

Identifying therapeutic drug targets for rare and common forms of short stature

Karol Estrada*¹, Steven Froelich¹, Arthur Wuster¹, Christopher R. Bauer¹, Teague Sterling¹, Wyatt T. Clark¹, Yuanbin Ru¹, Marena Trinidad¹, Hong Phuc Nguyen¹, Amanda R. Luu¹, Daniel J. Wendt¹, Gouri Yogalingam¹, Guoying Karen Yu¹, Jonathan H. LeBowitz¹, Lon R. Cardon¹

¹BioMarin Pharmaceutical Inc., Novato, CA 94949, USA

*Corresponding author

Email: karol.estrada@bmrn.com

Abstract

While GWAS of common diseases has delivered thousands of novel genetic findings, prioritizing genes for translation to therapeutics has been challenging. Here, we propose an approach to resolve that issue by identifying genes that have both gain of function (GoF) and loss of function (LoF) mutations associated with opposing effects on phenotype (Bidirectional Effect Selected Targets, BEST). Bidirectionality is a desirable feature of the best targets because it implies both a causal role on the phenotype in one direction and that modulating the target activity might be safe and therapeutically beneficial in the other.

We used height, a highly heritable trait and a model of complex diseases, to test our approach. Using 34,231 individuals with exome sequence data and height, we identified five genes (*IGF1R*, *NPPC*, *NPR2*, *FGFR3*, and *SHOX*) with evidence for bidirectional effects on stature. Rare protein-altering variants significantly increased risk for idiopathic short stature (ISS) (OR=2.75, $p=3.99 \times 10^{-8}$). These genes are key members of the only two pathways successfully targeted for short stature: Growth Hormone/Insulin-like growth factor 1 axis and C-type Natriuretic peptide (CNP) for Achondroplasia, a monogenic form of dwarfism. We assayed a subset of *NPR2* mutations and identified those with elevated (GoF) and diminished (LoF) activity and found that a polygenic score for height modulates the penetrance of pathogenic variants. We also demonstrated that adding exogenous CNP (encoded by *NPPC*) rescues the *NPR2* haploinsufficiency molecular phenotype in a CRISPR-engineered cell line, thus validating its potential therapeutic treatment for inherited forms of short stature. Finally, we found that these BEST targets increase the probability of success in clinical trials above and beyond targets

with other genetic evidence. Our results show the value of looking for bidirectional effects to identify and validate drug targets.

Introduction

Drug discovery has recently focused on target-based approaches to treat human disease instead of phenotypic screens. There are several ways in which potential targets are identified, the most common of which is the exploration of targets that have shown a mechanistic effect in either cell or animal models. The main disadvantage of these strategies is the high risk of failure when drug efficacy is tested in humans. Human genetic evidence for a target can improve the odds of success. A target with GWAS evidence has a two-fold higher probability of success compared to those targets with no evidence, and this is increased to 8-fold if the target is a gene that causes a Mendelian disorder.¹ GWAS studies have begun to produce hundreds of hits for complex phenotypes,² leading to challenges in prioritizing the large number of potential targets for those most likely to be amenable to therapeutic intervention. There are different approaches to prioritizing these genetic targets.³⁻⁵ The use of an allelic series of genetic variants that associate with the outcome of interest and mimic a dose-response curve can provide a direct indication that a promising target has been selected.⁵ It can also help to quantify the magnitude of the required modulation needed to achieve efficacy, and even inform us of potential adverse events.⁵

Some recent drug targets with strong human genetic evidence show an additional characteristic: they have allelic series that span a bidirectional effect. Gain-of-function (GoF) mutations are experiments of nature that might mimic the effect of a drug that increases the activity of a particular gene. Conversely, Loss-of-function (LoF, predicted LoF i.e., stop-gained, splice site disrupting, frameshift) mutations mimic a drug decreasing the activity of a gene

product. The presence of mutations driving a phenotype into both supra- and sub- physiological levels supports the idea that a drug mimicking that effect may have similar clinical outcome and a tolerable safety profile. This also suggests that an entire disease population could be treated, even those patients with a different genetic etiology such as individuals with a high polygenic risk score for a complex disease. Therefore, the allelic series should ideally span both extremes to make a stronger case for druggability. For example, *PCSK9* GoF variants associate with higher LDL levels and higher cardiovascular disease (CVD) risk.⁶ Conversely, *PCSK9* LoF have the exact opposite effect (Fig 1. lower LDL levels, lower CVD risk).⁷ This observation led to the therapeutic hypothesis that anti-*PCSK9* neutralizing antibodies would lower LDL levels in all patients regardless of their *PCSK9* carrier status and also decrease risk for CVD. This strategy was eventually validated in successful Phase III randomized clinical trials.^{8,9} Similarly, GoF and LoF variants with bidirectional effects on two members of the WNT pathway (*LRP5* and *SOST*)¹⁰ led to the creation of a new therapy for osteoporosis using antibodies against *SOST*.¹¹ This property of bidirectionality illustrated by the examples above should be generalizable to other complex phenotypes. We decided to test this hypothesis by using height as an example of a complex trait.

Height is a highly heritable (80%) and easily measured quantitative trait that has historically been used as a model to understand the genetic architecture of complex traits and diseases.¹² At the low end of the height distribution, Idiopathic Short Stature (ISS, defined as having a height at least two standard deviations (SDs) below the mean)¹² affects ~ 2% of the general

population. Mutations in several genes have been described in studies of familial ISS as well as in cohorts of ISS patients.¹³

Mendelian forms of extreme short stature (height at least 5 SDs below the mean) such as Achondroplasia, Acromesomelic dysplasia type Maroteaux (ADMD) and Osteogenesis Imperfecta type III, are less frequent (1/15,000, 1/1 million, unknown respectively).¹⁴⁻¹⁶ There is no approved treatment for these extreme forms of short stature, and while growth hormone has been used for idiopathic short stature, response to treatment is variable.¹⁷ Thousands of loci have been associated with height using GWAS.¹⁸⁻²³ Linking those association signals to genes, and those genes to mechanisms that may ultimately yield a new drug target has been challenging despite large scale efforts to map regulatory regions.^{24,25}

Here, we aimed to prioritize height-associated genes to identify a subset of genes most likely to be good therapeutic targets by selecting those genes exhibiting bidirectional effects on height (Bidirectional Effect Selected Targets, BEST). The attributes of these BEST targets included: having significant GWAS signals, known bidirectional effects on rare protein coding variants with an extreme phenotype, and population-level evidence of association with the trait using rare variant association tests.

To test this approach, we used data for adult height from published and unpublished sources and tested its relevance in 33,204 adults with exome sequence data from the UK Biobank project. We functionally validated genetic variants and compared the predicted effect of these

coding variants with their functional impact at a cellular level as well as with a clinical readout.

We then conducted experiments to validate the use of a therapeutic approach in a cell model to rescue the molecular phenotype.

Finally, to determine whether the bidirectionality approach was truly generalizable, we curated a large database of known mutations to identify 98 genes with other bidirectional effects on other human diseases. We were then able to show that drugs targeting these genes are more likely to be approved in clinical trials. This trend was consistent across multiple stages of regulatory approval and a wide range of disease indications.

Results

Growth genes

First, we obtained summary statistics from the largest published meta-analysis for height based on imputed GWAS data from ~700,000 individuals.¹⁸ In total, the 2067 closest genes to genetic signals (within 1-MB window) were obtained (Supplementary Tables 1 and 2, Methods). Other approaches to map GWAS signals to genes such as co-localization using eQTL data were not suitable since they require gene expression data in the tissue most relevant to the phenotype under study. For height, this would be limb bone growth plates and there are currently no gene expression datasets available for this tissue.^{26,27}

To identify genes that are regulators of growth, we analyzed the intersections of five gene lists including the list of genes from the GWAS. Growth regulators would be the most likely to contain rare coding mutations with bidirectional effects (i.e. short stature or skeletal dysplasia AND tall stature or overgrowth). First, we queried the Human Gene Mutation Database (HGMD version v2019_2) for genes associated with short or tall stature.²⁸ There were 47 genes annotated with at least one pathogenic variant reported in the literature to cause “short stature”. Only 20 genes were annotated as tall stature or overgrowth genes (Fig. 1 and Supplementary Table 2). Secondly, we used a manually curated list of 258 OMIM genes (248 short, 20 tall) which was created using the keywords: short stature, overgrowth, skeletal dysplasia, brachydactyly.²⁰ Third, we looked at the intersection of these four lists with the list of genes from GWAS (Fig. 1 and Supplementary Table 2). We found at the intersection of these five lists contained three genes known to be associated with height (Set 1: *IGF1R*, *NPPC*, *NPR2*).

Two additional genes were present in the intersection of all lists except GWAS (Set 2: *FGFR3*, *SHOX*). *FGFR3* was not in Set 1 as it was not the closest gene in the region of any GWAS signal and *SHOX* was not tested in the GWAS study as it is located on the X-chromosome (Supplementary Figs. 1-4). Thus, at least for height, at this point there was no added value on using the intersection with the list of genes coming from GWAS. The *NPPC/NPR2* and the *IGF1R*/growth hormone pathways have been implicated in GWAS of height.^{21,22}

Literature on extreme phenotypes such as short or tall stature is susceptible to publication bias. A randomly ascertained study population is required to obtain an unbiased assessment of the effects of rare coding variation of these genes on height on the general population. For that purpose, we used exome sequence data from 33,204 white British individuals from the UK Biobank project of which 572 (1.67%) reached the criterion of ISS (Supplementary Table 3. See Methods). After quality control, we identified 245,083 rare (allele frequency (AF) < 0.01%) missense and LoF variants in 1,964 genes with at least one rare protein coding variant. 1,918 genes with a mean cumulative allele count (AC) ≥ 20 remained after removing those with a cumulative AC < 20 (See methods).

Gene set analysis

To evaluate if any of the annotations (GWAS signal adjacent, short stature, tall stature, both short and tall, *etc.*) had a significant association with height, gene set analysis was performed using the combined effect of any rare protein modifying variants using the SNP-set (Sequence) Kernel Association Test (SKAT).²⁹ There were 32 possible intersections of the 5 annotations

described above and 25 of those contained at least one gene (Fig. 1 and Supplementary Table 4). Of these, surprisingly, only the two aforementioned gene sets (Set 1 and Set 2) showed significant association with height. They both contained bidirectional effects as reported by both HGMD and OMIM, namely: Set 1: *NPR2*, *NPPC*, *IGF1R* ($p=6.82 \times 10^{-7}$) and Set 2: *FGFR3* and *SHOX* ($p=1.75 \times 10^{-5}$) (Table 1, Fig. 1 and Supplementary Table 4). Given that both sets of genes share the property of having bidirectional effects on height, we combined all 607 independent rare protein-altering variants for the five genes in Set 1 or Set 2 using burden tests. The new group of five genes significantly decreased height ($\beta=-0.20$, $p=4.04 \times 10^{-11}$) and significantly increased risk for Idiopathic Short Stature (ISS) (OR=2.75, 95%CI [1.92 - 3.96], $p = 3.99 \times 10^{-8}$). Interestingly, this set of bidirectional genes has a much stronger risk for ISS as compared to the sets of genes reported for solely short stature or tall stature (Fig. 1C). Restricting the test to rare LoF variants resulted in an association with a larger decrease in height ($\beta=-0.85$, $p=2.49 \times 10^{-6}$). However, there were insufficient LoF carriers to calculate a precise odds ratio for ISS. Of the 607 variants, 158 were singletons and 192 (31.63%) were not reported in any of the public human databases and thus may represent novel pathogenic variants (Supplementary Table 5).

Associated genes

While most of these mutations are too rare to be significantly associated individually, each of the five associated genes (*FGFR3*, *IGF1R*, *NPPC*, *NPR2* and *SHOX*) were still nominally ($P < 0.05$) associated with height when considered individually, reflecting that no single gene was driving the gene set signal (Table 1, Online Methods). The direction of effects was consistent with LoF variants having Bonferroni-adjusted association with shorter stature ($\beta=-0.97$, $P=0.001$, for

NPR2) and nominal associations in the same direction for *IGF1R* and *NPPC*. Rare missense variants were associated with shorter height ($\beta=-0.22$, $P=0.002$; $\beta=-0.18$, $P=0.003$; and $\beta=-0.18$, $P=0.001$) for *NPR2*, *IGF1R* and *FGFR3* respectively. Combined LoF and missense variants in *NPR2* and *IGF1R* were also associated with increased risk for ISS (OR=3.31, $P=0.001$, OR= 2.85, $P=0.002$, respectively) (Table 1).

Entire gene deletions and/or mutations causing loss of protein function in *SHOX*, *IGF1R*, *NPPC*, *NPR2* have been reported in familial short stature with various degrees of severity (Table 1, Supplementary Table 6).^{15,30-33} Conversely, duplications, deletions of repressor regions, translocations and missense mutations leading to GoF in these genes were reported in individuals with tall stature or overgrowth.³⁴⁻³⁸ GoF mutations are hard to identify from in-silico predictions and usually require functional validation. GoF mutations in *FGFR3* cause Achondroplasia, the most common form of dwarfism.³⁹ For both *NPPC* and *IGF1R*, only large translocations and duplications have been reported with the overgrowth phenotype. In contrast, at least four independent *NPR2* missense mutations have been reported to be GoF.^{32,36,37,40} Importantly, mutations in these genes in animal and cell models also show bidirectional effects that are consistent with the directions of effects observed in humans. (Supplementary Table 7).

Polygenic Risk Scores

The polygenic nature of complex traits suggests that the combined effect of thousands of variants may be an independent risk factor to rare variants with strong effects possibly

modifying the overall risk. To test this hypothesis, we calculated polygenic risk scores (PRS) for height using the largest published GWAS meta-analysis for height that did not include any samples from the UK Biobank project (see Methods).²⁰ We divided our cohort into five equally-sized ($n=6,824$) PRS quintiles (PRS 1 being the lowest height, PRS 5 the tallest height). As expected, there was a dose-dependent relationship between increasing PRS score and mean height ($\beta=0.30$ per each PRS quintile increase) (Fig. 2A). Carriers of LoF variants in the five genes were consistently shorter than non-carriers across the five different PRS backgrounds (Fig. 2A and Supplementary Figs 5-9). Our data suggest that the combined effect of PRS and rare protein variants is consistent with an additive model: polygenic effects modulated height in both carriers and non-carriers. The mean effect of having a rare protein coding variant in any of the five genes was similar across PRS subgroups ($l^2 = 30\%$, $P=0.21$; $l_2=0$, $P_{het}=0.62$ for missense and LoF variants respectively) (Fig. 2A, Supplementary Figs. 10-16). We then calculated the risk for ISS across PRS groups using individuals at PRS3 who were non-carriers of mutations in the five genes as a reference. Non-carriers of the lowest PRS group were associated with increased risk for ISS and non-carriers of the highest PRS group with a decreased risk (Fig. 2B; OR=5.81, $P=2.74 \times 10^{-32}$; OR=0.25, $P=6.09 \times 10^{-6}$ for PRS 1 and PRS 5 respectively). We then evaluated the effect of rare coding variants of the five genes for ISS stratified by PRS group. Carriers of any of the five genes were at increased risk for ISS in the first three quintiles (OR=15.39, $P=4.44 \times 10^{-25}$; OR=4.76, $P=1.34 \times 10^{-4}$; OR=5.96, $P=1.21 \times 10^{-6}$; OR=2.16, $P=0.19$ Fig. 2B). There were not enough ISS individuals in the highest PRS group (14 ISS for PRS 5, all non-carriers).

Functional assessment of protein coding variants

A C-type natriuretic peptide analog (CNP, encoded by the *NPPC* gene which was also on the list of genes with bidirectional effects, Tables 1-2) is being investigated in a phase III Randomized Clinical Trial for Achondroplasia.⁴¹ Binding of CNP to its receptor (*NPR2*), triggers endochondral and skeletal growth via cGMP production.⁴² Furthermore, even wild-type rhesus monkeys exhibit increased growth velocity when treated with CNP.⁴³ Loss-of-function mutations in *NPR2* are responsible for dwarfism in mice and a lack of an intracellular cGMP response to CNP in cultured chondrocytes.⁴⁴

In order to explore the effect of gain of function mutations in *NPR2* on height, which should mimic the effect of adding exogenous CNP, we assessed the function of multiple missense mutations across the full spectrum of *NPR2* activity. We selected 39 variants for functional analysis: 9 *NPR2* protein-altering mutations that were present in UKBiobank ISS individuals, 7 *NPR2* missense mutations present in UKBiobank individuals with height > 1.2 SD, 5 *NPR2* variants in individuals with a mean height around 0 (in SDs), and 4 additional *NPR2* mutations with higher carrier number to complement the allelic series. Additionally, we tested 13 *NPR2* mutations that were reported in the literature for ISS and/or a more severe recessive form of short stature (Methods and Supplementary Table 8). To distinguish functional vs. neutral *NPR2* missense variants, we performed a series of experiments in cellular models (Methods and Fig. 3A). Constructs were tested in triplicate and the mean and SD of the functional readout (cGMP) was reported (Methods and Supplementary Table 8). We first confirmed the dynamic range of

the assay: 10 out of 11 previously reported pathogenic variants had mean cGMP < 0.20 relative to wildtype (wt), 2 out of 3 previously reported gain of function variants showed a mean cGMP of > 1 relative to wt (Supplementary Table 8). From the list of ISS patients found in our study, 10 out of 12 tested mutations were confirmed as low activity (mean cGMP < 0.2 relative to wt). (Fig. 3B, Supplementary Table 8). There was a significant association with increasing cGMP levels and increasing height ($\beta=0.91$ per 100% change in cGMP relative to wt, $P=2.7 \times 10^{-7}$; Fig. 3C). This assay data served as a better classifier for height than a predictor (Combined Annotation Dependent Depletion or CADD⁴⁵, ($\beta=-0.02$, $p=0.04$ for CADD unit change score; Fig. 3D).

One interesting outlier was identified carrying the *NPR2* variant R110G. While this variant showed a modest loss of function (cGMP = 0.70), the individual carrying it was very tall (height=3.10 SDs). A closer examination of this individual revealed that they were in the 99.85th percentile of the PRS for height (Supplementary Fig. 17). This prompted us to examine PRS groups in *NPR2* carriers of the variants we functionally tested. We grouped *NPR2* carriers by their PRS group and their residual *NPR2* activity. Despite having a small number of individuals in each group, as shown before there was an additive relationship between PRS and *NPR2* functional variants on height (Fig. 3E, Supplementary Table 9). This additive relationship was consistent with the combination of multiple risk factors of PRS and rare variants with strong effects

Phenotype rescue

In order to test whether a CNP-analog can rescue a LoF mutation in *NPR2*, we examined the response to CNP of cells engineered to have heterozygous or homozygous LoF mutations. We measured the amount of cGMP produced by CNP stimulation in CRISPR/Cas9-mediated knockout chondrocytes (Methods and Fig. 3F). cGMP production by NPR2 triggers a signaling cascade via activation of protein kinase cGMP-dependent 2 (*PRKG2*). Previous activation data reports cGMP EC₅₀ in the range of 40 to 360nM for activation of *PRKG2*.^{23,46,47} In the heterozygous *NPR2* knockout cells, a CNP dose >0.163nM is able to achieve an intracellular concentration exceeding the EC₅₀ range for PRKG2 activation (Fig. 3G). Whereas, in wild-type cells a CNP dose of 0.040nM is able to achieve the same cGMP concentration. These results demonstrate that CNP supplementation can achieve the cGMP levels necessary for PRKG2 activation and growth in cells with loss-of-function mutations of *NPR2*.

Generalizability to other diseases

Next, we wanted to assess the generalizability of our BEST strategy for other human traits or diseases. To address this, we manually identified 98 genes in HGMD that showed evidence of bidirectional effects for different gene variants (Online Methods, Supplementary Table 10). We then calculated the odds of clinical trial approval for those BEST target-indication pairs using a recently published database of clinical trial outcomes.⁴⁸ Finally, we compared those odds to the odds of targets with purely unidirectional evidence from OMIM and targets with evidence from GWAS (Online Methods, Supplementary Table 11). We found that overall, the BEST target-

indications that have reached Phase I have higher probability of successful transition to Approval as compared to other non-BEST target-indications across all MeSH similarity thresholds (Fig. 4A) and it reached a maximum OR=3.81 95%CI [2.81 - 4.74] at MeSh \geq 0.83. Importantly, that probability was significantly higher than that of other sources of genetic evidence (Fig. 4B, OR=2.12 95%CI [1.73 - 2.52], OR=1.12 95%CI [0.82 - 1.46] for OMIM unidirectional and GWAS targets respectively). BEST target-indications showed significant improvement in the probability of successful transition at all steps of the clinical development as compared to non-BEST target-indications (Fig. 4B). Interestingly, the phase where BEST target-indications showed the strongest effect was the transition from Phase II to Phase III (OR=1.46 95%CI [1.25 - 1.66]), a phase that is generally used as proof of concept or initial clinical efficacy.

Conclusions

In this study, our goal was to identify genes most likely to be drug targets for ISS by showing mutation bidirectionality (GoF and LoF), as well as GWAS effects with moderate effect sizes. We identified a set of five height genes with bidirectional effects and characterized the effect of these mutations across different height PRS backgrounds. Height has been used as model for common complex traits and diseases and its genetics are characterized by a large polygenic architecture. Despite this polygenic nature, there are critical genes where single LoF/GoF variants have large effects with magnitudes approaching the aggregated effects of thousands of common variants with small effects (as calculated by PRS). We demonstrate that genetic insults to these five genes contribute not only to rare skeletal malformations (Table 1, Supplementary

Table 6), but also to common forms of short stature. Additional evidence from both human and animal models support the notion that the *NPPC-NPR2-FGFR3-SHOX* pathway is a key modulator of growth (Table 2). In the same way, *IGF1R*, a key member of the growth hormone receptor pathway, points toward a second group of genes that would include IGF-I which is currently a short stature therapy for Growth Hormone non-responsive patients.⁴⁹

The additive effects of common genetic variation (i.e. PRS), predicted 20.1% of the variance in height in our dataset. These effects appeared to have similar magnitudes on the carriers of rare coding variation of genes as compared to non-carriers. This observation indicates that PRS may be a strong contributor in the differences in penetrance of rare pathogenic variants (especially in models of haploinsufficiency such as the ones described here). Supporting this idea, we saw that two of eight *NPR2* variant carriers with low *NPR2* activity had a low-normal height. One of them was in the 4th PRS quintile and the other was in the 5th PRS quintile, the remaining 6 were in quintiles 1st, 2nd and 3rd (Fig. 3E, Supplementary Table 9). This data suggests that most ISS individuals possessing mutations in *NPR2* may also have a polygenic background that made them more susceptible to the pathogenic effect of losing *NPR2* activity. This observation could be interpreted as supporting for the omnigenic model,^{50,51} where core genes would define a module that when disrupted, has large effects. Similarly, the omnigenic model predicts that the effect of core genes is modulated by multiple weaker common genetic variants driving regulatory networks. However, it is evident that in our data PRS has a strong effect on height even in non-carriers of the five genes we identified.

It is possible that that other genes will appear as key contributors of height and ISS as sample sizes increases. Our study was 63% powered to identify a gene set association that explained the proportion of variance we observed in the set of five genes we describe. A follow-up study would require over 400,000 samples to have 80% power to identify individual genes at exome-wide significance levels (See Methods).

Our results have important implications not only for drug discovery but also for drug repurposing. The presence of genes driving disease in the general population suggests the possibility of tailoring therapies to the small fraction of individuals with gene defects. Also, the fact that such genes have allelic series with bidirectional effects and additive effects across PRS groups, suggests that such tailored therapies may be effective for any patient irrespective of the genetic lesion. Importantly, we have taken initial steps to show that having bidirectionality increases the probability of success (4-fold) in clinical trials above and beyond just having human genetic evidence (2-fold).

As mentioned earlier, CNP-analogs are being tested as a treatment for Achondroplasia (incidence 1 in 10,000).⁴¹ Here we demonstrate that carriers of variants in any of the five genes are at ~ 3-fold increased risk for ISS and account for 6.7% of the total ISS population.

Furthermore, we showed dose-dependent rescue of *NPR2* signaling in a cell model of *NPR2* haploinsufficiency after adding exogenous CNP. These results support the idea that these CNP-based treatments could be effective in *NPR2* haploinsufficient patient populations if the right dose can be delivered. Further, our results showing a significant correlation of cGMP levels and

height in *NPR2* carriers of the general population suggest that targeting this receptor with CNP analogs could be an effective therapy for all ISS individuals.

In summary, we highlight a strategy to identify genes driving disease by leveraging a combination of human genetic databases and validation using large-scale exome sequencing studies followed by functional studies. We also highlight potential opportunities for drug target discovery and repurposing in this type of genetic targets.

Online content

Supplementary Figures and Tables are available online.

Acknowledgements

We thank all participants of the UK Biobank project for sharing their data. We also thank Dr.

Pavel Krejčí (Masaryk University) for providing us access to the Rat Chondrosarcoma Cell line.⁵²

We are grateful to Drs. Emily King and Wade Davis for their advice and assistance calculating the odds of drug approval based on the support of genetic evidence.

Author contributions

KE, LC, JL designed the study. SF, MT, HN, AL, GY designed and executed the functional experiments. KY, KE, AW, CB, TS, WC obtained and worked with the exome sequence data. DW provided access to the rat chondrosarcoma cell line and provided input for the experiments. KE drafted the manuscript and all co-authors contributed to it.

Competing interests

KE, AW, CB, TS, WC, LC, JL, SF, MT, HN, AL and GY are employees of BioMarin Pharmaceuticals and hold common stocks.

URLS. Additional information

UK Biobank website and data access <http://ukbiobank.ac.uk/>

Genomics PLC wecall: <https://github.com/Genomicsplc/wecall>

Hail: <https://github.com/hail-is/hail/releases/tag/0.2.16>

SKAT: <https://www.hsph.harvard.edu/skat/>

Genetic-evidence-approval: <https://github.com/AbbVie-ComputationalGenomics/genetic-evidence-approval>.

Figure legends

Figure 1. Strategy followed to identify bidirectional genes. A. Graphical representation of a gene (PCSK9) showing bidirectional effects on LDL and Cardiovascular disease risk. B. Flow chart describing the steps followed to identify causal genes. PRS, Polygenic Risk Score computed using previous meta-analysis of common variants. GWAS, Genome-Wide Association Study; HGMD, Human Gene Mutation Database; OMIM, Online Mendelian Inheritance in Man; SKAT SNP-set (Sequence) Kernel Association Test. C. Groups of genes analyzed to identify five genes that regulate height using both common and rare variation. The color codes of the gene names or the number of genes in the gene set corresponds to the SKAT test p-value for the association with height on 34,284 UK Biobank exome sequences. C. Odds Ratio for Idiopathic Short Stature (ISS) on highlighted gene-sets. Box represent estimate and 95% CI.

Figure 2. Combined effect of PRS and rare coding variants on height. A. Effects on height as a quantitative trait. Samples were divided in five groups based on their PRS. Horizontal lines in violin plots represent the 25%, 50% and 75% percentile of height. Samples were grouped by carrying status of missense (green), loss of function (red) or None (orange) in any of the five core genes. B. Effect reflected on Odds ratios for “Idiopathic Short Stature” or ISS. Odds for ISS using Non-Carriers of Core 5 genes with PRS =3 as reference vs the other PRS groups. Box colors represent: orange, non-carriers of LoF/missense variants in the five genes; green, carriers of LoF/missense variants in the five genes.

Figure 3. Functional effects of NPR2 protein coding variants.

A) Schematic for evaluation of missense variants. The bicistronic plasmid with a CMV promoter driving production of NPR2 variants and an EF1a promoter driving production of red firefly luciferase was transfected into HEK293T cells. Three days post transfection cells were treated with IBMX, incubated with CNP, and analyzed for luciferase and cGMP activity. cGMP activity was normalized of variants relative to luciferase, n = 3 biologically independent samples.

B) Diagram of NPR2 protein altering variants identified in the UKBiobank study. X-axis indicates genomic position. Y-axis indicates effect size (β) of a linear regression of that variant on height in standard deviations. Each circle represents a different variant with the number of carriers inside the circle, amino acid change of the variant is labeled at the end of the circle. Color code indicates cGMP activity of variants relative to luciferase.

C) Violin plots representing height of carriers of the NPR2 variants tested in functional analyses. Box-plots inside violin-plots represent the median and inter-quantile range of height in the group. Whiskers are 95% CI. Color categories represent binned cGMP functional readout.

D) Violin plots representing height of carriers of all *NPR2* variants found in the study. Box-plots inside violin-plots represent the median and inter-quantile range of height in the group. Whiskers are 95% CI. Color categories represent binned computer algorithm scores (CADD).

E) Predicted measured height (Z-transformed) as a function of their PRS and the percent of wild-type cGMP observed for their *NPR2* variant. The color maps to height (blue= highest, red=lowest).

F) Schematic for generating NPR2 deficient cells using CRISPR/Cas9.

G) cGMP dose response to CNP treatment with *NPR2* knockout (red), heterozygous knockout (blue) or wild-type (black) cells, $n = 3$ biological independent samples; mean \pm s.d.

Figure 4. Generalizability of bidirectional effects for increasing probability of success on other indications. A. Estimated odds ratio for transition from Phase I to Approved. We calculated the odds ratio for transition from Phase I to approval and 95% confidence, for the subset of target-indications with genetic support from GWAS, OMIM associations with unidirectional effects, and bidirectional effects curated from HGMD. This process was repeated using all thresholds of MeSH term similarity from 0 to 1 by increments of 0.01. Plotted points represent the estimated odds ratios and the error bars represent the 95% confidence intervals. Bidirectional effect supported data is shown in red, OMIM unidirectional support is shown in green, and GWAS support is shown in blue. B. Estimated odds ratio for transitions between various phases at MeSH similarity ≥ 0.83 . We calculated the odds ratio for transition and 95% confidence, for the subset of target-indications with genetic support from GWAS, OMIM associations with unidirectional effects, and bidirectional effects curated from HGMD. Plotted points represent the estimated odds ratios and the error bars represent the 95% confidence intervals. Bidirectional effect supported data is shown in red, OMIM unidirectional support is shown in green, and GWAS support is shown in blue.

Tables

Gene set / Gene	Orthogonal evidence			Burden Height (Protein-altering)			Burden Height (LoF)			Burden Idiopathic Short Stature (ISS) (Protein-altering)					
	H	M	C	AC	Effect	P	AC	Effect	P	AC case	AC ctr	Cum AF case	Cum AF ctr	OR 95%CI	P
Set 1															
<i>IGF1R</i>	●	●	●	264	-0.20	0.001	7	-0.76	0.042	12	252	2.0%	0.7%	2.85 [1.48-5.47]	0.002
<i>NPPC</i>	●	●	●	32	0.02	0.895	2	-1.43	0.042	2	30	0.3%	0.1%	4.83 [0.99-23.65]	0.052
<i>NPR2</i>	●	●	●	237	-0.26	5.38E-05	12	-0.97	0.001	12	225	2.0%	0.7%	3.31 [1.68-6.51]	0.001
Set 2															
<i>FGFR3</i>	●	●	●	321	-0.19	8.27E-04	7	-0.43	0.252	11	310	1.8%	0.9%	2.2 [1.13-4.28]	0.020
<i>SHOX</i>	●	●	●	83	-0.32	0.003	2	-1.36	0.052	4	79	0.7%	0.2%	3.34 [1.08-10.31]	0.036
Set 1 + Set 2				937	-0.20	4.04E-11	30	-0.85	2.49E-06	41	896	6.7%	2.7%	2.75 [1.92-3.96]	3.99E-08

Table 1. Individual gene association statistics for Height and Idiopathic Short Status (ISS, height < -2 SD below the mean) on rare protein-altering or LoF variants. Bold p-values represent significant associations after adjusting for five tests. AC, Allele count; LoF, Loss of Function; Effect, Effect estimate from regression; P, Nominal P-value estimate; Green and Red dots on the orthogonal evidence column reflects the presence of gain-of-function and Loss of function mutations in the respective model with corresponding evidence of effect on growth (For more details, see Supplementary Table 7)

Methods

Human databases

GWAS. We extracted 2,067 non-repeating closest genes for each of the 3,290 independent genetic variants reported by a large GWAS meta-analysis of height using ~700,000 individuals (Supplementary Tables 1 and 2).¹⁸ We considered this the most comprehensive GWAS for height.

HGMD. We queried the “allmut” table from HGMD version v2019_2 looking for all pathogenic variants labelled as “DM” having either “short stature” and “tall stature or overgrowth” in the same genes (Supplementary Table 2).

OMIM. The list of OMIM genes related to growth disorders was previously described and was created using the keywords: short stature, overgrowth, skeletal dysplasia, brachydactyly(Supplementary Table 2).²⁰

Samples

UKBiobank. Data was accessed under application 41232. The demographics and patient characteristics for the 50,000 exome sequenced individuals has been previously described.⁵³ In short, individuals from the UK aged 45 to 75 were invited to participate. A baseline questionnaire and several measurements were taken. Electronic health records were also

available. Standing height (field id 50) was measured to all individuals as part of the baseline assessment (Supplementary Table 3). We used GNU parallel to rapidly extract individual phenotype fields to CSV files.⁵⁴

Sequencing and variant calling

Sequencing protocol and details are described elsewhere.⁵³ In short, exomes were captured using a modified version of the IDT xGen Exome Research Panel 81 v1.0 and sequenced in pooled multiplexed samples using 75 base pair paired-end reads with two 10 base pair index reads on the Illumina NovaSeq 6000 platform using S2 flow cells. Reads were aligned to the GRCh38 genome reference with BWA-mem2. The WeCall variant caller was used to generate gVCF files (See URLs). A PVCF was created using the GLnexus joint genotyping tool.⁵⁵

Software

Genetic data analysis was performed on Hail 0.2.16 running on Apache Spark version 2.3.4.

Exome sequencing data

We downloaded UK Biobank exome sequencing genotypes in Plink format for 49,960 individuals and 10,448,724 variants (“SBP” data) in May 2019.

We annotated and filtered individuals with data provided by UK Biobank. Specifically, we included individuals with the genetic ethnic grouping “Caucasian” (field 2206), no sex chromosome aneuploidies (field 22019) (effectively removing any classical Turner syndrome patients) and with a PCA-corrected heterozygosity between 0.17 and 0.21 (field 22019). After applying these filters, our dataset contained 41,190 individuals.

We removed related individuals by excluding samples meeting the criteria for relatedness and kinship set by UK Biobank (2nd degree relatives or closer) and that are listed in fields 22011, 22012 and 22013. After applying this filter, our dataset contained 35,990 individuals.

We computed sample and variant QC metrics and applied additional filters that excluded sample with more than 80,000 non-reference variants, more than 200 singletons, a het/hom ratio of less than 1.3 or more than 1.85, and transition/transversion ratio of more than 2.5, a call rate of less than 0.985 or a proportion of heterozygous sites in sequencing data that are also heterozygous in chip data of less than 0.98. We also excluded variants that had a call rate of less than 0.99 or a Hardy-Weinberg p value of less than 10^{-10} .

We annotated variants using the Variant Effect Predictor (VEP)⁵⁶ and extracted rare high impact variants. We defined those as variants with a minor allele frequency of less than 1% and whose impact as predicted by VEP is either high (protein truncation) or moderate (missense). We observed that a subset of those high-impact variants had an unexpectedly high allele frequency compared to the allele frequency of non-Finnish European (NFE) individuals in the genome aggregation database (gnomAD).⁵⁷ We removed those variants if the observed allele count in our dataset was significantly different to what would be expected based on the NFE allele frequency in gnomAD ($p < 10^{-7}$, binomial test). After applying those filters, our dataset contained 581,392 variants. The assignment of variants to genes was accomplished by using canonical VEP gene models.

Quality control

After applying the filters described above, our exome sequencing dataset contained 34,284 individuals and 10,173,231 variants. There was no apparent clustering of the post-filtering European ancestry individuals by genotyping eigenvectors (Supplementary Figs. 18-19). The sample call rate was between 0.999941 and 1.0 (median: 0.999984) and the variant call rate was between 0.989909 and 1.0 (median: 1.0) (Supplementary Figs. 20-21).

We observed between 37 and 111 (median: 68) rare high-impact variants per individual (Supplementary Figure 22) and between 0 and 36 (median: 9) singletons per individual (Supplementary Figure 23). Most individuals had a rare high-impact variant in less than 1% of genes (Supplementary Figure 24).

Gene set analyses

We constructed a list of 2,355 genes (Supplementary Table 3) that had evidence from GWAS and/or HGMD/OMIM. To assign a GWAS signal to a gene, we used the latest meta-analysis for height and took the list of 3,290 predicted independent variants (Supplementary Table 1).¹⁸ Then we mapped the closest gene to each one of those lead variants and removed duplicates. That list of genes was compiled together with genes reported on HGMD and OMIM and genes were assigned one to five different annotations: GWAS, HGMD_SHORT, HGMD_TALL, OMIM-OVERGROWTH, OMIM-SHORT (Supplementary Table 2).

We then queried this list of genes into the UKBiobank exome dataset. Here 1,970 genes had at least 20 carriers (combined rare missense and Lof) and were considered for the analyses. Genes had a mean of 204 rare protein coding variants (Supplementary Fig. 20).

Statistical association analyses

Standing height measurements were first split by gender and then normalized to standard deviations adjusted for age and first five principal components using a linear regression implemented on the R package. Linear and logistic regression analyses for single variants were computed using Hail 0.2.

We grouped rare missense and loss of function variation per gene using GENCODE canonical transcripts. We applied SKAT and burden tests on missense variants and predicted protein truncating variants (PTVs) with a minor allele frequency < 0.01% using Hail 0.2.

For the association analysis on gene-sets and ISS we grouped genes in the following way:

Bidirectional: Genes present in HGMD_SHORT, HGMD_TALL, OMIM-OVERGROWTH, OMIM-SHORT. Tall: Genes present in HGMD_TALL or OMIM-OVERGROWTH but not in HGMD_SHORT or OMIM_SHORT; Short: Genes present in HGMD_SHORT and OMIM_SHORT but not in HGMD_TALL or OMIM-OVERGROWTH.

SNP chip data and polygenic risk scores

We used summary statistics from a large GWAS of human height to compute a height polygenic risk score (PRS) for UK Biobank participants.²⁰ Specifically, we downloaded UK Biobank imputed SNP chip genotypes in BGEN format for 487,409 individuals. We exported genotypes for 2,540,786 variants for which GWAS summary statistics were available to Plink format. We then used LDpred⁵⁸ version 1.0.7 to compute a height PRS for 49,796 individuals for which exome sequencing data was available. For the coordination step of LDpred we used a linkage disequilibrium reference panel of 2,000 randomly selected UK Biobank participants of European ancestry for which exome data was not available. For variant filtering, generating SNP weights and PRS scoring we used default LDpred parameters. Linear regression analyses were used to estimate the effect of PRS as a categorical variable (1-5) on height. Logistic regression analyses were used to estimate the effect of PRS as a categorical variable (1-5) on ISS, groups with low number of ISS cases (standard error > 100) were removed.

Selection of NPR2 variants for Functional assays

A total of 39 NPR2 protein coding variants were selected to investigate their functional effect at a cellular level (Supplementary Table 7). Fourteen of those variants had been previously described in the literature and aren't explicitly found in our study. Seven variants were found in our exome dataset in single individuals with height > 1 SD above the mean and were labeled as potential gain of function. Nine variants were found in individuals with height < 2 SD below the mean and were labeled as potential loss of function. Three variants with neutral effect on

height were tested. Four variants with MAF <0.1% (p.A164G, p.R601H, p.E609K, p.R358W) were added to complement the allelic series.

Transfection Functional assays and Assay for cGMP analyses

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin/streptomycin and Glutamax (Gibco) at 37 °C with 5% CO₂. Transfection was performed in triplicate using a 96-well Shuttle nucleofector (Lonza). Briefly, 250 ng of plasmid per 2×10^5 cells was transfected using Amaxa solution SF (Lonza) and program CA-138. Three days after transfection, cells were then incubated in DMEM containing 0.75 mM IBMX (3-isobutyl-1-methylxanthine) (Enzo life sciences) for 15 min. The cells were next treated with 6.67 nM of BMN 111 (CNP)⁴³ and incubated for another 15 min. The reaction was terminated with 40 μ l of lysis solution, and the cGMP concentration was measured by a competitive enzyme immunoassay (Molecular Devices, R8075). Luciferase levels were measured using the ONE-Glo Assay (Promega) and the results are presented as the cGMP/luciferase normalized to the wild-type plasmid transfected cells. For the dose-response curves, the CNP concentrations assayed in triplicate were 33.33, 6.67, 1.33, 0.27, 0.053, 0.0107, and 0.002133 nM. The cGMP concentrations were assayed as above and cellular concentrations calculated assuming 1.6E5 cells/well with a diameter of 35 μ m.

PRS and NPR2 functional data modeling

For all individuals carrying a rare variant in *NPR2*, for which functional assay data was available, we fit a generalized additive model to predict measured height (Z-transformed) as a function of

their PRS for height (Z-transformed) and the percent of wild-type cGMP observed for their *NPR2* variant (log base 2 transformed). Models were fit using the `gam()` function from package `mgcv` v1.8-31 in R version 3.6.1. A gamma parameter of 0.3 was chosen to minimize the amount of smoothing. Values smaller than this resulted in unstable models that produced unrealistic height predictions.

CRISPR/Cas9 knockout of *NPR2* in rat chondrosarcoma chondrocytes

The swarm rat chondrosarcoma (RCS) cell line⁵² was cultured with DMEM containing 10% fetal bovine serum, penicillin/streptomycin, and Glutamax (Gibco) at 37 °C with 5% CO₂. RCS cells were harvested with TrypLE Express Enzyme (Life Technologies) and transfected with RNPs formed with guide RNA containing ATTO 550 labelled tracrRNA (IDT 1075928) and crRNA targeting *NPR2*, with PAM in bold:

gRNA1 5'- CCCACTACTTCACCATCGAG**GGG**-3'

gRNA2 5'- CATTACACGGGCACTTCAAT**TGG**-3'

Briefly, equimolar ratios of crRNA and tracrRNA were annealed and 12pmol was complexed with 104pmol HiFi Cas9 (IDT 1081061) according to the manufacturers protocol. 2E5 cells were transfected with 500ng of GFP mRNA (TriLink Biotechnologies, cat# L-7201) or RNP using a 96-well Shuttle nucleofactor (Lonza) with Amaxa solution SF (Lonza) and program CA-150. Post transfection, each well was split and seeded into two wells of a 96well plate. Three days after transfection, single cells were sorted into 96 well plates containing 200µl of media using a FACSMelody cell sorter (BD Biosciences); the plates were monitored biweekly by imaging (Cell Metric, Solentim) when confluent cells were expanded into 6 well plates and genotyped.

Genomic DNA was extracted using Quanta Extracta solution (Quanta Biosciences, cat #95091).

DNA amplicons covering the gRNA target region were amplified and analyzed by PacBio sequencing following the manufacturer's instructions (Pacific Biosciences, 101-791-800 Version 01). Amplicon libraries were prepared using primers with universal sequences and the locus specific primers FW 5'-CATCCCTGCTACTGGTGGTG-3' and REV 5'-CTCCACCACCAAACCTGAACT-3'. Sequencing was performed using the Sequel System using the SMRT Cell 1M v3 and acquiring 10h movies. Demultiplexing and circular consensus sequence analysis and FASTQ generation was performed using the SMRTlink 6.0 software (Pacific Biosciences). To analyze amplicon sequencing, Geneious Prime 2019.0.4 software was used for quality assessment, quality filtering, reference alignment, and calculating indels (Supplementary Fig. 21).

Power analysis

We used the PAGEANT⁵⁹ method to estimate the power of the current study using the SKAT method on the following two scenarios: a. Our study design involved an $\alpha=0.002$ (adjusting for 25 gene set tests), we considered a variance explained of 0.09%. b. A study would require 400,000 samples to identify single genes at $\alpha=2.5e-6$, a maximum of 1000 causal genes, and considering 5% of variance explained for protein-altering variants genome-wide.

Bidirectional screen for broader indications

The HGMD allmut table v2016.4 was used as the source for associations to identify bidirectional effect associated genes. The table was grouped by each combination of gene and disease. Any genes with less than two unique disease associations were excluded from further analysis. The

remaining dataset contained 23,754 distinct gene-disease pairs between 4427 genes and 12302 diseases. This table was manually curated to identify genes with at least 2 associations that appeared to different directions of effect on a common underlying phenotype. For example, PSCK9 was identified as a bidirectional effect gene due to associations with “High LDL cholesterol” and “Low LDL cholesterol”. KCNJ2 was identified as a bidirectional effect gene due to associations with “Long QT syndrome” and “Short QT syndrome”. We also required that the disease field represent an actual disease or disease associated endophenotype such as a blood metabolite. Thus, associations between genes and molecular, cellular, or benign phenotypes (e.g. increased and decreased enzyme activity, cell permeability, or pigmentation) were also excluded.

Probability of Drug Approval Analysis

Analysis of the odds of drug target-indication approval was conducted in R/3.5.2 following the methods of King et al.⁴⁸ Since some OMIM disease genes were also associated with bidirectional phenotypic effects, we first reran their analysis after removing those genes to estimate the odds of approval for drugs that target OMIM genes with unidirectional effects. We then repeated this analysis with our set of gene-trait associations with bidirectional effects replacing the OMIM genes in the original King et al. data set. We called the function `annotate_target_indication_table_with_genetic_evidence()` to match the set of target-indications with the best supporting genetic evidence. We then used the function `replicate_table1()` to estimate the odds of transition between different phases of clinical

development using a range of MeSH term similarity thresholds (0 to 1 by increments of 0.01) to link drug indications with genetically associated diseases or traits.

The maximum odds of transition from Phase I through approval occurred for indications with MeSH similarity greater than 0.83.

References

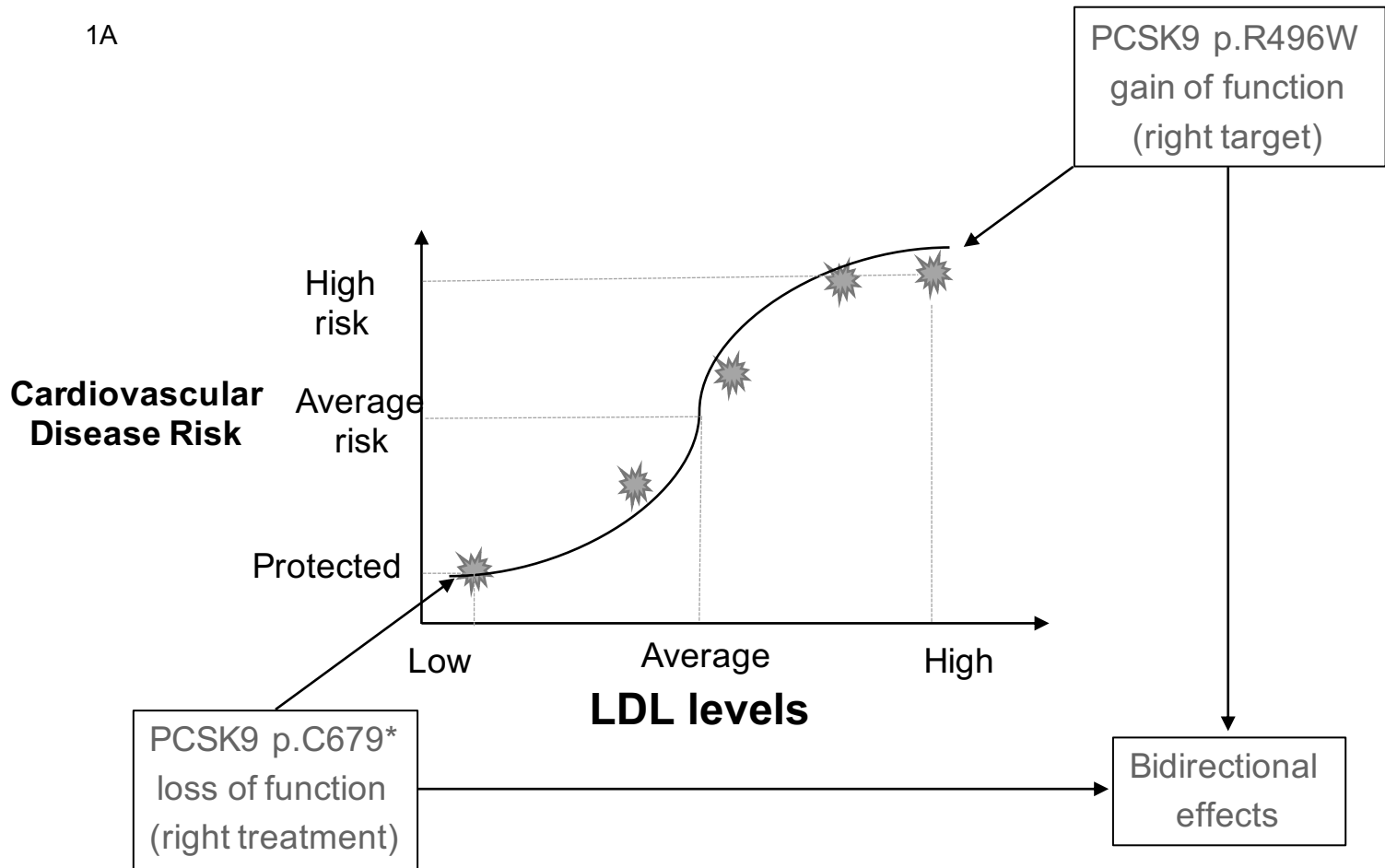
- 1 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-860, doi:10.1038/ng.3314 (2015).
- 2 Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179-189, doi:10.1038/s41586-019-1879-7 (2020).
- 3 Fang, H. *et al.* A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet* **51**, 1082-1091, doi:10.1038/s41588-019-0456-1 (2019).
- 4 Plenge, R. M. Priority index for human genetics and drug discovery. *Nat Genet* **51**, 1073-1075, doi:10.1038/s41588-019-0460-5 (2019).
- 5 Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**, 581-594, doi:10.1038/nrd4051 (2013).
- 6 Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* **34**, 154-156, doi:10.1038/ng1161 (2003).
- 7 Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161-165, doi:10.1038/ng1509 (2005).
- 8 Sabatine, M. S. *et al.* Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med* **376**, 1713-1722, doi:10.1056/NEJMoa1615664 (2017).
- 9 Athyros, V. G. *et al.* Drugs that Mimic the Effect of Gene Mutations for the Prevention or the Treatment of Atherosclerotic Disease: From PCSK9 Inhibition to ANGPTL3 Inactivation. *Curr Pharm Des* **24**, 3638-3646, doi:10.2174/1381612824666181009100517 (2018).
- 10 Baron, R. & Rawadi, G. Targeting the Wnt/beta-catenin pathway to regulate bone formation in the adult skeleton. *Endocrinology* **148**, 2635-2643, doi:10.1210/en.2007-0270 (2007).
- 11 Cosman, F. *et al.* Romosozumab Treatment in Postmenopausal Women with Osteoporosis. *N Engl J Med* **375**, 1532-1543, doi:10.1056/NEJMoa1607948 (2016).
- 12 Pedicelli, S., Peschiaroli, E., Violi, E. & Cianfarani, S. Controversies in the definition and treatment of idiopathic short stature (ISS). *J Clin Res Pediatr Endocrinol* **1**, 105-115, doi:10.4008/jcrpe.v1i3.53 (2009).
- 13 Hauer, N. N. *et al.* Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. *Genet Med* **20**, 630-638, doi:10.1038/gim.2017.159 (2018).
- 14 Shiang, R. *et al.* Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* **78**, 335-342, doi:10.1016/0092-8674(94)90302-6 (1994).

- 15 Bartels, C. F. *et al.* Mutations in the transmembrane natriuretic peptide receptor NPR-B impair skeletal growth and cause acromesomelic dysplasia, type Maroteaux. *Am J Hum Genet* **75**, 27-34, doi:10.1086/422013 (2004).
- 16 Sillence, D. O. *et al.* Osteogenesis imperfecta type III. Delineation of the phenotype with reference to genetic heterogeneity. *Am J Med Genet* **23**, 821-832, doi:10.1002/ajmg.1320230309 (1986).
- 17 Dahlgren, J. Growth outcomes in individuals with idiopathic short stature treated with growth hormone therapy. *Horm Res Paediatr* **76 Suppl 3**, 42-45, doi:10.1159/000330158 (2011).
- 18 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).
- 19 Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190, doi:10.1038/nature21039 (2017).
- 20 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186, doi:10.1038/ng.3097 (2014).
- 21 Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838, doi:10.1038/nature09410 (2010).
- 22 Estrada, K. *et al.* A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum Mol Genet* **18**, 3516-3524, doi:10.1093/hmg/ddp296 (2009).
- 23 Vaandrager, A. B. *et al.* Endogenous type II cGMP-dependent protein kinase exists as a dimer in membranes and can be functionally distinguished from the type I isoforms. *J Biol Chem* **272**, 11816-11823, doi:10.1074/jbc.272.18.11816 (1997).
- 24 Gamazon, E. R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967, doi:10.1038/s41588-018-0154-4 (2018).
- 25 Consortium, E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640, doi:10.1126/science.1105136 (2004).
- 26 Ndungu, A., Payne, A., Torres, J. M., van de Bunt, M. & McCarthy, M. I. A Multi-tissue Transcriptome Analysis of Human Metabolites Guides Interpretability of Associations Based on Multi-SNP Models for Gene Expression. *Am J Hum Genet* **106**, 188-201, doi:10.1016/j.ajhg.2020.01.003 (2020).
- 27 Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592-599, doi:10.1038/s41588-019-0385-z (2019).
- 28 Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665-677, doi:10.1007/s00439-017-1779-6 (2017).
- 29 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93, doi:10.1016/j.ajhg.2011.05.029 (2011).
- 30 Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet* **16**, 54-63, doi:10.1038/ng0597-54 (1997).

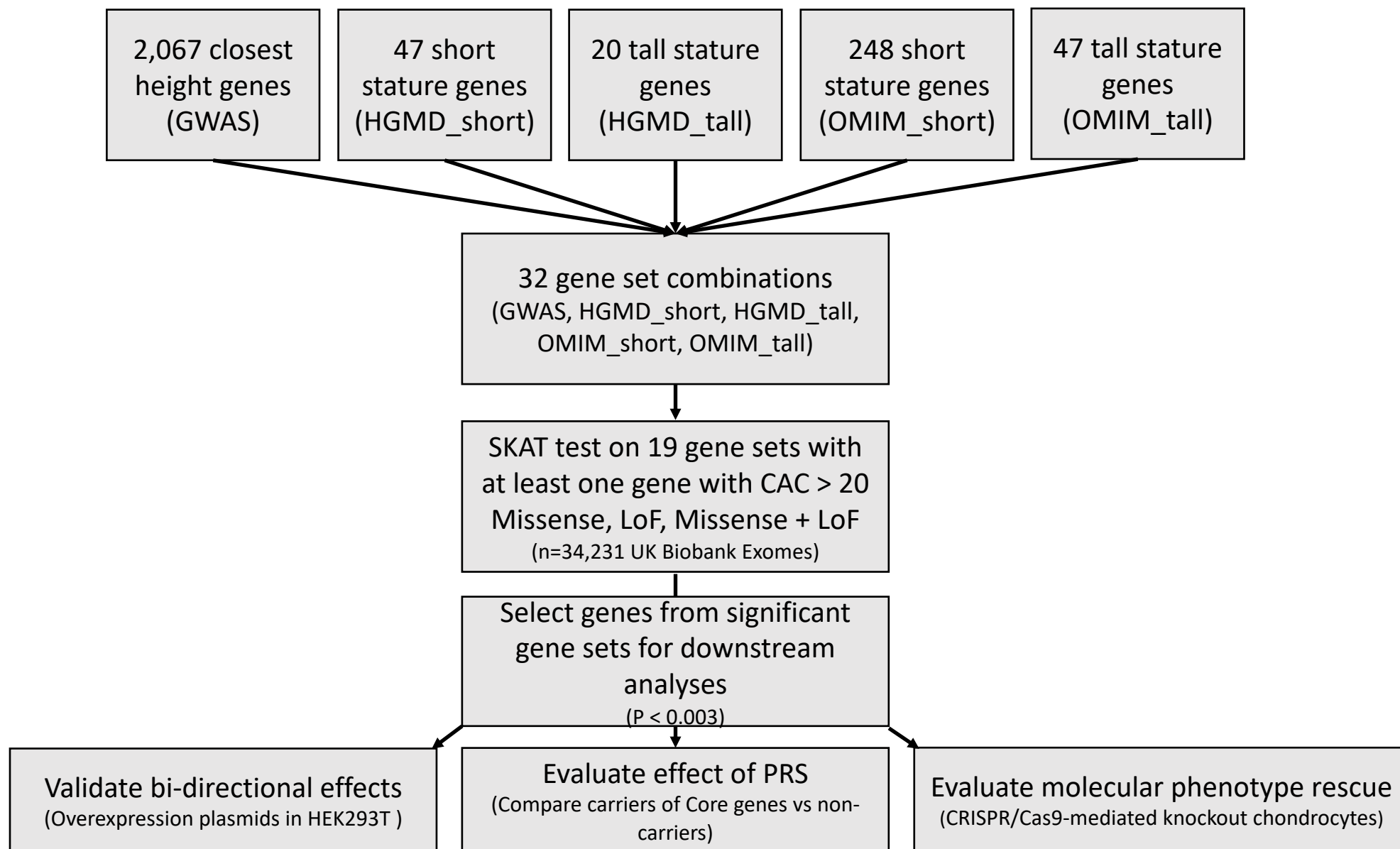
- 31 Tassano, E. *et al.* Genotype-Phenotype Correlation of 2q37 Deletions Including NPPC Gene Associated with Skeletal Malformations. *PLoS One* **8**, e66048, doi:10.1371/journal.pone.0066048 (2013).
- 32 Wang, S. R. *et al.* Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature. *Hum Mutat* **36**, 474-481, doi:10.1002/humu.22773 (2015).
- 33 Keselman, A. C. *et al.* A Homozygous Mutation in the Highly Conserved Tyr60 of the Mature IGF1 Peptide Broadens the Spectrum of IGF1 Deficiency. *Eur J Endocrinol*, doi:10.1530/EJE-19-0563 (2019).
- 34 Ogata, T. *et al.* Short stature homeobox-containing gene duplication on the der(X) chromosome in a female with 45,X/46,X, der(X), gonadal dysgenesis, and tall stature. *J Clin Endocrinol Metab* **85**, 2927-2930, doi:10.1210/jcem.85.8.6745 (2000).
- 35 Bocciardi, R. *et al.* Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. *Hum Mutat* **28**, 724-731, doi:10.1002/humu.20511 (2007).
- 36 Miura, K. *et al.* Overgrowth syndrome associated with a gain-of-function mutation of the natriuretic peptide receptor 2 (NPR2) gene. *Am J Med Genet A* **164A**, 156-163, doi:10.1002/ajmg.a.36218 (2014).
- 37 Miura, K. *et al.* An overgrowth disorder associated with excessive production of cGMP due to a gain-of-function mutation of the natriuretic peptide receptor 2 gene. *PLoS One* **7**, e42180, doi:10.1371/journal.pone.0042180 (2012).
- 38 Kant, S. G. *et al.* Tall stature and duplication of the insulin-like growth factor I receptor gene. *Eur J Med Genet* **50**, 1-10, doi:10.1016/j.ejmg.2006.03.005 (2007).
- 39 Toydemir, R. M. *et al.* A novel mutation in FGFR3 causes camptodactyly, tall stature, and hearing loss (CATSHL) syndrome. *Am J Hum Genet* **79**, 935-941, doi:10.1086/508433 (2006).
- 40 Hannema, S. E. *et al.* An activating mutation in the kinase homology domain of the natriuretic peptide receptor-2 causes extremely tall stature without skeletal deformities. *J Clin Endocrinol Metab* **98**, E1988-1998, doi:10.1210/jc.2013-2358 (2013).
- 41 Savarirayan, R. *et al.* C-Type Natriuretic Peptide Analogue Therapy in Children with Achondroplasia. *N Engl J Med* **381**, 25-35, doi:10.1056/NEJMoa1813446 (2019).
- 42 Lorget, F. *et al.* Evaluation of the therapeutic potential of a CNP analog in a Fgfr3 mouse model recapitulating achondroplasia. *Am J Hum Genet* **91**, 1108-1114, doi:10.1016/j.ajhg.2012.10.014 (2012).
- 43 Wendt, D. J. *et al.* Neutral endopeptidase-resistant C-type natriuretic peptide variant represents a new therapeutic approach for treatment of fibroblast growth factor receptor 3-related dwarfism. *J Pharmacol Exp Ther* **353**, 132-149, doi:10.1124/jpet.114.218560 (2015).
- 44 Tsuji, T. & Kunieda, T. A loss-of-function mutation in natriuretic peptide receptor 2 (Npr2) gene is responsible for disproportionate dwarfism in cn/cn mouse. *J Biol Chem* **280**, 14288-14292, doi:10.1074/jbc.C500024200 (2005).
- 45 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).
- 46 Campbell, J. C. *et al.* Structural Basis of Analog Specificity in PKG I and II. *ACS Chem Biol* **12**, 2388-2398, doi:10.1021/acscchembio.7b00369 (2017).

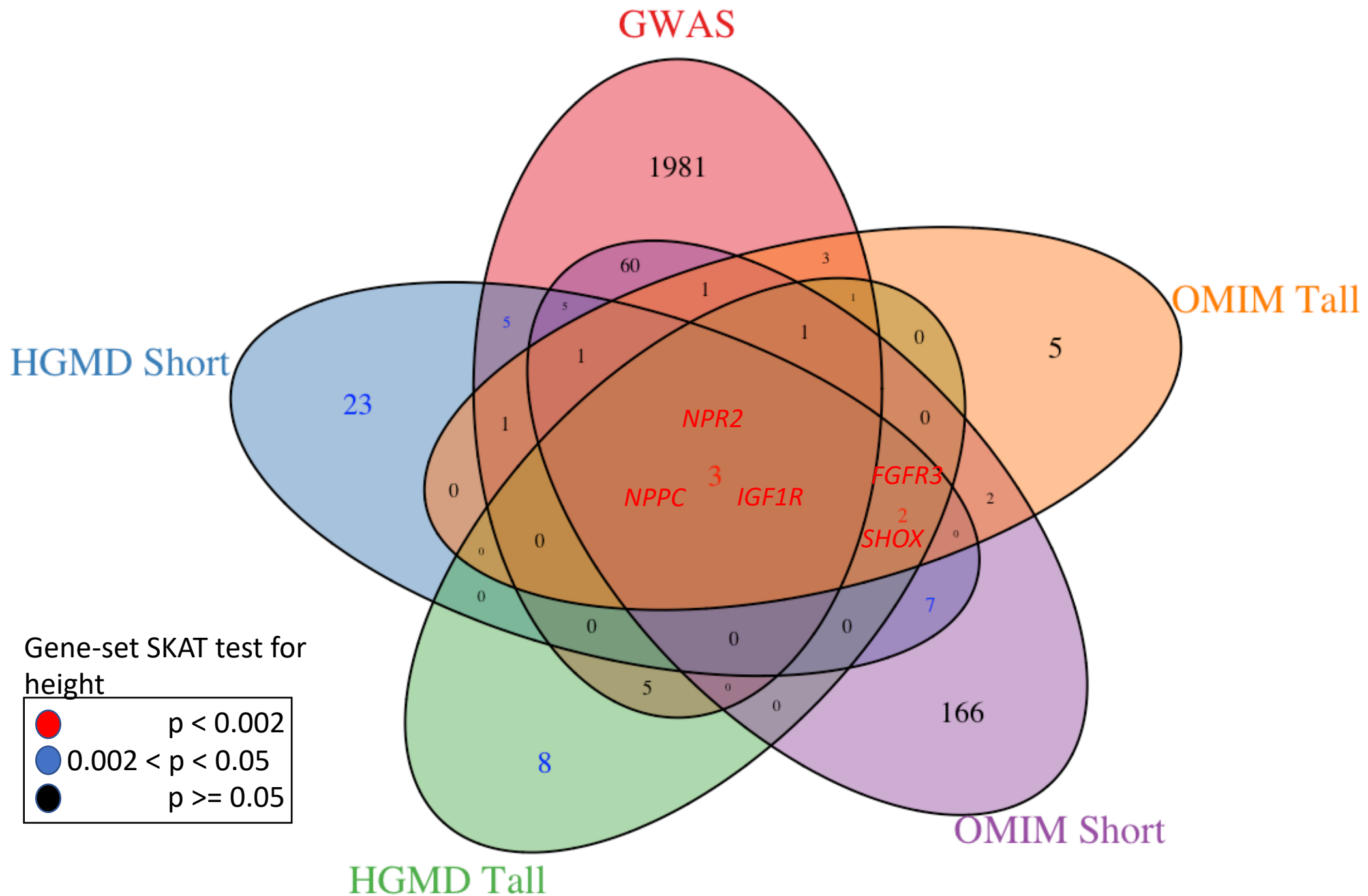
- 47 Pohler, D. *et al.* Expression, purification, and characterization of the cGMP-dependent protein kinases I beta and II using the baculovirus system. *FEBS Lett* **374**, 419-425, doi:10.1016/0014-5793(95)01168-e (1995).
- 48 King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet* **15**, e1008489, doi:10.1371/journal.pgen.1008489 (2019).
- 49 Keating, G. M. Mecasermin. *BioDrugs* **22**, 177-188, doi:10.2165/00063030-200822030-00004 (2008).
- 50 Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034 e1026, doi:10.1016/j.cell.2019.04.014 (2019).
- 51 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).
- 52 Mukhopadhyay, K. *et al.* Use of a new rat chondrosarcoma cell line to delineate a 119-base pair chondrocyte-specific enhancer element and to define active promoter segments in the mouse pro-alpha 1(II) collagen gene. *J Biol Chem* **270**, 27711-27719, doi:10.1074/jbc.270.46.27711 (1995).
- 53 Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv*, 572347, doi:10.1101/572347 (2019).
- 54 Tange, O. *GNU Parallel 2018*. (2018).
- 55 Lin, M. F. *et al.* GLnexus: joint variant calling for large cohort sequencing. *bioRxiv*, 343970, doi:10.1101/343970 (2018).
- 56 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
- 57 Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210, doi:10.1101/531210 (2019).
- 58 Vilhjalmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576-592, doi:10.1016/j.ajhg.2015.09.001 (2015).
- 59 Derkach, A., Zhang, H. & Chatterjee, N. Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics* **34**, 1506-1513, doi:10.1093/bioinformatics/btx770 (2018).

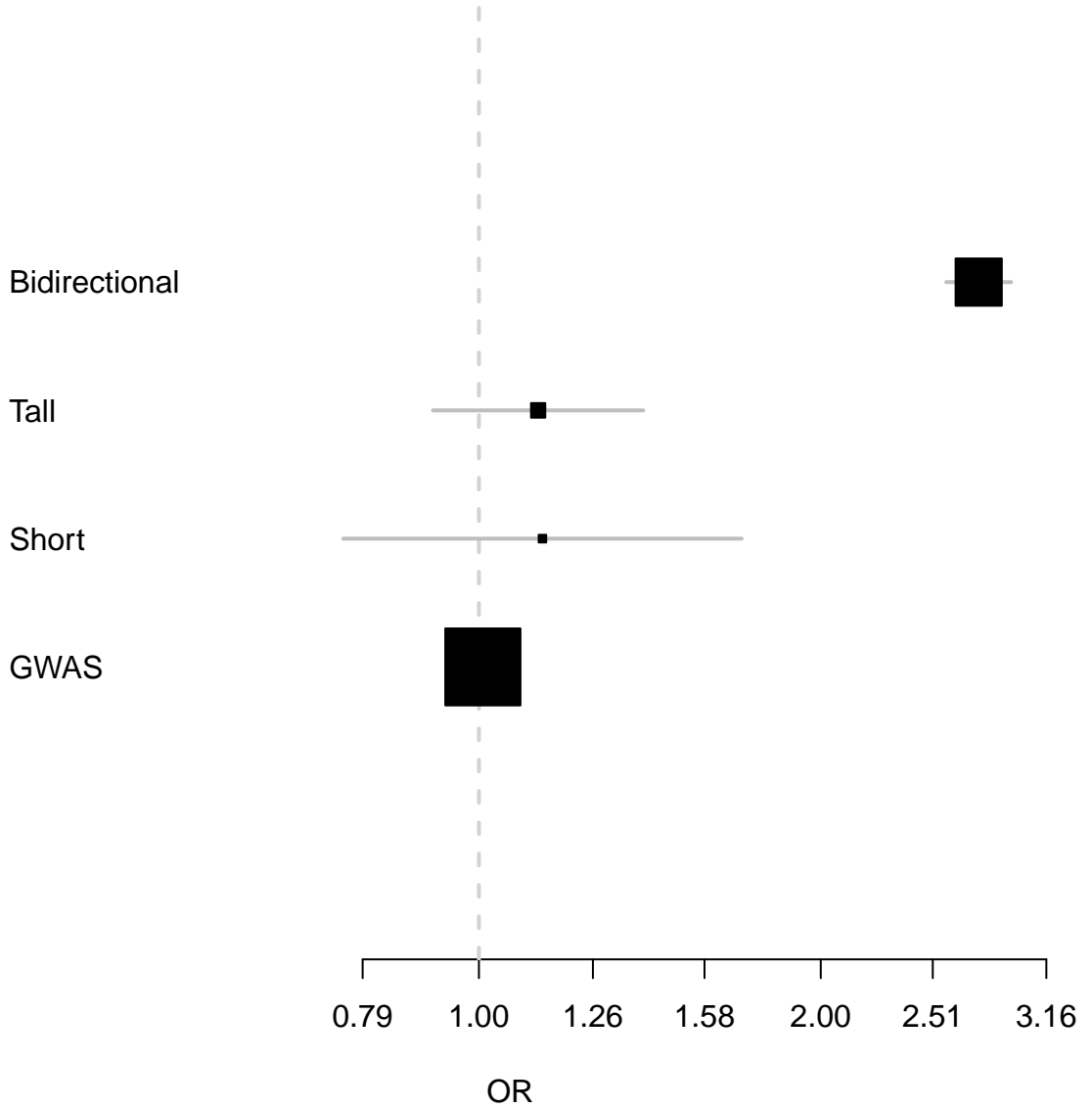
1A



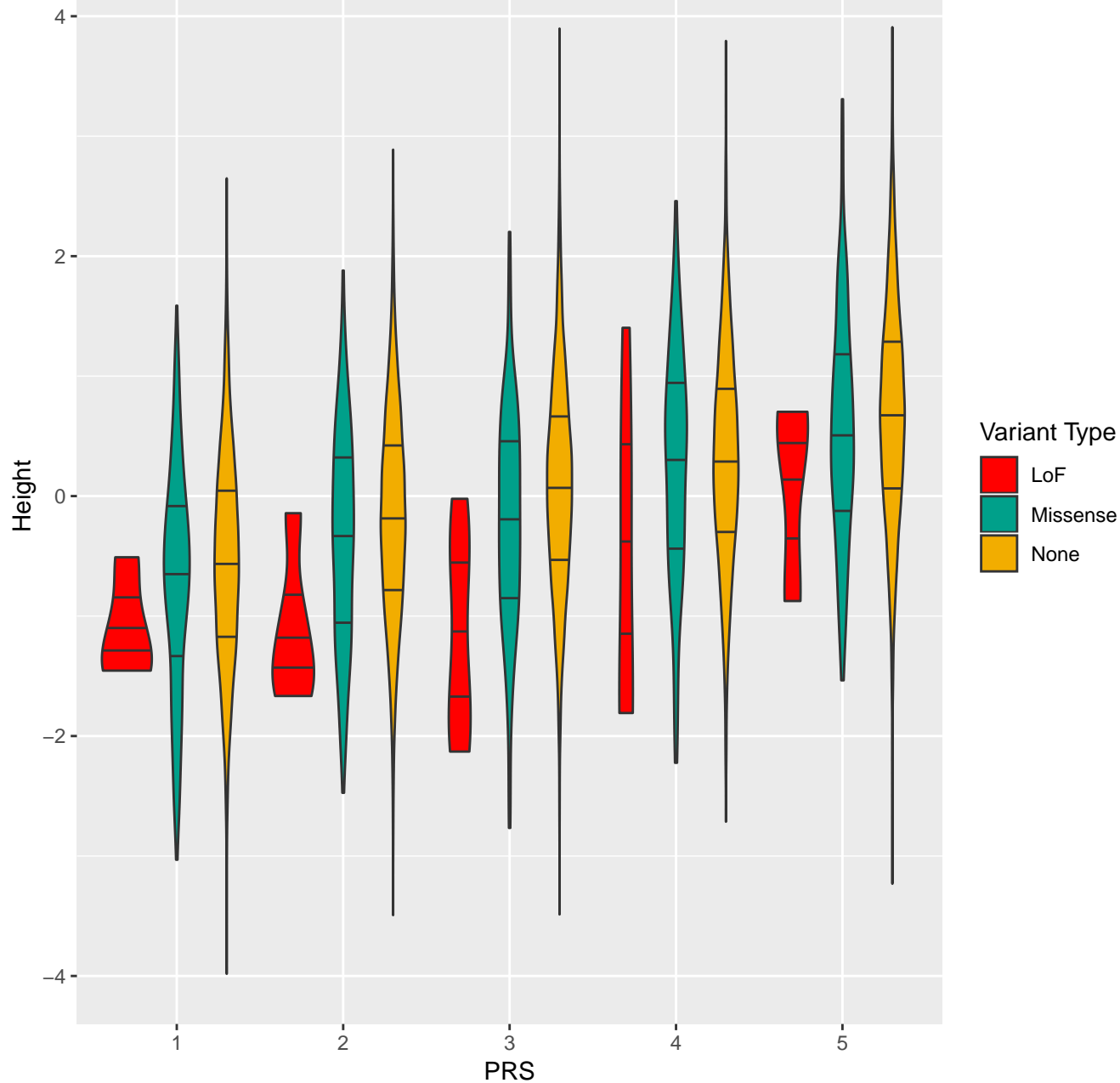
★ Specific *PCSK9* mutation

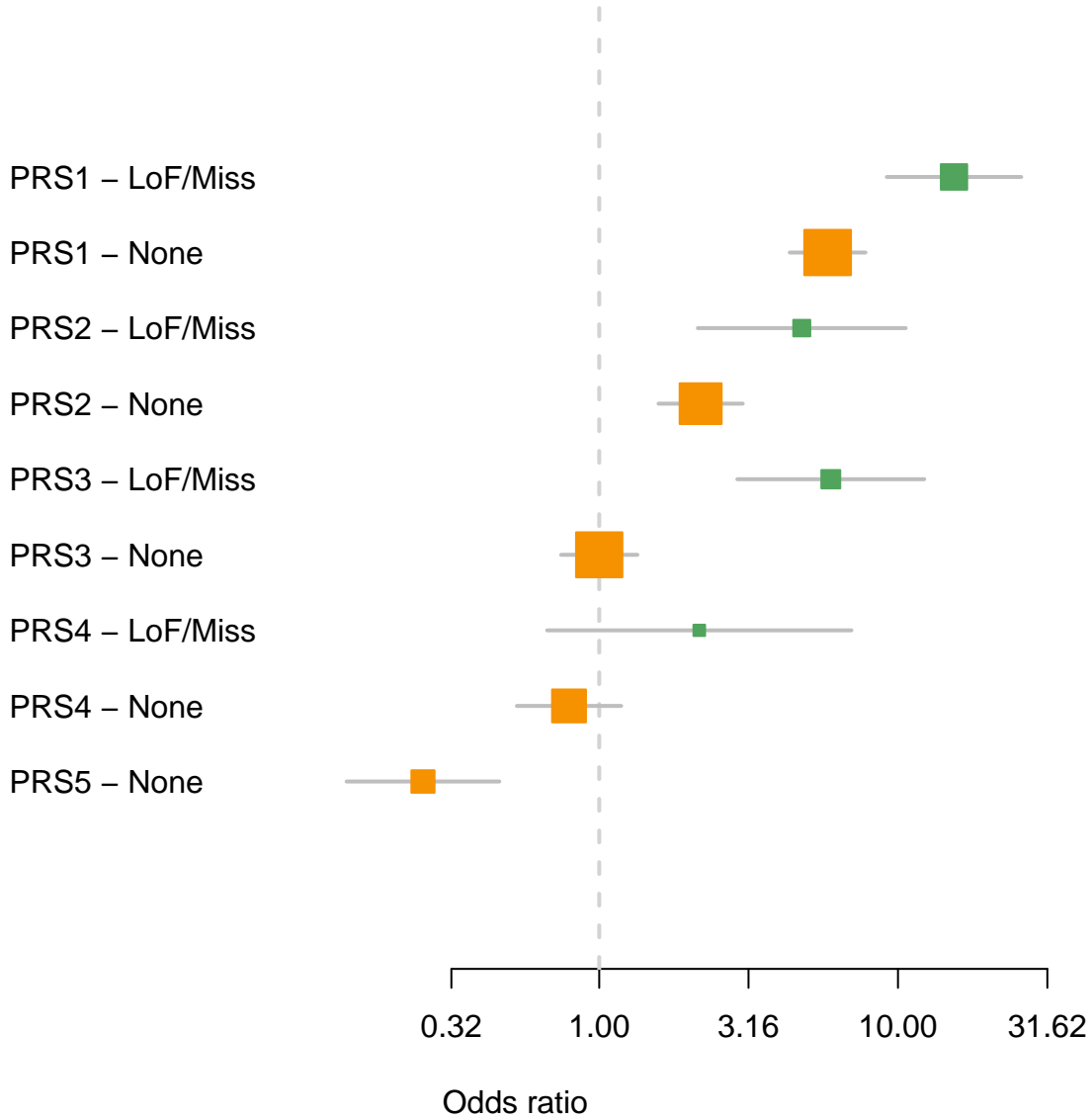




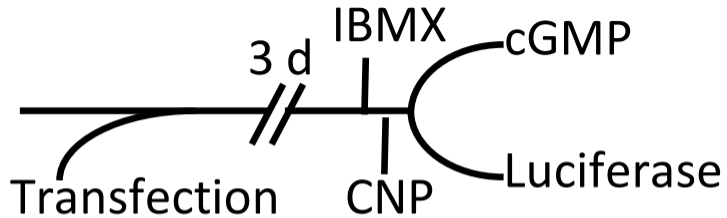
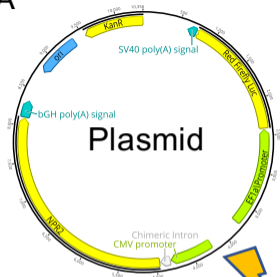


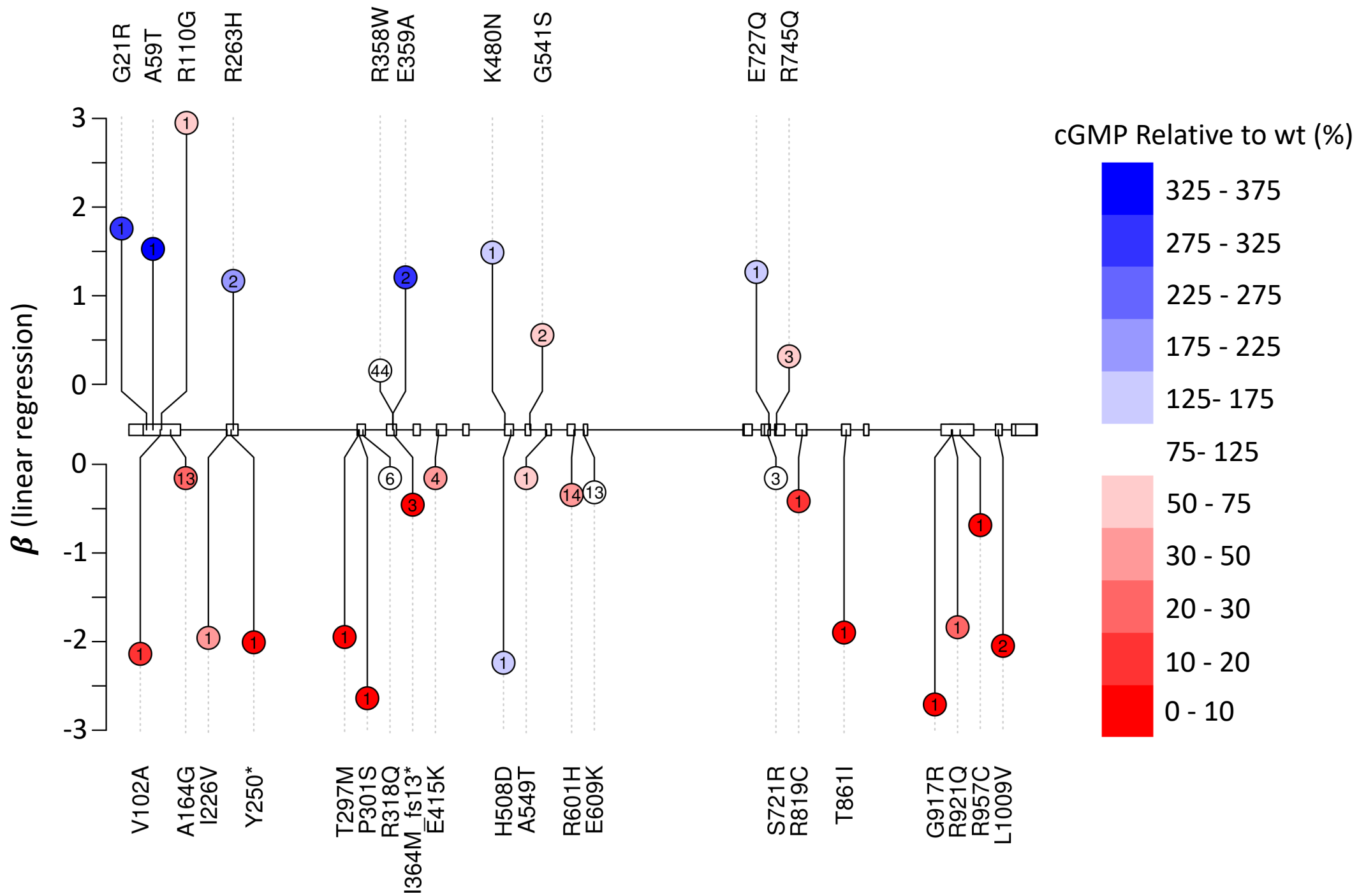
2A

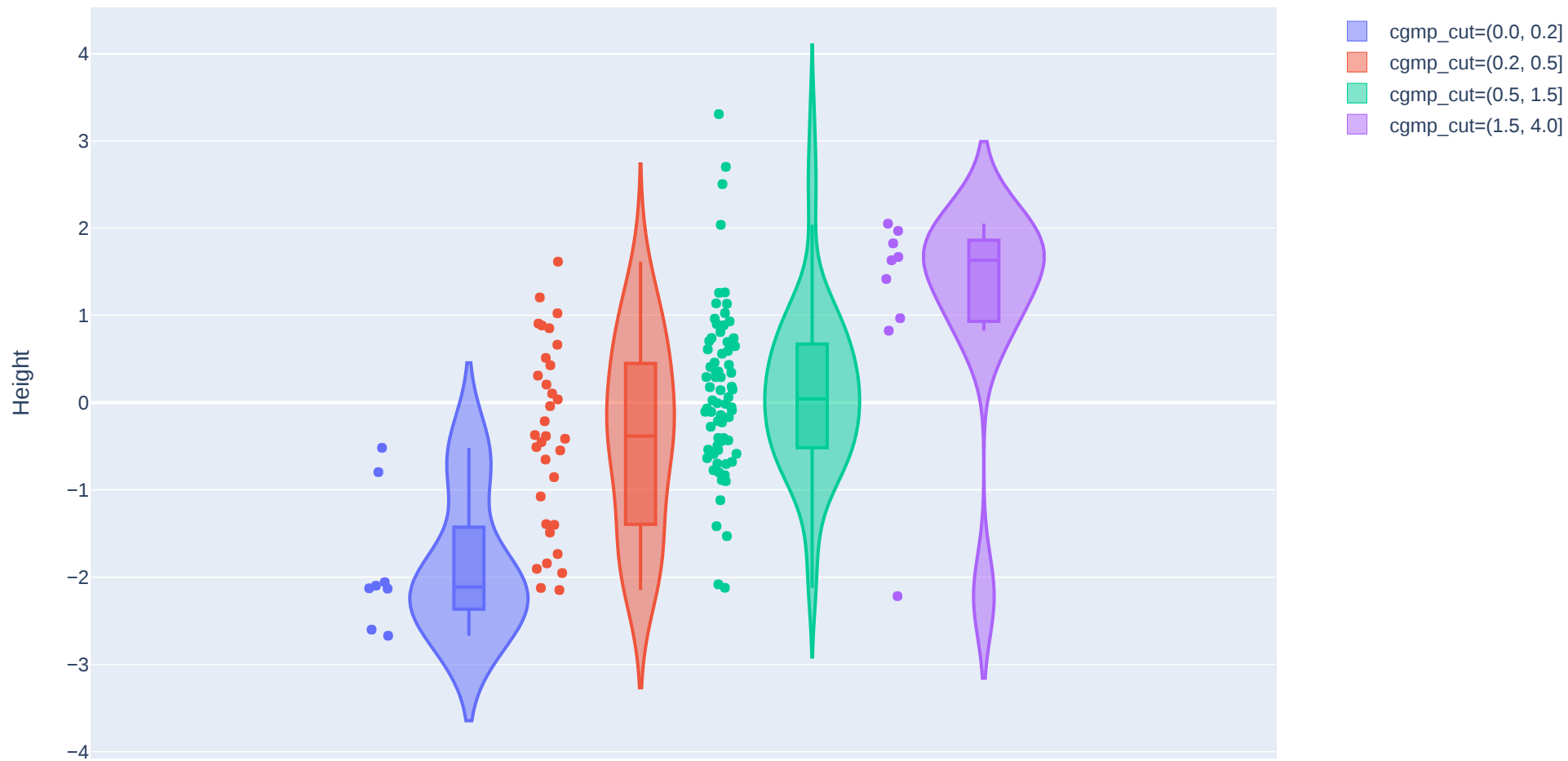


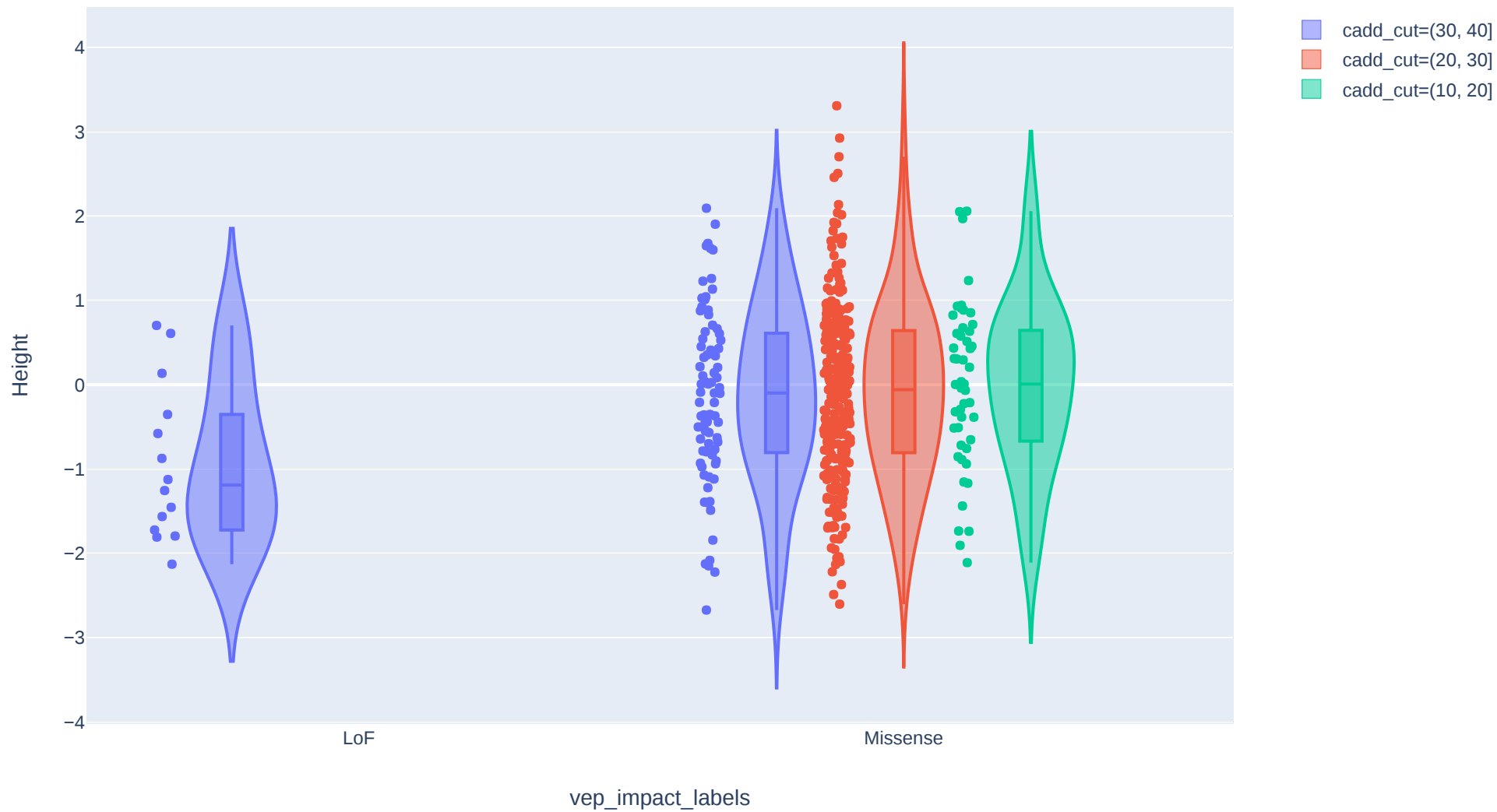


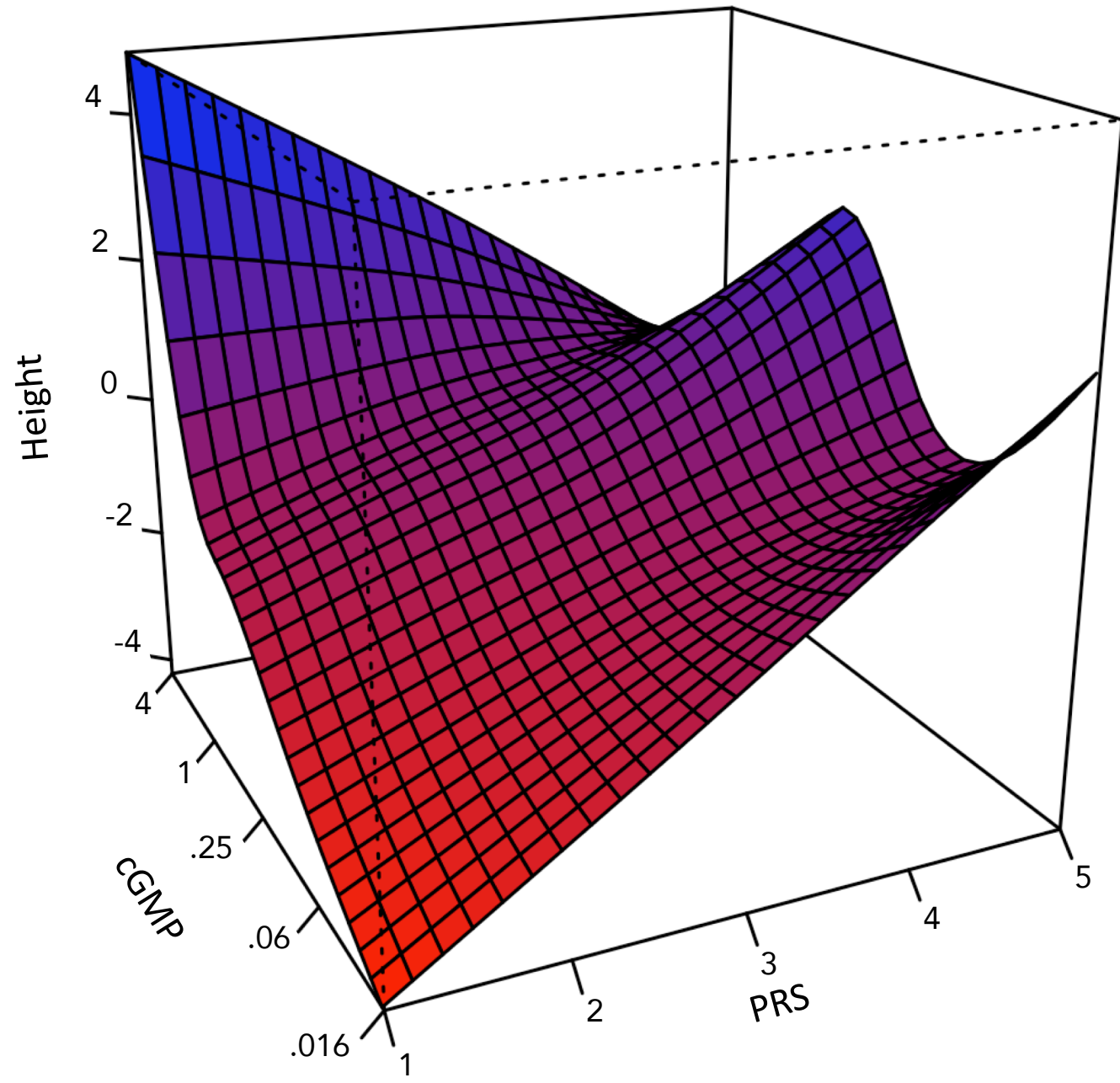
3A



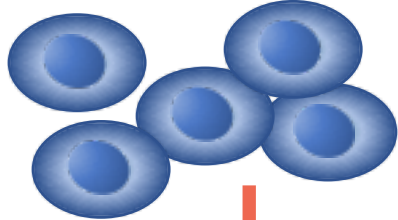






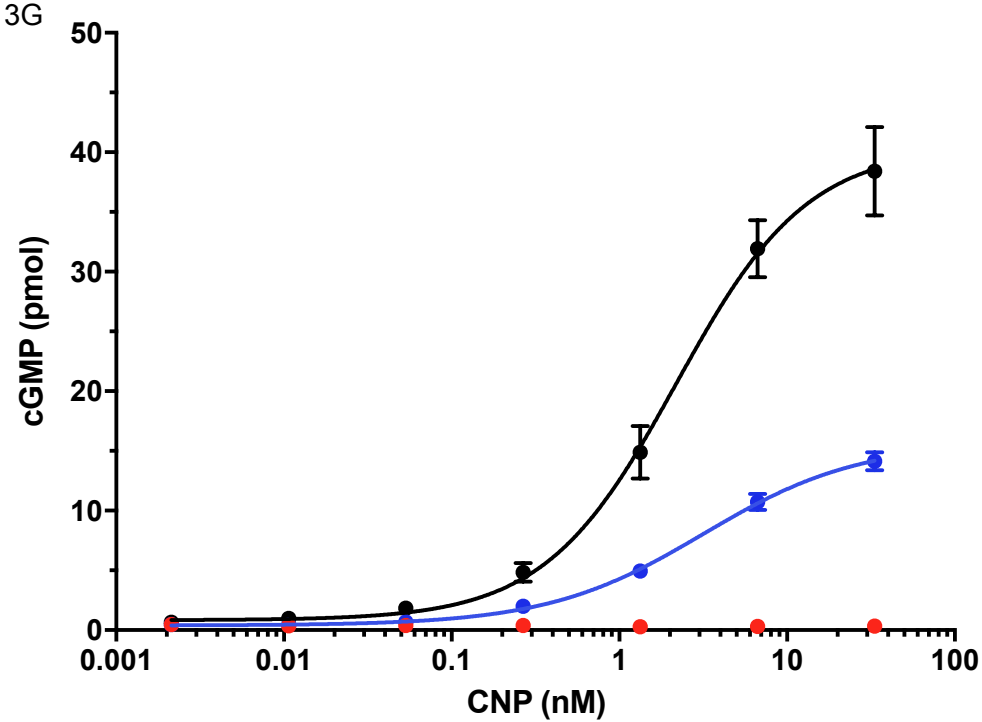


3F

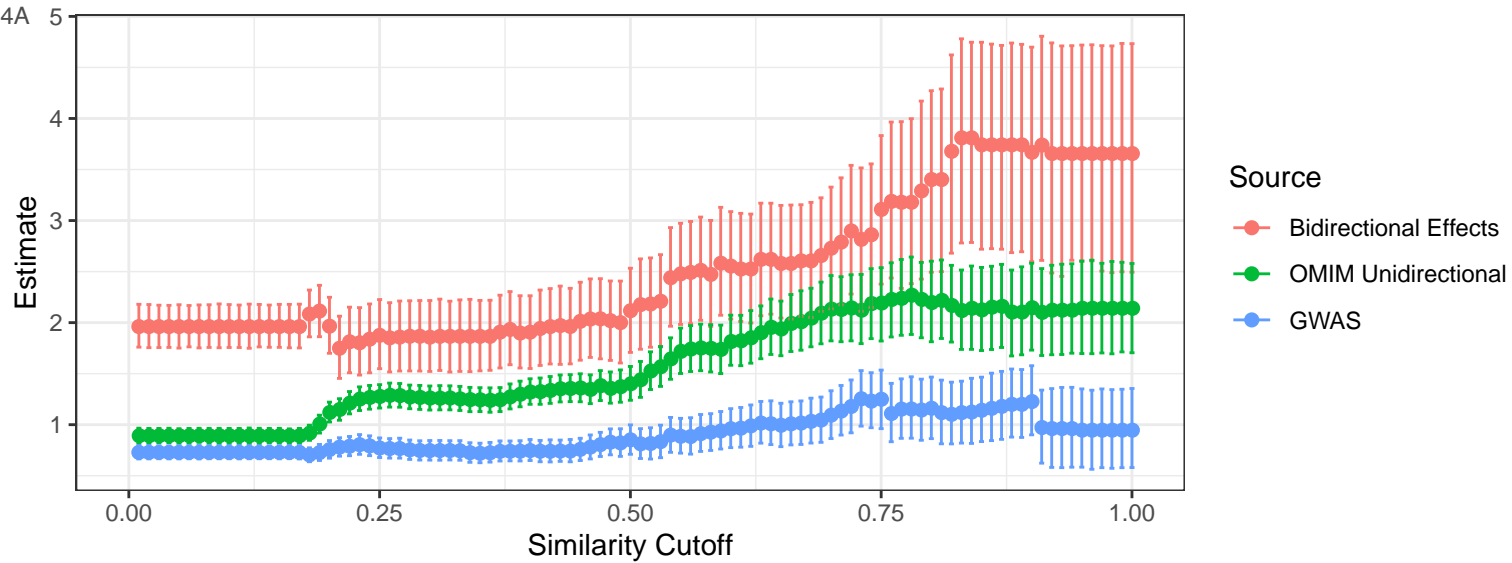


RCS Cells

NPR2 -
/-NPR2 -
/+NPR2
+/+



4A



4B

