Low-cost generation of clinical-grade layperson-friendly pharmacogenetic passports using oligonucleotide arrays

Authors and affiliations

Pauline Lanting[1,2,3*], Robert Warmerdam[1,3*], Jelle Slager[1], Harm Brugge[1], Taichi Ochi[4], Marloes Benjamins[1], Esteban Lopera-Maya[1], Soesma Jankipersadsing[1], Jody Gelderloos-Arends[1], Daphne Teuben[1], Dennis Hendriksen[1,6], Bart Charbon[1,6], Lennart Johansson[1,6], Thijs Oude Munnink[2], Nienke de Boer-Veger[5], Lifelines NEXT, LifeLines Cohort Study, Bob Wilffert[2,4 †], Morris Swertz[1,6], Daan Touw[2], Patrick Deelen[1,3], Nine Knoers[1], Jackie Dekens[1,7#], Lude Franke[1,3#]

[*] Contributed equally
[†] Deceased
[#] To whom correspondence should be addressed

1. Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
2. Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
3. Oncode Institute, Utrecht, The Netherlands
4. Groningen Research Institute of Pharmacy, Unit of PharmacoTherapy, Epidemiology & Economics, University of Groningen, Groningen, The Netherlands
5. Service Apotheek Boterdiep, Groningen, The Netherlands
6. Genomics Coordination Center, Groningen, The Netherlands
7. Center for Development and Innovation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Abstract

Pharmacogenomic (PGx) information is essential for precision medicine, enabling drug prescriptions to be personalized according to an individual's genetic background. Almost all individuals will carry a genetic marker that affects their drug response, so the ideal drug prescription for these individuals will differ from the population-level guidelines. Currently, PGx information is often not available at first prescription, reducing its effectiveness. Pharmacogenetic information is most often obtained using special assays, making it expensive and time-consuming to generate. We therefore hypothesized that we could also use genome-wide oligonucleotide genotyping arrays to generate comprehensive PGx information (PGx passports), thereby decreasing the cost and time required for PGx testing, and lowering the barrier to generating PGx information prior to first prescription.

Taking advantage of existing genetic data generated in two biobanks, we developed and validated Asterix, a low-cost clinical-grade PGx passport pipeline for 12 PGx genes. In these biobanks we performed and clinically validated genetic variant calling and statistical phasing and imputation. In addition, we developed and validated a novel *CYP2D6* copy number variant calling tool, foregoing the need to use separate PCR-based copy number detection. Ultimately, we returned 1227 PGx passports to biobank participants via a layperson-friendly app, improving knowledge of PGx among citizens. Our study demonstrates the feasibility of a low-cost clinical-grade PGx passport pipeline that could be readily implemented in clinical settings to enhance personalized healthcare, ensuring that patients receive the most effective and safe drug therapy based on their unique genetic makeup.

Keywords

Pharmacogenetics, pharmacogenomics, PGx profile, PGx passport, clinical pharmacogenomics

1. Introduction

There is now extensive knowledge available on pharmacogenes, and clinical guidelines are available for personalized drug treatment using genetic data (1). However, the application of pharmacogenetics (PGx) in clinical practice remains limited (2) and is mostly done reactively to understand why patients reacted poorly to prescribed drugs. However, a pre-emptive pharmacogenetic passport would allow results to be stored in (electronic) medical files for future medical use (3), ensuring that PGx data are available at first prescription, thereby enabling optimized treatment and preventing adverse effects (4,5). Currently, the major impediments to such pre-emptive testing are cost and speed: PGx information is most commonly obtained using dedicated assays that target a limited subset of haplotypes of a single gene. This makes the generation of a complete, multi-gene profile time-consuming and costly.

In theory, PGx data can also be obtained from genetic data that is routinely used in both research and clinical genetics. Single nucleotide polymorphism (SNP) microarrays in particular have proven essential for detection of Copy Number Variations (CNVs) in clinical genetics and genome-wide association studies (GWAS) in research (6,7). However, whether this data can be used to provide a clinical-grade PGx passport to individuals is largely unknown, and there are a number of challenges that need to be resolved before this data can be effectively used for PGx.

Firstly, SNP arrays do not consider every variant and do not capture on which haplotype specific information is located, requiring the use of genotype phasing and imputation. Secondly, *CYP2D6*, an important gene in pharmacogenomics (8) is highly polymorphic, and it is also affected by CNVs, making the genotyping of this gene non-trivial. To deliver on the promise of personalized medicine, we need to bridge these gaps in order to translate available genetic data to clinically relevant PGx data.

In parallel with the decreasing cost of genetic testing over the past decade, the number of large-scale biobank initiatives has also increased. Jointly, these biobanks have now generated genotype data for millions of individuals (9–12). This provides an opportunity to repurpose readily available genetic data to develop and validate the use of SNP array data for generating PGx passports. This would allow researchers to return clinical-grade PGx results to biobank participants, thereby decreasing the barriers to applying pre-emptive PGx testing and facilitating personalized medicine in clinical settings.

Another problem is that the disclosure of PGx information so far has mainly been limited to researchers, clinicians, and pharmacists. Yet sharing relevant PGx information with citizens is important to improve knowledge and understanding of PGx testing and its related health outcomes. Digital health technologies, including mobile phone applications, have increased citizen engagement in the last 10 years (13,14), and they allow patients to be engaged in the healthcare decision-making process. Implementation of digital health technology in optimizing medication adherence in patients with depression, respiratory diseases, and cancers has been

shown to improve patient outcomes. However, the use of digital health technologies to provide PGx information to patients has been limited. However, with growing populations throughout the world, governments are now actively stimulating the development and use of such tools. For instance, the recent Integral Care Agreement in the Netherlands highlights the importance of digitalization to promote self-reliant digital healthcare (15).

In the northern Netherlands, the Lifelines Cohort Study, a multigenerational, longitudinal, dynamic cohort and biobank, presents a unique opportunity to leverage available genetic data to facilitate healthy aging (16). This biobank has been genotyping its participants for multiple years, as has the separate Lifelines NEXT prospective birth cohort biobank. In this study, we developed and clinically validated 1) an automated software tool for the interpretation of array-based genotype data to generate reliable PGx profiles, 2) a layperson-friendly translation of Dutch Pharmacogenetics Working Group (DPWG) guidelines, and 3) a digital infrastructure to communicate these to biobank participants. By developing this tool, we aim to improve the lives of patients and the general population, starting with the population of the northern Netherlands and expanding to more individuals in the future.

2. Methods

2.1 Datasets and quality control

We used four different datasets to develop and validate the automated generation of PGx passports, divided over two different versions of Illumina Genotyping arrays. We used 38,030 samples from the UMCG Genetics Lifelines Initiative (UGLI) cohort of the Lifelines prospective follow-up biobank (17) to develop and validate an

automated PGx passport pipeline using the Infinium Global Screening Array® MultiEthnic Disease (GSA-MD) version 1.0 array. In parallel, we used three datasets to develop the pipeline for the GSA-MD version 3.0 array: 2473 samples from our Genome Diagnostics patient cohort (UMCG), 77 validation samples with known PGx genotypes, and another 574 samples from the Lifelines NEXT cohort (18).

The samples of the Lifelines UGLI cohort were genotyped using GSA-MD version 1.0. We performed quality control on both samples and markers, including removal of samples and variants with a low genotyping call rate (<97%), variants showing deviation from Hardy-Weinberg equilibrium ($p < 1\times10^{-6}$), excess of Mendelian errors in families (>1% of the parent–offspring pairs), and samples with very high or low heterozygosity (4 standard deviations (SD) above or below the mean for autosomes). We further revised and removed samples that did not show consistent information between reported sex and genotypes on the X chromosome, or between reported familial information and observed identity-by-descent sharing with family members. All other samples were genotyped using GSA-MD version 3.0 and were quality-controlled similarly to those in Lifelines UGLI. In total, 36,373 Lifelines UGLI samples, 470 Lifelines NEXT samples, 2285 validation samples of the Genome Diagnostics patient cohort, and 76 validation samples with known PGx genotypes passed quality control.

2.2 Selecting individuals to receive pharmacogenetic results

Quality control measures in research settings are commonly less stringent than in diagnostics. Since Lifelines UGLI samples were genotyped in a research setting, this dataset could potentially contain sample swaps. Therefore, we used Idéfix (19) to

154 select a high-quality subset of 5,765 samples suitable for returning personalized
155 results. In short, Idéfix selects those samples for which the actual phenotypes
156 sufficiently match the predicted phenotypes based on genetic predispositions. These
157 genetic predispositions were determined by calculating polygenic scores (PGSs) per
158 phenotype. These calculations aggregate the effects of common risk-decreasing or -
159 increasing alleles identified by GWAS. Here, we reran Idéfix using an updated list of
160 PGSs based on recent GWAS. Quality control, PGS calculation, and running Idéfix
161 were performed as described in Warmerdam et al. (19). Finally, we returned PGx
162 results to 996 active Lifelines participants who gave consent to receive their PGx
163 passport. For Lifelines NEXT, we were able to manually resolve sample swaps
164 based on family data and duplicated array runs. We returned PGx results to 231
165 mothers registered as active Lifelines NEXT participants who gave consent to
166 receive their PGx passport.
167
168 *2.3 Variant calling*
169 Variants were called in two separate stages. In general, we used the Opticall
170 software (20) to do genotype-calling on the majority of variants. However, Opticall is
171 not suitable for genotyping variants covering frequent CNVs. Since CYP2D6 is
172 affected by relatively frequent CNVs, it is not trivial to genotype variants within and
173 near this gene. At the same time, inclusion of CYP2D6 is essential to return a
174 comprehensive PGx passport because it metabolizes a large proportion of
175 medication. We therefore developed a novel algorithm for CNV-calling of the
176 CYP2D6 gene using the data obtained from a SNP array. In addition, this method
177 also performs genotype-calling of individual variants in the CYP2D6 gene.
178

179 *2.3.1 CYP2D6*

180 SNP arrays have been used extensively to study both SNPs and CNVs. Modern
181 SNP arrays genotype at hundreds of thousands of assays (probes designed to
182 genotype one particular variant) simultaneously by measuring the color and intensity
183 of light emitted when one of four labeled nucleotides hybridizes at the assays. In
184 addition to enabling study of which alleles are present in an individual through
185 comparison of the colors of the emitted light, it also allows study of CNVs by
186 combining the total signal intensity. We combined these two approaches in our
187 *CYP2D6* genotyping methodology, as described below.

188

189 For genotyping *CYP2D6*, we first selected a number of independent, high-quality
190 variants from the quality-controlled genetic data (missingness < 0.01, minor allele
191 frequency (MAF) > 0.05, Hardy-Weinberg equilibrium p-value > 0.05, *HLA* excluded,
192 minimum distance of 100kb between variants, $R^2 < 0.4$). Using these variants, our
193 method performs a principal component analysis of the intensities of the selected
194 variants (summed across the two alleles). As the first principal components (PCs)
195 likely tag batch effects, we chose to regress out the first 100 PCs from the genomic
196 region of interest – the *CYP2D6* gene – using ordinary linear regression (OLS). Both
197 PC loadings and OLS regression parameters can be estimated in one dataset, and
198 later applied in another. The methodology for correcting SNP array intensities follows
199 the approach previously established by Lude Franke et al. (21).

200

201 After correction for batch effects, one variant is selected to assign naive CNV-calls
202 (rs1135824 for GSA-MDv1, rs1423323203 for GSA-MDv3). Naive, preliminary CNV
203 assignments are made by binning the total signal intensity of the selected variant so

204 that the proportions of CNV genotypes match the expected frequency of CNV-calls.
205 The expected frequencies (*5 = 0.0295, Duplication = 0.0275) are obtained from
206 PharmGKB (https://www.pharmgkb.org/page/cyp2d6RefMaterials, European
207 population). Using the means and SDs of the signal intensities for each bin, a naive,
208 preliminary CNV call is assigned to every individual based on which CNV bin has the
209 highest likelihood at the signal intensity of the specific variant. If multiple GSA probes
210 are available for a variant, a majority voting ensemble classifier is used to obtain a
211 consensus of the various assays. Next, genotyping is performed using a manually
212 selected set of variants. These variants are selected based on whether the individual
213 signal intensities show clusters of genotypes in accordance with what can be
214 expected for a biallelic variant with possible deletions and duplications (i.e., at least
215 six visible intensity clusters rather than the three intensity clusters expected when
216 dealing with a biallelic variant outside a CNV).
217
218 For each variant and possible copy number, a nearest-neighbors classifier is used to
219 obtain a set of highly confident samples for every variant. Then, to perform
220 genotyping, the theta of every sample is calculated based on its intensities. For this,
221 the median of the intensities for the samples naively assigned to the group of
222 homozygous deletions is used as the origin. Thereafter, samples are binned into k
223 bins, where k corresponds to the number of genotypes that can be expected for a
224 biallelic variant at each copy number. Bins are assigned genotypes so that they best
225 match Hardy-Weinberg principles. Per variant, naive genotype clusters are then
226 iteratively improved in a gaussian mixture model algorithm adapted from *scikit-learn*
227 (22). The maximization step in the gaussian mixture model algorithm was changed to
228 also weight genotypes according to Hardy-Weinberg principles, which were

229 expanded to also take deletions and duplications into account. Per variant, this
230 procedure is performed while accounting for the intensity measures of all assays
231 tagging that specific variant.
232
233 For each combination of a variant, sample and copy number, a total probability is
234 calculated by summing the probabilities across the genotypes corresponding to that
235 copy number. Subsequently, the most likely copy number is assigned to each
236 variant-sample combination by selecting the copy number with the largest total
237 probability among the possible copy numbers. The CNV status of *CYP2D6* as a
238 whole is determined by summing the probabilities of variants over all possible copy
239 numbers, and rescaling these to sum to one. Genotype probabilities are then
240 calculated by rescaling the possible genotype probabilities in the assigned copy
241 number status to one.
242
243 Genotypes of samples with 1 or 2 *CYP2D6* copies (across both the maternal and
244 paternal chromosome) are written so that genotypes of samples with single-gene
245 deletions are homozygous for the reanalyzed variants. Genotypes for variants of
246 samples that are homozygous for whole-gene deletions do not exist. In human
247 genetic datasets, the representation of genotypes for variants of samples with a
248 *CYP2D6* duplication is not trivial, making the imputation and phasing non-trivial.
249 Therefore, these samples are excluded. *CYP2D6* variants in chromosome 22
250 genotyped using Opticall are replaced with the newly genotyped *CYP2D6* variants.
251 This is subsequently used in the phasing and imputation pipeline. The *CYP2D6*
252 CNV-calling algorithm is included as the first part of the *Asterix* software that we
253 developed to generate PGx profiles.

254 *2.4 Phasing and imputation pipeline*

255 Before phasing and imputation, PLINK files are lifted from Human genome build 37
256 to build 38 and alleles of insertions and deletions are recoded to match the notation
257 in the reference dataset. Then, variants for which the strands do not match the
258 reference panel are removed, and variants with >5% missingness or a Hardy-
259 Weinberg p-value < $1x10^{-6}$ are removed. Phasing and imputation are implemented
260 using Eagle v2.4.1 (23) and Minimac4 (24)(version 1.0.2). The reference dataset for
261 phasing and imputation is derived from the 1000 Genomes project 30x coverage
262 whole genome sequencing dataset (25). Finally, we replaced the imputed genotypes
263 with the original genotypes for the variants that were used as input in the imputation.

264

265 *2.5 Crucial backbone*

266 A minimum set of variants that should be included in our genotyping panel was
267 determined per gene (see Table S1). We used available 'tier 1' recommendations for
268 *CYP2C9*, *CYP2C19*, *CYP2D6, CYP3A4, CYP3A5,* and *VKORC1* (26–29). For the
269 other genes in our PGx passport, we assessed the variants/alleles that were
270 mentioned in DPWG background information and those that are included in our
271 validation dataset. We included a variant in the crucial backbone when: 1) the variant
272 results in an allele known to alter the drug response phenotype, and 2) the MAF in
273 the (North-Western) European population is > 0.001.

274

275 *2.6 Validation*

276 We validated GSA-MD versions 1 and 3 and *Asterix* following the clinical genome
277 diagnostic standards in our center. The validation parameters and corresponding
278 acceptance criteria are shown in Table 1.

279 *Table 1: Validation parameters and acceptance criteria for the clinical validation of the Infinium Global Screening*
280 *Array MultiEthnic Disease versions 1 and 3, and the software tool* Asterix. *‡For all variants available in both*
281 *gnomAD-NFE and GoNL datasets, the distribution of mutant allele frequency differences between those two*
282 *datasets was recorded. Z represents the standard deviation of that distribution. Variants that deviated more than*
283 *3*Z from the gnomAD-NFE mutant allele frequency failed validation.*

Topic	Validation parameter	Acceptance criterium
Quality of the array data	Call rate	³0.97
Phasing and imputation quality	Concordance with trio-data	³0.97
	R^2	³0.97
Result interpretation	Absolute number of mutant alleles detected	³5
	Mutant allele frequency	<3*Z‡
	Concordance with known genotypes	100%

284

285 We created a benchmark set of haplotypes using data of 621 (GSA-MD v1) and 135
286 (GSA-MD v3) parent–offspring trios to validate phasing performance. Using
287 *SHAPEIT5* (30) we then compared the results of our statistical phasing with the
288 benchmark haplotypes and computed the switch error rates (SER) of all genetic
289 variants that were directly typed. The call rate and imputation quality are validated
290 based on the output of our quality control procedures. Variants were removed if: 1)
291 they were not directly genotyped nor imputed, 2) they were not in the crucial
292 backbone and had a low imputation probability ($R^2 < 0.97$), 3) the mutant allele was
293 observed < 5 times in the full dataset, 4) the mutant allele frequency deviated more
294 than 3*Z from the gnomAD-non-Finnish Europeans (NFE) mutant allele frequency (Z
295 represents the SD of differences in mutant allele frequency between the GoNL and
296 gnomAD-NFE datasets), 5) the phasing SER was > 0%, or 6) the concordance with
297 known genotypes of validation samples was < 100%.

298

299 When a variant was in the crucial backbone, but the imputation probability was low,
300 we used haploid-level dosage (HDS) to assess the variant call confidence level in an
301 individual sample.
302
303 In addition to variant calling, we also validated our CNV-calling algorithm. To that
304 end, we selected samples with a confidence score of > 0.97 (1x zero copies, 14x one
305 copy, and 16x two copies) and samples with a confidence score of with a CNV-
306 calling confidence score of > 0.9, and < 0.97 (4x one copy, 3x two copies and 5x
307 three copies). DNA of these samples was analyzed with a dedicated TaqMan real-
308 time PCR assay specifically targeting exon 9 of the *CYP2D6* gene (Certe,
309 Groningen, the Netherlands). It is worth noting that this assay cannot detect gene
310 duplication or hybrid alleles. However, as we report *CYP2D6* to be not available
311 when >2 copies of *CYP2D6* are detected, these calls do not currently require
312 validation.
313
314 Finally, we confirmed that the software we developed correctly translated 1) the
315 detected mutations to the corresponding star alleles and, subsequently, 2) the
316 combination of haplotypes to the corresponding contraindication (CI) and/or
317 predicted phenotype. This is described in more detail in section 2.9.3 and in the
318 Supplementary Information.
319
320 *2.7 Pruning of allele definition tables*
321 We used gene-specific allele definition tables to assign star alleles. Available tables
322 were downloaded from PharmVar (v5.2.14 for GSA-MD version 3; v6.0.5 for GSA-
323 MD version 1) for *CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP3A4, CYP3A5* and

324 *NUDT15* and from PharmGKB for *ABCG2* and *TPMT*
325 (https://www.pharmgkb.org/page/pgxGeneRef). Custom tables were constructed for
326 *MTHFR, SLCO1B1* and *VKORC1*. We curated all allele definition tables for the GSA-
327 MD versions 1 and 3 separately and removed all variants that did not pass
328 acceptance criteria (Tables S2a-b). We then pruned each table to prevent
329 ambiguous calls caused by haplotypes that we cannot distinguish based on the
330 variants included in our pipeline. The resulting allele definition tables are shown in
331 Tables S3a-x.
332
333 *2.8 Translation to pharmacogenetic contraindication*
334 To translate the diplotype to the PGx CI used in medication surveillance, we
335 designed custom translation tables (see Tables S4a-l). Allele function of individual
336 star alleles and their translation into predicted phenotype and PGx CI are based on
337 PharmVar, PharmGKB, and/or DPWG background information. All possible
338 diplotypes based on the pruned allele definition tables are included in the translation
339 tables. We made sure the PGx CIs correspond to those used in the DPWG
340 recommendations.
341
342 *2.9 Generating PGx profiles*
343 We developed the *Asterix* tool to produce PGx profiles from genotyping data from
344 the GSA-MD versions 1 and 3. *Asterix* is composed of four main tasks: 1)
345 determining the *CYP2D6* copy number status, 2) determining star alleles, 3)
346 translating these to predicted phenotypes and PGx CIs, and 4) constructing a
347 logging file.
348

349 *2.9.1 Determining star alleles, predicted phenotypes, and PGx CIs*

350 After correction of the *CYP2D6* genotypes based on the CNV status, star alleles for
351 all included genes are assigned using the gene-specific allele definition tables.
352 Subsequently, predicted phenotypes and PGx CIs are determined with the custom
353 translation tables. When determining star alleles present for *CYP2D6*, *Asterix* takes
354 into account whether whole-gene deletions (*CYP2D6*5*) were detected in CNV-
355 calling.

356

357 *2.9.2 Constructing a logging file*

358 To be able to check *Asterix* output, an extensive logging file is constructed that
359 contains information for several checkpoints in our pipeline. Parameters that are
360 logged include the estimated phased haploid-level dosage (HDS) when < 0.97 (only
361 for imputed genotypes), missing variants, and haplotypes that were not described in
362 translation tables. Identical haplotypes are merged together to simplify the validation
363 process.

364

365 *2.9.3 Approval and release of the PGx profiles*

366 The PGx profiles generated are to be released by a registered pharmacist. To
367 maximize process efficiency, the approval and release of the PGx profiles is
368 performed on the level of unique combinations of detected mutations and unreliable
369 calls. When a sample contains only previously approved haplotypes, it is
370 automatically released. Each decision is logged to maintain an audit trail. The
371 procedure for approving and releasing the profiles is described in more detail in the
372 Supplementary Information.

373

374 *2.10 Building layperson-friendly PGx passports*

375 The output of *Asterix* is a PGx profile that is suitable for application in clinical practice
376 by healthcare practitioners. However, (future) patients should not be expected to
377 understand the meaning or implications of such a profile. We therefore developed a
378 layperson-friendly digital PGx passport to fulfill the informational needs of (future)
379 patients.

380

381 *2.10.1 Developing layperson-friendly communication materials*

382 We aimed to develop communication materials that explain the impact of drug–gene
383 interactions (DGIs) that are understandable for 90% of the study population. We
384 developed layperson-friendly texts based on the DPWG guidelines for healthcare
385 practitioners (see Table S5), written at Common European Framework of Reference
386 Common Reference Level B2 (31). One important difference that we had to address
387 is that the DPWG guidelines do not contain an explanation for normal metabolizers
388 because standard-of-care applies for these patients. However, texts for the (future)
389 patient do need to cover this topic because it is valuable to explain that standard-of-
390 care applies. The texts are constructed using three building blocks (Figure 1).

391

Block 1: Predicted influence of your DNA. *The generic drug name is used and the nature of the DGI is explained. The PGx result, including possible implications such as increased or decreased chance of ADRs and/or efficacy and recommendations for drug use is explained. A reference to the physician or pharmacist for more information is included for actionable DGIs.*	**Predicted influence of your DNA** How well *amitriptyline* is processed by the body varies from person to person. This is, in part, determined by your DNA. Your genes show that *amitriptyline* is processed much more slowly than average by your body. The normal dosage can, therefore, lead to a higher chance of side effects. You will probably need a lower dose of *amitriptyline*. Your physician or pharmacist can tell you more about the appropriate dose.
Block 2: Advice. *Statement when sharing the PGx result with the physician or pharmacist would be wise. An explanation that a dose currently in use may already be tailored sufficiently is included, as well as a reminder to never change drug use without consulting their treating physician or pharmacist.*	**Advice** If you are using or, in the future, get prescribed *amitriptyline*, it would be wise to share the information in your Pharmacogenetic Passport with your physician or pharmacist, so they can take this into account. Your physician and pharmacist also consider factors such as age, sex, and other medication you might be using. Based on that complete picture, it is decided whether *amitriptyline* and the standard dosage is the best choice for you. It is possible that your physician or pharmacist has established that your medication and dosage are suitable for you. **Attention: never adjust the dosage or stop the use of your medication yourself. Always consult with your physician or pharmacist.**
Block 3: This is what the drug does. *General information about the drug and link to a reliable website where more layperson friendly information about this drug can be found.*	**This is what the drug does** *Amitriptyline* is an antidepressant and painkiller. It improves mood and reduces anxiety and pain. It also improves complaints such as worrying, sleeping problems, shaking and irritability. *Amitriptyline* is prescribed for depression, migraine, neuropathic pain, and anxiety disorders. It is also sometimes prescribed for bedwetting, persistent hiccups, and to relieve symptoms of fibromyalgia and ALS. *Read more about amitriptyline at https://www.apotheek.nl/medicijnen/amitriptyline.*

Figure 1: Building blocks used to construct layperson-friendly texts to explain Drug–Gene interactions (left) and the corresponding text for CYP2D6 PM-amitriptyline (translated from Dutch, right).

396 *2.10.2 Digital infrastructure*

397 We created a DGI table that contains all the CIs produced by *Asterix*, a list of known
398 DGIs currently in the DPWG guidelines, and the layperson-friendly texts associated
399 to each combination of these. The relevant texts for each participant are selected
400 based on their PGx profile and known DGIs. For example, the text explaining the
401 consequences for amitriptyline in Figure 1 will be selected for the individuals who are
402 listed as CYP2D6 PM in their PGx profile. The PGx profile and corresponding texts
403 are subsequently unlocked through an Application Programming Interface (API)
404 following the Health Level Seven Fast Healthcare Interoperability Resources (HL7-
405 FHIR) format. Participants can view their PGx passport in the mobile application
406 (app) *Gen en Geneesmiddel,* which we developed with Synappz Digital Healthcare
407 and which is available for iOS and Android (see Figure 5).

408

409 *2.10.3 Additional features*

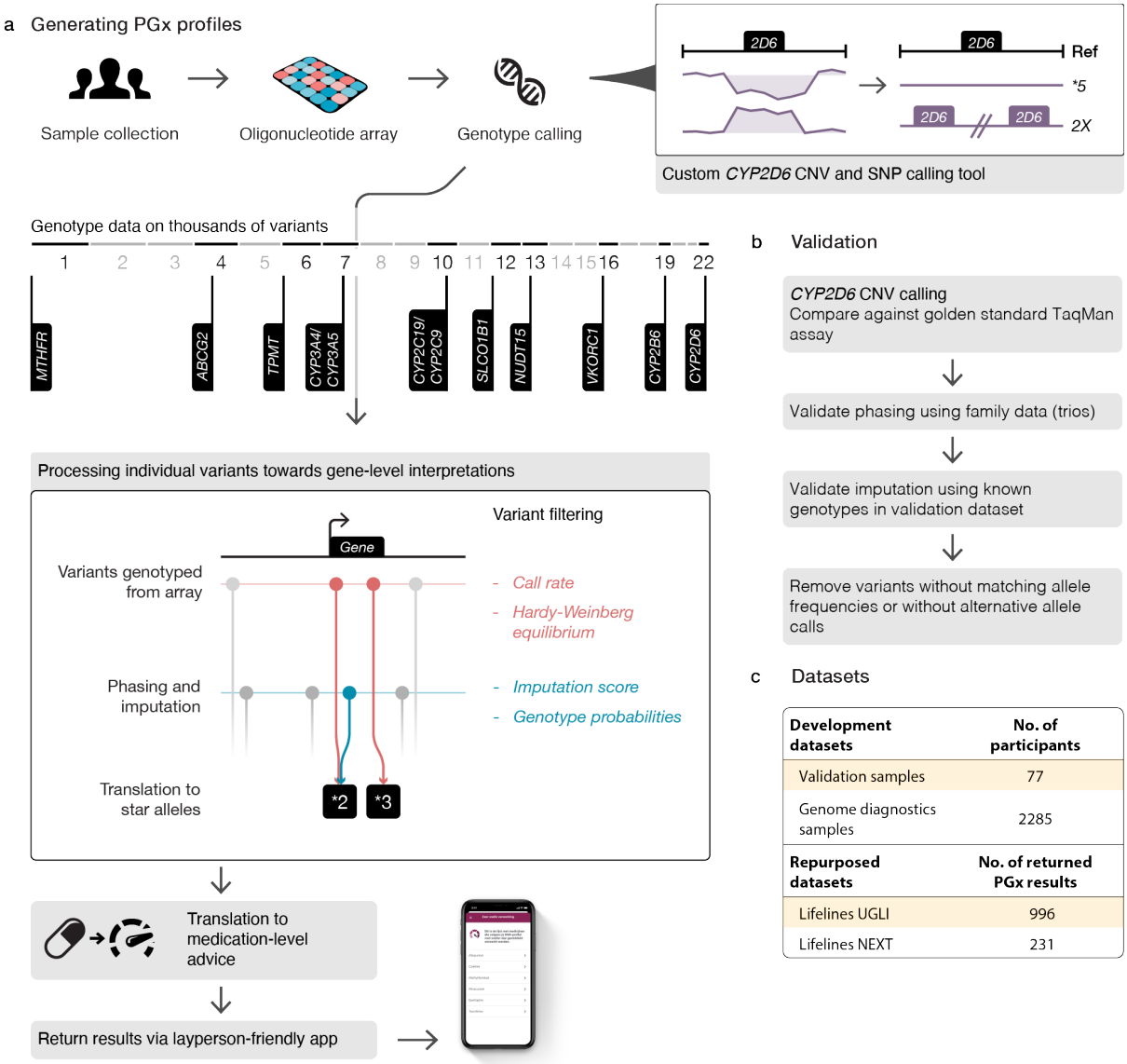
410 Upon installing the app, the user is shown a series of information screens that
411 explain PGx and the content and use of the app. This information can be reviewed
412 within the app after completion of this initial onboarding process. Furthermore,
413 frequently asked questions (FAQ) are included in the app to address several
414 anticipated questions. The FAQ-section is supplemented by new questions from
415 users and associated answers.

416

417 The layperson-friendly texts displayed in the mobile application are accompanied by
418 articles that describe the PGx CI that interacts with the drug being viewed (see Table
419 S5). When multiple PGx CIs are relevant for the drug being viewed, all the relevant
420 articles are shown. The articles provide interested participants with more detailed

421 information and facilitate communication between (future) patients and healthcare
422 practitioners.
423
424 Finally, a technical report aimed at healthcare practitioners can be downloaded from
425 the app. This report contains the PGx profile and accompanying information,
426 including contact information for a helpdesk and instructions on how to interpret and
427 process the results.
428
429 **3. Results**
430 *3.1. Pharmacogenomic profiles and passports*
431 To validate the use of a layperson-friendly, clinical-grade PGx passport using
432 oligonucleotide arrays, we have developed a tool (*Asterix*) to produce PGx profiles
433 from genotyping data from the GSA-MD versions 1 and 3. As described in Figure 2a,
434 our tool works by determining the *CYP2D6* copy number, phasing and imputing
435 genetic variants, determining the star alleles of the samples provided, and translating
436 diplotypes to PGx CIs. Instructions on the installation and use of *Asterix* are
437 described on the GitHub repository (https://github.com/molgenis/asterix).
438
439 *3.2. Validation*
440 A total of 34 variants, in 12 genes, were included in the crucial backbone of the PGx
441 pipeline, and subsequently underwent validation (Table 2). For *CYP2B6* we initially
442 also included rs3211371 (**5*). However, since the DPWG withdrew CI definitions
443 differentiating *CYP2B6*5* from **1*, we removed this SNP from the crucial backbone.
444 Three variants in *CYP2C9* (rs28371686 (**5*), rs9332131 (**6*) and rs7900194 (**8*))
445 were initially included in the crucial backbone based on their multiethnic allele

frequency (26). However, since each of these variants had a MAF < 0.001 in the NFE population, we subsequently excluded them from the crucial backbone for GSA-MD v1. It is important to note that the Lifelines population is known to be reflective of



Development datasets	No. of participants
Validation samples	77
Genome diagnostics samples	2285
Repurposed datasets	**No. of returned PGx results**
Lifelines UGLI	996
Lifelines NEXT	231

Figure 2: Overview of the work in this paper. (a) Schematic of the pharmacogenetic (PGx) pipeline developed to generate PGx profiles. Data collection involves obtaining DNA samples from participants and genotyping using an oligonucleotide array. We ascertain CYP2D6 *copy numbers (CN) using a custom CNV calling approach. This method estimates CN status at multiple points within the* CYP2D6 *gene and integrates these estimates to determine the overall gene copy number status. Twelve genes contribute to the generation of PGx profiles. The combined effects of these genes are aggregated into medication-level recommendations, which are communicated to participants via a layperson-friendly app. (b) Broad overview of the methodology used for*

457 *validating the PGx pipeline. (c) Overview of the datasets used for development and validation of the PGx*
458 *passport (Development datasets), and those for which we also returned PGx-profiles (Repurposed datasets).*

459

460 the NFE population. For the same reason, we excluded rs41303343 (*CYP3A5*7*).
461 For *TPMT* variant rs1800462 (**3A*/**3B*), no mutant alleles were detected on GSA-MD
462 v3. As we could not ensure the associated GSA probe was functional, *TPMT* was
463 excluded from PGx profiles generated with GSA-MD v3. All other variants in the
464 crucial backbone passed validation criteria. In case of an imputation score (R^2) <
465 0.97, imputation quality was judged from sample-specific HDS. When the HDS <
466 0.97, a registered pharmacist decided during approval and release whether a result
467 could be released for the associated gene.

468

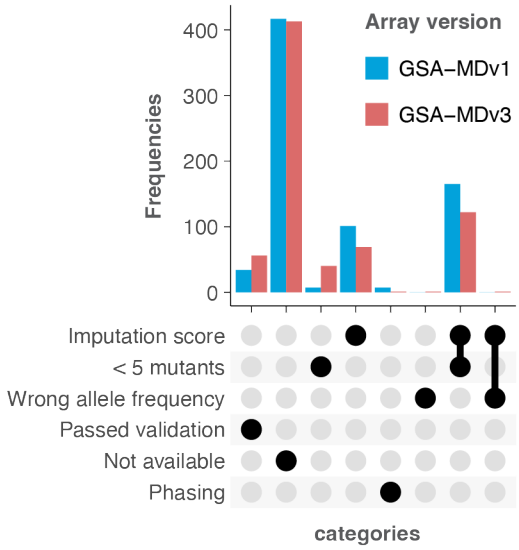
469 *Table 2: Validation results of crucial backbone variants.*
470 *[a]rs28371686, rs9332131, and rs7900194 were left out of crucial backbone for GSA-MD v1, based on known*
471 *ethnic composition of Lifelines*
472 *[b]rs41303343 was left out of the crucial backbone for both GSA-MD v1 and GSA-MD v3, based on known ethnic*
473 *composition of Lifelines*
474 *[c]Reference allele frequency in non-Finnish Europeans (NFE)*

			GSA-MD v1		**GSA-MD v3**		
Gene	**rsID**	**Alt. allele frequency[c]**	**method**	**R^2**	**method**	**R^2**	**failed validation**
ABCG2	rs2231142	0.1092	direct		direct		
CYP2B6	rs3745274	0.2372	direct		direct		
CYP2C19	rs4244285	0.1480	direct		imputation	0.99	
	rs4986893	0.0002	direct		direct		
	rs12248560	0.2200	direct		direct		
CYP2C9	rs1799853	0.1266	imputation	0.99	imputation	1.00	
	rs1057910	0.0662	direct		direct		
	rs28371686	0.0000	(*imputation*)[a]	(*0.02*)	direct		
	rs9332131	0.0001	(*imputation*)[a]	(*0.00*)	direct		
	rs7900194	0.0005	(*imputation*)[a]	(*0.01*)	imputation	0.90	

	rs28371685	0.0029	direct		direct		
CYP2D6	rs1135840	0.5588	imputation	0.87	imputation	0.93	
	rs16947	0.3347	imputation	0.89	imputation	0.96	
	rs35742686	0.0170	imputation	0.79	imputation	0.82	
	rs3892097	0.1986	imputation	0.91	imputation	0.92	
	rs5030655	0.0117	imputation	0.61	imputation	0.27	
	rs5030656	0.0285	imputation	0.67	imputation	0.70	
	rs1065852	0.2225	imputation	0.91	imputation	0.95	
	rs28371706	0.0011	direct		imputation	0.86	
	rs61736512	0.0006	imputation	0.04	imputation	0.97	
	rs59421388	0.0006	imputation	0.04	imputation	0.97	
	rs1058164	0.5525	imputation	0.86	imputation	0.93	
	rs28371725	0.0955	imputation	0.84	direct		
CYP3A4	rs35599367	0.0492	direct		direct		
CYP3A5	rs776746	0.9311	direct		direct		
	rs10264272	0.0009	direct		direct		
	rs41303343	0.0004	(*imputation*)[b]	*(0.96)*	*n/a*[b]		
MTHFR	rs1801133	0.3373	direct		direct		
NUDT15	rs116855232	0.0029	direct		direct		
SLCO1B1	rs4149056	0.1587	direct		direct		
TPMT	rs1800462	0.0024	imputation	0.46	direct		<5 mutants detected
	rs1800460	0.0382	imputation	0.96	direct		
	rs1142345	0.0423	imputation	1.00	direct		
VKORC1	rs9923231	0.3779	direct		direct		

We considered 731 and 703 variants outside the crucial backbone for inclusion in the translation tables of GSA-MD v1 (PharmVar tables v6.0.5) and GSA-MD v3 (PharmVar tables v5.2.14), respectively. Results of the validation process are summarized in Figure 3 and are fully available in Tables S2a-b. Many variants were removed due to unavailability on the genotyping arrays and in the imputation panel. Beyond that, poor imputation score and/or low allele frequency were the leading reasons for exclusion. Phasing quality was verified by comparison to phasing results using parent–offspring trios. Poorly phased variants were retained when they were

484 the only variant of interest in that gene, rendering phasing irrelevant. Importantly, 37
485 variants were typed in 76 validation samples, using a traditional DNA amplification
486 genotyping method (Leiden University Medical Center). For all 37 variants, the
487 genotypes derived from GSA-MD v3 (19 directly typed, 18 imputed) were a 100%
488 match to the traditional assays.
489



490
491 *Figure 3: Validation status of variants from the PharmVar translation tables outside the crucial backbone. The*
492 *UpSet plot displays the frequency of variants that either meet or fail to meet the validation criteria. Columns with*
493 *a single marked circle represent the frequency of variants failing exclusively that criterion, while columns with*
494 *multiple marked circles indicate variants that failed multiple corresponding criteria.*

495 Within *Asterix*, we embedded a custom *CYP2D6* copy number (CN) calling
496 algorithm. To validate the *Asterix*-derived *CYP2D6* CN calls, a total of 43 samples
497 were selected, representing a CN probability of at least 0.9 and each of the four
498 possible CN calls (zero copies, one copy, two copies, and three copies). The CN
499 status of the aforementioned samples was determined using a dedicated TaqMan
500 assay targeting exon 9. To compare the results, we initially evaluated whether the
501 overall CN calls with a moderate probability of over 0.9 matched the calls reported by
502 the TaqMan assay (Figure 4a). Of the 43 samples, *Asterix* identified three as having

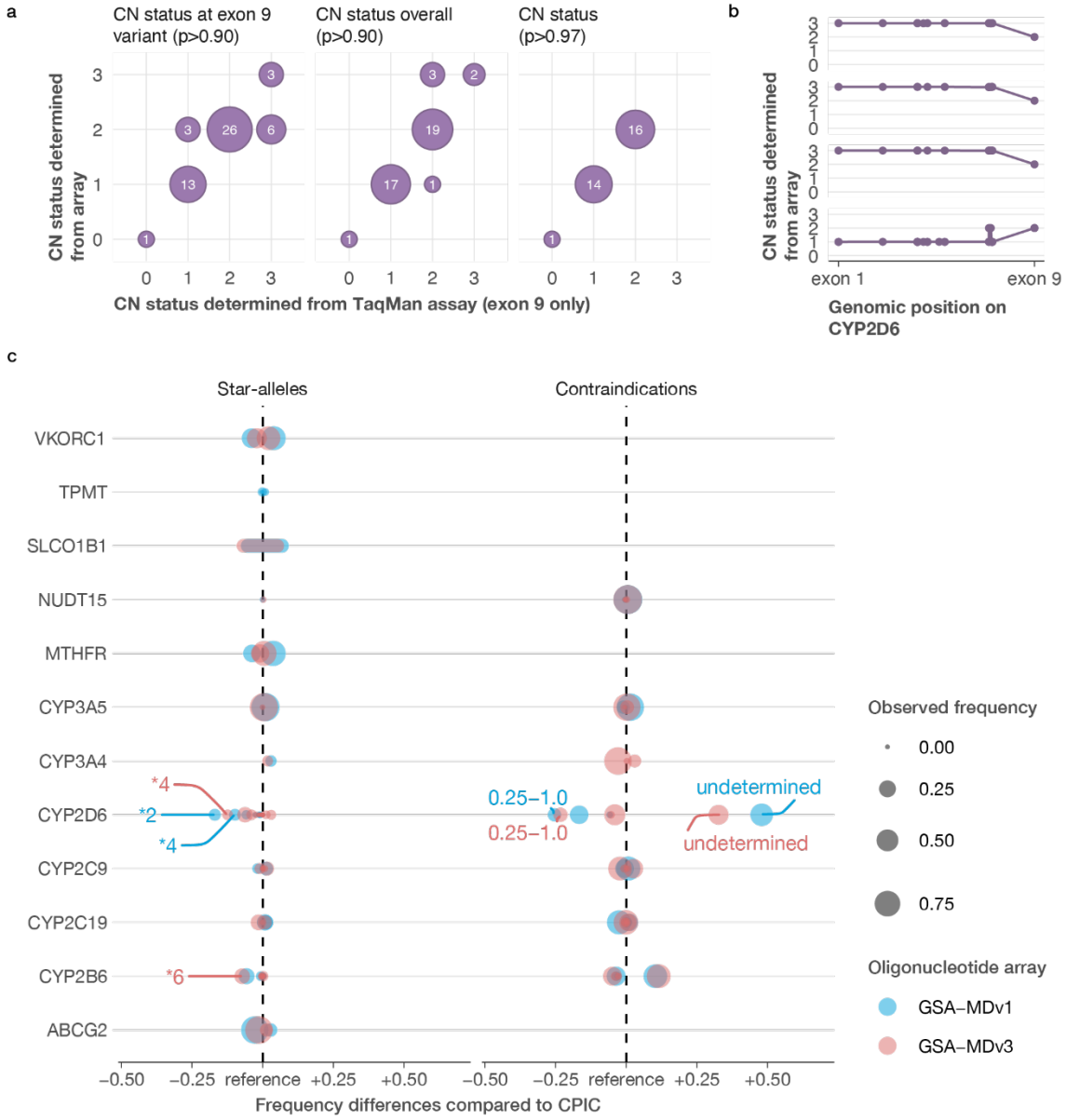


Figure 4: Validation results of CYP2D6 CNV-calling (a, b). a) correspondence between a standard CYP2D6 CNV-calling procedure using a TaqMan assay (x-axis) and the CNV-calling results determined using GSA-MD v3 and our specially designed algorithm. Only exon 9 was assessed in the TaqMan assay. The y-axis for the three graphs shows the copy number (CN) status determined from the array for a variant at exon 9 only (left), for the overall CN status at a probability threshold of 0.90 (middle), and for the overall CN status at a probability threshold of 0.97 (right). b) the CN status for the four deviating samples with a probability threshold of 0.90 over all variants included in the CNV-calling algorithm. c) Haplotype frequency differences compared to the frequencies reported by CPIC (left) and contraindication frequency differences compared to the frequencies reported by CPIC (right).

513 a CYP2D6 duplication (three copies), whereas the TaqMan assay reported these
514 samples as homozygous reference (two copies). In one instance, *Asterix* reported a
515 heterozygous deletion (one copy) while the TaqMan assay returned a homozygous
516 reference call. However, for these four samples, *Asterix* consistently reports a
517 homozygous reference call at exon 9, which is consistent with the call reported by
518 the TaqMan assay. For the majority of the other variants, *Asterix* reports calls
519 matching that of the overall aggregated estimate (Figure 4b). This suggests that
520 these samples are likely to possess a hybrid allele or tandem arrangement, resulting
521 in a distinct CN call at exon 9. It is worth noting that the TaqMan assay is not
522 capable of differentiating between regular copy numbers and hybrid alleles. At a
523 probability of at least 0.97, all 31 samples exhibit matching CN calls.
524
525 Based on the validated variants, final translation tables were created to first translate
526 variants to haplotypes (Tables S3a-x) and then to translate diplotypes to CI (Tables
527 S4a-l). These tables were then used as input for *Asterix* to create preliminary PGx
528 profiles for 996 Lifelines participants and 470 Lifelines NEXT participants. Approval
529 (or adjustment) and release by a registered pharmacist resulted in the final PGx
530 profiles. Haplotype frequencies were compared to reference frequencies in a
531 European population (source: Clinical Pharmacogenetics Implementation
532 Consortium (CPIC)) or NFE (source: DPWG/gnomAD) (Figure 4 and Tables S6a-b).
533
534 Finally, the technical PGx profiles were shared with the biobank participants in the
535 smartphone application *Gen en Geneesmiddel* (Figure 5). The app links the
536 technical information in the PGx profile with associated explanatory texts, yielding a
537 layperson-friendly PGx passport. Rather than using genes as the basic unit of

538 information, the app categorizes drugs into five categories (Figure 5a): poor
539 metabolism (‘very slow’), intermediate metabolism (‘slow’), normal metabolism
540 (‘average’), ultrarapid metabolism (‘very fast’), and non-metabolic drug–gene
541 interactions (‘other processes’). By selecting one of the categories, users can see a
542 list of all the medications in that category, based on their pharmacogenetic profile
543 (Figure 5b). More information can be found on specific medication (Figure 5c), i.e.,
544 what is it used for and how is it affected by the user’s PGx profile. Interested users
545 can get more information on the associated CIs by proceeding to a linked article.
546 Finally, the technical information (PGx profile) can be downloaded and/or shared
547 directly from the app (Figure 5d).
548
549 Based on the PGx profiles we generated using *Asterix*, 92% and 90% of the Lifelines
550 and Lifelines NEXT populations, respectively, carried at least one actionable PGx CI
551 (Tables S6a-b list frequencies of PGx haplotypes and CIs). We have returned these
552 profiles to the biobank participants through the *Gen en Geneesmiddel* app, enabling
553 them to get optimized (future) drug prescriptions with the help of their physician and
554 pharmacist.
555

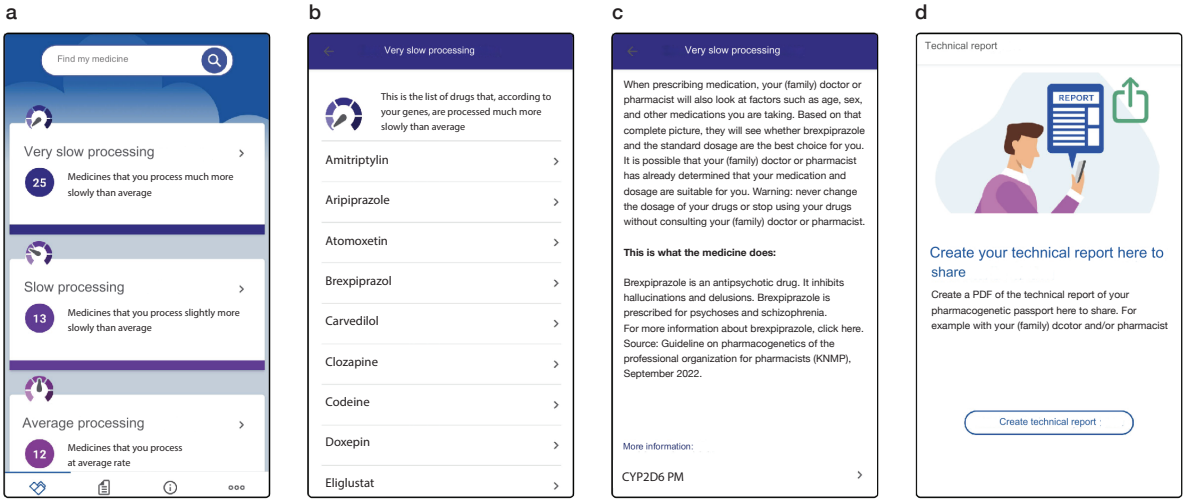


Figure 5: Overview of the 'Gen en Geneesmiddel' app (translated from Dutch). a) Main menu, where five categories can be selected: 'poor', 'intermediate', 'normal' and 'ultrarapid' metabolism and 'other processes'. b) A personalized list of medication in the 'poor metabolism' category. c) Explanation of the predicted metabolism of brexpiprazole. d) The pharmacogenetic profile can be saved or shared from the app.

4. Discussion

In this study, we use oligonucleotide genotyping arrays to develop a low-cost PGx passport that repurposes existing genotype data from the Lifelines and Lifelines NEXT cohorts for clinical use. We clinically validated the genotyping method by aligning validation parameters and corresponding acceptance criteria to the local clinical diagnostic center and by comparing genotyping results with results from classical PCR-based assays. Using parent–offspring trios, we could also verify the accuracy of haplotype phasing, information that is unavailable when using PCR-based PGx assays. We showed that we could confidently include most of the recommended 'tier 1' PGx variants for the genes included in our study (26–29), with only a few exceptions. *TPMT* was excluded for the GSA-MDv3 oligonucleotide array because variant rs1800462 did not meet validation criteria. In addition, three *CYP2C9* variants classified as tier 1 (mutant allele frequency > 0.001) were excluded from the crucial backbone on the grounds of their much lower allele frequencies in

576 the European population. To apply our pipeline in a broader context, these variants
577 would need to be validated in a population with higher MAFs.
578
579 Furthermore, to address previously identified barriers to use of PGx passports by
580 laypeople, we implemented communication of relevant PGx information through a
581 newly developed smartphone application. As the uptake of the PGx passport
582 increases, support will be provided for healthcare professionals to properly apply
583 PGx in their daily practice.
584
585 *4.1 Producing PGx profiles*
586 As the generation of clinically valid PGx profiles from array-based genotype data is
587 the cornerstone of this study, we discuss this first. We did not include *DPYD, FV,*
588 *HLA-A, HLA-B, CYP1A2,* and *COMT* in the PGx passport because: 1) *DPYD*
589 genotyping is already pre-emptively performed in combination with phenotyping in
590 routine care, 2) the DPWG guidelines for mutations in *FV* were retracted, 3) there is
591 no allele definition table for determining *HLA* alleles, so we could not use the same
592 pipeline for this region, and 4) *CYP1A2* and *COMT* are part of DPWG guidelines but
593 currently do not contain DGIs.
594
595 An important limitation in our passport production is that we could not provide a
596 complete passport for every individual. The reason is twofold: First, we used
597 imputation quality (R^2) to select imputed variants of sufficient quality. If a variant with
598 insufficient imputation quality ($R^2 < 0.97$) was in the crucial backbone, the variant
599 was not removed from the translation tables, as that would lead to the complete
600 exclusion of the corresponding gene for all individuals. Instead, we use haploid

601 dosage scores to select a subset of individuals with imputed genotypes of sufficient
602 quality. We could thereby still report results for a subset of individuals. When
603 applying our method in a diagnostic setting, genotypes should be determined by a
604 back-up analysis if no reliable result can be extracted from the array data. Second,
605 our *CYP2D6* CNV-calling is currently aimed at detecting whole gene deletions and
606 duplications, ignoring potential hybrid alleles. While this is an important limitation, our
607 method is already able to determine CNV-status on the level of single variants as
608 opposed to the full locus. So far, we have used this to our advantage to ensure that
609 CNV-status across the entire gene is consistent, before a whole gene deletions or
610 duplications is called. However, this could also be used to call hybrid alleles, allowing
611 us to assess structural variation on a much higher resolution than currently achieved
612 in diagnostic settings, without the need for more expensive sequencing technologies.
613 Within the scope of this study however, we were unable to develop this.
614
615 We confirm that actionable PGx variants are present in >90% of the general
616 population (Table S7), comparable to the percentages reported in literature. In this
617 implementation study, we were unable to evaluate the impact of these actionable
618 variants on drug efficacy and adverse effects, due to the scale of our study and the
619 absence of high-quality prescription records for our study populations. In the future,
620 we aim to further investigate this by linking prescription data from an external
621 database through the PharmLines initiative (28).
622
623 *4.2 Layperson-friendly communication of PGx passports*
624 Engaging patients with PGx testing is not limited to increasing their understanding of
625 genetic information, it is also meant to enable their participation in prescription

626 decisions. Patients who are well-informed about the implications for their prescription
627 management are better equipped to collaborate with their healthcare providers in
628 developing personalized treatment plans that align with their needs and preferences
629 (32). Furthermore, this engagement builds a sense of empowerment and ownership
630 over one's health, which has been shown to improve adherence to treatment
631 regimens and to lead to better health outcomes (33).
632
633 The PGx passport, as a digital health tool, facilitates patient engagement by
634 providing a platform to exchange PGx information and medication management to
635 enhance patient understanding. By providing layperson-friendly communication
636 within the PGx passport, it may in the future be used to support patient education
637 initiatives to broaden the understanding and implementation of pharmacogenomics
638 across the Netherlands.
639
640 We aimed to develop communication materials that are understandable for 80% of
641 the general population. Since the Lifelines and Lifelines NEXT cohorts are more
642 highly educated than the general population, we speculate that the material should
643 be understandable for around 90% of the study population. In an ongoing
644 implementation study, we will evaluate the level of comprehension among
645 participants, so that texts can be better adapted to citizens' needs.
646
647 Several practical decisions were made in the development of the smartphone
648 application. For example, *CYP3A5* non-expressers (*3/*3) were placed in the app
649 category that represents normal metabolism, while both hetero- and homozygous
650 expressers are placed in the ultrarapid metabolism category. In the Netherlands,

651 *CYP3A5* non-expressers are most prevalent, so the standard of care applies for
652 people with this metabolizer status.
653
654 *4.3 Future perspectives*
655 The completion of the PGx passport and smartphone application represents another
656 step forward on the road toward true personalized medicine. By demonstrating to
657 healthcare professionals how incorporating PGx into daily practice works to
658 streamline prescribing practices through the added value of reduced adverse effects,
659 the PGx passport aims to lower the barrier to making key prescription decisions.
660
661 Although clinical implementation of PGx has picked up pace in the past years, it has
662 not yet been integrated into routine care. One of the main barriers here is likely to be
663 the lack of strong evidence that its application effectively improves healthcare and
664 reduces healthcare costs, which are both prerequisites for reimbursement (34).
665 Other important potential barriers are the complexities associated with integrating
666 genetic data into electronic health records and the need to ensure that healthcare
667 professionals have the requisite expertise and resources to comprehend and utilize
668 PGx insights (35,36). Previous research at our center has shown that, although
669 these barriers exist, they can be overcome (2).
670
671 In this study, we have demonstrated that PGx passports can be generated much
672 more cheaply than the currently prevailing methodologies. However, the efficacy and
673 cost-effectiveness of pre-emptive PGx testing still needs to be more thoroughly
674 proven, and this will require the inclusion of large numbers of participants from the
675 general population. Only then will the numbers of drug prescriptions dispensed

based on the information in a PGx passport be sufficient to perform quantitative analysis of adverse effects and/or treatment effectiveness. As the PGx passport is currently limited to the Netherlands, the findings on cost-effectiveness would only be applicable to the Netherlands. However, with the push to harmonize the European Health Data Space, demonstrating cost-effectiveness in the Netherlands will enrich the changing digital health space. Lastly, as the initial data was derived from the Lifelines biobank, incorporating *Asterix* into future investigations in Lifelines would enable more in-depth research into the interplay between genetic variation, environment, population health, and pharmacotherapy.

4.4 Concluding remarks

This study provides a method to produce low-cost, clinically validated PGx profiles for use by healthcare professionals from genotype array data. As more and more individuals are genotyped, both in academic and healthcare settings, translating available genetic data towards clinical decision-making will be essential for personalized care. Furthermore, we have developed a smartphone application to ensure that the results from our pipeline, *Asterix,* reach those whose health may be impacted. To break down barriers to patient engagement, the application provides a novel approach to understandably communicate PGx results to laypeople. The reception of this information by users will be reported in an upcoming publication. Future steps include the evaluation of the costs and effectiveness of the pre-emptive PGx passport in the general population and further research into the interplay between genetic variation and population health and drug response phenotypes.

References

1. Abdullah-Koolmees H, Keulen AM, Nijenhuis M, Deneer VHM. Pharmacogenetics Guidelines: Overview and Comparison of the DPWG, CPIC, CPNDS, and RNPGx Guidelines. Front Pharmacol. 2021 Jan 25;11:595219.

2. Lanting P, Drenth E, Boven L, Hoek A, Hijlkema A, Poot E, et al. Practical Barriers and Facilitators Experienced by Patients, Pharmacists and Physicians to the Implementation of Pharmacogenomic Screening in Dutch Outpatient Hospital Care-An Explorative Pilot Study. J Pers Med. 2020 Dec 21;10(4):293.

3. Duarte JD, Dalton R, Elchynski AL, Smith DM, Cicali EJ, Lee JC, et al. Multisite investigation of strategies for the clinical implementation of pre-emptive pharmacogenetic testing. Genet Med. 2021 Dec;23(12):2335–41.

4. Swen JJ, Wouden CH van der, Manson LE, Abdullah-Koolmees H, Blagec K, Blagus T, et al. A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-label, multicentre, controlled, cluster-randomised crossover implementation study. The Lancet. 2023 Feb 4;401(10374):347–56.

5. Skokou M, Karamperis K, Koufaki MI, Tsermpini EE, Pandi MT, Siamoglou S, et al. Clinical implementation of preemptive pharmacogenomics in psychiatry. eBioMedicine [Internet]. 2024 Mar 1 [cited 2024 Aug 27];101. Available from: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(24)00044-6/fulltext

6. Mason-Suares H, Kim W, Grimmett L, Williams ES, Horner VL, Kunig D, et al. Density matters: comparison of array platforms for detection of copy-number

721 variation and copy-neutral abnormalities. Genetics in Medicine. 2013 Sep
722 1;15(9):706–12.

723 7. Momtaz R, Ghanem N m., El-Makky N m., Ismail M a. Integrated analysis of
724 SNP, CNV and gene expression data in genetic association studies. Clinical
725 Genetics. 2018;93(3):557–66.

726 8. Taylor C, Crosby I, Yip V, Maguire P, Pirmohamed M, Turner RM. A Review of
727 the Important Role of CYP2D6 in Pharmacogenomics. Genes. 2020
728 Nov;11(11):1295.

729 9. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort
730 Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu.
731 International Journal of Epidemiology. 2015 Aug 1;44(4):1137–47.

732 10. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank:
733 An Open Access Resource for Identifying the Causes of a Wide Range of
734 Complex Diseases of Middle and Old Age. PLOS Medicine. 2015
735 mrt;12(3):e1001779.

736 11. Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for
737 cardiovascular research. European Heart Journal. 2019 Apr 7;40(14):1158–66.

738 12. Minegishi N, Nishijima I, Nobukuni T, Kudo H, Ishida N, Terakawa T, et al.
739 Biobank Establishment and Sample Management in the Tohoku Medical
740 Megabank Project. The Tohoku Journal of Experimental Medicine.
741 2019;248(1):45–55.

13. Mahajan S, Lu Y, Spatz ES, Nasir K, Krumholz HM. Trends and Predictors of Use of Digital Health Technology in the United States. Am J Med. 2021 Jan;134(1):129–34.

14. Shen N, Jankowicz D, Strudwick G. Patient and Family Engagement Approaches for Digital Health Initiatives: Protocol for a Case Study. JMIR Research Protocols. 2021 Jul 21;10(7):e24274.

15. Integraal Zorg Akkoord. 2022 Sep;

16. Sijtsma A, Rienks J, Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. Int J Epidemiol. 2022 Oct 13;51(5):e295–302.

17. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. International Journal of Epidemiology. 2015 Aug 1;44(4):1172–80.

18. Warmink-Perdijk WDB, Peters LL, Tigchelaar EF, Dekens JAM, Jankipersadsing SA, Zhernakova A, et al. Lifelines NEXT: a prospective birth cohort adding the next generation to the three-generation Lifelines cohort study. Eur J Epidemiol. 2020 Feb;35(2):157–68.

19. Warmerdam R, Lanting P, Lifelines Cohort Study, Deelen P, Franke L. Idéfix: identifying accidental sample mix-ups in biobanks using polygenic scores. Bioinformatics. 2022 Feb 15;38(4):1059–66.

762 20. Shah TS, Liu JZ, Floyd JAB, Morris JA, Wirth N, Barrett JC, et al. optiCall: a
763 robust genotype-calling algorithm for rare, low-frequency and common variants.
764 Bioinformatics. 2012 Jun 15;28(12):1598–603.

765 21. Franke L, el Bannoudi H, Jansen DTSL, Kok K, Trynka G, Diogo D, et al.
766 Association analysis of copy numbers of FC-gamma receptor genes for
767 rheumatoid arthritis and other immune-mediated phenotypes. European Journal
768 of Human Genetics. 2016 Feb;24(2):263–70.

769 22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.
770 Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
771 2011;12(85):2825–30.

772 23. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et
773 al. Reference-based phasing using the Haplotype Reference Consortium panel.
774 Nat Genet. 2016 Nov;48(11):1443–8.

775 24. statgen/Minimac4 [Internet]. Center for Statistical Genetics; 2024 [cited 2024 Aug
776 27]. Available from: https://github.com/statgen/Minimac4

777 25. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-
778 coverage whole-genome sequencing of the expanded 1000 Genomes Project
779 cohort including 602 trios. Cell. 2022 Sep 1;185(18):3426-3440.e19.

780 26. Pratt VM, Del Tredici AL, Hachad H, Ji Y, Kalman LV, Scott SA, et al.
781 Recommendations for Clinical CYP2C19 Genotyping Allele Selection. The
782 Journal of Molecular Diagnostics. 2018 May;20(3):269–76.

783 27. Pratt VM, Cavallari LH, Del Tredici AL, Hachad H, Ji Y, Kalman LV, et al.
784 Recommendations for Clinical Warfarin Genotyping Allele Selection. The Journal
785 of Molecular Diagnostics. 2020 Jul;22(7):847–59.

786 28. Pratt VM, Cavallari LH, Del Tredici AL, Gaedigk A, Hachad H, Ji Y, et al.
787 Recommendations for Clinical CYP2D6 Genotyping Allele Selection. The Journal
788 of Molecular Diagnostics. 2021 Sep;23(9):1047–64.

789 29. Pratt VM, Cavallari LH, Fulmer ML, Gaedigk A, Hachad H, Ji Y, et al. *CYP3A4*
790 and *CYP3A5* Genotyping Recommendations. The Journal of Molecular
791 Diagnostics. 2023 Sep 1;25(9):619–29.

792 30. Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant
793 phasing of whole-genome and whole-exome sequencing data in the UK Biobank.
794 Nat Genet. 2023 Jul;55(7):1243–9.

795 31. Kennedy S. Fluent in 3 Months. 2023 [cited 2024 Aug 28]. CEFR Levels: What
796 They Are and How to Test Yourself. Available from:
797 https://www.fluentin3months.com/cefr-levels/

798 32. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging patients in decision-making
799 and behavior change to promote prevention. Stud Health Technol Inform.
800 2017;240:284–302.

801 33. Merker VL, Hyde JK, Herbst A, Solch AK, Mohr DC, Gaj L, et al. Evaluating the
802 Impacts of Patient Engagement on Health Services Research Teams: Lessons
803 from the Veteran Consulting Network. J Gen Intern Med. 2022 Apr;37(Suppl
804 1):33–41.

34. Ministerie van Volksgezondheid, Welzijn en Sport. Stand van zaken inzet DNA-technologie voor zorg en preventie [Internet]. Ministerie van Algemene Zaken; 2023 [cited 2024 Aug 27]. Available from: https://www.rijksoverheid.nl/documenten/kamerstukken/2023/10/25/kamerbrief-over-stand-van-zaken-inzet-dna-technologie-voor-zorg-en-preventie

35. Allison M, Patterson J, Burton S, Patel D, Brassil KJ. Digital engagement preferences of individuals aged 65 and older with cancer. JCO. 2020 Oct 10;38(29_suppl):289–289.

36. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. J Med Internet Res. 2021 Mar 24;23(3):e24387.

Acknowledgements

The authors thank the UMCG Genomics Coordination Center, the UG Center for Information Technology and their sponsors BBMRI-NL & TarGet for storage and compute infrastructure. They thank Maaike van der Lee and Jesse Swen from the department of Clinical Pharmacy and Toxicology of the Leiden University Medical Center for providing the validation samples with known genotypes. The authors thank Certe Groningen for their assistance in genotyping the CYP2D6 gene in a number of cases and the GIC of the KNMP for providing their drug information texts to generate layperson-friendly information. They also thank past and present members of the ELSI-team of the UMCG for valuable feedback on the developed communication materials.

Lifelines Cohort Study

The Lifelines Biobank initiative has been made possible by funding from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG the Netherlands), University of Groningen and the Northern Provinces of the Netherlands. The generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the UMCG Genetics Lifelines Initiative (UGLI). The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

Raul Aguirre-Gamboa (1), Patrick Deelen (1), Lude Franke (1), Jan A Kuivenhoven (2), Esteban A Lopera Maya (1), Ilja M Nolte (3), Serena Sanna (1), Harold Snieder (3), Morris A Swertz (1), Judith M Vonk (3), Cisca Wijmenga (1)

(1) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands

(2) Department of Pediatrics, University of Groningen, University Medical Center Groningen, The Netherlands

(3) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

Funding

This work was supported by SNN (Samenwerkingsverband Noord Nederland) [OPSNN0229 PCH Ecosysteem WP4.4]; by Lifelines NEXT, which has been made possible by funds from a grant from the UMCG Hereditary Metabolic Diseases Fund; by the grants awarded to A.Z. (Netherlands Organization for Scientific Research [NWO] VIDI Grant [016.178.056], a European Research Council [ERC] starting grant [ERC-715772], and Gravitation program EXPOSOME-NL (NWO grant number 024.004.017)), and the Lifelines Biobank initiative, which has been made possible by funds from FES (Fonds Economische Structuurversterking), SNN and REP (Ruimtelijk Economisch Programma).

L.F. is supported by grants from NWO [ZonMW-VIDI 917.14.374, ZonMW-VICI 09150182010019], by an ERC Starting grant [637640 ImmRisk] and through a Senior Investigator grant from the Oncode Institute. P.D. is funded by The Netherlands Organisation for Health Research and Development (ZonMw) [VENI 09150161910057].

Author contributions

PL, BW, MS, NK, JD and LF: conceptualization,

PL, RW, HB, TO, JS, MB, ELM, TOM, PD, JD and LF: methodology,

RW, HB, BC, JS, LJ and LF: software,

PL, RW, HB, TO, JS and MB: validation,

PL, RW, HB, TO, JS, MB, ELM and SJ: formal analysis,

SJ, JGA and DTe: investigation,

LLN, LLCS, PD, NK, JD and LF: resources,

PL, RW, HB, JS, ELM, SJ, BC and LJ: data curation,

PL, RW, HB, TO and JS: manuscript – original draft,

RW and JS: visualization,

BW, DTo, PD, NK, JD and LF: supervision,

PL, JS and JD: project administration,

NK, JD and LF: funding acquisition.

Data availability

Due to privacy reasons the individual-level data cannot be made publicly available but can be made available upon reasonable request. The individual-level data from Lifelines that support the findings in this publication were obtained from the Lifelines biobank under project application number ov21_0355. This request should be directed to the Lifelines Research Office through email (research@lifelines.nl) or by using the application form on their website (https://www.lifelines.nl/researcher/how-to-apply/apply-here). Access to the Lifelines NEXT Project data will be granted to all qualified researchers and will be governed by the provisions laid out in the LLNEXT Data Access Agreement: https://groningenmicrobiome.org/?page_id=2598. This access procedure is in place to ensure that the data is requested solely for research and scientific purposes, in compliance with the informed consent signed by Lifelines NEXT participants. Scripts and other software central to this manuscript is available on GitHub (https://github.com/molgenis/PGx-passport-pilot/ and https://github.com/molgenis/asterix). Other analyses of genotype data and publicly available reference data is performed using standard bioinformatics practices, for which the code is made available upon request.

Conflicts of interest

903 NdBV and DTo are members of the Dutch Pharmacogenetics Working Group

904 (DPWG). All other authors report no conflict of interest.