Spatial Covariant Management of Thermodynamics Contributes to Protein Fold Function through Quality Assurance

Frederic Angles

The Scripps Research Institute

Chao Wang

The Scripps Research Institute

William Balch (✉ webalch@scripps.edu)

The Scripps Research Institute

Article

Keywords: protein fold, genome variations, Spatial CoVariance

Posted Date: July 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-717846/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1

2

3

Spatial Covariant Management of Thermodynamics

4

Contributes to Protein Fold Function through Quality Assurance

5

6

Frédéric Anglès¹, Chao Wang¹ and William E. Balch^{1,#}

7

8

Scripps Research, Department of Molecular Medicine¹

9

10550 North Torrey Pines Rd

10

La Jolla, CA, 92037 USA

11

12

13

[#]To whom correspondence should be directed: webalch@scripps.edu

14

ABSTRACT

Although the impact of genome variation on the thermodynamic properties of the protein fold has been studied *in vitro*, it remains a challenge to assign these relationships across the entire polypeptide sequence *in vivo*. Using the Gaussian process regression-based principle of Spatial CoVariance (SCV), we globally assign on a residue-by-residue the biological thermodynamic properties contributing to the functional fold in the cell using CFTR as an example. We demonstrate the existence of a thermodynamically sensitive region of the CFTR fold involving the interface between NBD1 and ICL4 that contributes to the endoplasmic reticulum (ER) export. At the cell surface a new set of residues contribute uniquely to the management of channel function. These results support a general 'quality assurance' (QA) view of global protein fold management as an SCV principle describing the differential pre- and post-ER residue interactions contributing to compartmentalization of the energetics of the protein fold for function.

Achieving the native physiologic functional state of the protein fold *in vivo* (Anfinsen, 1973) is a complex process that reflects the sum of the multidimensional information defined among other things by the genome sequence, translation speed, specific interactions and the state of the proteostasis network in the cell (HSPs level and other components modulating folding/degradation) in response to the variable environments of the cell and its surroundings. Genetic variation in the population differentially affects protein function in complex cell, tissue and host environments, resulting in a range of functional to dysfunctional phenotypes that contribute to human disease (Wang and Balch, 2018). The different phenotypes of each of these variants reflect the impact of maladaptive folding intermediates (Roth et al., 2014), variants that we have previously proposed can be used to understand the relationships between genome sequence information and the protein fold that shapes function (Wang and Balch, 2018; Wang et al., 2019a; Wang et al., 2019b).

By using variants as reporters of sequence-to-function-to-structure relationships (Wang and Balch, 2018), we have developed variation spatial profiling (VSP), a Gaussian process regression (GPR) based method that can analyze and predict how each residue defining the polypeptide chain sequence impacts the folding and function of any protein (Wang and Balch, 2018; Wang et al., 2019a; Wang et al., 2019b) (**Fig. 1A**). VSP is based on the principle of spatial covariance (SCV) that uses a sparse collection of known variants found in the world-wide population and their associated biological features to quantitatively assign functional value (protein traits) as a matrix to all uncharacterized residues spanning the polypeptide sequence (**Fig. 1A**) thereby defining the role of each residue in the WT-fold on a residue-by-residue basis.

Understanding the role of the protein triage system in making decisions as to when to generate and maintain the fold for stability or degrade a protein in the cell, and how it links to protein function is of critical importance to our knowledge of the differential biology in health and disease impacting each cell type. In the eukaryotic cell exocytic pathway, chaperone assisted protein folding and endoplasmic reticulum (ER) associated degradation (ERAD) have been shown to play an important role in triaging proteins for export by the COPII export machinery (Bethune and Wieland, 2018; Lord et al., 2013; Routledge et al., 2010; Wiseman et al., 2007). The ER has been generically referred to as 'quality control (QC)' system (Ellgaard and Helenius, 2003; Hammond et al., 1994; Hanson et al., 2009) that only exports proteins with native conformation (Ellgaard and Helenius, 2003) competent for function in downstream compartments. However, this assumption is challenged in many diseases where the key proteins harbor multiple variants that have no effect on ER export yet can be completely dysfunctional in downstream compartments. An example is that of the cystic fibrosis transmembrane conductance regulator (CFTR) which, when mutated, contributes to CF. Whereas the F508del variant in the NBD1 domain (**Fig. 1A**) results in near total degradation by the ERAD (Lukacs et al., 1993; Wang et al., 2008; Wang et al., 2006), the G551D variant in the same domain as F508 is delivered with WT efficiently to the cell surface where it is functionally inactive as a chloride channel triggering severe disease (Kotha and Clancy, 2013). These observations pose a major challenge in understanding how every residue contributes as part of a complex of direct and indirect physical-chemical interactions in the polypeptide chain that commit the fold to stability versus degradation, and ultimately to the native functional conformation(s).

To address the global features managing the biological fold *in vivo*, we now apply VSP (Wang and Balch, 2018; Wang et al., 2019a; Wang et al., 2019b) to understand the thermodynamic

design of the entire CFTR fold on a residue-by-residue basis that contributes to trafficking from the ER and chloride channel function at the cell surface. Using temperature-shift to study the energetics of the CFTR fold *in vivo*, we quantitatively assign, for the first time, the role of each residue of the CFTR polypeptide sequence required in different cellular compartments to achieve either stability, trafficking and/or function in the context of non-equilibrium biological thermodynamics (Fang and Wang, 2020; Ornes, 2017). We demonstrate the existence of a SCV defined thermodynamically sensitive region of the CFTR fold involving the interfaces between nucleotide binding domain 1 (NBD1) and the fourth intracellular loop 4 (ICL4) of transmembrane (TM) domain 2 (TMD2) that is primarily responsible for the energetics of ER export (Gürkan et al., 2006; Nishimura et al., 1999; Peotter et al., 2019; Spear et al., 2015; Wang et al., 2004b). At the cell surface a new set of residues contribute uniquely to management of channel function. These results suggest an uncoupling of the energetics contributing to residue function for nascent synthesis, export, and trafficking versus downstream function at the cell surface- considerably expanding our understanding of triage in the living cell as dynamic collective system driving the physiological state of the fold.

We propose that application of SCV principles (Wang and Balch, 2018) to understand the biological thermodynamic properties of the fold *in vivo* provides an universal quantitative tool to assess the conserved rules dictating covariant relationships fundamental to functional protein design. We now propose a new view of triage in the cell from an energetic perspective as a more comprehensive 'quality assurance' (QA) program as part of a quality system (QS) platform (<https://asq.org/quality-resources/quality-assurance-vs-control>) (Box and Narasimhan, 2010) that embraces the continuum of processes that ultimately contribute to the value of the functional protein fold in the cell.

RESULTS

TrIdx and ClCon phenotypes of CFTR variants at 37°C

To address the role energetics in managing the folding, stability, trafficking and function of proteins migrating through the exocytic pathway, we turned our attention to CFTR as a prototypical example of the impact of variation on protein fold and function trajectories given that around 1,000 disease-causing variants impacting the structural fold have been reported in the clinic (www.genet.sickkids.on.ca; www.CFTR2.org) (Masica et al., 2015; Sharma and Cutting, 2019; Sosnay et al., 2013a). CF is a life-threatening Mendelian disease caused by genetic variation in the CFTR gene (Sosnay et al., 2013a; Wang and Balch, 2018). CFTR codes for an apically localized, chloride channel responsible for ionic and fluid homeostasis of diverse tissues. The loss of a functional CFTR channel disrupts chloride ion homeostasis resulting in increased mucus viscosity in the airway of the lungs and ductal systems of the pancreas and liver, as well as decreased hydration of the intestinal tract (Cutting, 2015; Hoffmann et al., 2007). Structurally, CFTR is composed of two nucleotide binding domains (NBDs), a regulatory domain (R) and two transmembrane domains (TMDs), each composed of six transmembrane (TM) helices connected by extra- and intra-cellular loops (ECL and ICL, respectively) (Liu et al., 2017) (Fig. **1B**). The biogenesis of CFTR involves nascent synthesis in the endoplasmic reticulum (ER), the first step in the exocytic pathway, followed by trafficking through the Golgi to reach its destination at the apical cell surface of epithelial cells where it functions as a gated chloride channel (Moran, 2017; Saint-Criq and Gray, 2017).

To characterize the features of the fold contributing to function of the CFTR polypeptide sequence, we characterized at physiological temperature (37°C) the properties of 64 most frequent

CFTR missense variants in the CF population that are dispersed along the entire polypeptide sequence dictating the organization of the structural fold (Cutting, 2015; Sosnay et al., 2013a; Wang and Balch, 2018) (**Fig. 1B** and **Fig. S1** and **Fig. S2A**, upper panel). These 64 variants have diverse relationships between their trafficking index (TrIdx), a metric that represents the export from the ER to the cell surface defined as a ratio of ER exported CFTR relative to total CFTR in the cell, and cell surface localized Cl^- conductance (ClCon). They include variants with a deficient TrIdx and no ClCon, as well as a wide range of TrIdx values associated with partial to more normal ClCon. This library represents the most frequent variants in the CF population and also accurately describes the distribution of pathogenic/benign mutations of the >200 variants clinically characterized to date (Landrum et al., 2015) (<https://www.ncbi.nlm.nih.gov/clinvar/>). Different from other mutation-base modeling approaches (Wang and Balch, 2018), VSP considers only mutations present in the population that are biologically relevant to onset and progression of human disease in order to focus on the key features of the fold that are important for the evolution of healthspan in the human population.

To calculate each variant TrIdx, we transiently transfected HEK293 cells each of the 64 variants in our library of native mutations responsible for the majority of CF in the clinic, and quantitated the relative abundance of the ER-localized core-glycosylated high mannose containing CFTR band B and the post-Golgi localized band C glycoforms that migrate uniquely on SDS-PAGE in response to trimming and modification by Golgi-localized enzymes (**Fig. S1, S2A** upper panel) (Anglès et al., 2019). A comparison of the normalized TrIdx obtained from HEK293 cells (**Fig. S2A**, upper panel) to the TrIdx calculated from previously published data using Fischer rat thyroid (FRT) (**Fig. S2A**, middle panel) and HeLa cells (Sosnay et al., 2013a; Van Goor et al., 2014) (**Fig. S2A**, lower panel) revealed a Pearson correlation coefficient (PCC) of 0.94 and 0.92,

respectively (**Fig. S2B** and **Fig. S2C**), suggesting that the TrIdx of common CFTR variants does not differ significantly between highly divergent heterologous cell lines reflecting conserved rules of protein design. The small difference between the datasets reported is likely due to a variable leakiness of variant export from the ER that is on average slightly higher in HEK293 cells when compared to the other cell lines.

To measure the ClCon of each variant at the cell surface, we adapted a fluorescence membrane potential (FMP) assay (Maitra et al., 2013) and normalized the fluorescence of each variant to that of WT-CFTR to obtain a normalized functional score for each CFTR variant (**Fig. S3**). A comparison of the normalized function between HEK293 cells and FRT cells (Sosnay et al., 2013a) (**Fig. S4A**) yielded a PCC score of 0.80 (**Fig. S4B**). In addition, while TrIdx studies of our variant library have not been reported in HBE cells, a ClCon study of around 60 variants was performed in HBE cells and the variants in-common with our studies in HEK293 (R334W, I336K, T338I, S341P, F508Del and G551D) behave in an identical fashion to what we have seen in HEK293 cells (Han et al., 2018), demonstrating that the basic functionality of human variants found in the population in different cellular environments is largely conserved.

SCV relationships spanning the entire polypeptide chain

To understand the relationships between TrIdx and ClCon across the entire polypeptide chain, we applied VSP, a GPR based approach we have pioneered (Wang and Balch, 2018) that uses sparse variants information to predict the full-spectrum of sequence-to-function-to-structure relationships based on the principle of SCV for any protein (**Fig. 1A**). The method relies on the regression principle of weighted proximity, relating a sparse collection of variant positions in the sequence (patients genotype) to phenotypic traits, such as TrIdx and ClCon used herein, to generate

163 predicted values for the entire CFTR sequence (Wang and Balch, 2018). To apply VSP, we
164 position variants on the x-axis by their normalized position in the linear polypeptide sequence,
165 where the full-length chain is assigned a value of '1', referred to as the variant sequence position
166 (VarSeqP) (**Fig. 1A**). VarSeqP defines our first coordinate of a molecular spatial state encoded by
167 the 1D sequence of the genome. Given the importance of cellular location in CFTR along the
168 exocytic pathway that contributes to function, we assigned the value of each variant's TrIdx (**Fig.**
169 **S2A**, upper panel) to the y-axis (**Fig. 1A**). The TrIdx is the normalized ratio of the fraction of
170 variant exported from the ER relative to the total variant level in the cell, where the value of the
171 WT CFTR TrIdx is assigned a value of 1.0. The TrIdx provides a second coordinate in the
172 molecular VSP strategy, thereby positioning the variant sequence information in a specified
173 biological context. To correlate the sequence position and TrIdx values to CFTR function at the
174 cell surface (ClCon), the measured ClCon value for each variant is normalized to WT ClCon (set
175 to 1.0) (**Fig. S4A**, upper panel). ClCon is assigned to the z-coordinate (**Fig. 1A**). To integrate
176 sequence variation (VarSeqP) with biological features (TrIdx and ClCon) in 3D space, we
177 calculate the 2D functional distance values based on VarSeqP and TrIdx, as well as the variance
178 of ClCon for all possible 2080 (based on 64 variants) variant pairwise combinations (**Fig. 1A**).
179 These spatial relationships can be viewed and modeled in a molecular variogram (**Fig. 1A**) (Chiles
180 and Delfiner, 2009; Wang and Balch, 2018). The molecular variogram defines the distance range
181 based on the differences of variant sequence position and TrIdx value where the known variants
182 co-vary with each other to generate the ClCon. The molecular variogram provides a platform to
183 apply GPR to predict amino acid function across the entire polypeptide sequence that can be
184 projected as a 3D phenotype landscape (**Fig. 1A**). The 3D phenotype landscape reveals the
185 covariant TrIdx and ClCon properties of the CFTR fold for each residue position in the context of

the full-length polypeptide. When the 3D landscape (**Fig. 1A**) is projected as a 2D heatmap of the z-coordinate ClCon values (**Fig. 1A** and **Fig. 2A**, left panel), a confidence interval (uncertainty) that is assigned to every data point through GPR can be delineated by contour lines in the 2D projection map (**Fig. 1A** and **Fig. 2A**, left panel). The regions surrounded by more variants will have higher confidence value (i.e., low uncertainty) associated with the prediction than the regions with fewer input variants in the landscape reflecting the natural distribution of variants in the population (Chiles and Delfiner, 2009; Isaaks and Srivastava, 1989; Wang and Balch, 2018; Wang et al., 2019a).

SCV based functional-structures

To address sequence-to-function-to-structure SCV relationships in our CFTR example, we project the generated ClCon-phenotype landscape values with the highest confidence for all CFTR residues onto the structure of CFTR (Liu et al., 2017; Zhang et al., 2017) to generate a ClCon 'functional-structure' (**Fig 2A**, right panel) (Wang and Balch, 2018). The confidence value is strictly correlated with the variant distribution in the landscape, and the predictions with highest confidence assigned to the functional-structure focus on the key features of the fold that are reported by genetic variants in the patient population representing an important evolutionary information contributing to healthspan and disease. Residues likely to convey value close to that of WT ClCon activity supporting normal lung function ($>\sim 0.5$ of the ClCon value for WT found at 37° C) are 'grayed' out, allowing us to focus on the effect of residues predicted to compromise function at 37°C (**Fig. 2B**). The ClCon-phenotype landscape observed in HEK293 cells and the derived ClCon functional-structure is similar to that observed in HeLa and FRT cells (Wang and Balch, 2018), an expected result given the high correlation for these phenotype properties (**Fig.**

S4B). These results emphasize the conservation of CFTR fold when examining the fundamental SCV properties of trafficking from the ER and ClCon at the cell surface across divergent eukaryotic cell lines at 37°C (**Fig. S2** and **Fig. S4**).

While the ClCon-phenotype landscape conveys the contribution of each residue in the polypeptide chain to channel activity, by transposing the y and z axes found in the ClCon-phenotype landscape we can build a TrIdx-phenotype landscape (**Fig. 2C**, left panel) and the derived TrIdx functional-structure to display the contribution on residue-by-residue basis the TrIdx of CFTR from the ER (**Fig. 2C**, right panel). We used a leave-one-out cross-validation analysis given it is the best or among the best validation approaches for small sample size modeling (Zhang and Yang, 2015) which yields a significant PCC of 0.57 and 0.79 with a $p = 5 \times 10^{-6}$ and 7×10^{-12} (ANOVA test, with null hypothesis with the coefficient equal to zero), for ClCon-phenotype and the TrIdx-phenotype landscapes, respectively (Table S1). To further validate the output of the ClCon- and TrIdx-phenotype landscapes (**Fig. 2A, 2C**), we used a different dataset of diverse CF variants (Van Goor et al., 2014; Yu et al., 2012) (**Fig. S5A**) not included in the training dataset. These validation variants have diverse relationships between TrIdx and ClCon (**Fig. S5B, D**). They include variants with deficient TrIdx and no ClCon, as well as a wide range of TrIdx values associated with partial to more normal ClCon (**Fig. S5A,B,D**). The variants in the validation dataset are distributed across the entire CFTR sequence. Validation reveals a strong correlation (**Fig. S5C,E**; PCCs of 0.67 (p value = 0.8×10^{-3}) and 0.93 (p value = 1.1×10^{-9}) for ClCon and TrIdx, respectively) between all the experimentally measured values and the predicted values that define the output phenotype landscape.

We grayed out any value higher than 0.5 found in the TrIdx functional-structure at 37°C to illustrate residues in the fold specifically contributing to ER export (**Fig. 2D**). The TrIdx functional-structure shows that residues required for a proper fold of the protein responsible for ER export are not distributed evenly throughout the polypeptide, but rather form specific clusters containing known and previously uncharacterized residues with severe trafficking defects. These residues form a region of the fold that contributes to the integrated TrIdx functional-structure properties of NBD1 with ICL4, TM11 and TM1 (**Fig. 2C**, left panel and **Fig. 2D**). The overall impact of residues on TrIdx are highly variable from mild to severe (**Fig. 2C**). For example, SCV reveals the critical role of TM1 on the TrIdx (**Fig. 2C**, left panel; **Fig. 2D**). TM1, the first membrane spanning helix, has been shown to be important for the membrane integration of TMD1 (Patrick et al., 2011). Furthermore, both ICL4 and TM11 in TMD2 have been shown to be essential for the assembly of the fully folded and functionally exportable CFTR protein (Wang and Balch, 2018).

Known and predicted residues with mild to no trafficking defects in the NBD2 domain reflect the fact that the folding of NBD2 is first influenced by the assembly of the NBD1-TMD1 (TM1) and TMD2 (ICL4/TM11) ER export sensitive residues (Du et al., 2005; Farinha et al., 2013) (**Fig. 2C**, left panel and **Fig. 2D**). This observation is consistent with the fact that the Δ NBD2 CFTR protein matures and has similar stability as the full-length WT CFTR during the time-frame measured to conclude that catalytically crucial inter-NBD associations are not required to satisfy ER export (Cui et al., 2007). However, it is interesting to note that among the 64 variants used in the study, the TrIdx of the N1303K is the less similar compared to other cell lines (0.5 of WT in HEK293 cells compared to 0.1 of WT in HeLa cell (Sosnay and Cutting, 2014), indicating that

specific NBD2 properties encompassing the N1303K SCV cluster could be differentially modulated depending on specific biological features of the tissue.

Thus, in general, SCV relationships dictating the basic functional properties of CFTR appear similar across all cell lines suggesting that unifying thermodynamic-based SCV principles are driving universal features the functional fold *in vivo*.

Biological thermodynamic SCV relationships dictating ER export

The allele frequency of CF-causing variants reveals that ~90% of the patient population carries at least one copy of the phenylalanine deletion at position 508 (F508del) in both homozygous and heterozygous states (Cutting, 2015; Sosnay et al., 2013a; Veit et al., 2016a). The F508del variant disrupts the proper folding of CFTR leading to its failure to be exported from the ER and clearance by ER-associated degradation (ERAD) (Ahner et al., 2007; El Khouri et al., 2013; Okiyoneda et al., 2010). While the impact of F508del on the thermodynamics of the NBD1 domain fold has been studied *in vitro* (Aleksandrov et al., 2012; Meng et al., 2017; Protasevich et al., 2010; Thomas et al., 1995) and the energetic response of the F508del variant has been captured by temperature shift (37°C to 27°C) *in vivo* (Bagdany et al., 2017; Denning et al., 1992; Wang et al., 2008), it remains completely unknown how reduced temperature impacts the entire CFTR polypeptide sequence on a residue-by-residue basis as a SCV based matrix, a global paradigm for understanding the thermodynamics of protein folding *in vivo*.

To address the role of energetics in CFTR fold synthesis, stability and function-to-structure relationships using SCV, we transiently expressed WT-CFTR and each of the 64 variants and monitored the steady state distribution of the band B and C glycoforms using SDS-PAGE

following incubation at 27°C for 24 h (**Fig. 3A**, upper right inset and **Fig. S1**). To compare the responsiveness of the variants to reduced temperature, we calculated the WT-normalized TrIdx for each variant at 37°C (orange) and 27°C (purple) (**Fig. 3A**, left panel). Surprisingly, most variants tested appear to be refractory to reduced temperature correction (**Fig. 3A**, left panel). Those variants with a WT-normalized TrIdx below 0.4 exhibited improved yet highly variable trafficking responses to reduced temperature. For example, in contrast to F508del CFTR, which is the most highly responsive non-trafficking variant at 37°C to reduced temperature (**Fig. 3A**, right inset), the NBD1 S492F-CFTR variant that exhibits a defect of TrIdx of equal severity to that of the F508del variant at 37°C, showed nearly no TrIdx improvement at 27°C (**Fig. 3A**, right inset). These results suggest that S492F may confer a dominant negative effect on the thermodynamics of the fold, for example, potentially energetically destabilizing the folded state of NBD1 thereby preventing ER export (Proctor et al., 2015). Residues outside of the NBD1 domain that showed trafficking defects largely failed to be temperature corrected for export from the ER (**Fig. 3A**).

To expand the analysis of the effect of reduced temperature on select variants to the entire fold, we applied VSP to generate a temperature sensitive (TS) delta (Δ) (Stride)-phenotype landscape that uses TrIdx at 37°C as the y-axis (**Fig. 3B**, left panel) to predict the delta response of TrIdx between 27°C and 37°C (Δ TrIdx) as the z-axis (**Fig. 3B**, left panel) for all residues. The TS Δ TrIdx-phenotype landscape (**Fig. 3B**, left panel) showing the temperature correctability on a residue-by-residue basis was mapped to the CFTR structure to generate a TS Δ TrIdx functional-structure (**Fig. 3B**, right panel). We observed that variants having a TrIdx > 0.4 (**Fig. 4B**, yellow to orange) exhibited less than a 20% increase in their TrIdx for ER export upon incubation at 27°C suggesting they are already thermodynamically favored for efficient export despite having disease causing function at the cell surface. Indeed, most residues were not responsive to reduced

temperature with the exception of 3 regions- two in the NBD1 (region I & II) and one in the TMD2 domain comprising ICL4 (region III) (**Fig. 3B**, left panel, red to purple) possibly reflecting energetic coupling between these domains.

To focus on energetic coupling, we generated a gray-out view of the TSΔTrIdx functional-structure (**Fig. 3C**) which shows in color only those residues having a >50% response to temperature shift (**Fig. 3C**). The gray-out models reveal that clusters defining regions I-III converge in the context of the entire CFTR fold design at the NBD1-ICL4 interface (**Fig. 3C**). For example, variants with the highest TrIdx correction in response to reduced temperature, including F508 and R560 (NBD1), L1065 (ICL4) (**Fig. 3C**, insert), show a critical role for these residues in stabilizing the thermodynamics of the fold to promote ER export and/or to prevent CFTR targeting for degradation. These results are consistent with the observation that residues 563 to 567 function as a di-acidic motif to confer COPII-dependent export from the ER as previously reported (Nishimura and Balch, 1997; Nishimura et al., 1999; Wang et al., 2004b; Zhang et al., 2011) (**Fig. 3D**), a conclusion consistent with mass spectrometry analysis of residues undergoing methylation and ubiquitination where ubiquitination of di-acidic motif residues targets CFTR for degradation when challenged through misfolding by F508del (Pankow et al., 2019).

Thus, thermodynamic sensitive SCV relationships that capture the energetic properties across the entire CFTR nascent fold in response to temperature shift *in vivo* reveals for the first time an integrated thermodynamic sensitive region comprising the NBD1-ICL4 interface that dictates SCV tolerance (Wang and Balch, 2018) of the fold for either recognition by the COPII vesicle machinery for export (Gürkan et al., 2006; Peotter et al., 2019; Spear et al., 2015; Wiseman et al., 2007) or for degradation through ERAD (Ahner et al., 2013; Ahner et al., 2007; Gong et al.,

2016; Gong et al., 2019; Hutt et al., 2018c; Schmidt et al., 2009) at the first step of the exocytic pathway.

SCV relationships defining cell surface stability

To assess the biological function-based thermodynamic properties of the CFTR fold at the cell surface using VSP, each variant was expressed at 37°C and subsequently incubated at 27°C for 24 h to promote stabilization and export of all reduced temperature responsive variants to the PM (**Fig. 4A**). The cells were then returned to physiological temperature (37°C) in the presence of cycloheximide (CHX) to block *de novo* protein synthesis (Ennis and Lubin, 1964) (**Fig. 4A**, insert). To measure the post-ER stability of expressed CFTR variants, we calculated the % of band C remaining after a 120 min chase in the presence of CHX at 37°C and normalized this percentage to that observed for WT-CFTR to obtain a normalized band C stability score (**Fig. 4A** and **Fig. S1**). The variants were ordered from the most stable to the least stable at 37°C with the N1303K and R560T variants serving as examples of variants with WT-like stability and those that show significant destabilization, respectively. We observed that ~ 50% of the tested variants exhibited stability of band C that was less than that of WT-CFTR (**Fig. 4A**, red box).

Using the TrIdx as the y-axis and the band C stability score as the z-axis values in our VSP analysis, we were able to build a stability band C (StabC)-phenotype landscape (**Fig. 4B**, left panel) that can predict the role of all residues with uncertainty impacting the post-ER thermodynamic stability of the CFTR polypeptide. These predictions were used to generate a StabC functional-structure (**Fig. 4B**, right panel and **Fig. 4C**). Interestingly, the post-ER stability of CFTR variants is highly correlated with their TrIdx at 37°C (**Fig. S7**). Variants with a TrIdx greater than 0.5 of WT-CFTR all had a band C stability score of at least 0.5 of that seen with WT-CFTR (**Fig. 4B**,

left panel, red to dark purple). Furthermore, we noted that select variants at any known or predicted residue based on SCV in the first half of the protein (NT, TMD1, NBD1, R domains) that reduced the TrIdx of CFTR at 37°C (TrIdx < 0.5), also reduced the post-ER stability of the polypeptide. These results illustrate that the SCV core linking NBD1 to TM1 and TM11 that determines the efficiency of ER trafficking (**Fig. 2D**) can in some cases contribute to cell surface stability (**Fig. 4C**). These results suggest the occurrence of energetic features used to maintain the function of the protein fold that are independent of cellular compartmentalization. However, we also noted that select regions in the C-terminal half of the protein that significantly impacted TrIdx at 37°C (TrIdx < 0.5), such as the cluster defined by N1303K found in NBD2, had only a minor impact on stability at 37°C indicating a unique role in ER export (**Fig. 4B**, left panel).

In general, the impact of NBD2 variants on protein stability is more dependent on the cellular localization, possibly reflecting genome to proteome design that contributes to differential function in distinct subcellular environments stressing the individual assignment of residues features in managing of the fold. Below we assess the impact of energetics on ClCon to correlate divergent requirements of sequence-to-function-to-structure design that link trafficking to channel function.

Thermodynamic SCV relationships driving ClCon

To assess impact of energetics on ClCon, variants were incubated at 27°C for 24 h, then ClCon was measured immediately following shift to 37°. The activity of each variant was normalized to that of WT-CFTR at 37°C (**Fig. 5A**). We observed a striking variant specific diversity of ClCon responses to reduced temperature pre-incubation. Some residues, including

N1303K in NBD2 (**Fig. 5A**) failed to respond, while others including F508del, exhibited a significant improvement in channel activity at 27°C (**Fig. 5A**).

To understand the relationships between TrIdx and ClCon at 27°C for all residues contributing to the CFTR fold, we built a ClCon-phenotype landscape at 27°C based on the TrIdx at 27°C (**Fig. 5B**, left panel). The molecular variogram analysis when compared with that observed for ClCon-phenotype landscape at 37°C based on the TrIdx at 37°C (**Fig. 2A**, left panel and **Fig. S8**) reveals that the plasticity of ClCon is significantly increased at 27°C (**Fig. S8A**, $p = 0.04$, right margin; **Fig. S8B**, plateau value), while the spatial distance reflecting the difference in TrIdx measurements is reduced by low temperature (**Fig. S8A**, $p = 5 \times 10^{-6}$, top margin; **Fig. S8B**, range value). These results suggest that distinct thermodynamic sensitive mechanisms impact TrIdx and ClCon in response to temperature shift. The ClCon-phenotype landscape and the corresponding functional-structure observed at 27°C (**Fig. 5B**, C) generally have higher predicted ClCon values when compared with that at 37°C (**Fig. 2A**, B), suggesting that the reduced temperature restoring ER export of temperature-sensitive residues can also restore channel activity.

To gain further insight into the coupling between TrIdx and ClCon, we examined the relationship of TrIdx at 37°C to the Δ of ClCon achieved upon temperature shift. For this purpose, we used the TrIdx at 37°C (**Fig. 5D**, left panel, y-axis) and the Δ temperature shift of ClCon (TS Δ ClCon) reflecting rescue of function upon shift from 37°C to 27°C (**Fig. 5D**, left panel, z-axis) to generate a TS Δ ClCon-phenotype landscape. The computed values were converted to the predicted TS Δ ClCon functional-structure (**Fig. 5D**, right panel) and its gray-out model (**Fig. 5E**) that highlights only those residues with >50% maximal Δ ClCon. Here, we see that many residues in the functional-structure exhibit a relative >50% of maximal Δ ClCon in response to reduced

temperature (**Fig. 5E**, right panel, red to purple), a response that is largely independent of their trafficking ability (**Fig. 5D**, left panel, y-axis). Furthermore, the TSΔClCon functional-structure (**Fig. 5E**) is dramatically different from the TSΔTrIdx functional-structure (**Fig. 3C**). These results suggest that reduced temperature impacts the SCV relationships of the CFTR polypeptide fold beyond simply stabilizing the fold for exit from the ER (Bagdany et al., 2017; Roth et al., 2014; Wang et al., 2008; Wang et al., 2004a; Wang et al., 2006).

In contrast to SCV predicted regions of the fold that show improved ClCon in response to reduced temperature, a number of regions have little to no change reflecting a minimal impact on biologically relevant thermodynamic features of the fold (**Fig. 5D**, clusters 1-3). Clusters 1-3 encompass regions of the CFTR polypeptide where residues have near WT trafficking and a minimal effect on channel activity at 37°C. Exposure to reduced temperature condition does not improve the near WT activity and therefore has a minimal impact on CFTR activity yielding a low TSΔClCon score (**Fig. 5D**, yellow). Moreover, clusters 4-7 encompass regions of the CFTR polypeptide fold where residues have no activity at physiological temperature (**Fig. S3** and **Fig. S4**, upper panel) also do not respond to temperature shift to 27°C (**Fig. 5D**). The combined results revealed by clusters 1-7 likely reflect the differential limits of thermodynamic plasticity of the fold. In contrast to these non-responsive regions to temperature shift, we see specific clusters where residues exhibit improved channel activity upon temperature shift. For example, a region defined by cluster 8, which is located proximally to D110H variant at the N-terminus (N) that traffics normally at 37°C (**Fig. 1A** and **B**), exhibits a weak basal ClCon that is dramatically increased in response to reduced temperature. Moreover, clusters 9 and 10 which are localized around measured L997 and S945 positions in the TMD2 domain, respectively, exhibit little to no activity at 37°C (**Fig. 5A**; **Fig. S3**) yet are responsive to reduced temperature correction of ClCon (**Fig. 5D**).

In general, SCV relationships capturing features of the fold responsive to temperature shift demonstrate that reduced temperature exposure differentially affects the biological thermodynamic properties of the CFTR polypeptide on a residue-by-residue basis that extends well beyond stabilization of the NBD1-ICL4 interface required to export CFTR from the ER. These results emphasize the differential plasticity of thermodynamic relationships dictating genome-based fold design of CFTR for ER export versus differential downstream function in post-ER compartments in response to its local folding and interaction environment (Hutt et al., 2018a; Pankow et al., 2019) indicating the limitations of the QC concept.

Functional thermostability *in vivo* is poorly correlated with *in silico* thermostability predictions

Given the results of the temperature shift-based phenotype landscapes projecting the folding energetics linked to CFTR TrIdx and ClCon, we examined the relationship between these fundamental energetic features dictating fold function *in vivo* to energetic thermostability values ($\Delta\Delta G^\circ$) calculated for each variant included in the study using the thermostability predictive program FoldX (Buss et al., 2018a; Buss et al., 2018b; Delgado et al., 2019). FoldX calculates the $\Delta\Delta G^\circ$ of a residue based on force field algorithm in response to a specific variation to capture the overall thermostability of protein structure (**Fig. S8A**). Strikingly, comparing the FoldX predicted $\Delta\Delta G^\circ$ between WT-CFTR and each of the CFTR variants included in the study with either the measured TrIdx (**Fig. S2A**, upper panel) or ClCon (**Fig. S4A**, lower panel) at 37° *in vivo* revealed that neither stability of the fold for exit from the ER as measured by the TrIdx at 37°C (**Fig. S6B**, PCC = 0.245) nor function at the cell surface as measured by ClCon at 37°C (**Fig. S6C**, PCC = 0.157) correlate with *in silico* predicted folding stability using FoldX (Buss et al., 2018a; Buss et

al., 2018b; Delgado et al., 2019) likely reflecting the non-equilibrium thermodynamic process defining protein folding in biology.

SCV based energetic code for the CFTR fold

To address the functional non-equilibrium thermodynamic differences driving CFTR function in health and disease on a residue-by-residue basis across the entire polypeptide sequence contributing to the TrIdx and ClCon response to temperature shift, we first normalized the predicted ΔTrIdx (**Fig. 3B**) and ΔClCon (**Fig. 5D**) values for each residue relative to WT (**Fig. 6A**, upper and lower bar graphs), and then calculated the difference between the normalized ΔTrIdx and ΔClCon values (**Fig. 6A**, ' $\Delta\text{TrIdx}-\Delta\text{ClCon}$ ' in the bottom panel and **Table S2**) to generate a 'differential energetics based correction' (DEC) plot. The DEC plot allows us to codify on a residue-by-residue basis regions in the fold that are differentially contributing to the non-equilibrium energetics that are critical for the proper trafficking of CFTR from the ER or channel function at the cell surface. For example, some regions that display a negative DEC value (**Fig. 6A**, bottom panel, $y < 0$) reflect a high ΔClCon value and a low ΔTrIdx such as the region encompassing the D110H variant (**Fig. 6A**).

The DEC plot allows us to determine if the correction observed in the ER is causally correlated with correction of ClCon at the cell surface. For example, the increase of 26% of WT level of TrIdx in response to temperature shift observed in the ICL4 region between the position 1061 and 1065 is associated with a similar ClCon correction which results in a DEC value close to 0 (**Fig. 6A**, lower DEC plot; Table S2). We observe a similar effect for the region contributing to the function of the F508 residue (**Fig. 6A**, lower DEC plot; Table S2), consistent with the observation that single channel patch clamp experiments reveal that temperature shift has no effect on current flow through the F508del channel in the full open state (Bose et al., 2019) but can

modulate the frequency and duration of F508del-CFTR channel opening with a maximum open probability at 30°C (Wang et al., 2018). While we did not monitor single channel activity nor quantify the absolute number of channels at the cell surface, SCV analyses suggests that the increased ClCon in response to residues F508 and the region define by 1061 to 1065 is largely due to the amount of variant trafficked to the surface (**Fig. 6A**, **Fig. 6B**, and **Table S2**). In contrast, SCV relationships dictated by residues between 512-536 and the di-acidic exit motif (YKDAD (residues 563-567)) have a strongly positive DEC value (**Fig. 6A**, bottom panel, NBD1 zoom; **Fig. 6B**) because the TrIdx for these regions are significantly improved in response to temperature shift, but ClCon is not (**Fig. 6A**, upper and lower bar graphs).

Combined, our analysis of CFTR TrIdx and ClCon relationships based on energetic insights derived from temperature shift reveals that the NBD1-ICL4 interface that includes F508del and residues 1061-1065 reflects the need for balance in stability of the fold for export versus plasticity for downstream function. While the interface provides a thermodynamic sensitive platform that contributes to both TrIdx and ClCon, there are many other regions such as residues 512-536 and the YKDAD di-acidic motif (**Fig. 6A**, bottom panel and **Fig. 6B**) that have differential energetic contributions to either ER export and/or downstream ClCon indicating a divergence in genome encoded fold design reflecting the need for compartment specialization of sequence-to-function-to-structure relationships contributing to human health and disease.

Discussion

The mechanism by which the biological thermodynamic properties of the polypeptide sequence participate in folding trajectories *in vivo* to achieve function for any protein in response to different environmental conditions operating in the cell remains an enigma. This is particularly

evident in eukaryotic cells where compartmentalization and membrane trafficking from and to specialized compartments provide a major advance to more precisely tune the fold for function (Powers and Balch, 2013).

Herein, we have applied the principle of SCV to a sparse collection of variants found in the population to define how the biological thermodynamic properties of each residue generate function. We used CFTR as a prototypical example of the impact of variation on protein fold and function trajectories in response to the environment given the extensive number of variants contributing to disease in the CF population, but whose roles in disease remains largely mechanistically and clinically uncharacterized (Sharma and Cutting, 2019). Export of F508del from the ER in response to reduced temperature is causally coupled to the co-translational program that, in response to the reduced rate of CFTR synthesis at 27°C (Farinha et al., 2013; Hutt et al., 2018b; Oliver et al., 2019; Veit et al., 2016b), has been shown to provide a biological thermodynamic advantage to its folding (Hutt et al., 2018a; Meriin et al., 2012; Oliver et al., 2019; Veit et al., 2016a) an observation which we have now extended to every residue in the polypeptide chain. Our results are consistent with the evidence that reduced temperature energetically stabilizes the NBD1 domain *in vitro* (Mendoza et al., 2012; Qu and Thomas, 1996; Rabeh et al., 2012; Soya et al., 2019; Vernon et al., 2017). We have now defined the covariant energetic relationships to show that a thermodynamic foundation for generating the CFTR fold in the ER is largely defined by the residues that contribute to the design of the NBD1-ICL4 interface facing the cytosol stressing the integration of ER and cytosolic proteostasis pathways. These key energetic folding events are coordinated allosterically with features of the TM1/TM11 helices linked to ICL4 and ultimately, to more downstream domains in the coding sequence completing the assembly of the fold for function (Dawson et al., 2013; Rabeh et al., 2012). Our results are consistent with the fact

that the folding of the NBD1-ICL4 region is the region primarily responsible for COPII engagement for trafficking from the ER via the di-acidic exit motif (Nishimura et al., 1999; Wang et al., 2021; Wang et al., 2004b; Zhang et al., 2011) and to prevent CFTR degradation by ERAD (Ahner et al., 2007; El Khouri et al., 2013; Gong et al., 2016; Pankow et al., 2019) and autophagy (Hutt et al., 2018c).

While the FoldX algorithm is well established in protein engineering and prediction of protein thermostability based on physical-chemical considerations (Buss et al., 2018a; Buss et al., 2018b; Delgado et al., 2019), it failed to predict CFTR genotype to phenotype traits highlighting the fact that CFTR biogenesis leading to function *in vivo* is a multidimensional process that is not only defined by the genome encoded protein sequence but also by other parameters such as the translation speed, specific interactions and the state of the proteostasis network (Balch et al., 2008; Jayaraj et al., 2020; Sala et al., 2017) that manages protein folding dynamics. This observation is also consistent with the fact that CFTR trafficking and function require multiple conformational changes of flexible regions to couple folding to channel function. In general, SCV analyses focused on the global features of the protein fold suggests that theoretic thermostability predictions generated *in vitro* may have limited relevance to predict the dynamic biology of the functional fold *in vivo*, consistent with the now recognized highly dynamic and likely disordered states in protein design responsible for human biology.

While we showed that the fold required for the stability of the protein and export from the ER can contribute to stability at the PM (**Fig. S7**), DEC codification reveals that a set of SCV-based biological thermodynamic relationships dictating export from the ER have no impact on the function at the cell surface. In addition, a number of residues are exclusively used to manage ClCon

at the cell surface. Thus, SCV demonstrates that ER export is not designed strictly, as currently assumed, to perform a quality control (QC) role related to acquisition of function. Instead, we propose the ER provides a more ancient and fundamental role in genome to proteome design based on SCV matrix principles (Wang and Balch, 2018), serving to construct a minimal core fold that balances the demands of nascent co-translational stability with the need for future plasticity to allow the cell to adapt function in response to the local environment. This view reflects the features controlling the dynamics of the fold inherent in the Anfinsen view (Anfinsen, 1973) that "the protein fold....exists under conditions similar to those for which it was selected- the so-called physiological state"- in which it is the environment, not the snapshot (Liu et al., 2017) nor computed structure (Senior et al., 2020), that is the key to understanding the energetics based functionality of the folding state.

Given the genetic diversity captured by GnomAD (<https://gnomad.broadinstitute.org/>) and numerous other genome sequencing and phenotyping efforts (<https://allofus.nih.gov/>; <https://www.ukbiobank.ac.uk/>; <https://www.ncbi.nlm.nih.gov/clinvar/>), application of SCV principles through GPR-ML can be tailored to any protein to independently assess stability or function, or any other feature, through DEC codification of variation found in the population. DEC provides a mechanism to generate a more comprehensive and unified understanding of the numerous genotypes contributing to health and disease (Claussnitzer et al., 2020; Torkamani et al., 2017; Wang and Balch, 2018) (**Fig. 6C**). Furthermore, since the GPR predicted functional value is based on the SCV principle that considers the covariance matrix information of a variant in the context of other variants, future development of the approach may enable the use of the SCV-based DEC codification to understand the complexity of heterozygosity or co-allelic polymorphism in rare disease, or variant heterogeneity in complex disease such as cancer.

Therefore, combining sequence with function and structure through GPR-ML provides a new paradigm to assess meaningful biological responses to impact of variation in the population.

To reframe SCV based energetic principles responsible for fold management, we now hierarchically organize the cell as a 'quality system' (QS) (<https://asq.org/quality-resources/quality-assurance-vs-control>) (Box and Narasimhan, 2010) (**Fig. 6D**). At the most fundamental level is the concept of quality control (QC) that includes features contributing to the energetics of fold stability independent of function (**Fig. 6D**). Consistent with this view, the ER generates around one third of eukaryotic proteome without knowledge of downstream functionality. Many variants in rare disease are efficiently exported from the ER that lack downstream function (Wang and Balch, 2018). QC is based on unknown SCV relationships defined on a residue-by-residue by energetics alone that dictate stability versus degradation that we have posited provides an evolutionary conserved SCV set-point for fold management (Wang and Balch, 2018). QC is only a subset of a more comprehensive quality assurance (QA) program that manages different energetic features of the fold required to achieve function (**Fig. 6D**). QA manages *function* of the fold in response to energetics influenced by physiological state of the cell (Anfinsen, 1973) (**Fig. 6D**) taking into account the compartmentalized environments in the cell that we have referred to as SCV 'tolerance' (Wang and Balch, 2018). QA is flexibly managed by proteostasis dependent folding and stress responsive signaling pathways (Balch et al., 2008; Balchin et al., 2016; Jayaraj et al., 2020; Sala et al., 2017) including the unfolded protein response (UPR) (Grandjean et al., 2019; Karagoz et al., 2019), the integrated stress response (ISR) (Costa-Mattioli and Walter, 2020) and heat shock response (HSR) (Li et al., 2017), among others. QS, defined by the genetic and epigenetic features of *genome design* of each cell type, manages globally the energetic tolerance of the fold for function *in response to the environment* (**Fig. 6D**)-

balancing synthesis, stability, function and degradation contributing to development and aging (Morimoto, 2019; Sala et al., 2017). We posit that SCV principled relationships based on genomic variation in the population capturing the functional thermodynamic features of the protein fold can reveal the fundamental role of diversity managed by a QC-QA-QS system in driving complex biology found in the individual (**Fig. 6D**).

Materials and Methods

Cell culture and transfection. HEK293 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 2 mM L-glutamine and 10% FBS at 37°C (regular medium). Plasmid transfections were performed as followed. Lipofectamine 3000 (Invitrogen), 100 ng of pBI-CMV2 empty vector or vector containing WT-CFTR or any of the 64 CFTR variants, were mixed with a ratio of 1/2/3 (DNA/Reagent/Lipofectamine) in OPTI-MEM. Optimization of the level of transfected DNA for WT and F508del was performed to define conditions for the optimal production CFTR in the cell (**Fig. S9**). Using 96-well plates, 80% confluent HEK293 were transfected by adding the transfection mixture to regular medium. Cells were incubated at 37°C for 48 h and were then washed and incubated either at 37°C or 27°C for an additional 24 h in regular medium.

GFP fluorescence measurement. Transfection efficiency was validated quantitatively using immunoblotting following transfection of HEK293 cells with 0, 12, 25, 50 100 or 200 ng of pBI-CMV2-WT-CFTR or pBI-CMV2-F508del-CFTR after 24, 48 or 72 h (**Fig. S9A**) and by monitoring GFP fluorescence using a Synergy H1 hybrid reader (Biotek) (**Fig. S9B** and **Fig. S9C**). We observe a similar level of GFP-fluorescence at a given time-point when HEK293 cells were

transfected with the same quantity of either pBI-CMV2-WT-CFTR or pBI-CMV2-variant-CFTR. In all cases, the level of GFP- fluorescence and the level of CFTR generated by either the pBI-CMV2-WT-CFTR or pBI-CMV2-variant-CFTR using immunoblotting were directly comparable (**Fig. S9A-C**). These data validate the use of GFP fluorescence as a metric to ensure uniform transfection efficiencies across the range of variants tested.

Immunoblotting. HEK293 cells lysates were prepared in 50 mM Tris-HCl, 150 mM NaCl, 1% Triton X-100 and 15 µg of total protein were separated using 7% SDS-PAGE. Proteins were transferred to nitrocellulose and probed with 3G11 antibody for CFTR detection followed by a rat secondary antibody, and detection was performed by chemiluminescence. GFP fluorescence was monitored using Living Colors A.v. Monoclonal antibody (JL-8; Cat No. 632380; Clontech) followed by a rabbit secondary antibody and detection performed using chemiluminescence.

CFTR stability assay. HEK293 cells were transfected with pBI-CMV2 empty vector or vectors containing WT-CFTR or any of the 64 CFTR variants tested and incubated for 48 h at 37°C. The cells were subsequently incubated at 27°C for 24 h. After 24 h at 27°C, HEK293 cells transiently expressing CFTR were washed and incubated at 37°C in regular medium supplemented with 50 µM of cycloheximide (CHX) final concentration for 0, 10, 30, 60, 120 min. CFTR variant stability was quantified by immunoblotting based on the difference between the measurements performed at 0 and 120 min of incubation post CHX addition.

Fluorescence Membrane Potential assay. The FMP assay was performed as previously described (Maitra et al., 2013). Briefly, HEK293 cells transiently expressing CFTR were washed twice in phosphate-buffered saline and incubated for 30 min with the blue membrane potential dye dissolved in a modified PBS at 37°C. The plates were read in a fluorescence plate reader at 37°C

following a dye equilibration period of 30 min. CFTR was stimulated with an activation cocktail (10 μ M Forskolin and 50 μ M genistein; final concentration) diluted in the same buffer. Fluorescence was monitored every 0.5 second for a total of 30 sec. CFTR-mediated depolarization of the membrane was monitored as an increase in fluorescence and hyperpolarization or repolarization as a decrease.

Variation Spatial Profiling. The VSP analysis of the data was performed as previously described (Wang and Balch, 2018; Wang et al., 2019a; Wang et al., 2019b). VSP uses a biological adaptation of geostatistics (Chiles and Delfiner, 2009; Isaaks and Srivastava, 1989) based on a GPR-ML approach that provides a means to predict the distribution of a large range of known and unknown features using sparse sampling methods. GPR-based ML is an interpolation/regression method providing unbiased prediction based on the modeled SCV relationships generated by a molecular variogram. Below, we briefly describe how we apply classical GPR-ML through VSP to understand globally the role of variation in human disease. See reference 1 (Wang and Balch, 2018) for further details.

We first positioned the known variants by their sequence positions in the polypeptide chain on the ‘x’ axis coordinate and measurements of a biological parameter of these known variants on the ‘y’ axis coordinate to another biological parameter along the ‘z’ axis coordinate. For this purpose, suppose the i^{th} (or j^{th}) observation in a dataset consists of a value z_i (or z_j) at coordinates x_i (or x_j) and y_i (or y_j). The distance h between the i^{th} and j^{th} observation is calculated by:

$$h_{i,j} = \sqrt{((x_i - x_j)^2 + (y_i - y_j)^2)} \quad (1)$$

The $\gamma(h)$ -variance for a given distance (h) is defined by:

$$\gamma(h) = \frac{1}{2}(z_i - z_j)^2 \quad (2)$$

where (h) -variance is the spatial variance (i.e., the degree of dissimilarity) of the z value between the two observations, which is also the whole variance of z value for one observation at the given separation distance h . The distance (h) and $\gamma(h)$ -variance for all the data pairs are generated by the equations (1) and (2). Then, the average values of $\gamma(h)$ -variance for different distance intervals are calculated to plot $\gamma(h)$ versus h . A molecular variogram is generated from these calculations to define quantitatively the correlation between the spatial variance of z changes and the separation distance defined by the x and y coordinates based on known variants. The distance where the model first flattens out is known as the range. Locations separated by distances closer than the range are spatially correlated, whereas locations farther apart than the range are not. The spatial covariance (SCV) at the distance (h) is calculated by $C(h) = C(0) - \gamma(h)$, where $C(0)$ is the covariance at zero distance representing the global variance of the data points under consideration (i.e., the plateau of the variogram). Based on the molecular variogram defined by the known variants we use a GPR-ML based regression approach referred to as ordinary Kriging that has the least assumptions and is the most commonly used form to assign uncertainty to every unknown residue in the sequence (Oliver and Webster, 2015). Essentially, SCV can predict the unknown value by local weighted averaging the surrounding known values, where the weight associated with the known value is determined according to their positions both in relation to the unknown point and to one another. According to the values found in the molecular variogram, observations within close distance are usually highly correlated and have more weight for prediction. To solve the optimum and unbiased weights of SCV relationships, molecular geostatistics aims to minimize the variance associated with the prediction of the unknown value at location u , which is generated according to the expression:

$$\sigma_u^2 = E[(z_u^* - z_u)^2] = \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j C_{i,j} - 2 \sum_{i=1}^n \omega_i C_{i,u} + C_{u,u} \quad (3)$$

where z_u^* is the prediction value while z_u is the true but unknown value, $C_{i,j}$ and $C_{i,u}$ are SCV between data points i and j, and data points i and u, respectively, and $C_{u,u}$ is the SCV within location u. ω_i is the weight for data point i. The SCV is obtained from the above molecular variogram analysis. To ensure an unbiased result, the sum of weight is set as one:

$$\sum_{i=1}^n \omega_i = 1 \quad (4)$$

Equations (3) and (4) not only solved the set of weights associated with input observations, but also provided the minimized ‘molecular variance’ at location u which can be expressed as:

$$\sigma_u^2 = C_{u,u} - \left(\sum_{i=1}^n \omega_i C_{i,u} + \mu \right) \quad (5)$$

where $C_{u,u}$ is the SCV within location u, ω_i is the weight for data point i, and $C_{i,u}$ are SCV between data points i and u. μ is the Lagrange Parameter that is used to convert the constrained minimization problem in Eq. (3) into an unconstrained one. The resulting minimized molecular variance provides a weighted SCV score that represents the confidence for using the SCV relationships both within the input data points and in relation to the unknown locations. The confidence level is tightly linked with the distance range in the variogram. The shorter distance between the unknown point to the input data points, the higher confidence for using the SCV relationships for prediction. With the solved weights W, we can calculate the prediction of all unknown values to generate the complete phenotype landscape by the equation:

$$z_u^* = \sum_{i=1}^n \omega_i z_i \quad (6)$$

where z_u^* is the prediction value for the unknown data point u , ω_i is the weight for the known data point, and z_i is the measured value for data point i .

In the final prediction, for each unknown position, minimal 5 nearby variants and maximal 15 nearby variants were included in the weighted averaging calculation (i.e., Eq. (6)). All the procedures were performed in the ordinary Kriging module in GS+, version 10 from gammadesign software (<https://geostatistics.com/>).

To map the function predictions onto structure, we assign the prediction value with highest confidence to each residue to generate a functional-structure that illustrates all values interpolated from the sparse collection variants used to generate the phenotype landscape. PDB:5UAK is used for CFTR structure mapping(Liu et al., 2017). All the structural presentations were produced by the software of PyMOL.

VSP prediction validation

The statistical validation methods to assess the performance of the VSP strategy used in this study include a leave-one-out cross-validation and a validation by an external dataset when possible. The default validation method is leave-one-out cross-validation because of small sample size modeling. In the leave-one-out cross-validation, all data are initially used to build the molecular variogram and geostatistical models. We remove each data point, one at a time and use the rest of the data points to predict the missing value. We repeat the prediction for all data points and compare the prediction results to the measured value to generate the Pearson's PCC-value and its associated p value (ANOVA test, with null hypothesis with the coefficient equal to zero).

For the external dataset validation of ClCon and TrIdx predictions at 37°C (**Fig. S5**), we considered the results of 21 CF variants from a separate study (Van Goor et al., 2014; Yu et al.,

2012) that were not used for training. Predicted z values were generated by feeding the model with x- and y- values, and subsequently compared to the observed values by Pearson's correlation analysis and p value calculation (ANOVA test).

SCV-based functional insights are able to achieve a significantly more accurate prediction to inform on novel genotype to phenotype relationships as described (Wang and Balch, 2018). These relationships cannot be defined using other predictive model such as PolyPhen-2, SIFT, CADD, M-CAP, Envision (Gray et al., 2018), Grantham, as well as POSE (Masica et al., 2015) that were trained with the same dataset (Wang and Balch, 2018).

Figure Legends

Fig. 1. VSP workflow and CFTR organization. (A) VSP workflow. *Variant library:* VSP uses a sparse collection of variants with known position and phenotypical traits. *Spatial organization of the biological data:* We position each of the variant on the x-axis by their normalized position in the linear polypeptide sequence, where the full-length chain is assigned a value of 1.0, referred to as the variant sequence position (VarSeqP). We assigned the value of each variant's feature (in this the TrIdx, see **Results**) to the y-axis. The TrIdx is the normalized ratio of the fraction of variant exported from the ER relative to the total variant level in the cell, where the value of the WT CFTR TrIdx is assigned a value of 1.0. To correlate the sequence position and TrIdx values to CFTR ClCon, the normalized ClCon is assigned to the z-coordinate, where the value of the WT CFTR ClCon is assigned a value of 1.0. *Spatial relationship analysis:* We calculate the 2D distance values based on VarSeqP and TrIdx, as well as the variance of ClCon for all possible 2080 (based on 64 variants) variant pairwise combinations. *Variogram modeling:* The molecular variogram shows the calculated spatial variance and distance values for each comparison. Such modeling quantitatively defines the linear sequence range where the variants co-vary with each other for a given set of functions defined by the y- and z-axis coordinates (TrIdx and ClCon, respectively) until it reaches a plateau that we refer to as the 'molecular range'. Variants with distance relationships extending beyond the range are generally not correlated (Wang and Balch, 2018). *Predictive phenotypic landscape:* In the example shown, GPR uses the molecular variogram to predict all uncharacterized impacts of amino acid residue variation on the ClCon (z-axis) based on the residue position in the polypeptide sequence (x-axis) and the TrIdx (y-axis). *2D mapping of the 3D phenotype landscape:* The 3D phenotype landscape is projected as a 2D heatmap of the z-coordinate ClCon values to generate a 2D phenotype landscape. A confidence interval

(uncertainty) can be assigned to every data point that can be delineated by contour lines in the 2D projection map. *3D functional-structure*: We project the generated ClCon-phenotype landscape values with the highest confidence for all CFTR residues to create a ClCon functional-structure displaying how amino acid variation at each position is impacting the functional-structure relationships. **(B)** Upper panel: domain organization of CFTR. Middle panel: visualization of the 64 variants (dark purple) used in the current study on the open (ATP-bound) structure of CFTR (PDB: 5UAK). Transmembrane domain (TMD)1 (purple), nucleotide binding domain (NBD)1 (orange), R domain (RD), TMD2 (red) and NBD2 (yellow). Lower panel: the 64 listed variants used in the study are distributed as indicated by the domain color (ovals).

Fig. 2. VSP-based ClCon- and TrIdx-phenotype landscapes at 37°C. **(A)** (Left panel) ClCon-phenotype landscape using the VSP-predicted values for TrIdx (y-axis) and ClCon (z-axis) in response to sequence position (VarSeqP, x-axis). ClCon-phenotype landscape overlaid with 5, 10, 25, 50, and 75% uncertainty contour intervals (left pane, gray lines). Uncertainty values less than 25% are of high confidence and cover 95% of known and predicted residues. (Right panel) ClCon functional-structure (PDB: 5UAK) mapped using the highest confidence values for each residue found in the ClCon-phenotype landscape (left panel). **(B)** ‘Grayout’ of panel A ClCon functional-structure showing only residues with less than 50% of WT measured value. The grayout structure highlights in a color gradient the known and predicted unknown residue functions that are likely to contribute to disease due to reduced ClCon. **(C)** (Left panel) trafficking index (TrIdx)-phenotype landscape using the VSP-predicted values for ClCon (y-axis) and TrIdx (z-axis) in response to sequence position (VarSeqP, x-axis) overlaid with the confidence contour intervals. (Right panel) TrIdx-phenotype landscape defined functional-structure (PDB: 5UAK). **(D)** ‘Grayout’ of panel C TrIdx functional-structure showing only residues with less than 50% of WT measured value.

Fig. 3. SCV relationships contributing to the energetics of ER export and trafficking. (A)

TrIdx of variants relative to WT at 37°C (orange) and 27°C (blue) (upper left inset). Variants are ordered from the lowest to the highest TrIdx at 37°C (left panel). Zoom view of variants with trafficking defects <0.8 of WT value (right lower panel). SDS-PAGE of F508del and S492F variants after 24 h at 37°C or 27°C showing the difference between F508del (rescued) and S492F (not rescued) in response to reduced temperature (upper right insets). **(B)** (Left panel) SCV predicted temperature shift (TS) delta (Δ TrIdx (TS Δ TrIdx) phenotype landscape using the VSP-predicted TrIdx at 37°C (y-axis) values of the Δ TrIdx between 27°C and 37°C (z-axis) in response to sequence position (VarSeqP, x-axis) overlaid with the confidence contour intervals (left panel). (Right panel) The TS Δ TrIdx-phenotype landscape defined functional-structure. **(C)** Grayout of panel B TS Δ TrIdx functional-structure showing only residues with more than 50% of relative correction. The grayout structure highlights in a color gradient the known and predicted unknown residue functions that are likely to display TrIdx correction in response to reduced temperature. **(D)** Close-up view of the TrIdx correctability at the interface formed between NBD1/ICL4 including F508, Y563, R560, and D567 (NBD1) and T1064, R1066 (ICL4) in the TS Δ TrIdx functional-structure.

Fig. 4: SCV relationships defining stability at the cell surface. (A)

The % of band C stability relative to WT of each variant at 37°C was assessed after 24 h incubation at 27°C followed by 120 min at 37°C in the presence of CHX to inhibit protein synthesis (upper left, inset). Variants are ordered from the most to the least stable. The red box highlights variants that are destabilized relative to WT stability. SDS-PAGE of N1303K (lower left insert) and R560T variants (upper right insert) following treatment with CHX for 0, 10, 30, 60 and 120 min are shown as examples of WT-like stability and low-stability variants, respectively. **(B)** (Left panel) Stability of band C (StabC)-

phenotype landscape using the VSP-predicted values of band C stability relative to WT (z-axis) relating to TrIdx (y axis) to the sequence position (VarSeqP, x-axis) overlaid with the confidence contour intervals. (Right panel) The StabC-phenotype landscape defined functional-structure. (C) Grayout of panel B, StabC functional-structure showing only residues with less than 50% of WT measured value. The grayout structure highlights in a color gradient the known and predicted unknown residue functions that are likely to contribute to disease due to reduced stability at the cell surface.

Fig. 5. Contribution of SCV to CFTR energetics driving function at the cell surface. (A) % of variant function relative to WT at 37°C (orange) and 27°C (blue); the variants are ordered from the lowest to the highest function measured at 37°C (left panel). Representative FMP traces of N1303K, F508Del, L997F and R75Q variants after 24h at 37°C (orange) or 27°C (blue) (right panel). (B) (Left panel) ClCon-phenotype landscape at 27°C using the VSP-predicted values relating TrIdx (y-axis) and ClCon (z-axis) in the context of sequence position (VarSeqP, x-axis) overlaid with the confidence contour intervals. (Right panel) ClCon-associated functional-structure at 27°C. (C) Grayout of panel B, ClCon functional-structure at 27°C showing only residues with less than 50% of WT measured value. The grayout structure highlights in a color gradient the known and predicted unknown residue functions that are likely to contribute to reduced ClCon at 27°C. (D) (Left panel) Δ of the temperature shift of ClCon (TS Δ ClCon) phenotype landscape using the VSP-predicted values of the Δ ClCon between 27°C and 37°C (z axis) relating to the TrIdx at 37°C in the context of sequence position (VarSeqP, x-axis) overlaid with the confidence contour intervals. (Right panel) TS Δ ClCon-phenotype landscape defined functional-structure. (E) Grayout of panel B, TS Δ ClCon functional-structure showing only residues with more than 50% of relative correction. Maximal Δ ClCon (70% of WT) is set as 1.0

(Fig. 5D, E). The grayout structure highlights in a color gradient the known and predicted unknown residue functions that are likely to display ClCon correction in response to reduced temperature.

Fig. 6: SCV code scoring global response to folding energetics across the entire CFTR fold.

(A) SCV-based linear code representation of the temperature correctability of CFTR TrIdx (ΔTrIdx , upper), ClCon (ΔClCon , middle) and the differential energetics based correction (DEC) plot displayed as the difference value ($\Delta\text{TrIdx}-\Delta\text{ClCon}$) between the predicted ΔTrIdx and ΔClCon phenotype landscape values *relative* to that observed for WT values in response to temperature (Table S2). (B) CFTR DEC functional-structure highlighting regions such as the YKDAD motif that have a higher energetic contribution to the ER export than to the ClCon. (C) In general, combining the energetic features of the fold with functional assays using only a sparse collection of variants in the population and GPR-ML based VSP for any protein (generic example illustrated), we can determine the DEC for any residue relative to that observed for the WT protein as an SCV principled description of the physiological protein fold. The color scales reflect the differential response of the energetics of each residue to either stability or function which together contribute to the biology of the fold. (D) An SCV-based QS system dictating the biology of cell through genetic and epigenetic design of the genome in response to cell autonomous and non-autonomous signaling pathways. QS operates hierarchically with QC focused on the energetics of fold stability based on the physical-chemical properties of genome-encoded protein sequence that are independent of function. QA assesses the function of the fold throughout the cell in the context of proteostasis and physical-chemical environments unique to its compartmentalized organization, whereas QS encompasses the signaling pathways in genome design required to differentially manage biology in response to the environment. The proposed hierarchy of covariant relationships

816 are defined by an ancient SCV principled language base built on biological non-equilibrium
817 thermodynamic physical-chemical principles (black solid circle at base) (Wang and Balch, 2018).

818

819

Supplementary Figure Legends

Fig. S1. SDS-PAGE blots of variant response to temperature shift and CHX chase. SDS-PAGE blots showing distribution of variant band B and C glycoforms. HEK293 cells were transfected with pBI-CMV2 empty vector or vectors containing WT-CFTR or any of the 64 CFTR variants tested and incubated for 48 h at 37°C. The cells were subsequently incubated at 27°C for 24 h. After 24 h at 27°C, HEK293 cells transiently expressing CFTR were washed and incubated at 37°C in regular medium supplemented with 50 μ M of cycloheximide (CHX) final concentration for 0, 10, 30, 60, 120, 240 min. Raw data used to generate the TrIdx at 37°C and 27°C and the stability of band C (StabC) at 37°C of each variant are shown.

Fig. S2. Comparison of TrIdx measured in FRT, HeLa (McCague et al., 2019; Sharma et al., 2018; Sosnay and Cutting, 2014; Sosnay et al., 2013b) and HEK293 cells for each variant at 37°C. **(A)** Shown is the % of TrIdx (band C/total (B + C) ratio) in HEK293, FRT and HeLa cells relative to WT for the 64 variants included in the study. **(B)** Comparison of TrIdx for each variant in FRT and HEK293 cells. **(C)** Comparison of TrIdx for each variant in HeLa and HEK293 cells. A comparison of the normalized TrIdx obtained from HEK293 cells to the TrIdx calculated from previously published data from FRT and HeLa cells revealed PCCs of 0.94 and 0.92, respectively, suggesting that the TrIdx of CFTR variants does not differ significantly between highly divergent heterologous cell lines.

Fig. S3. FMP measurement of variants function at 37°C and 27°C. Representative FMP traces of the null (-/-), WT and the 64 variants at 37°C and 27°C. HEK293 cells were transfected with pBI-CMV2 empty vector or vectors containing WT-CFTR or any of the 64 CFTR variants tested and incubated for 48 h at 37°C. The cells were subsequently incubated at 27°C for 24 h. After 24

h at 27°C, HEK293 cells transiently expressing CFTR were washed twice in phosphate-buffered saline and incubated for 30 min with the blue membrane potential dye dissolved in a modified PBS at 37°C. The plates were read in a fluorescence plate reader at 37°C following a dye equilibration period of 30 min. CFTR was stimulated with an activation cocktail (10 μ M forskolin and 50 μ M genistein; final concentration) diluted in the same buffer. Fluorescence was monitored every 0.5 second for a total of 30 sec. CFTR-mediated depolarization of the membrane was monitored as an increase in fluorescence and hyperpolarization or repolarization as a decrease. Raw data used to generate the ClCon at 37°C and 27°C of each variant.

Fig. S4. Comparison of function in FRT(McCague et al., 2019; Sharma et al., 2018; Sosnay and Cutting, 2014; Sosnay et al., 2013b) and HEK293 (HEK) cells for each variant at 37°C. (A) Shown is the % of the ClCon for each variant in HEK293 and FRT cells relative to WT. (B) Comparison of the ClCon for each variant observed in HEK293 and FRT cells. A comparison of the normalized ClCon between HEK293 cells and FRT cells(Sosnay et al., 2013a) yielded a PCC score of 0.80, demonstrating that the basic functionality of human variants is conserved between highly divergent heterologous cell lines.

Fig. S5. VSP prediction validation by external datasets. (A) Table representing the 21 variants from two separate studies (Van Goor et al., 2014; Yu et al., 2012) used for validation and their measured and VSP predicted phenotype traits (i.e., TrIdx and ClCon). (B, D) Position of the 21 external variants on the ClCon and TrIdx landscapes, respectively. These validation variants have diverse relationships between TrIdx and ClCon. They include variants with a deficient TrIdx and no ClCon, as well as a wide range of TrIdx values associated with variable ClCon. The variants in the validation dataset are distributed across the entire CFTR sequence. (C, E) PCC-value, and the

p-value (ANOVA test) with null hypothesis with the coefficient equal to zero between the measured and predicted ClCon and TrIdx, respectively.

Fig. S6. Correlation between $\Delta\Delta G^\circ$ and CFTR trafficking and function. (A) $\Delta\Delta G^\circ$ between WT-CFTR and each CFTR variant included in the study at 37°C. $\Delta\Delta G^\circ$ is calculated using FoldX (Buß et al., 2018; Buss et al., 2018b; Delgado et al., 2019). (B) Correlation between the FoldX calculated $\Delta\Delta G^\circ$ and the *in vivo* measured CFTR TrIdx. (C) Correlation between the FoldX calculated $\Delta\Delta G^\circ$ and the *in vivo* CFTR ClCon.

Fig. S7. Correlation between variant trafficking defect from the ER and stability at the cell surface. Representation of variant TrIdx (x-axis) and stability at the cell surface (y-axis) of the 64 variants included in the study. A comparison of the normalized TrIdx obtained from HEK293 cells (Fig. 2A, upper panel) to the StabC (Fig. 4A), both calculated from Fig. S1, revealed a Pearson correlation coefficients (PCC) of 0.87, suggesting that the TrIdx from the ER and the stability at the cell surface of CFTR variants is highly correlated.

Fig. S8. Molecular variograms describing the behavior of CFTR variants at 37°C and 27°C. The spatial relationships of all possible variant pairwise combinations representing the spatial variance of ClCon and the distance value defined by VarSeqP and TrIdx are plotted (left panels, black dots for 37°C and red dots for 27°C) and then binned by a distance interval of 0.01 (left panels, green dots for 37°C and blue dots for 27°C, mean $\pm$ SEM). Statistical comparisons of the global spatial variation and global distance are shown as box and whisker plots at the right and top margins respectively (box = 25th and 75th with median line inside the box, whisker length = outmost data point in the inner fence, square = mean, asterisks = min and max; p-value by Student's two tailed t-test). Variograms (right panels, black (37°C) and red (27°C) lines) were modeled by

using the binned spatial variance values. The plateau and the range values of the variograms are indicated. In the variograms correlating the distance of VarSeq and TrIdx to the spatial variance of ClCon (A), the plateau value of 27°C is higher than that in 37°C (right panel), reflecting the fact that the global diversity of ClCon value is increased in 27°C (left panel, $p = 0.04$) as the ClCon of many variants at 37°C are clustered at '0'. Shifting the temperature to 27°C not only rescues many of these variants with '0' ClCon, but also increases the functional value of some variants with ClCon. For example, the ClCon of G1069 is increased from 113% of WT at 37°C to 151% of WT at 27°C, which expands the range and diversity of ClCon at 27°C. The higher ClCon diversity at 27°C compared to that at 37°C suggests that a lower energetic state can expand the ClCon spectrum of CFTR variants, probably due to changes in stability or folding dynamics involved in ClCon. In contrast, both the global distance based on VarSeq and TrIdx (left panel, $p = 5 \times 10^{-6}$) and the local distance range of VarSeq-TrIdx relationships correlating spatial variance of ClCon (right panel, 0.05 at 27°C versus 0.08 at 37°C) are shorter at 27°C than that observed at 37°C, indicating the distance values of TrIdx between variants are compacted by low temperature. Unlike ClCon where many variants clustered at '0' under 37°C, TrIdx distribution at 37°C is broader since there are variants with '0' ClCon but have significantly expanded TrIdx. Shifting the temperature to 27°C improves the TrIdx for some variants with low TrIdx at 37°C but did not change the variants with high TrIdx, so the overall distance or diversity of TrIdx between variants is reduced by the low temperature.

Fig. S7. Optimization of variant expression and production. (A) Relative GFP fluorescence monitored after transfection of HEK293 cells with 0, 12, 25, 50 100 or 200 ng of pBI-CMV2-WT-CFTR or pBI-CMV2-F508del-CFTR after 24, 48 or 72 h. (B and C, upper panel). SDS-PAGE blots showing distribution of WT and F508del band B and C glycoforms as well as GFP. HEK293

cells were transfected with 0, 12, 25, 50, 100, 200 ng of pBI-CMV2 vectors containing WT-CFTR or F508del-CFTR and incubated for 48 h at 37°C. (**B** and **C**, lower panel) Quantification of the TrIdx (band C/Total (band B + band C)) (C/B+C) ratio relative to WT). We observe a similar level of GFP-fluorescence at a given time-point when HEK293 cells were transfected with the same quantity of either pBI-CMV2-WT-CFTR or pBI-CMV2-F508Del-CFTR. In all cases, the level of GFP- fluorescence and the level of CFTR generated by either the pBI-CMV2-WT-CFTR or pBI-CMV2-F508del-CFTR using immunoblotting were directly comparable. These data validate the use of GFP fluorescence as a metric to ensure uniform transfection efficiencies across the range of variants tested.

Acknowledgements: Grant support provided by NIH HL095524, DK051870 and HL141810 to WEB. FA is a recipient of a CFFT Postdoctoral Fellowship. CW is a past recipient of a CFFT Postdoctoral Fellowship. We thank P. Thomas (UT Southwestern Medical School) for providing vectors encoding CFTR2 variants. **Author contributions:** FA designed and performed the experiments; FA, CW and WEB contributed to concepts and wrote the manuscript. **Competing interests:** The authors declare no competing interests. The authors declare no advisory, management, or consulting positions. **Data Availability:** All trafficking and functional data for each CFTR variants described in this project can be found in the Supplementary Information. All phenotype landscape outputs are available as a Zip file named “Landscapes_Output.Zip”. **Code availability:** A demo input data, the R-codes for generating phenotype landscape and instructions were previously published³ and are available as a Zip file named “ExampleCode.Zip”, which is uploaded to <https://doi.org/10.17632/ycw667nv5f.1>.

References Cited

- Ahner, A., Gong, X., Schmidt, B.Z., Peters, K.W., Rabeh, W.M., Thibodeau, P.H., Lukacs, G.L., and Frizzell, R.A. (2013). Small heat shock proteins target mutant cystic fibrosis transmembrane conductance regulator for degradation via a small ubiquitin-like modifier-dependent pathway. *Mol Biol Cell* 24, 74-84.
- Ahner, A., Nakatsukasa, K., Zhang, H., Frizzell, R.A., and Brodsky, J.L. (2007). Small heat-shock proteins select $\Delta F508$ -CFTR for endoplasmic reticulum-associated degradation. *Molecular biology of the cell* 18, 806-814.
- Aleksandrov, A.A., Kota, P., Cui, L., Jensen, T., Alekseev, A.E., Reyes, S., He, L., Gentzsch, M., Aleksandrov, L.A., and Dokholyan, N.V. (2012). Allosteric modulation balances thermodynamic stability and restores function of $\Delta F508$ CFTR. *Journal of molecular biology* 419, 41-60.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230.
- Anglès, F., Hutt, D.M., and Balch, W.E. (2019). HDAC inhibitors rescue multiple disease-causing CFTR variants. *Human molecular genetics* 28, 1982-2000.
- Bagdany, M., Veit, G., Fukuda, R., Avramescu, R.G., Okiyoneda, T., Baaklini, I., Singh, J., Sovak, G., Xu, H., Apaja, P.M., *et al.* (2017). Chaperones rescue the energetic landscape of mutant CFTR at single molecule and in cell. *Nat Commun* 8, 398.
- Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. *science* 319, 916-919.
- Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2016). In vivo aspects of protein folding and quality control. *Science* 353, aac4354.
- Bethune, J., and Wieland, F.T. (2018). Assembly of COPI and COPII Vesicular Coat Proteins on Membranes. *Annu Rev Biophys* 47, 63-83.
- Bose, S.J., Bijvelds, M.J.C., Wang, Y., Liu, J., Cai, Z., Bot, A.G.M., de Jonge, H.R., and Sheppard, D.N. (2019). Differential thermostability and response to cystic fibrosis transmembrane conductance regulator potentiators of human and mouse F508del-CFTR. *Am J Physiol Lung Cell Mol Physiol* 317, L71-L86.
- Box, G., and Narasimhan, S. (2010). Rethinking statistics for quality control. *Quality Engineering* 22, 60-72.
- Buss, O., Muller, D., Jager, S., Rudat, J., and Rabe, K.S. (2018a). Improvement in the Thermostability of a beta-Amino Acid Converting omega-Transaminase by Using FoldX. *Chembiochem* 19, 379-387.

963 Buß, O., Rudat, J., and Ochsenreither, K. (2018). FoldX as protein engineering tool: better than
964 random based approaches? *Computational and structural biotechnology journal* 16, 25-33.

965 Buss, O., Rudat, J., and Ochsenreither, K. (2018b). FoldX as Protein Engineering Tool: Better
966 Than Random Based Approaches? *Comput Struct Biotechnol J* 16, 25-33.

967 Chiles, J.-P., and Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty*. 497.

968 Claussnitzer, M., Judy H. Cho, Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., Kathiresan,
969 S., Kenny, E.E., Lindgren, C.M., MacArthur, D.G., *et al.* (2020). A brief history of human disease
970 genetics. *Nature* 577, 179-189.

971 Costa-Mattioli, M., and Walter, P. (2020). The integrated stress response: From mechanism to
972 disease. *Science* 368.

973 Cui, L., Aleksandrov, L., Chang, X.B., Hou, Y.X., He, L., Hegedus, T., Gentzsch, M.,
974 Aleksandrov, A., Balch, W.E., and Riordan, J.R. (2007). Domain interdependence in the
975 biosynthetic assembly of CFTR. *J Mol Biol* 365, 981-994.

976 Cutting, G.R. (2015). Cystic fibrosis genetics: from molecular understanding to clinical
977 application. *Nature Reviews Genetics* 16, 45-56.

978 Dawson, J.E., Farber, P.J., and Forman-Kay, J.D. (2013). Allosteric coupling between the
979 intracellular coupling helix 4 and regulatory sites of the first nucleotide-binding domain of CFTR.
980 *PLoS One* 8, e74347.

981 Delgado, J., Radusky, L.G., Cianferoni, D., and Serrano, L. (2019). FoldX 5.0: working with RNA,
982 small molecules and a new graphical interface. *Bioinformatics* 35, 4168-4169.

983 Denning, G.M., Anderson, M.P., Amara, J.F., Marshall, J., Smith, A.E., and Welsh, M.J. (1992).
984 Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-
985 sensitive. *nature* 358, 761.

986 Du, K., Sharma, M., and Lukacs, G.L. (2005). The $\Delta F508$ cystic fibrosis mutation impairs domain-
987 domain interactions and arrests post-translational folding of CFTR. *Nature Structural and*
988 *Molecular Biology* 12, 17.

989 El Khouri, E., Le Pavec, G., Toledano, M.B., and Delaunay-Moisan, A. (2013). RNF185 is a novel
990 E3 ligase of endoplasmic reticulum-associated degradation (ERAD) that targets cystic fibrosis
991 transmembrane conductance regulator (CFTR). *Journal of Biological Chemistry* 288, 31177-
992 31191.

993 Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nat Rev Mol*
994 *Cell Biol* 4, 181-191.

995 Ennis, H., and Lubin, M. (1964). Cycloheximide: aspects of inhibition of protein synthesis in
996 mammalian cells. *Science* *146*, 1474-1476.

997 Fang, X., and Wang, J. (2020). Nonequilibrium Thermodynamics in Cell Biology: Extending
998 Equilibrium Formalism to Cover Living Systems. *Annu Rev Biophys* *49*, 227-246.

999 Farinha, C.M., King-Underwood, J., Sousa, M., Correia, A.R., Henriques, B.J., Roxo-Rosa, M.,
1000 Da Paula, A.C., Williams, J., Hirst, S., and Gomes, C.M. (2013). Revertants, low temperature, and
1001 correctors reveal the mechanism of F508del-CFTR rescue by VX-809 and suggest multiple agents
1002 for full correction. *Chemistry & biology* *20*, 943-955.

1003 Gong, X., Ahner, A., Roldan, A., Lukacs, G.L., Thibodeau, P.H., and Frizzell, R.A. (2016). Non-
1004 native conformers of cystic fibrosis transmembrane conductance regulator NBD1 are recognized
1005 by Hsp27 and conjugated to SUMO-2 for degradation. *Journal of Biological Chemistry* *291*, 2004-
1006 2017.

1007 Gong, X., Liao, Y., Ahner, A., Larsen, M.B., Wang, X., Bertrand, C.A., and Frizzell, R.A. (2019).
1008 Different SUMO paralogues determine the fate of wild-type and mutant CFTRs: biogenesis versus
1009 degradation. *Mol Biol Cell* *30*, 4-16.

1010 Grandjean, J.M.D., Plate, L., Morimoto, R.I., Bollong, M.J., Powers, E.T., and Wiseman, R.L.
1011 (2019). Deconvoluting Stress-Responsive Proteostasis Signaling Pathways for Pharmacologic
1012 Activation Using Targeted RNA Sequencing. *ACS Chem Biol* *14*, 784-795.

1013 Gray, V.E., Hause, R.J., Luebeck, J., Shendure, J., and Fowler, D.M. (2018). Quantitative
1014 Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. *Cell Syst* *6*, 116-124
1015 e113.

1016 Gürkan, C., Stagg, S.M., LaPointe, P., and Balch, W.E. (2006). The COPII cage: unifying
1017 principles of vesicle coat assembly. *Nature Reviews Molecular Cell Biology* *7*, 727.

1018 Hammond, C., Braakman, I., and Helenius, A. (1994). Role of N-linked oligosaccharide
1019 recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc Natl*
1020 *Acad Sci U S A* *91*, 913-917.

1021 Han, S.T., Rab, A., Pellicore, M.J., Davis, E.F., McCague, A.F., Evans, T.A., Joynt, A.T., Lu, Z.,
1022 Cai, Z., Raraigh, K.S., *et al.* (2018). Residual function of cystic fibrosis mutants predicts response
1023 to small molecule CFTR modulators. *JCI Insight* *3*.

1024 Hanson, S.R., Culyba, E.K., Hsu, T.L., Wong, C.H., Kelly, J.W., and Powers, E.T. (2009). The
1025 core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances
1026 stability. *Proc Natl Acad Sci U S A* *106*, 3131-3136.

1027 Hoffmann, N., Lee, B., Hentzer, M., Rasmussen, T.B., Song, Z., Johansen, H.K., Givskov, M., and
1028 Høiby, N. (2007). Azithromycin blocks quorum sensing and alginate polymer formation and
1029 increases the sensitivity to serum and stationary-growth-phase killing of *Pseudomonas aeruginosa*

1030 and attenuates chronic *P. aeruginosa* lung infection in *Cftr*^{-/-} mice. *Antimicrobial agents and*
1031 *chemotherapy* *51*, 3677-3687.

1032 Hutt, D.M., Loguercio, S., Campos, A.R., and Balch, W.E. (2018a). A proteomic variant approach
1033 (ProVarA) for personalized medicine of inherited and somatic disease. *Journal of molecular*
1034 *biology* *430*, 2951-2973.

1035 Hutt, D.M., Loguercio, S., Roth, D.M., Su, A.I., and Balch, W.E. (2018b). Correcting the F508del-
1036 CFTR variant by modulating eukaryotic translation initiation factor 3-mediated translation
1037 initiation. *J Biol Chem* *293*, 13477-13495.

1038 Hutt, D.M., Mishra, S.K., Roth, D.M., Larsen, M.B., Angles, F., Frizzell, R.A., and Balch, W.E.
1039 (2018c). Silencing of the Hsp70-specific nucleotide-exchange factor BAG3 corrects the F508del-
1040 CFTR variant by restoring autophagy. *J Biol Chem* *293*, 13682-13695.

1041 Isaaks, E.H., and Srivastava, M.R. (1989). *Applied geostatistics*.

1042 Jayaraj, G.G., Hipp, M.S., and Hartl, F.U. (2020). Functional Modules of the Proteostasis Network.
1043 *Cold Spring Harb Perspect Biol* *12*.

1044 Karagoz, G.E., Acosta-Alvear, D., and Walter, P. (2019). The Unfolded Protein Response:
1045 Detecting and Responding to Fluctuations in the Protein-Folding Capacity of the Endoplasmic
1046 Reticulum. *Cold Spring Harb Perspect Biol* *11*.

1047 Kotha, K., and Clancy, J.P. (2013). Ivacaftor treatment of cystic fibrosis patients with the G551D
1048 mutation: a review of the evidence. *Ther Adv Respir Dis* *7*, 288-296.

1049 Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J.,
1050 Hoffman, D., and Hoover, J. (2015). ClinVar: public archive of interpretations of clinically
1051 relevant variants. *Nucleic acids research* *44*, D862-D868.

1052 Li, J., Labbadia, J., and Morimoto, R.I. (2017). Rethinking HSF1 in Stress, Development, and
1053 Organismal Health. *Trends Cell Biol* *27*, 895-905.

1054 Liu, F., Zhang, Z., Csanády, L., Gadsby, D.C., and Chen, J. (2017). Molecular structure of the
1055 human CFTR ion channel. *Cell* *169*, 85-95. e88.

1056 Lord, C., Ferro-Novick, S., and Miller, E.A. (2013). The highly conserved COPII coat complex
1057 sorts cargo from the endoplasmic reticulum and targets it to the golgi. *Cold Spring Harb Perspect*
1058 *Biol* *5*.

1059 Lukacs, G.L., Chang, X.B., Bear, C., Kartner, N., Mohamed, A., Riordan, J.R., and Grinstein, S.
1060 (1993). The delta F508 mutation decreases the stability of cystic fibrosis transmembrane
1061 conductance regulator in the plasma membrane. Determination of functional half-lives on
1062 transfected cells. *J Biol Chem* *268*, 21592-21598.

1063 Maitra, R., Sivashanmugam, P., and Warner, K. (2013). A rapid membrane potential assay to
1064 monitor CFTR function and inhibition. *Journal of biomolecular screening* 18, 1132-1137.

1065 Masica, D.L., Sosnay, P.R., Raraigh, K.S., Cutting, G.R., and Karchin, R. (2015). Missense
1066 variants in CFTR nucleotide-binding domains predict quantitative phenotypes associated with
1067 cystic fibrosis disease severity. *Hum Mol Genet* 24, 1908-1917.

1068 McCague, A.F., Raraigh, K.S., Pellicore, M.J., Davis-Marcisak, E.F., Evans, T.A., Han, S.T., Lu,
1069 Z., Joynt, A.T., Sharma, N., Castellani, C., *et al.* (2019). Correlating Cystic Fibrosis
1070 Transmembrane Conductance Regulator Function with Clinical Features to Inform Precision
1071 Treatment of Cystic Fibrosis. *Am J Respir Crit Care Med* 199, 1116-1126.

1072 Mendoza, J.L., Schmidt, A., Li, Q., Nuvaga, E., Barrett, T., Bridges, R.J., Feranchak, A.P.,
1073 Brautigam, C.A., and Thomas, P.J. (2012). Requirements for efficient correction of DeltaF508
1074 CFTR revealed by analyses of evolved sequences. *Cell* 148, 164-174.

1075 Meng, X., Clews, J., Kargas, V., Wang, X., and Ford, R.C. (2017). The cystic fibrosis
1076 transmembrane conductance regulator (CFTR) and its stability. *Cellular and Molecular Life*
1077 *Sciences* 74, 23-38.

1078 Meriin, A.B., Mense, M., Colbert, J.D., Liang, F., Bihler, H., Zaarur, N., Rock, K.L., and Sherman,
1079 M.Y. (2012). A novel approach to recovery of function of mutant proteins by slowing down
1080 translation. *Journal of Biological Chemistry* 287, 34264-34272.

1081 Moran, O. (2017). The gating of the CFTR channel. *Cellular and molecular life sciences* 74, 85-
1082 92.

1083 Morimoto, R.I. (2019). Cell-Nonautonomous Regulation of Proteostasis in Aging and Disease.
1084 *Cold Spring Harb Perspect Biol.*

1085 Nishimura, N., and Balch, W.E. (1997). A di-acidic signal required for selective export from the
1086 endoplasmic reticulum. *Science* 277, 556-558.

1087 Nishimura, N., Bannykh, S., Slabough, S., Matteson, J., Altschuler, Y., Hahn, K., and Balch, W.E.
1088 (1999). A di-acidic (DXE) code directs concentration of cargo during export from the endoplasmic
1089 reticulum. *J Biol Chem* 274, 15937-15946.

1090 Okiyoneda, T., Barrière, H., Bagdány, M., Rabeh, W.M., Du, K., Höhfeld, J., Young, J.C., and
1091 Lukacs, G.L. (2010). Peripheral protein quality control removes unfolded CFTR from the plasma
1092 membrane. *Science* 329, 805-810.

1093 Oliver, K.E., Rauscher, R., Mijnders, M., Wang, W., Wolpert, M.J., Maya, J., Sabusap, C.M.,
1094 Kesterson, R.A., Kirk, K.L., Rab, A., *et al.* (2019). Slowing ribosome velocity restores folding and
1095 function of mutant CFTR. *J Clin Invest* 129, 5236-5253.

1096 Oliver, M.A., and Webster, R. (2015). Basic steps in geostatistics: the variogram and kriging, Vol
1097 106 (Springer).

1098 Ornes, S. (2017). Core Concept: How nonequilibrium thermodynamics speaks to the mystery of
1099 life. *Proc Natl Acad Sci U S A* *114*, 423-424.

1100 Pankow, S., Bamberger, C., and Yates, J.R. (2019). A posttranslational modification code for
1101 CFTR maturation is altered in cystic fibrosis. *Sci Signal* *12*, eaan7984.

1102 Patrick, A.E., Karamyshev, A.L., Millen, L., and Thomas, P.J. (2011). Alteration of CFTR
1103 transmembrane span integration by disease-causing mutations. *Molecular biology of the cell* *22*,
1104 4461-4471.

1105 Peotter, J., Kasberg, W., Pustova, I., and Audhya, A. (2019). COPII-mediated trafficking at the
1106 ER/ERGIC interface. *Traffic* *20*, 491-503.

1107 Powers, E.T., and Balch, W.E. (2013). Diversity in the origins of proteostasis networks--a driver
1108 for protein function in evolution. *Nat Rev Mol Cell Biol* *14*, 237-248.

1109 Proctor, E.A., Kota, P., Aleksandrov, A.A., He, L., Riordan, J.R., and Dokholyan, N.V. (2015).
1110 Rational Coupled Dynamics Network Manipulation Rescues Disease-Relevant Mutant Cystic
1111 Fibrosis Transmembrane Conductance Regulator. *Chem Sci* *6*, 1237-1246.

1112 Protasevich, I., Yang, Z., Wang, C., Atwell, S., Zhao, X., Emtage, S., Wetmore, D., Hunt, J.F., and
1113 Brouillette, C.G. (2010). Thermal unfolding studies show the disease causing F508del mutation in
1114 CFTR thermodynamically destabilizes nucleotide-binding domain 1. *Protein Science* *19*, 1917-
1115 1931.

1116 Qu, B.H., and Thomas, P.J. (1996). Alteration of the cystic fibrosis transmembrane conductance
1117 regulator folding pathway. *J Biol Chem* *271*, 7261-7264.

1118 Rabeh, W.M., Bossard, F., Xu, H., Okiyoneda, T., Bagdany, M., Mulvihill, C.M., Du, K., di
1119 Bernardo, S., Liu, Y., Konermann, L., *et al.* (2012). Correction of both NBD1 energetics and
1120 domain interface is required to restore DeltaF508 CFTR folding and function. *Cell* *148*, 150-163.

1121 Roth, D.M., Hutt, D.M., Tong, J., Bouchecareilh, M., Wang, N., Seeley, T., Dekkers, J.F.,
1122 Beekman, J.M., Garza, D., Drew, L., *et al.* (2014). Modulation of the maladaptive stress response
1123 to manage diseases of protein folding. *PLoS Biol* *12*, e1001998.

1124 Routledge, K.E., Gupta, V., and Balch, W.E. (2010). Emergent properties of proteostasis-COPII
1125 coupled systems in human health and disease. *Mol Membr Biol* *27*, 385-397.

1126 Saint-Criq, V., and Gray, M.A. (2017). Role of CFTR in epithelial physiology. *Cellular and*
1127 *Molecular Life Sciences* *74*, 93-115.

1128 Sala, A.J., Bott, L.C., and Morimoto, R.I. (2017). Shaping proteostasis at the cellular, tissue, and
1129 organismal level. *J Cell Biol* 216, 1231-1241.

1130 Schmidt, B.Z., Watts, R.J., Aridor, M., and Frizzell, R.A. (2009). Cysteine string protein promotes
1131 proteasomal degradation of the cystic fibrosis transmembrane conductance regulator (CFTR) by
1132 increasing its interaction with the C terminus of Hsp70-interacting protein and promoting CFTR
1133 ubiquitylation. *J Biol Chem* 284, 4168-4178.

1134 Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Zidek, A., Nelson,
1135 A.W.R., Bridgland, A., *et al.* (2020). Improved protein structure prediction using potentials from
1136 deep learning. *Nature* 577, 706-710.

1137 Sharma, N., and Cutting, G.R. (2019). The genetics and genomics of cystic fibrosis. *J Cyst Fibros.*

1138 Sharma, N., Evans, T.A., Pellicore, M.J., Davis, E., Aksit, M.A., McCague, A.F., Joynt, A.T., Lu,
1139 Z., Han, S.T., Anzmann, A.F., *et al.* (2018). Capitalizing on the heterogeneous effects of CFTR
1140 nonsense and frameshift variants to inform therapeutic strategy for cystic fibrosis. *PLoS Genet* 14,
1141 e1007723.

1142 Sosnay, P.R., and Cutting, G.R. (2014). Interpretation of genetic variants. *Thorax* 69, 295-297.

1143 Sosnay, P.R., Siklosi, K.R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., Ramalho, A.S.,
1144 Amaral, M.D., Dorfman, R., and Zielenski, J. (2013a). Defining the disease liability of variants in
1145 the cystic fibrosis transmembrane conductance regulator gene. *Nature genetics* 45, 1160.

1146 Sosnay, P.R., Siklosi, K.R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., Ramalho, A.S.,
1147 Amaral, M.D., Dorfman, R., Zielenski, J., *et al.* (2013b). Defining the disease liability of variants
1148 in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 45, 1160-1167.

1149 Soya, N., Roldan, A., and Lukacs, G.L. (2019). Differential Scanning Fluorimetry and Hydrogen
1150 Deuterium Exchange Mass Spectrometry to Monitor the Conformational Dynamics of NBD1 in
1151 Cystic Fibrosis. *Methods Mol Biol* 1873, 53-67.

1152 Spear, J.M., Koborssy, D.A., Schwartz, A.B., Johnson, A.J., Audhya, A., Fadool, D.A., and Stagg,
1153 S.M. (2015). Kv1.3 contains an alternative C-terminal ER exit motif and is recruited into COPII
1154 vesicles by Sec24a. *BMC Biochem* 16, 16.

1155 Thomas, P.J., Qu, B.H., and Pedersen, P.L. (1995). Defective protein folding as a basis of human
1156 disease. *Trends Biochem Sci* 20, 456-459.

1157 Torkamani, A., Andersen, K.G., Steinhubl, S.R., and Topol, E.J. (2017). High-Definition
1158 Medicine. *Cell* 170, 828-843.

1159 Van Goor, F., Yu, H., Burton, B., and Hoffman, B.J. (2014). Effect of ivacaftor on CFTR forms
1160 with missense mutations associated with defects in protein processing or function. *Journal of*
1161 *Cystic Fibrosis* 13, 29-36.

1162 Veit, G., Avramescu, R.G., Chiang, A.N., Houck, S.A., Cai, Z., Peters, K.W., Hong, J.S., Pollard,
 1163 H.B., Guggino, W.B., Balch, W.E., *et al.* (2016a). From CFTR biology toward combinatorial
 1164 pharmacotherapy: expanded classification of cystic fibrosis mutations. *Mol Biol Cell* 27, 424-433.

1165 Veit, G., Oliver, K., Apaja, P.M., Perdomo, D., Bidaud-Meynard, A., Lin, S.T., Guo, J., Icyuz, M.,
 1166 Sorscher, E.J., Hartman, J.I., *et al.* (2016b). Ribosomal Stalk Protein Silencing Partially Corrects
 1167 the DeltaF508-CFTR Functional Expression Defect. *PLoS Biol* 14, e1002462.

1168 Vernon, R.M., Chong, P.A., Lin, H., Yang, Z., Zhou, Q., Aleksandrov, A.A., Dawson, J.E.,
 1169 Riordan, J.R., Brouillette, C.G., Thibodeau, P.H., *et al.* (2017). Stabilization of a nucleotide-
 1170 binding domain of the cystic fibrosis transmembrane conductance regulator yields insight into
 1171 disease-causing mutations. *J Biol Chem* 292, 14147-14164.

1172 Wang, C., Angles, F., and Balch, W.E. (2021). Spatial Covariant Design Shapes Information Flow
 1173 Through and Across Central Dogma. *Cell Submitted*.

1174 Wang, C., and Balch, W.E. (2018). Bridging genomics to phenomics at atomic resolution through
 1175 variation spatial profiling. *Cell reports* 24, 2013-2028. e2016.

1176 Wang, C., Scott, S.M., Subramanian, K., Loguercio, S., Zhao, P., Hutt, D.M., Farhat, N.Y., Porter,
 1177 F.D., and Balch, W.E. (2019a). Quantitating the epigenetic transformation contributing to
 1178 cholesterol homeostasis using Gaussian process. *Nature communications* 10, 1-15.

1179 Wang, C., Scott, S.M., Sun, S., Zhao, P., Hutt, D.M., Shao, H., Gestwicki, J.E., and Balch, W.E.
 1180 (2019b). Individualized management of genetic diversity in Niemann-Pick C1 through modulation
 1181 of the Hsp70 chaperone system. *Human molecular genetics*.

1182 Wang, X., Koulov, A.V., Kellner, W.A., Riordan, J.R., and Balch, W.E. (2008). Chemical and
 1183 biological folding contribute to temperature-sensitive DeltaF508 CFTR trafficking. *Traffic* 9,
 1184 1878-1893.

1185 Wang, X., Matteson, J., An, Y., Moyer, B., Yoo, J.-S., Bannykh, S., Wilson, I.A., Riordan, J.R.,
 1186 and Balch, W.E. (2004a). COPII-dependent export of cystic fibrosis transmembrane conductance
 1187 regulator from the ER uses a di-acidic exit code. *The Journal of cell biology* 167, 65-74.

1188 Wang, X., Matteson, J., An, Y., Moyer, B., Yoo, J.S., Bannykh, S., Wilson, I.A., Riordan, J.R.,
 1189 and Balch, W.E. (2004b). COPII-dependent export of cystic fibrosis transmembrane conductance
 1190 regulator from the ER uses a di-acidic exit code. *J Cell Biol* 167, 65-74.

1191 Wang, X., Venable, J., LaPointe, P., Hutt, D.M., Koulov, A.V., Coppinger, J., Gurkan, C., Kellner,
 1192 W., Matteson, J., Plutner, H., *et al.* (2006). Hsp90 cochaperone Aha1 downregulation rescues
 1193 misfolding of CFTR in cystic fibrosis. *Cell* 127, 803-815.

1194 Wang, Y., Cai, Z., Gosling, M., and Sheppard, D.N. (2018). Potentiation of the cystic fibrosis
 1195 transmembrane conductance regulator Cl(-) channel by ivacaftor is temperature independent. *Am*
 1196 *J Physiol Lung Cell Mol Physiol* 315, L846-L857.

1197 Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., and Balch, W.E. (2007). An adaptable
1198 standard for protein export from the endoplasmic reticulum. *Cell* *131*, 809-821.

1199 Yu, H., Burton, B., Huang, C.J., Worley, J., Cao, D., Johnson, J.P., Jr., Urrutia, A., Joubran, J.,
1200 Seepersaud, S., Sussky, K., *et al.* (2012). Ivacaftor potentiation of multiple CFTR channels with
1201 gating mutations. *J Cyst Fibros* *11*, 237-245.

1202 Zhang, X., Dong, C., Wu, Q.J., Balch, W.E., and Wu, G. (2011). Di-acidic motifs in the membrane-
1203 distal C termini modulate the transport of angiotensin II receptors from the endoplasmic reticulum
1204 to the cell surface. *J Biol Chem* *286*, 20525-20535.

1205 Zhang, Y., and Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *J*
1206 *Econometrics* *187*, 95-112.

1207 Zhang, Z., Liu, F., and Chen, J. (2017). Conformational Changes of CFTR upon Phosphorylation
1208 and ATP Binding. *Cell* *170*, 483-491 e488.

1209

A

Input: Variants library

Spatial organization of biological data

Spatial relationship analyses

3D-Functional structure

2D mapping of the landscape

Predictive phenotypic landscape

Variogram modeling

Output

B

Genetic map of the *SLC12A2* gene showing exons (ECL1-ECL6) and introns (ICL1-ICL4). The map is color-coded by domain: TMD1 (purple), NBD1 (orange), TMD2 (pink), NBD2 (yellow), and ECL1 (green). The protein structure is shown below the map, with domains and residues labeled. The structure is color-coded to match the genetic map: TMD1 (purple), NBD1 (orange), TMD2 (pink), NBD2 (yellow), and ECL1 (green). The structure is shown in a ribbon representation, with residues labeled with their amino acid codes and positions. The structure is shown in a ribbon representation, with residues labeled with their amino acid codes and positions. The structure is shown in a ribbon representation, with residues labeled with their amino acid codes and positions.

VSP workflow and CFTR organization.

Figure 2

VSP-based ClCon- and TrIdx-phenotype landscapes at 37°C.

Figure 3

SCV relationships contributing to the energetics of ER export and trafficking.

Figure 4

SCV relationships defining stability at the cell surface.

Figure 5

Contribution of SCV to CFTR energetics driving function at the cell surface.

Figure 6

SCV code scoring global response to folding energetics across the entire CFTR fold.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SuppTable1.xlsx](#)
- [SuppTable2.xlsx](#)
- [Supplementary.pdf](#)