RESEARCH

Check for updates

A comprehensive map of human glucokinase variant activity

Sarah Gersing[1], Matteo Cagiada[1], Marinella Gebbia[2,3,4], Anette P. Gjesing[5], Atina G. Coté[2,3,4], Gireesh Seesankar[2,3,4], Roujia Li[2,3,4,6], Daniel Tabet[2,3,4,6], Jochen Weile[2,3,4,6], Amelie Stein[1], Anna L. Gloyn[7,8], Torben Hansen[5], Frederick P. Roth[2,3,4,6*], Kresten Lindorff-Larsen[1*] and Rasmus Hartmann-Petersen[1*]

*Correspondence:
fritz.roth@utoronto.
ca; lindorff@bio.ku.dk;
rhpetersen@bio.ku.dk

[1] The Linderstrøm-Lang Centre
for Protein Science, Department
of Biology, University
of Copenhagen, Ole Maaløes Vej
5, 2200 Copenhagen, Denmark
[2] Donnelly Centre, University
of Toronto, Toronto, ON M5S 3E1,
Canada
[3] Department of Molecular
Genetics, University of Toronto,
Toronto, ON M5S 1A8, Canada
[4] Lunenfeld-Tanenbaum
Research Institute, Sinai Health,
Toronto, ON M5G 1X5, Canada
[5] Novo Nordisk Foundation
Center for Basic Metabolic
Research, Faculty of Health
and Medical Sciences, University
of Copenhagen, Copenhagen,
Denmark
[6] Department of Computer
Science, University of Toronto,
Toronto, ON M5T 3A1, Canada
[7] Division of Endocrinology,
Department of Pediatrics,
Stanford University School
of Medicine, Stanford, CA, USA
[8] Stanford Diabetes Research
Center, Stanford University,
Stanford, CA, USA

Abstract

Background: Glucokinase (GCK) regulates insulin secretion to maintain appropriate blood glucose levels. Sequence variants can alter GCK activity to cause hyperinsulinemic hypoglycemia or hyperglycemia associated with GCK-maturity-onset diabetes of the young (GCK-MODY), collectively affecting up to 10 million people worldwide. Patients with GCK-MODY are frequently misdiagnosed and treated unnecessarily. Genetic testing can prevent this but is hampered by the challenge of interpreting novel missense variants.

Result: Here, we exploit a multiplexed yeast complementation assay to measure both hyper- and hypoactive GCK variation, capturing 97% of all possible missense and nonsense variants. Activity scores correlate with in vitro catalytic efficiency, fasting glucose levels in carriers of GCK variants and with evolutionary conservation. Hypoactive variants are concentrated at buried positions, near the active site, and at a region of known importance for GCK conformational dynamics. Some hyperactive variants shift the conformational equilibrium towards the active state through a relative destabilization of the inactive conformation.

Conclusion: Our comprehensive assessment of GCK variant activity promises to facilitate variant interpretation and diagnosis, expand our mechanistic understanding of hyperactive variants, and inform development of therapeutics targeting GCK.

Keywords: Diabetes, Variants of uncertain significance, Deep mutational scanning

Background

Glucokinase (GCK) is the body's primary glucose sensor as it regulates glucose-stimulated insulin secretion. Variants that decrease GCK activity cause elevated fasting glucose levels, known as GCK-maturity-onset diabetes of the young (GCK-MODY, MIM# 125851) [1, 2]. Unlike most forms of diabetes, GCK-MODY does not require treatment, as glycemia remains unaltered [3], and patients do not suffer complications [4, 5]. However, patients with GCK-MODY are often misdiagnosed with either type 1 or type 2

diabetes [6–8], leading to unnecessary treatment and surveillance [3]. Correct diagnosis of GCK-MODY can therefore terminate pharmaceutical treatments and decrease surveillance [3], with both economic and lifestyle benefits. Diagnosis of GCK-MODY can be achieved by identifying pathogenic variants in the gene encoding glucokinase (GCK, hexokinase-4).

GCK regulates glucose levels by catalyzing the first step of glycolysis — the phosphorylation of glucose to form glucose-6-phosphate. Glucose phosphorylation is the rate-limiting step in insulin secretion in pancreatic β-cells [9–11] and glycogen synthesis in liver cells [12], and these processes are therefore regulated by GCK activity.

In addition to GCK, there are three other human hexokinases. These hexokinases have a high affinity for glucose and show hyperbolic kinetics. In contrast, GCK has a low affinity for glucose ($S_{0.5}$ 7.5–10 mM) [13] and sigmoidal kinetics, which together allow GCK to respond rapidly to changes in glucose levels in the physiological range. The sigmoidal kinetics of GCK are due to positive cooperativity with glucose (Hill coefficient 1.7) [14]. This positive cooperativity is unusual as GCK functions as a monomer and contains only one glucose binding-site. Instead, the sigmoidal response to glucose is caused by intrinsic protein conformational dynamics [15–17].

GCK is a dynamic 52-kDa enzyme consisting of 465 amino acid residues, which fold into a large and a small domain. Between the two domains is a cleft forming the active site where glucose binds. The orientation of the two domains is not static, as GCK exists in multiple conformational ensembles [18, 19]. These ensembles are often described as three conformations: the super-open, open and closed conformation [15]. At low glucose levels, GCK predominantly exists in the super-open conformation. Upon glucose binding, GCK shifts into the open conformation [20, 21], and subsequently catalysis takes place in the closed conformation. Due to slow conversion between the super-open and open states [22], the population of each of the three conformations depends on glucose levels. At low glucose levels, GCK shifts into the super-open conformation prior to binding a new glucose molecule following catalysis, which results in slow glucose turnover. Turnover increases with higher glucose levels as GCK bypasses the super-open state and cycles between the open and closed conformations. The ratio between the slow and fast catalytic cycles, dependent on glucose concentration, gives rise to the positive cooperativity that is the basis for GCK sigmoidal kinetics and function [15, 23].

Underscoring the importance of GCK function for glucose homeostasis, variants that alter GCK activity are associated with several diseases [24]. Gain-of-function variants that increase GCK activity cluster at an allosteric activator site [25, 26] and cause hyperinsulinemic hypoglycemia (HH, MIM# 601820), which is characterized by increased insulin secretion even at low blood glucose levels [25, 27]. In contrast, loss-of-function variants where GCK activity is eliminated or decreased cause hyperglycemia. Inactivation of both *GCK* alleles can result in the severe permanent neonatal diabetes mellitus (PNDM, MIM# 606176) [28, 29], while heterozygous mutations cause a mild form of diabetes known as GCK-MODY [1, 2]. GCK-MODY has an estimated population prevalence of 0.11%-0.21% [30, 31], suggesting about 10 million people worldwide have GCK-MODY. Patients have mild and stable fasting hyperglycemia within 6–8 mM that does not usually require treatment, in contrast to other types of diabetes. Due to the population prevalence of GCK-MODY, and given that early diagnosis can limit unnecessary

Gersing *et al. Genome Biology*   (2023) 24:97

Page 3 of 23

treatment and surveillance due to misdiagnosis [3], GCK has been proposed for inclusion in population screening programs [32]. Therefore, a comprehensive dataset of GCK activity and an increased understanding of GCK variants is highly warranted.

To address this challenge, we generated a variant effect map of GCK using a multiplexed assay of variant effects [33, 34]. We assessed GCK activity using functional complementation in yeast, scoring both hypo- and hyperactive variants. The variant effect map recapitulates both the active site and a known allosteric activator site and includes 9003 of the 9280 (97%) possible missense and nonsense variants. The activity scores correlate with previous in vitro measurements of the catalytic efficiency and fasting blood glucose levels in patients. Furthermore, the map correctly classifies 78% of functionally characterized pathogenic variants that were previously curated [24]. To substantiate the map more broadly, we analyzed evolutionary conservation and conformational free energies. Conservation analysis generally agreed with the activity scores but was unable to capture hyperactive variants. When we examined these variants mechanistically, we found that some hyperactive variants likely shift GCK towards the active conformation through differential destabilization of the super-open and closed conformations. In conclusion, we present a comprehensive map of GCK activity to aid in variant understanding and interpretation.

Results

Assessing human GCK activity using yeast complementation

To measure the activity of human GCK variants at large scale, we coupled yeast growth to human GCK activity using yeast complementation. To test complementation, we constructed a yeast strain deleted for all three yeast hexokinase genes (*hxk1Δ hxk2Δ glk1Δ*) that is unable to grow on glucose medium (Fig. 1A). This growth deficiency was rescued by expressing human pancreatic *GCK* (Fig. 1B), as previously shown [35].

Next, we examined whether the complementation assay could separate both hypo- and hyperactive pathogenic variants from non-pathogenic GCK variants. We selected an initial test set of eight variants. Two variants (p.D217N and p.E279Q) were used as non-damaging controls, as they were two of the three most common alleles in gnomAD [36], while a third variant (p.D205H) has previously been shown to be catalytically dead [15, 37] and was used as a loss-of-function control. The five remaining test-set variants were pathogenic variants from the ClinVar database [38], including both one variant (p.W99R) associated with HH [26] and four variants (p.F150S, p.R191W, p.G223S and p.G261R) with compelling evidence for linkage to GCK-MODY [39–42]. To test complementation of the eight variants, they were expressed in the *hxk1Δ hxk2Δ glk1Δ* yeast strain (Additional file 1: Fig. S1). On glucose medium, the two common variants showed growth that was similar to reference ('WT') GCK, while the catalytically dead variant (p.D205H) grew comparable to the vector control (Fig. 1B). The hyperactive variant (p.W99R) grew faster than WT, while the four disease-associated variants with reduced activity (p.F150S, p.R191W, p.G223S and p.G261R) grew slower than WT (Fig. 1B). The growth of the eight *GCK* variants was consistent with their level of activity, showing that the complementation assay could assess GCK activity.

GCK activity depends on the concentration of its substrate glucose. To find a glucose concentration that enabled detection of variants with both decreased and
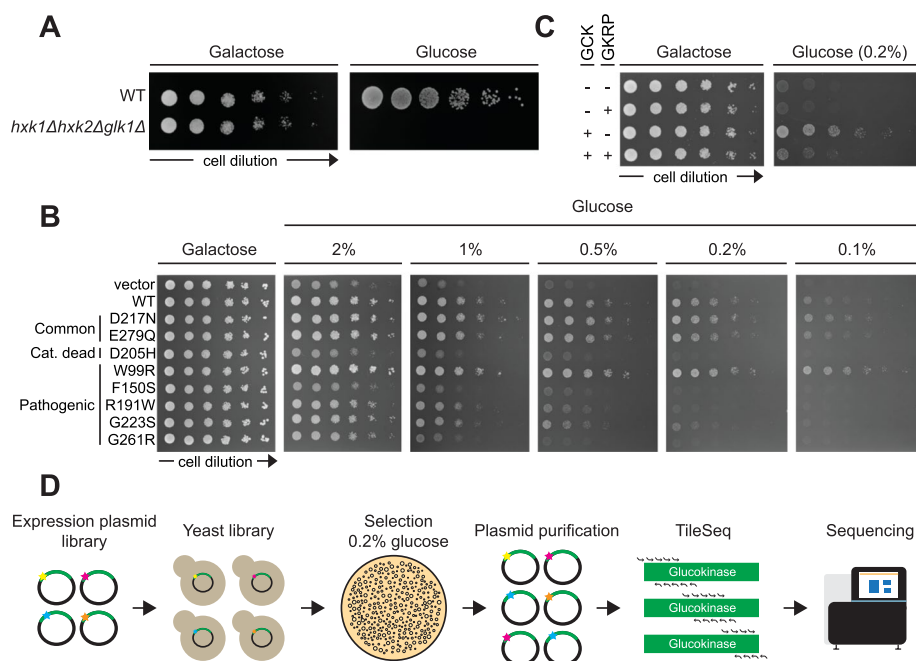
Fig. 1 Yeast complementation as a readout for human glucokinase variant activity. **A** Yeast growth assay of wild-type (WT) and *hxk1Δ hxk2Δ glk1Δ* yeast strains on galactose and glucose media. **B** The growth of the *hxk1Δ hxk2Δ glk1Δ* yeast strain expressing either a vector control, wild-type GCK (WT) or a GCK variant was compared on media containing galactose or varying concentrations of glucose. **C** Growth assay of different combinations of vectors (-), GCK and GKRP expressed in the *hxk1Δ hxk2Δ glk1Δ* yeast strain on galactose and glucose media. **D** Illustration of the multiplexed assay for GCK variant activity

increased glucose affinities, we next tested complementation on media with varying glucose concentrations (Fig. 1B). A concentration of 0.2% (11.1 mM) glucose enabled detection of the hyperactive variant (p.W99R), while retaining a good dynamic range between WT and variants with decreased activity (p.F150S, p.R191W, p.G223S and p.G261R) (Fig. 1B). This concentration is close to GCK's affinity for glucose ($S_{0.5} = 7.5–10$ mM) [13, 24].

To further establish the fidelity of the yeast system, we tested whether glucokinase regulatory protein (GKRP), a known inhibitor of hepatic GCK in humans [43, 44], could repress GCK activity in the yeast system. We co-expressed human GKRP and GCK in the *hxk1Δ hxk2Δ glk1Δ* yeast strain and examined growth on glucose medium (Fig. 1C). Co-expression of GKRP on glucose medium led to reduced growth (Fig. 1C). As GCK expression levels were comparable (Additional file 1: Fig. S2), human GKRP inhibited GCK activity in yeast, further validating the relevance of the yeast system to assay human GCK activity.

Having established an assay coupling yeast growth to GCK enzymatic activity, we used the assay to score the activity of a saturated library of human *GCK* variants produced by codon-randomization. For library construction, we divided the *GCK* sequence into three regions that were separately mutagenized, assayed and sequenced. Regional libraries were mutagenized using pools of oligos, each containing a central NNK-degeneracy targeting a codon within the targeted region. Subsequently, we cloned the mutagenized regional libraries *en masse* into a yeast expression vector
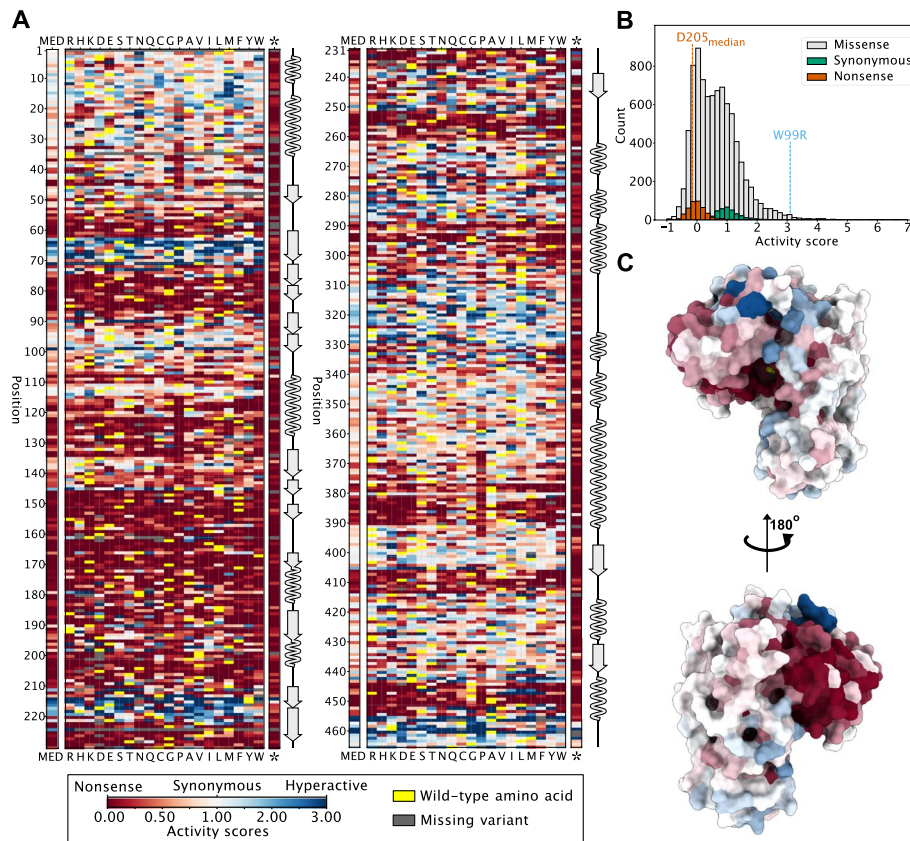
Gersing *et al. Genome Biology*   (2023) 24:97

Page 5 of 23

Fig. 2 Map of glucokinase variant activity. **A** Heatmap showing the GCK activity score for each substitution (x-axis) at every position (y-axis) based on the multiplexed functional complementation assay. A score of one (white) corresponds to the activity of wild-type-like synonymous variants, a score of zero (red) corresponds to total loss-of-function (nonsense) variants and a score above one (blue) corresponds to activity above synonymous variants. The median (MED) activity score at each position is included. Yellow indicates the wild-type amino acid and variants with missing data ('missing variants') are shown in grey. The secondary structure of GCK is shown next to the activity map with β-sheets shown as arrows and α-helices shown as waves. A map where missing variants have been imputed can be found in Additional file 1: Fig. S3. **B** Distributions of activity scores for synonymous variants (green), nonsense variants (red) and missense variants (grey). The activity score of the hyperactive variant p.W99R (blue) and the median activity score of the catalytic residue p.D205 (red) are shown. **C** The median activity score at each position was mapped onto the closed glucose-bound conformation of GCK (PDB 1V4S). The color scheme is the same as in panel **A**

and transformed the resulting plasmid libraries into the *hxk1Δ hxk2Δ glk1Δ* yeast strain. The yeast libraries were grown on 0.2% glucose media in duplicate to select for GCK activity. Before and after selection, plasmids were extracted from yeast libraries and each region was deeply sequenced in tiles of ~150 bp (~1.6 M-4.8 M reads per sequenced tile) (Fig. 1D) [45]. The resulting reads were used to quantify the relative frequency of each variant both before and after selection, and thus calculate an activity score and an associated measurement error for each variant. Activity scores were rescaled such that synonymous variants had a scores centered on one while nonsense variants had scores centered on zero. The resulting dataset contained scores for 9003 (97%) of the 9280 possible single amino acid GCK variants (including stop codons) (Fig. 2A), and most of the remaining variants could be imputed using the

Human Protein Variant Effect Map Imputation Toolkit [45, 46] (Additional file 1: Fig. S3AB).

The activity scores of nonsense and synonymous variants were well separated (Fig. 2B), and the distribution of missense variants spanned from total loss of function to above-wildtype activity (Fig. 2B). The activity of our test-set variants was similar to the low-throughput growth assays, such that the common variants (p.D217N and p.E279Q) had wild-type-like activity, the HH-associated variant (p.W99R) had increased activity and the GCK-MODY-associated variants (p.F150S, p.R191W, p.G223S and p.G261R) had decreased activity. Although the catalytic site variant p.D205H was not included in the map, all other variants at this position showed decreased activity as expected.

To examine variant effects structurally, we mapped the median activity score at each position onto the structure of glucose-bound GCK [15] (Fig. 2C, Additional file 1: Fig. S4AB). The resulting activity-colored structure was consistent with characteristics of proteins in general and GCK specifically. First, surface residues were generally resistant to mutations, while active site and buried residues were mutation-sensitive (Fig. 2C, Additional file 1: Fig. S4AB). Second, several positions where mutations on average increased activity clustered at a known allosteric activator site [25]. Together with the expected behavior of our test-set variants, these observations support that the map reflects human GCK activity.

Correlations with enzyme kinetics, fasting blood glucose levels, and clinical genetics

To examine what aspects of GCK activity the map reflects, we examined the correlation between our assay scores and previously published kinetic parameters of 38 variants characterized in vitro [47]. Assay scores correlated with the catalytic efficiency ($k_{cat}/S_{0.5}$) of GCK variants (Fig. 3A, $r_s = 0.76$, 95% CI [0.58, 0.88]), indicating that our yeast assay captures GCK catalytic efficiency with a dynamic range that includes both decreased and increased values.

Having established that our assay recapitulates in vitro GCK activity, we next examined the correlation with measures of fasting glucose in carriers of *GCK* variants. Samples were obtained from four different Danish populations and included a population-based cohort ($n = 6,058$ (men: 3,020; women: 3,038); Glümer et al. [50]), patients with newly diagnosed type-2 diabetes ($n = 2,855$ (men: 1,678; women: 1,177); Christensen et al. [51]), a population sample of Danish children ($n = 1,138$; (boys: 508; girls: 630); Kloppenborg et al. [52]) as well as ten patients diagnosed with GCK-MODY (men: 4; women: 6); Johansen et al. [53]). Our dataset included fasting glucose levels of 33 patients with known *GCK* variants and presenting with fasting glucose levels below 9 mM. Although fasting glucose levels likely depend on several genetic and environmental factors [54–58], activity scores correlated with patient glucose levels (Fig. 3B, $r_s = -0.58$, 95% CI [-0.20, -0.79]) suggesting that the yeast assay reflects the genetic contribution of GCK to fasting glucose in carriers of *GCK* variants.

We next examined whether the scores of previously classified GCK-MODY, HH, and benign variants separated into distinct classes. Our dataset included 71 variants with experimentally determined activity scores: 68 (60 GCK-MODY, 8 HH) pathogenic variants which have previously been functionally characterized [24] and three benign variants [38, 48, 49]. We were unable to generate a score for the benign variant p.G68D
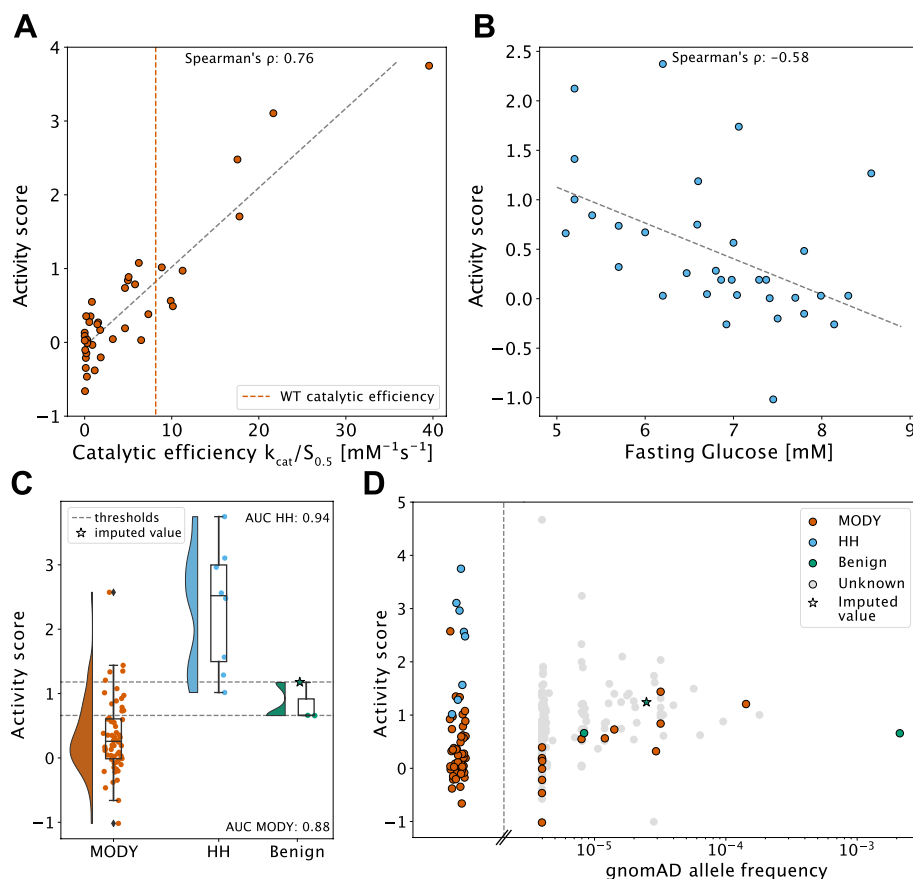
Fig. 3 Correlations with enzyme kinetics, fasting blood glucose levels, and clinical genetics. **A** Plot showing the correlation between the GCK activity scores and previously measured catalytic efficiency ($k_{cat}/S_{0.5}$) [47] of 38 variants with a Spearmans' ρ of 0.76. The red dotted line indicates WT catalytic efficiency [47] and the black dotted line shows the best fitting curve. **B** The activity score plotted against the fasting glucose level of 33 carriers with an identified single variant in the *GCK* gene. The black dotted line indicates the best fitting curve. Spearmans' ρ is -0.58. **C** Raincloud plot showing the distributions of activity scores for 60 functionally characterized glucokinase-maturity-onset diabetes of the young (GCK-MODY) variants [24], eight functionally characterized hyperinsulinemic hypoglycemia (HH) variants [24] and three benign variants [38, 48, 49]. Due to the limited number of benign variants, the imputed activity score of the benign variant p.G68D was used, as no experimental score was obtained for this variant. The black dotted lines show the thresholds for variants associated with GCK-MODY (0.66, AUC = 0.88) and HH (1.18, AUC = 0.94) based on receiver-operating characteristic (ROC) analyses. **D** Plot showing the activity scores and the associated gnomAD allele frequency [36] for GCK variants present in gnomAD. As in panel C, the imputed activity score of p.G68D was used. In addition, the pathogenic variants from panel **C** that were not present in gnomAD are plotted to the left of the stippled line

and, due to the already limited number of benign variants, used the imputed score for this variant obtained using the Human Protein Variant Effect Map Imputation Toolkit [45, 46]. Although the assay did not correctly classify all variants, variants belonging to each of the three classes showed distinct distributions centered on either scores comparable to synonymous mutations (benign variants), a hyperactive score (HH variants) or a decreased activity score (GCK-MODY variants) (Fig. 3C). Furthermore, variants with a high allele frequency ($> 10^{-4}$) in gnomAD [36] had WT-like scores while rarer variants displayed a wide range of activities (Fig. 3D). To determine threshold values to classify variants as either GCK-MODY or HH, we applied receiver-operating characteristic

Gersing *et al. Genome Biology*   (2023) 24:97

Page 8 of 23

(ROC) analyses to the set of previously classified pathogenic and benign variants, noting that the very small number of benign variants hampers a detailed analysis. The threshold for variants associated with GCK-MODY was 0.66 (AUC = 0.88), while variants with a score above 1.18 were predicted to be associated with HH (AUC = 0.94). Using these cutoff values, the activity assay correctly identifies 76.7% of the analyzed GCK-MODY variants and 87.5% of the analyzed HH variants.

Although our assay was able to detect the vast majority of pathogenic variants, some reported pathogenic variants appeared as benign. There were several causes for this misclassification, but as examples we include three variants (p.V62M, p.T65I, and p.H137R) that show different molecular mechanisms causing them to score as benign in our assay. The genetic evidence for p.V62M as a loss-of-function mutation is compelling but initial in vitro kinetic characterization demonstrated increased affinity for glucose. Subsequent functional studies demonstrated that it is thermally labile with evidence for defective regulation by both GKRP and allosteric activators [59]. Similarly, the HH variant p.T65I has an increased affinity for glucose (reduced $S_{0.5}$) and therefore induces insulin secretion at inappropriate glucose levels [26]. However, p.T65I also has a decreased turnover number, resulting in a WT-like catalytic efficiency [47], which is the aspect of GCK activity that our assay captures. Finally, the GCK-MODY variant p.H137R has been previously reported to have WT-like kinetics but mildly decreased thermal stability [60]. Although our assay was not able to detect the p.H137R instability variant, it could detect severely unstable variants such as p.E300K [61]. Thus, based on an evaluation of these three mutations, which display complex mechanisms for their pathogenicity, we note that our assay may not detect all pathogenic variants involving complex mechanisms, including modest instability, especially for those with WT-like intrinsic catalytic efficiency.

Evolutionary conservation predicts glucokinase variant effects

While our experiments probed nearly all possible GCK variants, only a limited number of variants have previously been characterized experimentally. We therefore examined GCK activity scores more broadly by analyzing evolutionary conservation across species. Conservation analysis can assess the mutational tolerance of each position in a protein [62, 63], and is widely used to assess the effect of protein variants [64–67]. We analyzed GCK evolutionary conservation using a sequence alignment of homologous proteins, evaluating the evolutionary distance between the GCK WT sequence and the single mutant variant, using an alignment that included both hexokinases and glucokinases more widely. The resulting evolutionary distance score (ΔE) quantifies the likelihood of a given substitution (Additional file 1: Fig. S5AB). Therefore, a score close to zero suggests that a substitution is compatible with function and does not affect the structural stability of the native conformations, while variants with a highly negative score are likely to be detrimental. Activity scores correlated weakly with ΔE ($r_p = 0.44$, 95% CI [0.43, 0.46], Additional file 1: Fig. S6A). However, the correlation increased when we compared the two datasets using the residue median score ($r_p = 0.64$, 95% CI [0.57, 0.68], Additional file 1: Fig. S6B), as averaging decreases noise in both datasets. Residue-averaged ΔE and activity scores agreed on regions where mutations severely decreased activity as well as regions that were tolerant towards mutations (Fig. 4A, Additional
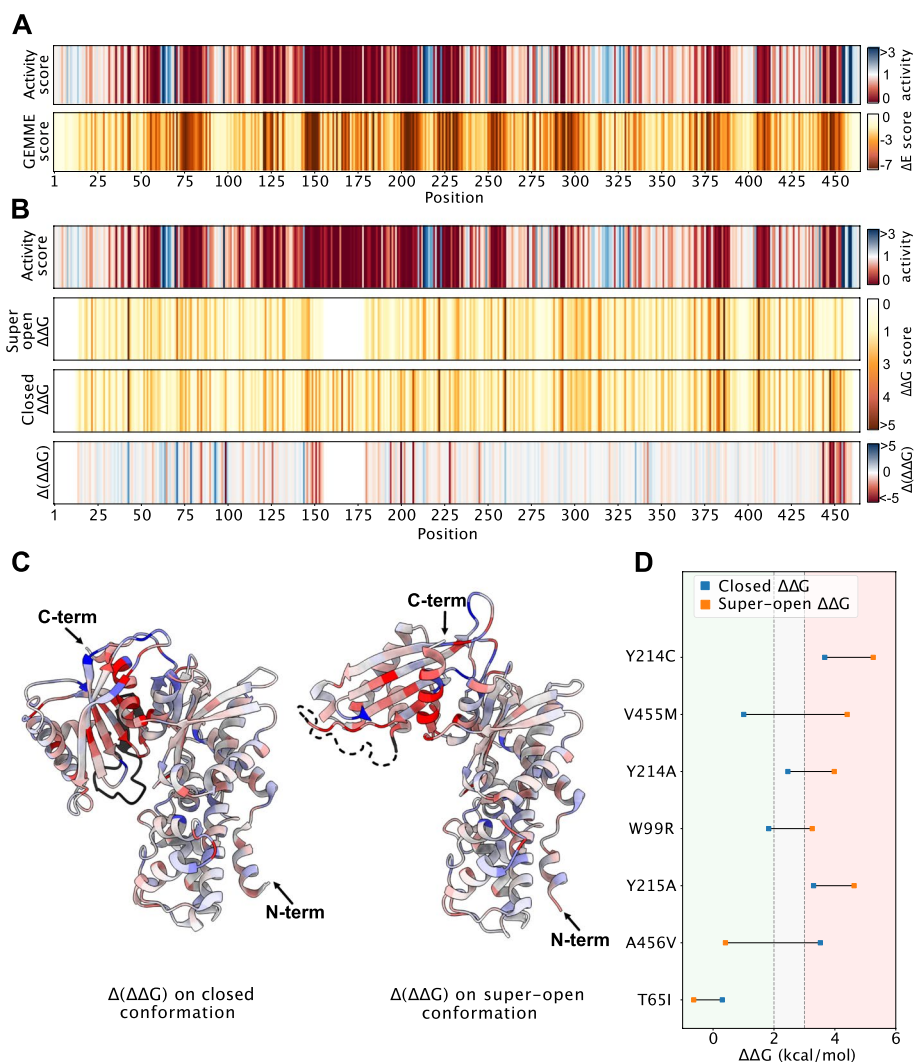
Fig. 4 Interpreting the glucokinase activity scores using analyses of evolutionary conservation and conformational free energies. **A** Plots showing the residue median activity scores and evolutionary conservation scores for GCK. For activity scores, a score of one corresponds to a wild-type-like activity, a score of zero corresponds to total loss of function, and a score above one indicates increased activity. For GEMME scores (ΔE score), a score of zero suggests compatibility with function, while a high negative score indicates that mutations are detrimental. **B** Plots showing the residue median of activity scores, $\Delta\Delta G$ of the super-open conformation, $\Delta\Delta G$ of the closed conformation, and the difference between the $\Delta\Delta G$ in the closed and super-open conformation ($\Delta\Delta G_{super-open} - \Delta\Delta G_{closed}$, $\Delta(\Delta\Delta G)$), respectively. For $\Delta\Delta G$ of the super-open and closed conformations, a score of zero corresponds to the same stability as wild-type, while a high negative score means decreased stability compared to wild-type GCK. For $\Delta(\Delta\Delta G)$, a score of zero indicates an equal destabilization of the two conformations. A negative score indicates a destabilization of the closed state relative to the super-open, while a positive score indicates a destabilization of the super-open conformation relative to the closed. Note that residues at positions 157–179 are absent from the structure of the super-open conformation, and therefore lack super-open $\Delta\Delta G$ and $\Delta(\Delta\Delta G)$ scores. **C** The two GCK conformations colored by residue median $\Delta(\Delta\Delta G)$. The 157–179 region is colored black in both structures. **D** Plot showing the $\Delta\Delta G$ values of seven previously characterized variants [22, 70]. The background coloring indicates whether the $\Delta\Delta G$ value is in the stable (green), intermediate (grey) or unstable (red) range. Variants that are predicted to shift the equilibrium towards the closed conformation will have the highest $\Delta\Delta G$ in the super-open state, and vice versa for variants that are predicted to shift the equilibrium towards the inactive state. PDBs: super-open (1V4T) and closed (1V4S)

file [1]: Fig. S6E). However, the evolutionary conservation analysis did not detect gain-of-function positions nor the mutational sensitivity of the ~150–200 region that is likely specific for GCK compared to other hexokinases (Fig. [4]A, Additional file [1]: Fig. S6E). Residues in this region include the 151–179 loop that undergoes a disorder to order transition upon glucose binding [68]. The glucose-induced conformational changes in GCK are the basis for the sigmoidal kinetics and low glucose affinity that distinguishe GCK from other hexokinases. Hence, the ~150–200 region and gain-of-function variants are likely not captured by the evolutionary analysis due to the broadness of the multiple sequence alignment (MSA). When we restricted our analysis to include only residues with a median assay score below 1.18, correlation with ΔE further increased ($r_p = 0.73$, 95% CI [0.69, 0.77], Additional file [1]: Fig. S6CD). The remaining deviation between ΔE and activity scores could stem from assay conditions buffering mutation effects, such as the high expression level, the temperature, or the short timeframe of selection compared to evolution, making subtle fitness effects difficult to discern [69]. In conclusion, analysis of evolutionary conservation supports the activity assay. However, our conservation analysis does not include effects that are GCK-specific, such as substitutions that increase activity or affect conformational regulation.

Mechanistic evaluation of variant effects

To explore variants with functional impacts that were not detected via conservation analysis, we examined GCK variant effects mechanistically. We speculated that some variants shift the equilibrium of the conformational ensemble towards the catalytically inactive (or active) conformation, thereby decreasing (or increasing) GCK activity. This shift might arise if a given variant differentially affects the stability of the two conformations. To identify such variants, we used Rosetta [71] to predict changes in thermodynamic protein stability ($\Delta\Delta G$) for both the super-open and closed GCK structure [15] (Additional file [1]: Fig. S7AB), and then calculated the difference between the two structures ($\Delta\Delta G_{\text{super-open}} - \Delta\Delta G_{\text{closed}}$, Additional file [1]: Fig. S7C). A negative score indicates that a given variant destabilizes the closed conformation relative to the super-open, while positive scores indicate variants that destabilize the super-open conformation relative to the closed. However, this shift is only relevant for variants that do not overly destabilize both structures, which would lead to disruption of the enzyme structure, likely degradation and loss of function.

Variants at most positions destabilized the two conformations equally (Fig. [4]B white lines, Additional file [1]: Fig. S8A). However, variants shifting the equilibrium towards the super-open inactive conformation were concentrated in two regions surrounding position 150 and 450 respectively (Fig. [4]B lower panel, Additional file [1]: Fig. S8A). The region surrounding position 450 mapped to an α-helix that is surface-exposed in the super-open state but buried in the closed state (Fig. [4]C), consistent with variants causing a greater destabilization of the closed state. Similarly, the residues surrounding position 150 were buried in the active conformation (Fig. [4]C). In addition, this region is N-terminal to residues 151–179 that form a disordered loop in the super-open conformation and therefore are not included in the stability calculations. The disordered loop folds into a β-hairpin in the closed conformation [15, 68]. Due to this disorder-order transition, mutations in the 151–179 loop/β-hairpin are likely to destabilize the closed state relative

to the super-open state. A conformational shift towards the inactive state is therefore a potential mechanism for mutations in the ~150–200 region. Positions where variants shifted the equilibrium towards the closed active conformation (Fig. 4B blue lines) were spread throughout the GCK sequence. However, these positions concentrated in the 3D structure at the allosteric activator site (Fig. 4C), which is also evident from the structure of GCK bound to a synthetic allosteric activator (Additional file 1: Fig. S8BC). Previously, several hyperactive variants were examined to determine their mechanism using kinetic analysis [22, 70]. By examining enzyme kinetics, Heredia et al. found that p.T65I, p.Y215A and p.A456V mainly accelerate glucose binding, while p.W99R, p.Y214A, p.Y214C and p.V455M predominantly increase the conformational preference towards the active state. Although p.T65I, p.Y215A and p.V455M affect the equilibrium mildly, all variants except p.A456V to some extent increased conversion to the closed state [70]. Consistently, p.A456V shifted the equilibrium towards the super-open conformation according to our stability predictions (Fig. 4D). In addition, except for the p.T65I variant, stability calculations predicted the remaining five variants (p.W99R, p.Y214A, p.Y214C, p.Y215A and p.V455M) to shift the equilibrium towards the active conformation (Fig. 4D). Thus, for six out of seven variants our stability predictions were consistent with prior kinetic and mechanistic analysis, thereby validating our mechanistic predictions. We next used the stability calculations to assess how many hyperactive variants potentially shift the equilibrium towards the active state. The supplementary data contains a list of 467 hyperactive variants (activity score > 1.18) which are predicted to both be stable ($\Delta\Delta G_{closed}$ < 2 kcal/mol) and promote increased populations of the closed active conformation ($\Delta(\Delta\Delta G)$ > 0.25 kcal/mol) (see Additional file 2).

Discussion

Recent developments in molecular biology and sequencing technologies have made it possible to perform deep mutational scanning experiments, in which thousands of gene variants are assayed in a single, multiplexed experiment. Here we applied this technology to *GCK*, a gene of central importance in metabolism and where missense and nonsense variants are associated with several diseases. The resulting activity map covers 97% of all possible single amino acid variants, and activity scores correlate with in vitro catalytic efficiency, fasting glucose levels in individuals carrying *GCK* variants, and evolutionary conservation.

There are a number of limitations to our study. First, due to our mutagenesis strategy, each variant in the library may contain multiple mutations within the mutagenized region. As each regional library is sequenced in tiles, only *in cis* mutations that occur within the same ~150 bp tile are detected. To limit the impact of potential background mutations that co-occur in the same clone(s), the regional libraries have a low mutational density, with an estimated average of 0.3–0.4 mutations per mutagenized region, and a high coverage so that each variant's score is based on a large number of independent clones. In this way, the impact of secondary variants within a clone can be averaged out. However, a few variant scores may be affected by background mutations, as potentially seen for synonymous variants. The activity score distributions of synonymous and nonsense variants show a slight overlap (Fig. 2B). For synonymous variants with an unexpectedly low activity score, the vast majority are associated with a high estimated

standard deviation, and the remaining variants might be explained by background mutations. For nonsense variants associated with an unexpectedly high activity score, the majority are associated with a high standard deviation or are in the extreme C-terminus where they appear compatible with activity. The remaining high-scoring nonsense variants could be due to stop codon readthrough, which is known to occur context-dependently in yeast [72, 73], and/or the presence of yeast cells transformed with vectors expressing multiple distinct variants, although we attempted to limit this by outgrowth for three days following transformation [74]. We also note that apparently functional nonsense variants have previously been observed in high throughput cell-based assays [45, 75–78]. A second limitation is that apparently-functional nonsense variants in our study should not be confidently inferred to provide normal function in humans, as our assay does not provide a faithful model for impacts due to nonsense-mediated decay in a human cell. A third limitation is that our mutagenesis libraries include only synonymous, missense and nonsense variants. Other types of pathogenic variation have previously been identified in the *GCK* gene, including insertions [79, 80], a promoter variant [81], and splice site variants. A map of missense variants is, however, still highly relevant. First, the effects of missense variants can be challenging to interpret without functional data. Second, missense variants constitute 437 out of the total 794 current *GCK* ClinVar entries [38] (accessed on February 20th 2023), and of the 201 *GCK* variants that have been annotated as variants of uncertain significance, 161 (80%) of these are missense variants. A fourth limitation of our study is that 22.1% of the analyzed pathogenic variants have scores that are similar to those of wild-type-like synonymous variants. The sensitivity of the assay could potentially be increased by repeating the assay using different expression levels and glucose concentrations as well as increasing the number of replicates. However, others have found that pathogenic variants, even when thoroughly examined in low-throughput assays, can appear wild-type-like [47, 82, 83]. Multiple assays characterizing different aspects of GCK function might therefore be necessary to capture all pathogenic variants.

The hyperactive variants identified in our assay clustered at a site distal from the active site, known as the allosteric activator site [25]. The allosteric site is being pursued as a target for treatment of type 2 diabetes using drugs known as glucokinase activators (GKAs) [84, 85]. The development of GKAs has faced several problematic effects including accumulation of triglycerides in the liver [86, 87] and hypoglycemia [88, 89]. However, recently both a dual-acting and a hepatoselective GKA have shown promising results [90, 91]. Most GKAs bind the allosteric activator site and mimic the effect of known hyperactive variants. Previously, a few hyperactive variants were found to increase GCK activity by favoring the conformational change to the closed active state [22, 70]. We extended the mechanistic analysis of hyperactive variants using predictions of protein stability for the super-open and closed conformations. Our findings are in accordance with prior kinetic analyses [22, 70] and we identify 467 hyperactive variants that are predicted to shift GCK towards the closed state. Furthermore, our results indicate that these variants facilitate isomerization by destabilizing the super-open conformation relative to the closed. In conclusion, our study provides a comprehensive mapping of the allosteric activator site and an improved mechanistic understanding of some hyperactive variants. Our results may aid in refining drug design and the

development of GKAs, specifically to address the problem of hypoglycemia resulting from GKA treatment.

In contrast to hyperactive variants we found that, in the region spanning residues ~ 150–200, nearly all substitutions substantially decreased activity. Variants previously identified in the region are associated with loss of activity and elevated fasting plasma glucose levels [24, 47], and this region is central in the conformational dynamics of GCK [17, 68]. Hence, substitutions in this region might interfere with the transitions required for activity. In addition, some variants might destabilize the closed active conformation relative to the super-open state, thereby leading to increased population of the inactive state. However, the mechanisms underlying the mutational sensitivity of this region are not entirely clear and require further studies.

Conclusions

Here, we provide the first comprehensive map of GCK variant activity, measuring the functional consequences of thousands of previously uncharacterized GCK variants. More than 1 in 1000 people are estimated to suffer from GCK-MODY, and although the functional evidence from our variant effect map cannot alone classify a variant as pathogenic, it increases our understanding of GCK variants including those causing GCK-MODY.

Methods

Buffers

TE Buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0. SDS sample buffer (4x): 250 mM Tris/HCl, 8% SDS, 40% glycerol, 0.05% bromophenol blue, 0.05% pyronin G, 2% β-mercaptoethanol, pH 6.8.

Plasmids

The pancreatic isoform of human GCK (Ensembl ENST00000403799.8) was codon optimized for yeast expression and cloned into pDONR221 (Genscript). The initial test set GCK variants were generated by Genscript. For yeast expression, WT GCK, test set variants and libraries were cloned into pAG416GPD-EGFP-ccdB (Addgene plasmid # 14,316; http://n2t.net/addgene:14316; RRID:Addgene_14316, [92]) using Gateway cloning (Invitrogen). Human GKRP (Ensembl ENST00000264717.7) was codon optimized for yeast expression and cloned into pDONR221 with an N-terminal HA-tag (Genscript). For yeast expression, GKRP was cloned into pAG415GPD-ccdB (Addgene plasmid # 14,146; http://n2t.net/addgene:14146; RRID:Addgene_14146, [92]) using Gateway cloning (Invitrogen).

Yeast strains

The *hxk1Δ hxk2Δ glk1Δ* strain used for GCK complementation assays was obtained in two steps. First, the following two strains were crossed to obtain a haploid *hxk1Δ hxk2Δ MATa* strain: *hxk1*::kanMX *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 MATa* and *hxk2*::natMX4 *can1Δ*::STE2pr-Sp_HIS5 *lyp1Δ0 his3Δ1 leu2Δ0 ura3Δ0* met15Δ0 LYS2 + MATα. This strain was then used to knock out *GLK1* using a HygroMX cassette. BY4741 was used as a wild-type control. Yeast cells were cultured in synthetic complete (SC) medium (2%

D-galactose, 0.67% yeast nitrogen base without amino acids, 0.2% drop out (Sigma), (0.0076% uracil, 2% agar)). To select for GCK activity, yeast cells were grown on SC medium containing various concentrations of D-glucose monohydrate as indicated in figures. The growth defect of the *hxk1Δ hxk2Δ glk1Δ* strain was tested on Yeast Extract-Peptone (YP) medium (2% D-glucose or 2% D-galactose, 2% tryptone, 1% yeast extract). Small scale yeast transformations were done as described in [93].

Yeast growth assays

For yeast growth assays on solid medium, cultures were grown overnight (30 °C, vigorous agitation) until reaching exponential phase. The yeast cells were washed once in sterile water (1200 g, 5 min, RT) and resuspended in sterile water. The OD_{600nm} of all cultures were adjusted to 0.4, and the cultures were used for a five-fold serial dilution in water. Serial-diluted cultures were spotted in 5 μL drops onto agar plates, which were briefly dried and incubated at 30 °C for 2-4 days.

Protein extraction from yeast cells

For extraction of proteins from yeast, cultures were grown overnight (30 °C, vigorous agitation) until reaching exponential phase, at which point $100–125 \times 10^6$ cells were harvested (1200 g, 5 min, 4 °C). The pelleted cells were washed in 25 mL ice-cold water (1200 g, 5 min, 4 °C). The cell-pellet was resuspended in 1 mL 20% ice-cold TCA, transferred to an Eppendorf tube and centrifuged (4000 g, 5 min, 4 °C). The supernatant was discarded and the pellet was resuspended in 200 μL 20% ice-cold TCA. The resuspension was transferred to a screw cap tube containing 0.5 mL glass beads (Sigma), and cells were lysed using a Mini Bead Beater (BioSpec Products) by three 15 s cycles with 5 min incubations on ice between each burst. Then, 400 μL ice-cold 5% TCA was added, the tubes were punctured at the bottom using a needle and transferred to a 15 mL Falcon tube containing a 1.5 mL Eppendorf tube without the lid. The sample was isolated from the glass beads by centrifugation (1000 g, 5 min, 4 °C). The Eppendorf tube containing the sample was centrifuged (10,000 g, 5 min, 4 °C) and the resulting pellet was washed in 500 μL 80% acetone (10,000 g, 5 min, 4 °C). The acetone was removed and the pellet was dried for 5 min before resuspension in 100 μL SDS sample buffer (1.5x) and 25 μL 1 M Tris/HCl, pH 9. The samples were boiled for 5 min and cleared by centrifugation (5000 g, 5 min, RT). The supernatant was transferred to an Eppendorf tube and was analysed by SDS-PAGE and Western blotting.

Electrophoresis and blotting

SDS-PAGE was done using 12.5% acrylamide gels. Following SDS-PAGE, 0.2 μm nitrocellulose membranes were used for the Western blotting procedure. After protein transfer, membranes were blocked in 5% fat-free milk powder, 5 mM NaN_3 and 0.1% Tween-20 in PBS.

Antibodies and their sources were: anti-HA (Roche, 15,645,900) and anti-GFP (Chromotek, 3H9 3h9-100). The secondary antibody was HRP-anti-rat (Invitrogen, 31,470).

Library mutagenesis

To construct a library of GCK variants, oligos containing a central degenerate NNK codon were designed for each codon in the GCK sequence. An online tool (http://llama. mshri.on.ca/cgi/popcodeSuite/main) [45] was used to obtain oligo sequences, and oligos were obtained from Eurofins. The *GCK* sequence was divided into three regions spanning aa 2–171 (region 1), 172–337 (region 2) and 338–465 (region 3). Oligos were pooled for each region. The three regional oligo pools were phosphorylated using T4 Polynucleotide Kinase (NEB) as recommended by the provider using 300 pmol of each oligo pool and incubation for one hour at 37 °C. Then, the phosphorylated oligos were annealed to the WT *GCK* sequence. For each region, the following was combined in a PCR tube: 25 fmoles pENTR221-GCK, 3 µL 10 µM SKG_1, 5.6 µL oligo pool, 10.1 µL nuclease-free water. The reactions were denatured (95 °C, 3 min) and cooled (4 °C, 5 min). Following template annealing, 5 µL of each reaction was combined with 5 µL Phusion Hot Start Flex $2 \times$ Master Mix (NEB) and sequences were extended (95 °C 3 min, 4 °C 5 min, 50 °C 120 min). To each reaction, the following was added: 1.5 µL Taq DNA ligase buffer (NEB), 0.5 µL Taq Ligase (NEB), 3 µL nuclease-free water, followed by incubation (45 °C 20 min). Next, 1 µL of each reaction was combined with the following: 2 µL 10 µM SKG_2, 2 µL 10 µM SKG_3, 25 µL Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), 20 µL nuclease-free water. The libraries were amplified using the following conditions: 98 °C 30 s, 20 cycles of 98 °C 15 s, 55 °C 30 s, 72 °C 150 s, followed by 72 °C 5 min and 4 °C hold. The resulting PCR products were used in a PCR to add Gateway attB sites: 25 µL Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), 1 µL PCR product, 2 µL 10 µM SKG_4, 2 µL 10 µM SKG_5, 20 µL nuclease-free water. The following PCR program was used: 98 °C 30 s, 5 cycles of 98 °C 15 s, 58 °C 30 s, 72 °C 150 s, followed by 12 cycles of 98 °C 15 s, 72 °C 150 s and finally 72 °C 5 min, 4 °C hold. The resulting PCR products containing Gateway attB-sites were gel purified and used for Gateway cloning.

Library cloning

Next, PCR products were cloned into pDONR221 to generate three regional pENTR221 libraries. For each of the three regions a 25 µL Gateway BP reaction was prepared: 114 ng PCR product, 375 ng pDONR221, 5 µL Gateway BP Clonase II enzyme mix (ThermoFisher), TE Buffer pH 8.0 to 25 µL. Reactions were incubated overnight at RT. The following day, 3.1 µL proteinase K was added and reactions were incubated for 10 min at 37 °C. For each region, 4 µL BP reaction was transformed into 100 µL of NEB 10-beta electrocompetent *E. coli* cells using electroporation. The cells were recovered in 3900 µL NEB 10-beta outgrowth medium in 50 mL Falcon tubes at 37 °C for 1 h. Then, cells were plated on LB containing kanamycin and were incubated at 37 °C overnight. A minimum of 500,000 colonies were obtained for each region. The following day, the cells were scraped from the plates using sterile water and plasmid DNA was extracted from 400 OD_{600nm} units (Nucleobond Xtra Midiprep Kit, Macherey–Nagel).

The resulting pENTR221-GCK libraries were used in large-scale Gateway LR reactions to clone the libraries into the pAG416GPD-EGFP vector. For each region, the following was mixed: 216.9 ng pENTR221-GCK library, 450 ng pAG416GPD-EGFP vector, 6 µL Gateway LR Clonase II enzyme mix (ThermoFisher), TE Buffer pH 8.0 to 30 µL.

The reactions were incubated at RT overnight. Next day, 3 µL proteinase K was added to each reaction and tubes were incubated at 37 °C for 10 min. The LR reactions were transformed into NEB 10-beta electrocompetent *E. coli* cells using electroporation of 4 µL reaction per 100 µL cells. Cells were recovered in NEB 10-beta outgrowth medium for 1 h at 37 °C, and were then plated on LB containing ampicillin for incubation at 37 °C overnight. A minimum of 500,000 colonies were obtained for each regional library. The following day, cells were scraped from plates using sterile water and plasmid DNA was extracted from 400 OD_{600nm} units (Nucleobond Xtra Midiprep Kit, Macherey–Nagel). The resulting plasmid DNA was used for yeast transformation.

Library yeast transformation

The GCK expression libraries were transformed into the *hxk1Δ hxk2Δ glk1Δ* strain as described in [94] using a 30 × scale-up and 30–60 µg of plasmid DNA. Small aliquots of each transformation were plated on SC-URA galactose in duplicate for colony counting. The rest of each transformation was diluted in SC-URA galactose to an OD_{600nm} of 0.2, and incubated at 30 °C with shaking until saturated. A minimum of 500,000 colonies were obtained for each regional library. For each regional library, 36 OD_{600nm} units of yeast cells were harvested in duplicate to serve as pre-selection samples. Pellets were stored at -20 °C before plasmid DNA extraction. In parallel with library transformations, a vector (pAG416GPD-EGFP-ccdB) and GCK WT control were transformed in small scale, and 36 OD_{600nm} units in duplicate were stored at -20 °C for DNA extraction.

Library selection

Yeast transformations were next used for selection of GCK activity. For each regional library, 20 OD_{600nm} units of yeast cells were harvested in duplicate to serve as two biological repeats. The cells were washed three times in sterile water and resuspended in 500 µL sterile water. The cells were then plated on large plates (500 cm^2) of SC-URA 0.2% glucose. Plates were incubated at 30 °C for three days. After selection, yeast cells were scraped off plates using 30 mL milliQ water and 36 OD_{600nm} units from each plate were harvested to serve as post-selection samples. Cell pellets were stored at -20 °C prior to plasmid DNA extraction. In parallel 2.6 OD_{600nm} units of vector (pAG416GPD-EGFP-ccdB) and GCK WT yeast transformations were washed, plated and incubated for three days at 30 °C in duplicate. After incubation, 36 OD_{600nm} units of cells were harvested and stored at -20 °C prior to plasmid DNA extraction.

Library sequencing

To determine the change in frequency of variants after selection for GCK activity, we sequenced the GCK ORF before and after growth on glucose medium. The GCK ORF was sequenced in 14 tiles, such that each tile could be sequenced on both strands to reduce base-calling errors. Region one (tile 1–5) and two (tile 6–10) both spanned five tiles, while region three was sequenced in four tiles (tile 10–14).

First, plasmid DNA was extracted from yeast cells for both duplicates of the regional libraries and the GCK WT control pre- and post-selection. Plasmid DNA was extracted from 9 OD_{600nm} units using the ChargeSwitch Plasmid Yeast Mini kit (Invitrogen). Plasmid DNA was adjusted to equal concentrations, and was then used for two rounds of

PCR to first amplify each tile and then add index sequences to allow for multiplexing. In the first PCR, each tile was amplified with primers containing a binding site for Illumina sequencing adapters. For each tiling PCR, the following was mixed: 20 µL Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 18 µL plasmid library template. The sequences of forward and reverse primers for each tile are listed in the supplemental material (SKG_tilenumber_fw/rev). Tiles were amplified using the following program: 98 °C 30 s, 21 cycles of 98 °C 10 s, 63 °C 30 s, 72 °C 60 s, followed by 72 °C 7 min and 4 °C hold.

In the next PCR, Illumina index adapters were added to all amplified tiles. Each PCR consisted of: 20 µL Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), 2 µL 10 µM i5 indexing adapter, 2 µL 10 µM i7 indexing adapter, 1 µL 1:10 diluted PCR product, 15 µL nuclease-free water. Tiles were amplified using the following program: 98 °C 30 s, 7 cycles of 98 °C 15 s, 65 °C 30 s, 72 °C 120 s, followed by 72 °C 7 min and hold at 4 °C. An equal volume of each PCR was then pooled and 100 µL were used for gel extraction from a 4% E-Gel EX Agarose Gel (Invitrogen). The fragment size and quality of the extracted DNA were tested using a 2100 Bioanalyzer system (Agilent), and DNA concentration was determined using Qubit (ThermoFisher). Finally, libraries were paired-end sequenced using an Illumina NextSeq 550.

Sequencing data analysis

The TileSeqMave (https://github.com/jweile/tileseqMave, version 0.6.0.9000) and TileSeq mutation count (https://github.com/RyogaLi/tileseq_mutcount, version 0.5.9) pipelines were used to process sequencing data to obtain variant activity scores.

Statistical analyses

The empirically determined standard deviations of the activity scores were likely imprecise, as they were based on only two replicates. To obtain more robust estimates of standard deviations, we used Bayesian regularization or refinement as described by Baldi and Long [95]. The prior estimate of the standard deviation was obtained by linear regression based on the fitness score and the read counts from the permissive condition. The prior was combined with the empirical standard deviation using Baldi and Long's original formula:

$$\sigma^2 = \frac{v_n * \sigma_n^2}{v_n - 2} = \frac{v_0 \sigma_0^2 + (n-1)s^2}{v_0 + n - 2}$$

Here, σ_0 is the prior estimate obtained from regression, v_0 represents the degrees of freedom assigned to the prior estimate, n is the number of experimental replicates, and s is the empirical standard deviation. The full code used for performing error regularization can be found on Github (https://github.com/jweile/tileseqMave, version 0.6.0.9000).

Confidence intervals (CIs) were obtained using the SciPy bootstrap function with 10,000 resamples.

Imputation of missing activity scores

The Human Protein Variant Effect Map Imputation Toolkit webserver [45, 46], was used to impute activity scores for missing variants. The webserver was run using standard parameters and with equal quality index on all variant scores. The original and imputed refined scores showed a Pearson's correlation of 0.985. The imputed scores were only used for Additional file 1: Fig. S3 except for the benign variant p.G68D. The imputed score of p.G68D was used for Fig. 3CD and receiver-operating characteristic (ROC) analyses, due to the limited number of benign variants.

Evolutionary conservation analysis

The HHblits suite [96, 97] and GEMME package [98] were used to evaluate evolutionary distances from the WT GCK sequence (Uniprot ID: P35557—isoform 1).

The MSA was generated using HHblits with an E-value threshold of 10^{-20} and using UniClust30 as sequence database. The resulting MSA contained 1179 sequences. Two additional filters were applied to the HHblits output MSA: First, all the columns that were not present in the WT GCK sequence were removed. Second, all the sequences with more than 50% gaps were removed, leaving 1079 sequences in the MSA. The GEMME package was run using standard parameters. Median scores were evaluated for each position using all the 19 substitutions.

Thermodynamic stability measurements (ΔΔG)

Rosetta package (GitHub SHA1 c7009b3115c22daa9efe2805d9d1ebba08426a54) with Cartesian ΔΔG protocol [71, 99] was used to predict changes in thermodynamic stability from the crystal structures [15] of super-open (PDB 1V4T) and closed (PDB 1V4S) conformations of GCK. The values obtained from Rosetta in internal Energy Unit were divided by 2.9 to convert the unit to kcal/mol [71]. Median scores were evaluated for each position using all the 19 substitutions.

Fasting plasma glucose study population

Variants in *GCK* were identified using sequencing. Samples were collected from a population-based cohort of 6,058 individuals both with and without diabetes [50], 2,930 patients with newly-diagnosed diabetes [51], patients diagnosed with *GCK*-MODY [53] and from a population of 1,146 Danish children [52]. Individuals were included if they carried one missense *GCK* variant according to transcript NM_000162 and if a measure of fasting plasma glucose was available. Measures of fasting plasma glucose were examined using a glucose oxidase method (Granutest; Merck, Darmstadt, Germany) in the population based cohort and in samples from patients with known *GCK*-MODY [50, 53], an enzymatic hexokinase method (Gluco-quant Glucose/HK, Roche Diagnostics) in newly diagnosed diabetes patients [51], and using a Dimension Vista® 1500 Analyzer (Siemens, Erlangen, Germany) in children [52]. Samples were excluded if fasting plasma glucose level exceeded 9 mM.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-023-02935-8.

Additional file 1: Fig. S1. Expression of GCK variants. **Fig. S2.** Co-expression of GKRP and GCK. **Fig. S3.** Imputed map of glucokinase variant activity. **Fig. S4.** Activity scores mapped onto GCK ribbon diagram. **Fig. S5.** Evolutionary analysis of GCK homologous sequences. **Fig. S6.** Correlations between evolutionary conservation and activity scores. **Fig. S7.** Rosetta ΔΔG heatmaps. **Fig. S8.** Positions predicted to shift GCK towards the closed conformation are enriched at the allosteric activator site.Please check additional files if captured correctly.Error during converting author query response. Please check the eproofing link or feedback pdf for details

Additional file 2.

Additional file 3.

Additional file 4. Review history.

Review history
The review history is available as Additional file 3.

Peer review information
Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions
S.G., M.C., M.G., G.S. and A.P.G. performed the experiments. S.G., M.C., M.G., A.G.C., A.P.G., G.S., R.L., D.T., A.S., J.W., A.L.G., T.H., F.P.R, K.L.-L. and R.H.-P. analyzed the data. F.P.R., K.L.-L. and R.H.-P. conceived the study. S.G. wrote the paper. The authors read and approved the final manuscript.

Availability of data and materials
The Illumina FASTQ files can be accessed at the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE198878 [100]. The activity scores have been deposited on MaveDB (https://www.mavedb.org) under accession number urn:mavedb:00,000,096-a, and are also available at Zenodo (https://doi.org/10.5281/zenodo.7636310select_t1_simple_aa.csv) [101]. All data generated or analyzed during this study are included in this published article and its supplementary information file.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
F.P.R. is ashareholder and advisor for SeqWell, Constantiam, BioSymetrics, and a shareholderof Ranomics, and his lab has received research support from Alnylam, Deep Genomics, Beam Therapeutics and Biogen, Inc. A.L.G.'s spouse holds stock options in Roche and is an employee of Genentech.

References

1. Froguel P, Vaxillaire M, Sun F, Velho G, Zouali H, Butel MO, et al. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature. 1992;356:162–4.
2. Hattersley AT, Turner RC, Patel P, O'Rahilly S, Hattersley AT, Patel P, et al. Linkage of type 2 diabetes to the glucokinase gene. Lancet. 1992;339:1307–10.
3. Stride A, Shields B, Gill-Carey O, Chakera AJ, Colclough K, Ellard S, et al. Cross-sectional and longitudinal studies suggest pharmacological treatment used in patients with glucokinase mutations does not alter glycaemia. Diabetologia. 2014;57:54–6.
4. Steele AM, Shields BM, Wensley KJ, Colclough K, Ellard S, Hattersley AT. Prevalence of vascular complications among patients with glucokinase mutations and prolonged mild hyperglycemia. JAMA. 2014;311:279–86.
5. Szopa M, Wolkow J, Matejko B, Skupien J, Klupa T, Wybrańska I, et al. Prevalence of retinopathy in adult patients with GCK-MODY and HNF1A-MODY. Exp Clin Endocrinol Diabetes. 2015;123:524–8.
6. Pihoker C, Gilliam LK, Ellard S, Dabelea D, Davis C, Dolan LM, et al. Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and Glucokinase: results from the SEARCH for Diabetes in Youth. J Clin Endocrinol Metab. 2013;98:4055–62.
7. Kleinberger JW, Copeland KC, Gandica RG, Haymond MW, Levitsky LL, Linder B, et al. Monogenic diabetes in overweight and obese youth diagnosed with type 2 diabetes: the TODAY clinical trial. Genet Med. 2018;20:583–90.
8. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia. 2010;53:2504–8.
9. German MS. Glucose sensing in pancreatic islet beta cells: the key role of glucokinase and the glycolytic intermediates. Proc Natl Acad Sci. 1993;90:1781–5.
10. Meglasson MD, Matschinsky FM. Pancreatic islet glucose metabolism and regulation of insulin secretion. Diabetes Metab Rev. 1986;2:163–214.
11. Meglasson MD, Matschinsky FM. New perspectives on pancreatic islet glucokinase. Am J Physiol. 1984;246:E1-13.
12. Ferre T, Riu E, Bosch F, Valera A. Evidence from transgenic mice that glucokinase is rate limiting for glucose utilization in the liver. FASEB J. 1996;10:1213–8.
13. Vinuela E, Salas M, Sols A. Preliminary communications J glucokinase and hexokinase in liver in relation to glycogen synthesis*. J Biol Chem. 1963;238:PC1175-7.
14. Storer AC, Cornish Bowden A. Kinetics of rat liver glucokinase. Co-operative interactions with glucose at physiologically significant concentrations. Biochem J. 1976;159:7–14.
15. Kamata K, Mitsuya M, Nishimura T, Eiki JI, Nagata Y. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. Structure. 2004;12:429–38.
16. Lin SX, Neet KE. Demonstration of a slow conformational change in liver glucokinase by fluorescence spectroscopy. J Biol Chem. 1990;265:9670–5.
17. Larion M, Hansen AL, Zhang F, Bruschweiler-Li L, Tugarinov V, Miller BG, et al. Kinetic cooperativity in human pancreatic glucokinase originates from millisecond dynamics of the small domain. Angew Chemie - Int Ed. 2015;54:8129–32.
18. Sternisha SM, Whittington AC, Martinez Fiesco JA, Porter C, McCray MM, Logan T, et al. Nanosecond-timescale dynamics and conformational heterogeneity in human GCK regulation and disease. Biophys J. 2020;118:1109–18.
19. Larion M, Salinas RK, Bruschweiler-Li L, Brüschweiler R, Miller BG. Direct evidence of conformational heterogeneity in human pancreatic glucokinase from high-resolution nuclear magnetic resonance. Biochemistry. 2010;49:7969–71.
20. Liu S, Ammirati MJ, Song X, Knafels JD, Zhang J, Greasley SE, et al. Insights into mechanism of glucokinase activation: observation of multiple distinct protein conformations. J Biol Chem. 2012;287:13598–610.
21. Zhang J, Li C, Chen K, Zhu W, Shen X, Jiang H. Conformational transition pathway in the allosteric process of human glucokinase. Proc Natl Acad Sci. 2006;103:13368–73.
22. Heredia VV, Thomson J, Nettleton D, Sun S. Glucose-induced conformational changes in glucokinase mediate allosteric regulation: transient kinetic analysis. Biochemistry. 2006;45:7553–62.
23. Sternisha SM, Miller BG. Molecular and cellular regulation of human glucokinase. Arch Biochem Biophys. 2019;663:199–213.
24. Osbak KK, Colclough K, Saint-Martin C, Beer NL, Bellanné-Chantelot C, Ellard S, et al. Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. Hum Mutat. 2009;30:1512–26.
25. Christesen HBT, Jacobsen BB, Odili S, Buettger C, Cuesta-Munoz A, Hansen T, et al. The second activating glucokinase mutation (A456V). Diabetes. 2002;51:1240–6.
26. Gloyn AL, Noordam K, Willemsen MAAP, Ellard S, Lam WWK, Campbell IW, et al. Insights into the biochemical and genetic basis of glucokinase activation from naturally occurring hypoglycemia mutations. Diabetes. 2003;52:2433–40.
27. Glaser B, Kesavan P, Heyman M, Davis E, Cuesta A, Buchs A, et al. Familial hyperinsulinism caused by an activating glucokinase mutation. N Engl J Med. 1998;338:226–30.
28. Njølstad PR, Sagen JV, Bjørkhaug L, Odili S, Shehadeh N, Bakry D, et al. Permanent neonatal diabetes caused by glucokinase deficiencyinborn error of the glucose-insulin signaling pathway. Diabetes. 2003;52:2854–60.
29. Njølstad PR, Søvik O, Cuesta-Muñoz A, Bjørkhaug L, Massa O, Barbetti F, et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. N Engl J Med. 2001;344:1588–92.
30. Ma Y, Han X, Zhou X, Li Y, Gong S, Zhang S, et al. A new clinical screening strategy and prevalence estimation for glucokinase variant-induced diabetes in an adult Chinese population. Genet Med. 2019;21:939–47.
31. Chakera AJ, Spyer G, Vincent N, Ellard S, Hattersley AT, Dunne FP. The 0.1% of the population with glucokinase monogenic diabetes can be recognized by clinical characteristics in pregnancy: the Atlantic diabetes in pregnancy cohort. Diabetes Care. 2014;37:1230–6.

32. Schiabor Barrett KM, Bolze A, Ni Y, White S, Isaksson M, Sharma L, et al. Positive predictive value highlights four novel candidates for actionable genetic screening from analysis of 220,000 clinicogenomic records. Genet Med. 2021;23:2300–8.

33. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11:801–7.

34. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7:741–6.

35. Mayordomo I, Sanz P. Human pancreatic glucokinase (GlkB) complements the glucose signalling defect of Saccharomyces cerevisiae hxk2 mutants. Yeast. 2001;18:1309–16.

36. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.

37. García-Herrero CM, Rubio-Cabezas O, Azriel S, Gutierrez-Nogués A, Aragonés A, Vincent O, et al. Functional characterization of MODY2 mutations highlights the importance of the fine-tuning of glucokinase and its role in glucose sensing. PLoS ONE. 2012;7:e30518.

38. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 2020;48:D835–44.

39. Hattersley AT, Beards F, Ballantyne E, Appleton M, Harvey R, Ellard S. Mutations in the glucokinase gene of the fetus result in reduced birth weight. Nat Genet. 1998;19:268–70.

40. Ellard S, Beards F, Allen LIS, Shepherd M, Ballantyne E, Harvey R, et al. A high prevalence of glucokinase mutations in gestational diabetic subjects selected by clinical criteria. Diabetologia. 2000;43:250–3.

41. Massa O, Meschi F, Cuesta-Munoz A, Caumo A, Cerutti F, Toni S, et al. High prevalence of glucokinase mutations in Italian children with MODY. Influence on glucose tolerance, first-phase insulin response, insulin sensitivity and BMI. Diabetologia. 2001;44:898–905.

42. Stoffel M, Froguel PH, Takeda J, Zouali H, Vionnet N, Nishi S, et al. Human glucokinase gene: isolation, characterization, and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. Proc Natl Acad Sci. 1992;89:7698–702.

43. van Schaftingen E, Veiga-da-Cunha M, Niculescu L. The regulatory protein of glucokinase. Biochem Soc Trans. 1997;25:136–40.

44. van Schaftingen E. A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. Eur J Biochem. 1989;179:179–84.

45. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, et al. A framework for exhaustively mapping functional missense variants. Mol Syst Biol. 2017;13:957.

46. Wu Y, Weile J, Cote AG, Sun S, Knapp J, Verby M, et al. A web application and service for imputing and visualizing missense variant effect maps. Bioinformatics. 2019;35:3191–3.

47. Gloyn AL, Odili S, Buettger C, Njølstad PR, Shiota C, Magnuson MA, et al. Glucokinase and the regulation of blood sugar A mathematical model predicts the threshold for glucose stimulated insulin release for GCK gene mutations that cause hyper-and hypoglycemia. Nov Ther Front Diabetes Basel, Karger. 2004;16:92–109.

48. Beer NL, Osbak KK, Van De Bunt M, Tribble ND, Steele AM, Wensley KJ, et al. Insights into the pathogenicity of rare missense GCK variants from the identification and functional characterization of compound heterozygous and double mutations inherited in Cis. Diabetes Care. 2012;35:1482–4.

49. Steele AM, Tribble ND, Caswell R, Wensley KJ, Hattersley AT, Gloyn AL, et al. The previously reported T342P GCK missense variant is not a pathogenic mutation causing MODY. Diabetologia. 2011;54:2202–5.

50. Glümer C, Jørgensen T, Borch-Johnsen K. Prevalences of diabetes and impaired glucose regulation in a Danish population the Inter99 study. Diabetes Care. 2003;26:2335–40.

51. Christensen DH, Nicolaisen SK, Berencsi K, Beck-Nielsen H, Rungby J, Friborg S, et al. Danish Centre for Strategic Research in Type 2 Diabetes (DD2) project cohort of newly diagnosed patients with type 2 diabetes: a cohort profile. BMJ Open. 2018;8:e017273.

52. Kloppenborg JT, Gamborg M, Fonvig CE, Nielsen TRH, Pedersen O, Johannesen J, et al. The effect of impaired glucose metabolism on weight loss in multidisciplinary childhood obesity treatment. Pediatr Diabetes. 2018;19:366–74.

53. Johansen A, Ek J, Mortensen HB, Pedersen O, Hansen T. Half of clinically defined maturity-onset diabetes of the young patients in Denmark do not have mutations in HNF4A, GCK, and TCF1. J Clin Endocrinol Metab. 2005;90:4607–14.

54. Wędrychowicz A, Tobór E, Wilk M, Ziółkowska-Ledwith E, Rams A, Wzorek K, et al. Phenotype heterogeneity in glucokinase–maturity-onset diabetes of the young (GCK-MODY) patients. J Clin Res Pediatr Endocrinol. 2017;9:246.

55. Cuesta-Muñoz AL, Tuomi T, Cobo-Vuilleumier N, Koskela H, Odili S, Stride A, et al. Clinical heterogeneity in monogenic diabetes caused by mutations in the glucokinase gene (GCK-MODY). Diabetes Care. 2010;33:290–2.

56. Simonis-Bik AMC, Eekhoff EMW, Diamant M, Boomsma DI, Heine RJ, Dekker JM, et al. The heritability of HbA1c and fasting blood glucose in different measurement settings. Twin Res Hum Genet. 2008;11:597–602.

57. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021;53:840–60.

58. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42:105–16.

59. Gloyn AL, Odili S, Zelent D, Buettger C, Castleden HAJ, Steele AM, et al. Insights into the structure and regulation of glucokinase from a novel mutation (V62M), which causes maturity-onset diabetes of the young. J Biol Chem. 2005;280:14105–13.

60. Davis EA, Cuesta-Muñoz A, Raoul M, Buettger C, Sweet I, Moates M, et al. Mutants of glucokinase cause hypoglycaemia- and hyperglycaemia syndromes and their analysis illuminates fundamental quantitative concepts of glucose homeostasis. Diabetologia. 1999;42:1175–86.

61. Burke CV, Buettger CW, Davis EA, McClane SJ, Matschinsky FM, Raper SE. Cell-biological assessment of human glucokinase mutants causing maturity-onset diabetes of the young type 2 (MODY-2) or glucokinase-linked hyperinsulinaemia (GK-HI). Biochem J. 1999;342:345.

62. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res. 2005;15:978–86.

63. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11:863–74.

64. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073–82.

65. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.

66. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.

67. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31:2745–7.

68. Larion M, Salinas RK, Bruschweiler-Li L, Miller BG, Brüschweiler R. Order-disorder transitions govern kinetic cooperativity and allostery of monomeric human glucokinase. PLOS Biol. 2012;10:e1001452.

69. Boucher JI, Bolon DNA, Tawfik DS. Quantifying and understanding the fitness effects of protein mutations: laboratory versus nature. Protein Sci. 2016;25:1219–26.

70. Heredia VV, Carlson TJ, Garcia E, Sun S. Biochemical basis of glucokinase activation and the regulation by glucokinase regulatory protein in naturally occurring mutations. J Biol Chem. 2006;281:40201–7.

71. Park H, Bradley P, Greisen P, Liu Y, Mulligan VK, Kim DE, et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. J Chem Theory Comput. 2016;12:6201–12.

72. Fearon K, McClendon V, Bonetti B, Bedwell DM. Premature translation termination mutations are efficiently suppressed in a highly conserved region of yeast Ste6p, a member of the ATP-binding cassette (ABC) transporter family. J Biol Chem. 1994;269:17802–8.

73. Bonetti B, Fu L, Moon J, Bedwell DM. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. J Mol Biol. 1995;251:334–45.

74. Scanlon TC, Gray EC, Griswold KE. Quantifying and resolving multiple vector transformants in S. cerevisiae plasmid libraries. BMC Biotechnol. 2009;9:1–10.

75. Sun S, Weile J, Verby M, Wu Y, Wang Y, Cote AG, et al. A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. Genome Med. 2020;12:1–18.

76. Weile J, Kishore N, Sun S, Maaieh R, Verby M, Li R, et al. Shifting landscapes of human MTHFR missense-variant effects. Am J Hum Genet. 2021;108:1283–300.

77. Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol. 2013;425:1363–77.

78. Ahler E, Register AC, Chakraborty S, Fang L, Dieter EM, Sitko KA, et al. A combined approach reveals a regulatory mechanism coupling Src's Kinase activity, localization, and phosphotransferase-independent functions. Mol Cell. 2019;74:393-408.e20.

79. García-Herrero CM, Galán M, Vincent O, Flández B, Gargallo M, Delgado-Alvarez E, et al. Functional analysis of human glucokinase gene mutations causing MODY2: exploring the regulatory mechanisms of glucokinase activity. Diabetologia. 2007;50:325–33.

80. Sayed S, Langdon DR, Odili S, Chen P, Buettger C, Schiffman AB, et al. Extremes of clinical and enzymatic phenotypes in children with hyperinsulinism caused by glucokinase activating mutations. Diabetes. 2009;58:1419–27.

81. Gašperíková D, Tribble ND, Staník J, Hučková M, Mišovicová N, Van De Bunt M, et al. Identification of a Novel β-Cell Glucokinase (GCK) Promoter Mutation (−71G>C) That Modulates GCK Gene Expression Through Loss of Allele-Specific Sp1 Binding Causing Mild Fasting Hyperglycemia in Humans. Diabetes. 2009;58:1929–35.

82. Miller SP, Anand GR, Karschnia EJ, Bell GI, LaPorte DC, Lange AJ. Characterization of glucokinase mutations associated with maturity-onset diabetes of the young type 2 (MODY-2): different glucokinase defects lead to a common phenotype. Diabetes. 1999;48:1645–51.

83. Sagen JV, Odili S, Bjørkhaug L, Zelent D, Buettger C, Kwagh J, et al. From clinicogenetic studies of maturity-onset diabetes of the young to unraveling complex mechanisms of glucokinase regulation. Diabetes. 2006;55:1713–22.

84. Grimsby J, Sarabu R, Corbett WL, Haynes NE, Bizzarro FT, Coffey JW, et al. Allosteric activators of glucokinase: potential role in diabetes therapy. Science. 2003;301:370–3.

85. Matschinsky FM. Assessing the potential of glucokinase activators in diabetes therapy. Nat Rev Drug Discov. 2009;8:399–416.

86. De Ceuninck F, Kargar C, Ilic C, Caliez A, Rolin JO, Umbdenstock T, et al. Small molecule glucokinase activators disturb lipid homeostasis and induce fatty liver in rodents: a warning for therapeutic applications in humans. Br J Pharmacol. 2013;168:339–53.

87. Rees MG, Gloyn AL. Small molecular glucokinase activators: has another new anti-diabetic therapeutic lost favour? Br J Pharmacol. 2013;168:335–8.

88. Bonadonna RC, Heise T, Arbet-Engels C, Kapitza C, Avogaro A, Grimsby J, et al. Piragliatin (RO4389620), a novel glucokinase activator, lowers plasma glucose both in the postabsorptive state and after a glucose challenge in patients with type 2 diabetes mellitus: a mechanistic study. J Clin Endocrinol Metab. 2010;95:5028–36.

89. Meininger GE, Scott R, Alba M, Shentu Y, Luo E, Amin H, et al. Effects of MK-0941, a novel glucokinase activator, on glycemic control in insulin-treated patients with type 2 diabetes. Diabetes Care. 2011;34:2560–6.

90. Zhu D, Gan S, Liu Y, Ma J, Dong X, Song W, et al. Dorzagliatin monotherapy in Chinese patients with type 2 diabetes: a dose-ranging, randomised, double-blind, placebo-controlled, phase 2 study. Lancet Diabetes Endocrinol. 2018;6:627–36.

91. Vella A, Freeman JLR, Dunn I, Keller K, Buse JB, Valcarce C. Targeting hepatic glucokinase to treat diabetes with TTP399, a hepatoselective glucokinase activator. Sci Transl Med. 2019;11:3441.

92. Alberti S, Gitler AD, Lindquist S. A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. Yeast. 2007;24:913–9.

93. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007;2:31–4.

94. Gietz RD, Schiestl RH. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007;2:38–41.

95.   Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinformatics. 2001;17:509–19.

96.   Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 2011;9:173–5.

97.   Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20:1–15.

98.   Laine E, Karami Y, Carbone A. GEMME: A Simple and Fast Global Epistatic Model Predicting Mutational Effects. Mol Biol Evol. 2019;36:2604–19.

99.   Frenz B, Lewis SM, King I, DiMaio F, Park H, Song Y. Prediction of protein mutational free energy: benchmark and sampling improvements increase classification accuracy. Front Bioeng Biotechnol. 2020;8:1175.

100.   Gersing S, Cagiada M, Gebbia M, Gjesing AP, Coté AG, Seesankar G, Li R, Tabet D, Weile J, Stein A, Gloyn AL, Hansen T, Roth FP, Lindorff-Larsen K, Hartmann-Petersen R. A comprehensive map of human glucokinase variant activity. GSE198878. Gene Expression Omnibus. 2023. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198878.

101.   Gersing S, Cagiada M, Gebbia M, Gjesing AP, Coté AG, Seesankar G, Li R, Tabet D, Weile J, Stein A, Gloyn AL, Hansen T, Roth FP, Lindorff-Larsen K, Hartmann-Petersen R. A comprehensive map of human glucokinase variant activity. Zenodo. 2023. https://doi.org/10.5281/zenodo.7636310

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.