Open

Targeted mutation screening panels expose systematic population bias in detection of cystic fibrosis risk

Regine M. Lim, MS¹, Ari J. Silver, BS^{1,2}, Maxwell J. Silver, BA¹, Carlos Borroto, MS¹, Brett Spurrier, PhD¹, Tanya C. Petrossian, PhD¹, Jessica L. Larson, PhD² and Lee M. Silver, PhD¹⁻⁴

Purpose: Carrier screening for mutations contributing to cystic fibrosis (CF) is typically accomplished with panels composed of variants that are clinically validated primarily in patients of European descent. This approach has created a static genetic and phenotypic profile for CF. An opportunity now exists to reevaluate the disease profile of *CFTR* at a global population level.

Methods: *CFTR* allele and genotype frequencies were obtained from a nonpatient cohort with more than 60,000 unrelated personal genomes collected by the Exome Aggregation Consortium. Likely disease-contributing mutations were identified with the use of public database annotations and computational tools.

Results: We identified 131 previously described and likely pathogenic variants and another 210 untested variants with a high probability of

causing protein damage. None of the current genetic screening panels or existing *CFTR* mutation databases covered a majority of deleterious variants in any geographical population outside of Europe.

Conclusions: Both clinical annotation and mutation coverage by commercially available targeted screening panels for CF are strongly biased toward detection of reproductive risk in persons of European descent. South and East Asian populations are severely underrepresented, in part because of a definition of disease that preferences the phenotype associated with European-typical *CFTR* alleles.

Genet Med advance online publication 16 April 2015

Key Words: carrier screening; cystic fibrosis; exome sequencing; ExAC; genetic testing; next-generation sequencing

INTRODUCTION

Cystic fibrosis (CF) (OMIM # 219700) is the most common life-limiting autosomal recessive disorder in people of European ancestry and is one of the most extensively studied diseases at the molecular level. CF symptoms occur as a consequence of homozygosity or compound heterozygosity for mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Deleterious variants in *CFTR* can disrupt function in multiple organ systems, including the lungs, intestines, male reproductive system, exocrine pancreas, and sweat glands.¹

Classic symptoms of CF include pancreatic insufficiency, congenital absence of the vas deferens in males, and progressive lung disease, which is the major cause of morbidity and mortality. However, the severity of symptoms can vary widely among different people and between organ systems in any one individual. As a means for reducing this complexity into positive and negative diagnostic categories, it is customary to establish a diagnosis of CF with the presence of one or more phenotypic symptoms and a positive sweat chloride test of 60 mmol/l or more.² A chloride concentration of less than 40 mmol/L is interpreted as a negative result. An alternative method by which a CF diagnosis can be established is with the presence of one or more phenotypic symptoms and the detection of two *CFTR* disease-contributing allelic variants.

In the absence of a positive CF diagnosis, a range of organ-limited conditions can also be associated with *CFTR* variants. These “non-classic” or “*CFTR*-related” diseases include isolated or combined symptoms of congenital absence of the vas deferens, idiopathic pancreatitis, bronchiectasis, allergic bronchopulmonary aspergillosis, and chronic rhinosinusitis.³⁻⁵ Classic CF and various *CFTR*-related diseases lie on a multi-dimensional phenotypic continuum that is unified through the primary causal agent of various *CFTR* genotypes.¹ Thus, at a higher level of biological understanding, diagnostic thresholds for distinguishing the two disease categories become arbitrary.

The American College of Medical Genetics and Genomics (ACMG) recommends population-wide screening for carriers of CF-associated mutations. For this purpose, a core panel of 23 variants has been assembled based on strict selection criteria of classic disease presentation and a CF patient chromosome frequency of more than 0.1.^{6,7} The ACMG also recommends that mutations selected for inclusion on any DNA-based panel should be representative of the existing “pan-ethnic United States population.”⁷ Since the establishment of the core ACMG panel, numerous commercial laboratories have produced expanded genetic panels of 90 or more *CFTR* variants, including those designed by the genetic testing companies Integrated Genetics, Counsyl, and Recombine.^{8,9}

The first two authors contributed equally to this work.

¹GenePeaks, Inc., New York, New York, USA; ²GenePeaks, Inc., Cambridge, Massachusetts, USA; ³Department of Molecular Biology, Princeton University, Princeton, New Jersey, USA; ⁴Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey, USA. Correspondence: Jessica L. Larson (jlarson@genepeaks.com)

Submitted 24 November 2014; accepted 4 March 2015; advance online publication 16 April 2015. doi:10.1038/gim.2015.52

Although DNA genotyping can be extraordinarily accurate from an analytical perspective, it is not sufficient by itself for determination of downstream effects of gene function and prediction of phenotypes in hypothesized homozygous or compound heterozygous individuals (see Cooper et al. for a review¹⁰). This is especially true for diseases with continuous phenotypes, like CF.

In recognizing the need for more functional and molecular data to inform variant disease contribution, a large systematic study of *CFTR* DNA sequences was recently reported by Sosnay and colleagues.¹¹ They analyzed genotypic, functional, and phenotypic data gathered from nearly 40,000 CF patients in the “CFTR2” database. Sosnay et al.¹¹ characterized 159 *CFTR* variants as CF-causing, “indeterminate,” or non-CF-causing. Together, these variants account for 96.4% of the *CFTR* mutation load in the CFTR2 database.

The CFTR2-curated set of 159 variants has been presented as the basis for a highly sensitive targeted mutation carrier screen.¹² The expected utility is dependent on two premises. First, it assumes that the CFTR2 database, which is 95% Caucasian, is representative of the global mix of genomes that now constitute the US population. Second, it assumes that carrier frequencies in a nonpatient population can be inferred from variant frequencies within a CF patient sample. This inference is problematic if variants associate with differently diagnosed effects in the context of homozygous and compound heterozygous genotypes. Ascertainment bias could result in overrepresentation of some mutations while leading to the dramatic underrepresentation of others.³

An opportunity to evaluate the global detection and representation issues came about with the completion and public archiving of the Exome Aggregation Consortium (ExAC), a collection of whole exome sequencing data from more than 60,000 individuals representing a diverse sample of human populations.¹³ In this article, we present an extensive analysis of likely disease-contributing *CFTR* variants in the ExAC data set. We have found evidence of strong superpopulation stratification of allele spectra and distinguish a low-complexity, high-frequency group of established disease-contributing variants that appear in homozygous adults without evidence of pediatric disease. Our results call into question the sensitivity and usefulness of targeted mutation carrier screens in reducing the reproductive risk for children being born with diseases caused by various *CFTR* genotypes. We do *not* believe that the disease-contributing variants we found should be added to an enhanced screening panel. We also do *not* believe that we have presented a preferred method for carrier detection. Rather, we assert that the traditional carrier screening framework needs to be updated based on the strengths of more modern tools.

MATERIALS AND METHODS

ExAC data acquisition and sequencing intervals

We recovered *CFTR* variant allele and genotype frequencies from the 60,706 sequenced exomes that were consolidated and processed by the Exome Aggregation Consortium (ExAC) and

made available through the Consortium’s website (version 0.3, <http://exac.broadinstitute.org>). ExAC subjects represent individual participants in several large-scale, disease-specific, and population genetic studies. Persons affected by severe pediatric disease were excluded from participation.

The ExAC cohort is composed of unrelated individuals from five geographically defined and genetically distinguishable superpopulations: Europe (non-Finnish), Finland, South Asia, East Asia, and Africa. A sixth superpopulation is composed of admixed individuals from South, Central, and North America; we call this group “the Americas.” An additional 454 subjects could not be assigned to one of these superpopulations; their results are included only in the global summary statistics presented in this report.

Regions of *CFTR* included in Illumina’s TruSight One Sequencing Panel defined the scope of our inquiry.

Clinically annotated variants

Annotation data were consolidated from two data sets in the ClinVar database made available at the National Center for Biotechnology Information website (<http://www.ncbi.nlm.nih.gov/clinvar>, accessed on 10 January 2015).¹⁴ The two data files—*clinvar.vcf* and *variant_summary.txt*—contain partially overlapping sets of variant records with clinical assertions of pathogenicity. All variants in *CFTR* that are reported as “pathogenic” or “likely pathogenic” in at least one ClinVar record (as of the date of accession) are included in a “Clinically Annotated As Pathogenic” (CAAP) variant list. If there were any “pathogenic” ClinVar variants within 50 basepairs from the edge of a TruSight One *CFTR* interval, then we extended the corresponding interval an additional 10 basepairs beyond that variant’s position.

CAAP variants were separated into two groups based on evidence for the absence of disease causation by at least some homozygous genotype. Variants were placed into group 1 if they were observed only in heterozygous ExAC genotypes (**Supplementary Table S1** online). Group 2 variants were observed in a homozygous state in at least one person (**Supplementary Table S2** online).

Gene-damage likelihoods

We used the HumDiv-trained PolyPhen-2 variant scoring program to compute the likelihoods of gene damage for each *CFTR* missense variant uncovered in our data sets. PolyPhen-2 operates exclusively on missense mutations with an algorithm that utilizes protein structural data and comparative evolutionary considerations to determine the likelihood of damage.^{15,16} PolyPhen-2 results are reported in the form of damage likelihoods on a scale of 0.0 to 1.0, with a score of at least 0.85 considered “probably damaging.”

Additionally, we tested each novel *CFTR* variant with PROVEAN, a second independently developed protein damage-assessment program.¹⁷ PROVEAN scores each variant based on cross-species multi-sequence alignment and evolutionary conservation at the polypeptide level by encompassing the variant’s phylogenetic origin. On their website, the

PROVEAN authors classify any variant with a score less than or equal to -2.5 as “deleterious.”

A third group of likely disease-contributing variants was built with the following four criteria: not previously reported as tested in the ClinVar database; the absence of observed homozygotes in the ExAC cohort; the absence of nondisease status in the CFTR2 data set; and either missense with a PolyPhen-2 likelihood score of at least 0.85 and a PROVEAN score less than or equal to -2.5 or a premature termination signal or functional change (nonsense, frameshift indels, and splice site–altering variants).¹¹

Cystic fibrosis mutation screening panels

Carrier screening was performed with four targeted mutation panels: the ACMG-recommended core panel with 23 mutations,⁷ the Integrated Genetics CFplus panel with 92 *CFTR* mutations,⁹ the Counsyl Family Prep Screen 1.0 panel with 103 *CFTR* mutations,⁸ and the Recombine CarrierMap panel with 108 *CFTR* mutations (<http://www.recombine.com>, accessed on 20 January 2015). In addition, mock screening was performed with the hypothetical targeted mutation panel representing the 159 variants characterized by Sosnay et al.¹¹ in genomes from the CFTR2 patient cohort.

The variant content of each panel, in HGVS format,^{18,19} was retrieved from available online resources. The identity of each individual variant was validated through mapping onto a table of ClinVar annotated loci. In instances of mapping failure, manual curation was used to identify the panel’s most likely intended variant. All variants were indexed on the human genome hg19 reference sequence to enable ease of comparison.

Figure 1 Predicted cystic fibrosis (CF)-contributing carrier detection rates based on group 1 variant detection potential from six sources: Group 1, American College of Medical Genetics and Genomics (ACMG), Counsyl, Integrated Genetics, Recombine, and CFTR2 stratified by superpopulations. Most platforms are able to capture the majority of previously published pathogenic variants for the European populations, but they perform poorly in other populations, especially East Asia. For each population, the systematic blue line at 100% represents the potential maximum coverage of any platform based on all group 1 carriers; the green line is the corresponding minimum based on the core ACMG panel. The values between the blue line and 100% represent the group 3 variants, which correspond to the many previously undescribed but likely disease-contributing variants.

Population-specific carrier detection rates were computed relative to CAAP group 1 variant carriers (Figure 1) or the total number of group 1 and 3 variant carriers (Figure 2).

RESULTS

Identification of 341 likely disease-contributing *CFTR* variants in the ExAC data set

We obtained population-specific *CFTR* variant information from 121,412 independent chromosome 7 in the newly released ExAC data set, which includes approximately 2,500 samples from phase III of the 1000 Genomes Project (Table 1). Each individual is assigned to one of five continental populations or the isolated population of Finland, with an overrepresentation of non-Finnish Europeans. Data from Finland were processed separately by ExAC because of the genetic uniqueness of its population, which is likely a result of its geographical isolation.^{20,21}

The ExAC database contains a total of 1,135 distinct *CFTR* variants within our intervals, including 990 in coding regions and 25 at intronic positions known to affect splicing. Within this subset, we identified 131 variants that have been designated “pathogenic” or “likely pathogenic” in at least one report accepted by the curators of the ClinVar database (Supplementary Tables S1 and S2 online).¹⁴

ClinVar-curated variants are separated further into two groups based on annotations and ExAC allele and genotype frequencies as described below. Finally, we identified a third group of 210 variants that have not been clinically annotated but are likely to be disease-contributing based on the properties of the

Figure 2 Predicted cystic fibrosis (CF)-contributing carrier detection rates based on combined group 1 and group 3 variant detection potential from six sources: Group 1, American College of Medical Genetics and Genomics (ACMG), Counsyl, Integrated Genetics, Recombine, and CFTR2 stratified by superpopulations. Most platforms are able to capture the majority of pathogenic variants on our combined list for the European populations, but they perform poorly in other populations, especially East Asia. As in Figure 1, for each population, the blue line represents the potential maximum coverage of any platform based on all group 1 carriers; the green line is the corresponding minimum based on the core ACMG panel. The values between the blue line and 100% represent the group 3 variants, which correspond to the many previously undescribed but likely disease-contributing variants.

predicted gene product or analysis by computational models of protein damage (**Supplementary Table S3** online).

The 329 variants in groups 1 and 3 are likely to be disease-contributing in the homozygous case and result in some form of *CFTR* disease. Group membership is data-driven, and any variant can change groups based on additional evidence.

Group 1 *CFTR* variants are classic Mendelian recessive mutations

Group 1 includes variants that are denoted as CAAP. This group's 119 variants have additional annotations indicating CF disease and no evidence contrary to disease contribution (**Supplementary Table S1** online). The summed CAAP allele frequency ranges from a high of 2.3% in non-Finnish Europe, with intermediate values of 1.4 and 1.6% in Africa and the Americas and low values between 0.4 and 0.7% in Finland, South Asia, and East Asia (**Tables 1** and **2**). The relative detection of CAAP variants by targeted mutation screening panels does not correlate with the population-specific mutational load (**Figure 1**). Consistent with the very low traditional "CF disease" incidence, the group 1 variants have a very low frequency in South and East Asia.

Group 2 *CFTR* variants display disease-contributing potential in compound heterozygotes

Twelve *CFTR* variants are denoted as disease-contributing in at least one curated ClinVar report but are homozygous in at least one adult member of the ExAC data set or phase III of the 1000 Genomes Project. As expected, these variants display relatively high frequencies and are predominantly confined to a single superpopulation. Because the absence of pediatric disease was a requirement for ExAC inclusion, group 2 variants are not fully penetrant CF mutations. However, group 2 variants can cause *CFTR*-related symptoms when present *in trans* in a group 1 variant in a compound heterozygous genotype.^{4,5,22–26} The partial damaging effects of group 2 variants on the *CFTR* gene product have been demonstrated in functional and biochemical analyses.

Group 3 *CFTR* variants have not been reported in disease genotypes but are likely deleterious

The remaining protein-coding variants in the ExAC data set were analyzed for likelihood of a deleterious effect on gene function or protein damage.²⁶ Included in this group are 11 nonsense, 26 frameshift, and 14 invariant splice-site variants. In addition, we tested each missense variant with both the PolyPhen-2 and PROVEAN tools. To derive a "protein damage" likelihood, PolyPhen-2 makes use of both structural and evolutionary parameters, whereas PROVEAN utilizes cross-species multi-sequence alignment. Our criteria for inclusion in group 3 were stringent for missense variants and required both a PolyPhen-2 "likely damaging" score and a PROVEAN "deleterious" score according to the respective authors' categorizations. One hundred fifty-nine missense variants that met these criteria were also added to group 3. Group 3 variants were individually rare and randomly distributed across all genomes (unpublished data). With expanded genetic testing of patients with all expressions of *CFTR*-related disease, we anticipate that some group 3 variants will be recategorized into group 1.

Sensitivity of simulated carrier screening

With the ExAC cohort as a proxy for the increasingly pan-ethnic American population, we sought to ascertain the efficacy of *CFTR* carrier screening with the use of tests that are based on defined panels of targeted mutations. Toward this goal, we performed mock screening with three commercial products and the CFTR2 variant set. We compared the outcomes with a minimal test represented by the ACMG recommended mutation panel and a maximum possible clinically validated test represented by the complete set of group 1 CAAP mutations.

Commercial carrier tests perform at their highest level in European populations with average sensitivities of 83% (compared with total group 1 mutations) in non-Finnish Europe and 97% in Finland (**Figure 1**). Test results were least consistent and showed generally poor coverage in East Asia (from 3 to 52%).

Table 1 *CFTR* pathogenic variant counts by group and population

	Europe (non-Finnish)	Finland	South Asia	Americas	Africa	East Asia	Global
Maximum number of chromosomes examined	66,740	6,614	16,512	11,578	10,406	8,654	121,412
Group 1 chromosomes	1,503	39	109	182	139	31	2,015
Group 2 chromosomes	4,287	456	500	378	765	624	7,068
Group 3 chromosomes	362	5	88	57	40	78	636
Group 1 variants	89	9	20	29	32	9	119
Group 2 variants	11	5	9	11	8	5	12
Group 3 variants	105	4	45	38	22	21	210
Total dysfunctional variants (groups 1 + 3)	194	13	65	67	54	30	329

Table 2 *CFTR* population frequencies of allele groups

	Europe (non-Finnish)	Finland	South Asia	Americas	Africa	East Asia	Global
Group 1	2.3%	0.6%	0.7%	1.6%	1.4%	0.4%	1.7%
Group 2	7.1%	7.2%	3.1%	3.6%	9.0%	7.3%	6.3%
Group 3	0.6%	0.1%	0.5%	0.6%	0.4%	0.9%	0.5%
Groups 1 and 3	2.8%	0.7%	1.2%	2.2%	1.8%	1.3%	2.2%

When carrier detection rates are evaluated in the context of all likely disease-contributing variants, the results are more striking (Figure 2). All tested commercial panels displayed detection rates less than 15% in East Asia and 33% or less in South Asia. Even in non-Finnish Europe, detection of carriers is less than 67%.

DISCUSSION

Systematic bias present in current CF carrier screening techniques

An analysis of the *CFTR* alleles in more than 60,000 unaffected individuals sampled from high-level human population groups allowed us to distinguish three temporally defined groups of likely disease-contributing variants: (1) clinically validated *CFTR* mutations that are not observed in homozygosity in our cohort; (2) clinically validated mutations where homozygosity is not necessarily associated with disease, and the most severe disease expression occurs in compound heterozygotes with mutations from the first group; and (3) previously uncharacterized, rare mutations that induce premature protein truncation or alternative splicing, or that are computationally modeled to be deleterious.

Population stratification of these *CFTR* alleles is striking. Previously described variants common to European populations are well covered by current tests. However, the commercially available genotyping platforms we explored, by design, do not recognize likely disease-contributing variants unless they have been reported in patients diagnosed with classic CF. These screening panels are particularly ineffective in detecting reproductive risk in our sampling of East Asian carriers. Thus, targeted mutation panels are likely to perform poorly in an increasingly pan-ethnic American population. Others have drawn attention to the low clinical utility of current targeted mutation panels outside of European populations, but, to our knowledge, this report is the first systematic, quantitative demonstration of population bias with a large, genetically diverse data set.^{5,27–29}

Our classification system and method of discovery have general utility for other recessive disease genes. The presence of an analytical bias in identifying *CFTR* mutations on carrier screening panels calls into question whether this bias is present among all disease-associated genes tested by current platforms, especially those not as well annotated as *CFTR*. Bias may exist with the selection of “relevant” genes for a given disease, as well as with the detection of particular damaging variants within the selected genes.

The method of discovery informing our group 3 variants demonstrates the potential of applying computational tools to undefined variants as a means for detecting reproductive risk in the absence of a clinically affected individual. By using a computational system trained on clinically validated variants, we take advantage of previously demonstrated clinical findings to assess the likelihood that any newly discovered variant will have similar disease-contributing effects.

Although clinical correlation is valuable in informing variant behavior *in vivo*, it is not always available. Approaching variant

discovery from a nondiseased pool of diverse genotypes facilitates a new way of thinking about variant behavior that is unrestricted by the limitations of CF disease identification.

Need for a shift in CF disease identification

The results presented here are consistent with the historical context of CF biomedical research. CF was recognized as a disease in 1938, long before the *CFTR* gene was discovered, and most CF studies have been conducted primarily in patients of European descent. As a consequence, a specific European allele spectrum is used to define “classic CF.” Based on the static European definition of classic CF, the disease was thought to be exceedingly rare in South and East Asia.^{27,29}

Many mutations that cause classic CF were identified in Europeans after the discovery of the *CFTR* gene in 1989 and the ensuing analysis of patient DNA sequences. At the same time, additional protein-changing *CFTR* variants were discovered in association with diseases that had been thought to have etiologies distinct from CF. Molecular studies have now shown that many *CFTR*-associated diseases (including classic CF) are caused by a disruption of ion transport in epithelial cells due to an altered expression or functionality of the *CFTR* protein.¹

Different populations have particular *CFTR* allele spectra that are associated with distinct manifestations of disease.^{4,27,29} A system for scoring *CFTR* disease liability needs to focus on the degree to which a defined mutation increases morbidity and mortality in certain ancestral backgrounds, irrespective of a narrow presentation of disease symptoms. Ours is not the first application of computational analysis to novel variants in the *CFTR* gene. Updated European recommendations, in the context of molecular genetic diagnosis of CF and *CFTR*-related disorders, recognize the utility of computational analysis in assisting determination of the potential pathogenicity of novel sequence changes.²⁶

Overcoming current limitations in carrier screening

The current study reveals that many likely disease-contributing mutations in *CFTR* are untested variants not defined in the literature. These variants are not detected by the standard screening panels and are currently excluded in traditional reproductive risk assessment.

Overcoming the disadvantages inherent in targeted mutation testing for genes such as *CFTR* begins with exome sequencing for the genes of interest as an analytical standard. Although next-generation exome sequencing is incorporated into some carrier screening panels, the resulting analyses are still subject to a traditional clinical interpretation. When deciding which targeted mutations to sequence and/or report, the designers of these tests typically restrict the scope to clinically observed and previously published disease-contributing mutations. The selection of a particular variant may increase a given panel’s mutation detection rate in one population but not in others (Figures 1 and 2).

These imposed limitations lead to widespread inadequacies of current screening practices. With the advancement of sequencing technology, along with its plummeting price tag, we now

have more comprehensive and accessible options for analyzing *CFTR* and other genes' contributions to reproductive risk.

Observing a recessive mutation in an affected individual is clinically valuable and actionable. However, the frequency at which rare recessive mutations are observed in nondiseased individuals is orders of magnitude higher than their expected appearance in diseased individuals. As a result, many likely deleterious mutations may never actually be clinically validated.

Recent advances in computational analyses have made it possible to critically assess the disease contributions of uncharacterized variants. Here, we select the integration of protein damage modeling and population statistics as useful revelatory tools in exposing the limitations of current carrier screens and reliance on clinical validation. On a larger stage, where previously uncharacterized variants have the potential to play a role in reproductive decision-making, protein damage modeling and population statistics become key components of a dynamic and complex analytic methodology.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

The authors thank Anne Morriss, Adam Cohn, and Gholson Lyon for their insightful comments with the development of the project and revisions of the manuscript. The authors also thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison (a full list of contributing groups can be found at <http://exac.broadinstitute.org/about>). The authors thank the scientists who produced the data and bioinformatics tools of the 1000 Genomes Project and the individual participants who contributed their genomes for analysis.

DISCLOSURE

R.M.L., A.J.S., M.J.S., C.B., B.S., T.C.P., and J.L.L. are currently employees at GenePeeks, Inc.

L.M.S. is currently a consultant for GenePeeks, Inc.

REFERENCES

- Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet* 2015;16:45–56.
- De Boeck K, Wilschanski M, Castellani C, et al.; Diagnostic Working Group. Cystic fibrosis: terminology and diagnostic algorithms. *Thorax* 2006;61:627–635.
- Noone PG, Knowles MR. 'CFTR-opathies': disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. *Respir Res* 2001;2:328–332.
- Ngiam NS, Chong SS, Shek LP, et al. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations in Asians with chronic pulmonary disease: a pilot study. *J Cyst Fibros* 2006;5:159–164.
- Sakamoto H, Yajima T, Suzuki K, Ogawa Y. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutation associated with a congenital bilateral absence of vas deferens. *Int J Urol* 2008;15:270–271.
- Grody W, Cutting GR, Klinger KW, et al. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* 2001;3:149.
- Watson MS, Cutting GR, Desnick RJ, et al. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med* 2004;6:387–391.
- Lazarin GA, Haque IS, Nazareth S, et al. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. *Genet Med* 2012;15:178–186.
- Rohlfes EM, Zhou Z, Heim RA, et al. Cystic fibrosis carrier testing in an ethnically diverse US population. *Clin Chem* 2011;57:841–848.
- Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* 2013;132:1077–1130.
- Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013;45:1160–1167.
- Castellani C, CFTR2 team. CFTR2: How will it help care? *Paediatr Respir Rev* 2013;14 Suppl 1:2–5.
- Reppell M, Koch E, Peter BM, Novembre J. Surfing waves of data in San Diego: sophisticated analyses provide a broad view of human genetic diversity. *Genome Biol* 2014;15:562.
- Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42(Database issue):D980–D985.
- Goldstein DB, Allen A, Keebler J, et al. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet* 2013;14:460–470.
- Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7:e46688.
- den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;15:7–12.
- Taschner PE, den Dunnen JT. Describing structural changes by extending HGVS sequence variation nomenclature. *Hum Mutat* 2011;32:507–511.
- Palo JU, Ulmanen I, Lukka M, Ellonen P, Sajantila A. Genetic markers and population history: Finland revisited. *Eur J Hum Genet* 2009;17:1336–1346.
- Nevanlinna HR. The Finnish population structure. A genetic and genealogical study. *Hereditas* 1972;71:195–236.
- Krasnov KV, Tzetis M, Cheng J, Guggino WB, Cutting GR. Localization studies of rare missense mutations in cystic fibrosis transmembrane conductance regulator (CFTR) facilitate interpretation of genotype-phenotype relationships. *Hum Mutat* 2008;29:1364–1372.
- Schneider A, Larusch J, Sun X, et al. Combined bicarbonate conductance-impairing variants in CFTR and SPINK1 variants are associated with chronic pancreatitis in patients without cystic fibrosis. *Gastroenterology* 2011;140:162–171.
- Strom CM, Redman JB, Peng M. The dangers of including nonclassical cystic fibrosis variants in population-based screening panels: p.L997F, further genotype/phenotype correlation data. *Genet Med* 2011;13:1042–1044.
- Oka H, Ishii H, Kishi K, et al. A Japanese adult case of forme fruste cystic fibrosis. *Intern Med* 2010;49:1251–1252.
- Dequeker E, Stuhmann M, Morris MA, et al. Best practice guidelines for molecular genetic diagnosis of cystic fibrosis and CFTR-related disorders—updated European recommendations. *Eur J Hum Genet* 2009;17:51–65.
- Lee JH, Choi JH, Namkung W, et al. A haplotype-based molecular analysis of CFTR mutations associated with respiratory and pancreatic diseases. *Hum Mol Genet* 2003;12:2321–2332.
- Chang MC, Chang YT, Wei SC, et al. Spectrum of mutations and variants/haplotypes of CFTR and genotype-phenotype correlation in idiopathic chronic pancreatitis and controls in Chinese by complete analysis. *Clin Genet* 2007;71:530–539.
- Prasad R, Sharma H, Kaur G. Molecular basis of cystic fibrosis disease: an Indian perspective. *Indian J Clin Biochem* 2010;25:335–341.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>