with the identified alleles (blue, red, yellow, and green slices). *CF* cystic fibrosis, *CFTR* cystic fibrosis transmembrane conductance regulator, *RS* Rio Grande do Sul

in 52.43% of samples, a very similar percentage to the NGS result, using an inexpensive methodology and simpler data analysis. In addition, based on sample analysis, the PPV and the NPV were 100% (95% CI 88.3–100) and 99.1% (95% CI 98.7–99.5), respectively.

Another point to consider is the panel design, which is based on the demand to complement the c.1521_1523delCTT (legacy: F508del) and current molecular protocol, which contains 11 pathogenic variants [14]. For RS, 311 alleles were identified (90%) by the current molecular screening [14], relative to a total number of 347 alleles (Table 1 and Fig. 2b). Similarly, our approach can detect 93% of all known mutated alleles in Brazil (Fig. 2b). This percentage is obtained from the division of 4424 alleles described from the Brazilian registry of CF (http://www.gbefc.org.br) by the total 4756 of all Brazilian mutated alleles (Fig. 2b). It is important to note that all Brazilian mutated alleles were previously analyzed from NGS and commercial diagnostic kits. In this sense, the updated protocol that combines the detection of c.1521_1523delCTT (legacy: F508del) plus 11 variants and 25 new variants can be applied to diagnose mutated alleles with cost–benefit, both regionally and at the national level. Considering the total number of variants, the molecular research for 37 variants in RS has a PPV of 100% (95% CI 90.2–100) and an NPV of 98.5% (95% CI 97.9–99.0). In Brazil, this panel has a PPV of 100% (95% CI 92.2–100) and an NPV of 93.7% (95% CI 92.6–94.7), demonstrating good predictive power of genotyping in both analyses, in addition to good coverage of alleles with identified variants.

We suggest that the mini-sequencing technique for the 37-variant targeted panel can be an alternative method for inclusion in the neonatal screening routines, after IRT assay and sweat test, in emerging countries with limited resources. The technique also delivers high-throughput molecular screening with lower cost, allowing the simultaneous analysis of several single-nucleotide polymorphisms (SNPs) in a short period [22]. Another great advantage of mini-sequencing is the ability to customize the target SNP group, and it is possible to adapt panels according to the frequency of variants in each country, and even in each state. This is a very important factor considering populations with great ethnic heterogeneity such as Brazil, where allelic frequencies may vary between regions of the country [17]. In addition, the high allele frequency of the c.1521_1523delCTT (legacy: F508del) variant (45.97% observed in Brazil [17] and 49.19% observed in RS; Table 1) and lower reagent costs show that molecular testing on this variant must be the first conducted, followed by molecular analysis of other variants (11-variant panel plus 25-variant panel) combined in a 36-variant analysis. These characteristics would help to incorporate the molecular strategy proposed in this work into existing Brazilian public health.

Despite all the advantages that this technique has, it fails to identify large deletions, other types of variants such as insertions, and some overlapping variants. Another limitation is that it is not possible to include all the known variants described within the same mini-sequencing panel.

Although this molecular strategy performs in a specific population, like that of Brazil, it can be adapted to any other population.

Currently, several Food and Drug Administration (FDA)-approved testing platforms are commercially available for molecular testing of the *CFTR* gene. According to the American College of Medical Genetics, a molecular panel must include a minimum of 23 variants. In addition, any testing platform should rely heavily on the testing equipment available and on the ethnic makeup of the patient population that will be screened. In our work, the proposed molecular panel meets these international requirements and is adapted to the economic reality of Brazil; for these reasons, its applicability is real.

4 Conclusion

In this work we employed a panel specially customized for the detection of the most frequent Brazilian CF variants. The inclusion of 25 complementary variants in molecular diagnostics for CF NBS increased the sensitivity of detection in patients with undefined alleles, allowing a wider scope of molecular diagnosis. In this sense, 93.01% of the mutated alleles for CF in the Brazilian population are covered using this new approach. In addition, this test can facilitate the implementation of molecular screening in developing countries, where the availability of financial resource is limited.

Compliance with Ethical Standards

Funding The study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Finance Code 001) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) (1260-2551/13-0).

Ethics approval This study was approved by the ethics committee of the Secretaria da Saúde do Estado do Rio Grande do Sul, Universidade Federal do Rio Grande do Sul, Hospital Materno Infantil Presidente Vargas, Hospital da Criança Santo Antônio, Hospital São Lucas da PUCRS, and Hospital de Clínicas de Porto Alegre.

Patient consent Obtained.

Conflict of interest The authors, Thaiane Rispoli, Grazielle M. Rodrigues, Simone M. de Castro, Mayara J. Prado, Cláudia M.D. da Silva, Tarciana Grandi, Gilberto B Fischer, Leonardo A. Pinto, Paulo J.C.

Maróstica, Laís C.R. Scortegagna, Helena T. Mocelin, José E. Vargas, and Maria L.R. Rossetti, have no conflict of interest to declare.

References

1. O'neal WK, Knowles MR. Cystic fibrosis disease modifiers: complex genetics defines the phenotypic diversity in a monogenic disease. Annu Rev Genom Hum Genet. 2018;19:201–22.
2. Drumm ML, Pope HA, Cliff WH, et al. Correction of the cystic fibrosis defect in vitro by retrovirus-mediated gene transfer. Cell. 1990;62:1227–33.
3. Collins FS, Drumm ML, Cole JL, et al. Construction of a general human chromosome jumping library, with application to cystic fibrosis. Science. 1987;235:1046–9.
4. Castellani C, Assael BM. Cystic fibrosis: a clinical view. Cell Mol Life Sci. 2017;74(1):129–40 **(Epub ahead of print)**.
5. De Boeck K, Vermeulen F, Dupont L. The diagnosis of cystic fibrosis. Presse Med. 2017;46:e97–108.
6. Sosnay PR, White TB, Farrell PM, et al. Diagnosis of cystic fibrosis in nonscreened populations. J Pediatr. 2017;181S:S52–S57.e2.
7. Sathe M, Houwen R. Meconium ileus in cystic fibrosis. J Cyst Fibros. 2017;16(Suppl 2):S32–9.
8. Castellani C, Massie J, Sontag M, et al. Newborn screening for cystic fibrosis. Lancet Respir Med. 2016;4:653–61.
9. Rodrigues R, Magalhaes PKR, Fernandes MIM, et al. Neonatal screening for cystic fibrosis in São Paulo State, Brazil: a pilot study. 2009. www.bjournal.com.br. Accessed 2 June 2019.
10. Athanazio RA, da Silva Filho LVRF, Vergara AA, et al. Brazilian guidelines for the diagnosis and treatment of cystic fibrosis. J Bras Pneumol. 2017;43:219–45.
11. Servidoni MF, Gomez CCS, Marson FAL, et al. Sweat test and cystic fibrosis: overview of test performance at public and private centers in the state of São Paulo, Brazil. J Bras Pneumol. 2017;43:121–8.
12. Taylor CJ, Hardcastle J, Southern KW. Physiological measurements confirming the diagnosis of cystic fibrosis: the sweat test and measurements of transepithelial potential difference. Paediatr Respir Rev. 2009;10:220–6.
13. Farrell PM, Rosenstein BJ, White TB, et al. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. J Pediatr. 2008;153:S4–14.
14. Rispoli T, Martins de Castro S, Grandi T, et al. A low-cost and simple genetic screening for cystic fibrosis provided by the Brazilian public health system. J Pediatr. 2018;199:272–7.
15. Faucz FR, Souza DAS, Olandoski M, et al. CFTR allelic heterogeneity in Brazil: historical and geographical perspectives and implications for screening and counseling for cystic fibrosis in this country. J Hum Genet. 2010;55:71–6.
16. Mousquer GT, Maciel LP, Pompeu Saraiva AC, et al. Validation of a quick and low-cost DNA extraction protocol applicable to long-stored blood samples. Anal Biochem. 2018;561–562:47–51.
17. Grupo Brasileiro de Estudos de Fibrose Cística, The Brazilian Cystic Fibrosis Patient Registry, 2016. http://www.gbefc.org.br. Accessed 2 June 2019.
18. Rodrigues R, Magalhaes PKR, Fernandes CS, et al. Neonatal screening for cystic fibrosis in São Paulo State, Brazil: a pilot study. Braz J Medical Res. 2009;42(10):973–8.
19. Sih T, Godinho R, Franco LP, et al. Cystic fibrosis: Brazilian ENT experience. Int J Otolaryngol. 2012;2012:204696.
20. Nunes LM, Ribeiro R, Niewiadonski VDT, et al. A new insight into CFTR allele frequency in Brazil through next generation sequencing. Pediatr Pulmonol. 2017;52:1300–5.
21. Salinas DB, Sosnay PR, Azen C, et al. Benign and deleterious cystic fibrosis transmembrane conductance regulator mutations identified by sequencing in positive cystic fibrosis newborn screen children from California. PLoS One. 2016;11:e0155624.
22. Latini FRM, Gazito D, Arnoni CP, et al. A new strategy to identify rare blood donors: single polymerase chain reaction multiplex SNaPshot reaction for detection of 16 blood group alleles. Blood Transfus. 2014;12(Suppl 1):S256–63.

Affiliations

Thaiane Rispoli[1,2] · Grazielle M. Rodrigues[2,3] · Simone M. de Castro[3,4] · Mayara J. Prado[1,2] · Cláudia M. D. da Silva[2] · Tarciana Grandi[2] · Gilberto B. Fischer[5] · Leonardo A. Pinto[6] · Paulo J. C. Maróstica[7] · Laís C. R. Scortegagna[8] · Helena T. Mocelin[4] · José Eduardo Vargas[9] · Maria L. R. Rossetti[1,10]

1 Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

2 Centro de Desenvolvimento Científico e Tecnológico, Secretaria da Saúde do Estado do Rio Grande do Sul (SES-RS), Porto Alegre, Rio Grande do Sul, Brazil

3 Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

4 Serviço de Referência em Triagem Neonatal, Hospital Materno Infantil Presidente Vargas, Porto Alegre, Rio Grande do Sul, Brazil

5 Hospital da Criança Santo Antônio, Irmandade da Santa Casa de Misericórdia de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil

6 Hospital São Lucas da PUCRS, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

7 Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil

8 Programa de Pós-Graduação em Ciências Pneumológicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

9 Instituto de Ciências Biológicas, Universidade de Passo Fundo, Passo Fundo, Rio Grande do Sul, Brazil

10 Programa de Pós-Graduação em Biologia Molecular Aplicada a Saúde, Canoas, Universidade Luterana do Brazil, Canoas, Rio Grande do Sul, Brazil