

Coverage Rate of ADME Genes: Application to CYP2C8, CYP2D6 and CYP3A Genes for Personalized Chloroquine Treatment Against Coronavirus

Nabil Zaid

Universite Mohammed V de Rabat

Loubna Khalki

Mohammed VI University of Health Sciences

Imane Hadri

Universite Cadi Ayyad

Jihane Toughza

Mohammed VI University of Health Sciences

Oussama Badad

Southern Illinois Healthcare

Khalid Sadki

Universite Mohammed V de Rabat

Amine Cheikh

Universite Internationale Abulcasis des Sciences de la Sante

Youness Limami

Universite Internationale Abulcasis des Sciences de la Sante

Younes Zaid (✉ younes_zaid@yahoo.ca)

Universite Internationale Abulcasis des Sciences de la Sante <https://orcid.org/0000-0001-8750-9106>

Primary research

Keywords: ADME genes, Chloroquine, Coronavirus, Covid-19, genotyping, enrichment

Posted Date: April 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-22808/v1>

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

[Read Full License](#)

Abstract

Background: The latest studies have shown the effectiveness of Chloroquine against Coronavirus. However, since the tolerance and effectiveness of statistical data must be taken into account before proposing treatment to a patient, these promising results are often lacking.

Since the CYP2C8, CYP2D6 and CYP3A Absorption, Distribution, Metabolism and Elimination (ADME) genes are involved in the drug response of Chloroquine, we are interested in studying the variations of these genes.

Methods: The purpose of this study is to make a comparison between the various current genotyping and enrichment platforms, to know which of them allows the best coverage.

Conclusions: This will allow us to carry out genome-wide association studies (GWAS) with the aim of finding new therapeutic targets against Coronavirus using Chloroquine.

Background

At this time there is no known specific, effective and proven, pharmacological treatment against Covid-2. Recent publications have brought attention to the possible benefit of Chloroquine, a worldwide used drug, in the treatment of patients infected by the novel emerged coronavirus [1, 2]. Chloroquine has been widely used to treat Malaria, amoebiasis, HIV and autoimmune diseases [3]. Chloroquine is known to block virus infection by increasing endosomal pH and by interfering with terminal glycosylation of the ACE2 receptor which may affect SARS-Cov-2 binding [3, 4]. In clinical trials conducted in China and a nonrandomized trial of COVID-19 patients in France, Chloroquine demonstrated superior clinical efficacy than controls [1, 5].

Additional information on the safety and effectiveness of Chloroquine is essential for doctors to take decisions and save lives in the global battle against COVID-19

The safety and efficacy of medication is of great clinical concern and several teams world-wide are investing concerted efforts toward the identification of the genetic causes of variable drug responses in the hopes of offering genetically-determined personalized therapy. The effective concentration of a drug at its effective site is a key determinant of the drug's safety and efficacy, which depends on the drug's Absorption, Distribution, Metabolism and Elimination (ADME) mechanisms [6]. Genetic variations in ADME genes can lead to large differences in drug exposure between individuals [7, 8]. Unlike other factors, genetic variants remain stable throughout a person's lifetime and provide a valuable means to predict drug response and prevent adverse drug reactions (ADRs) pre-emptively. Pharmacogenetics has achieved impressive progress toward the personalization of pharmaceutical treatment, with over one hundred drugs in the list of the U.S. Food and Drug Administration (FDA) recommendations for genetic testing to ensure drug safety and efficacy [9].

The interaction of these many genes and pathways are very complex, and current commercial platforms don't allow good coverage of the ADME variants [10, 11].

This analysis aims to calculate the coverage rate of CYP2C8, CYP2D6 and CYP3A ADME genes for each of the variants and amplicon lists from our collection. To this end, we used the chromosome and the position of each variant from the ADME lists according to the build Hg19, and by using scripts written in python, we calculated the coverage of our interest lists according to the formula:

$$\% \text{Coverage} = \frac{\text{Interest list} \cap \text{ADME list}}{\text{ADME list}} \times 100$$

We have also considered coverage of the markers of interest that can be achieved by markers that are in linkage disequilibrium (LD). To refine the work and make it more significant, we compared these theoretical results expected according to the targeted coverage of the different technologies with the practical results obtained. The result of this analysis allows us to know which interest lists cover the best ADME genes.

Methods

Data

We have two sets of lists, ADME lists and interest lists (Table 1) which we attend to calculate coverage rates.

Table 1
ADME and interest lists used in this study

ADME lists	Interest lists	
	Genotyping	Enrichment
ADME core (34 variants)	Omni (1S, 2.5S, 2.5 & 5S)	SureSelect V5 (~ 21 000 genes)
	Axiom (~ 85 500 variants)	HaloPlex (~ 21 000 genes)

ADME variants are extracted from genes that were determined to be associated with drug metabolism [12]. The 34 variants used in this work are extracted from CYP2C8, CYP2D6 and CYP3A genes, the main isoforms affected or involved in the metabolism of Chloroquine [13].

We focus in this study on genotyping lists (Omni and Axiom) and enrichment lists (SureSelect and HaloPlex). These platforms are likely to cover ADME genes.

Genotyping platforms

Omni coverage

The Omni lists contain variants obtained *via* the different Illumina genotyping chips [14]. We have 4 Omni lists (1S, 2.5S, 2.5 and 5.0) which were obtained from Illumina's FTP website:

<http://www.illumina.com/forms/ftp.ilmn>. These lists contain 1 million variants for the Omni 1S, 2 million variants for the Omni 2.5S and Omni 2.5, and 5 million variants for the Omni 5.0.

Axiom coverage

Axiom® is a solution designed by Affymetrix for the genotyping of large sample collections such as those screened at biobanks, genome centers, and core labs. The arrays incorporate multiple content categories, including a genomewide association study (GWAS) panel of markers for genomewide coverage in major ethnic groups, rare coding SNPs and indels for exome analysis, pharmacogenomic markers, eQTLs, and newly discovered loss-of-function variants, including sequence insertions and deletions from recent exome sequencing initiatives [15]. We considered using Axiom in our comparison since it includes pharmacogenomic markers.

Enrichment platforms

Sure Select coverage

The list SureSelect is a list of covered polymorphisms from the "Agilent Technologies" product for capture hybridization, which has the following website: <http://www.genomics.agilent.com/>. This list includes 554,751 amplicons likely to cover the variants of each one of the ADME lists.

HaloPlex coverage

HaloPlex technology provides outstanding performance, streamlined workflow, and low sample input requirements for next generation sequencing of human exomes. The HaloPlex Exome has been optimized to provide comprehensive coverage of the coding regions of the human genome [16].

Choice of programming language

Find a programming language to achieve a given project is not an easy task. Almost daily, new languages are created and old ones are updated. Improving the programming languages allows making programs more reliable, faster to develop and easier to maintain.

Among languages references, there is the C language, which dates from the 70 s and is still current. It combines the features of advanced languages associated to features related to assembly languages.

Compared with C, Python is a language relatively slower in terms of execution time. However, if we take into consideration the time required for programming, and biological information processing that is in text form, Python is much better than the C language as it is specifically used for its powerful text processing. Thus, since our project is based largely on the manipulation of files and database (extraction

and processing of information), we opted for the use of Python that facilitates this task compared to the C language.

Linkage disequilibrium

Linkage disequilibrium (LD), the non-random association of alleles from different loci, is often the basis for evaluating the association of genomic variation with human traits among unrelated subjects. If such an association is found between a particular marker locus and the phenotype, it suggests that either the variation at that marker locus affects the phenotype of interest, or that the variation of that marker locus is in LD with the true phenotype-related locus, which was not genotyped [17].

In order to see whether better coverage of the ADME variants of interest could be achieved with the genotyping platforms when taking into consideration the linkage disequilibrium between variants of interest and SNPs on the chips, we have proceeded with LD calculations.

In our case, we are interested only in the variants that can be caught, therefore the variants that are in LD and whose correlation is greater or equal to 0,8 ($r^2 \geq 0.8$), and that, only for the Omni 5.0 and Axiom that allow the best coverage rate for variants of core ADME.

To compute the LD, we used the program PLINK, with the command that allows the extraction of variants of 1000 genomes database that are in LD with variants of core ADME list:

```
plink --bfile mydata --show-tags mysnps.txt
```

Where “mysnps.txt” is a list of IDs of SNPs.

Results And Discussion

To calculate the coverage, we relied on the variants contained in each of the genotyping chips.

The physical coverage rates of the CYP2C8, CYP2D6 and CYP3A ADME genes by genotyping platforms detailed beforehand are recorded (Table 2).

Table 1
ADME and interest lists used in this study

ADME lists	Interest lists	
	Genotyping	Enrichment
ADME core (34 variants)	Omni (1S, 2.5S, 2.5 & 5S)	SureSelect V5 (~21 000 genes)
	Axiom (~85 500 variants)	HaloPlex (~21 000 genes)

Table 2
The physical coverage of CYP2C8, CYP2D6 and CYP3A pharmacogenetic variants by genotyping chips

	Omni 1S	Omni 2,5S	Omni 2,5	Omni 5	Axiome
Physical coverage	1/34 2.94%	4/34 11.76%	5/34 14.71%	8/34 23.53%	18/34 52.94%

As shown in the previous table, from the genotyping technologies, we see that the Axiom and the Omni 5.0 has the best coverage of the Core list.

We took into consideration the coverage that can be achieved by LD with the variants of the ADME lists (Table 3).

Table 3
The coverage of CYP2C8, CYP2D6 and CYP3A pharmacogenetic variants by genotyping chips when considering LD

	Omni 1S	Omni 2.5S	Omni 2.5	Omni 5	Axiome
COVERAGE With Id	2/34 5.88%	14/34 41.18%	10/34 29.41%	20/34 58.82%	30/34 88.24%

As shown in the table, even by taking into consideration the SNPs of Omni platforms which are in LD with the variants of ADME list, the coverage rates increase only slightly, and therefore stays sub-optimal. Concerning the Axiome platform, the coverage taking into account the LD increases significantly up to 88.24%, that is to say a coverage of 30 among the 34 variants of our genes of interest. Even by taking into consideration the markers that are in LD, this list's coverage rate has not reached 100% (Fig. 1).

Moreover, even by combining the five platforms, and taking LD into consideration, we will never be able to cover the two variants rs72549353 and rs72549357 (Table 4).

Genes	rsID	chr:pos	Physical coverage					Coverage with LD				
			Omni 1S	Omni 2.5S	Omni 2.5	Omni 5	Axiome	Omni 1S	Omni 2.5S	Omni 2.5	Omni 5	Axiome
CYP3A5	rs776746	7:99270539	0	0	1	1	1	0	0	1	1	1
CYP3A5	rs10264272	7:99262835	0	0	1	1	1	0	0	1	1	1
CYP3A5	rs41279854	7:99247772	0	0	0	0	1	0	0	0	0	1
CYP3A5	rs41303343	7:99250393	0	0	0	0	1	0	0	0	0	1
CYP3A5	rs55965422	7:99264573	0	1	0	1	1	0	1	0	1	1
CYP3A4	rs2242480	7:99361466	1	1	1	1	1	1	1	1	1	1
CYP3A4	rs4646438	7:99364034	0	0	0	0	0	0	1	0	1	0
CYP3A4	rs55785340	7:99365983	0	0	0	0	1	0	0	0	0	1
CYP3A4	rs67666821	7:99355806	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs1065852	22:42526694	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs1080985	22:42528382	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs3892097	22:42524947	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs5030655	22:42525086	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs5030656	22:42524178	0	0	0	0	0	0	0	1	0	1
CYP2D6	rs5030862	22:42526670	0	1	0	1	0	0	1	0	1	0
CYP2D6	rs5030863	22:42539583	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs5030865	22:42525035	0	0	0	0	0	1	1	1	1	1
CYP2D6	rs5030867	22:42523858	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs28371706	22:42525772	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs28371725	22:42523805	0	0	0	1	1	0	0	0	1	1
CYP2D6	rs35742686	22:42524244	0	0	0	0	1	0	0	0	0	1
CYP2D6	rs72549346	22:42523534	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs72549347	22:42537305	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs72549349	22:42523843	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs72549351	22:42524206	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs72549352	22:42524214	0	0	0	0	0	0	0	0	0	0
CYP2D6	rs72549353	22:42537956	0	0	0	0	0	0	1	0	1	1
CYP2D6	rs72549354	22:42524820	0	0	0	0	0	0	1	1	1	1
CYP2D6	rs72549357	22:42526657	0	0	0	0	0	0	0	0	0	0
CYP2C8	rs1058930	10:96818119	0	0	1	1	1	0	0	1	1	1
CYP2C8	rs10509681	10:96798749	0	1	0	0	1	0	1	0	0	1
CYP2C8	rs11572103	10:96818106	0	0	1	1	1	0	0	1	1	1
CYP2C8	rs72558195	10:96824643	0	0	0	0	0	0	0	1	1	1
CYP2C8	rs72558197	10:96826973	0	0	0	0	0	0	0	1	1	1
Total			1	4	5	8	18	2	14	10	20	30

Table 4
Summary
table of
genotyping
platforms
coverage

For enrichment platforms, we relied on probes contained in these lists.

The coverage rates of the ADME genes by HaloPlex and SureSelect enrichment platforms are recorded in the following summary tables (Tables 5–6).

Table 5
coverage rate of CYP2C8, CYP2D6 and CYP3A pharmacogenetic variants by HaloPlex

Target ID	Regions	Coverage	High Coverage (>= 90%)	Low Coverage (< 90%)
CYP2C8	5	100%	5	0
CYP2D6	20	85%	17	3
CYP3A4	4	75%	3	1
CYP3A5	5	100%	5	0

Table 6
coverage rate of CYP2C8, CYP2D6 and CYP3A pharmacogenetic variants by SureSelect

Target ID	Regions	Coverage	High Coverage (>= 90%)	Low Coverage (< 90%)
CYP2C8	5	100%	5	0
CYP2D6	20	95%	19	1
CYP3A4	4	100%	4	0
CYP3A5	5	100%	5	0

As we can infer, from enrichment lists previously described, the SureSelect platform allows the best coverage of the ADME variants, up to 33 among the 34 variants of our genes of interest (97.05%), which is sufficiently to conduct pharmacogenomics studies with this tool.

Due to technological limitations, complex genomic regions, including certain ADME genes, are generally excluded from high-throughput genotyping and sequencing chips [18].

Although the quality of clustering is generally good for the vast majority of genetic variants present on such commercial platforms, it is not uncommon for clustering to function poorly in highly homologous genomic regions such as those of several ADME genes, including, but not limited to, CYP2C8, CYP2D6, CYP3A4 or CYP3A5 [8].

Previous studies have shown the limitations of genome-wide methods for pharmacogenomic testing. Gamazon *et al.* [19] focused on one set of genes most important in pharmacogenomics and personalized medicine, using only genotyping platforms. Their results demonstrated that even taking into account the SNPs that are in LD, the rate of coverage of these genes by genotyping platforms is sub-optimal.

In another study [7], they also evaluated the sequencing platforms. The HaloPlex enrichment platform enabled the best coverage of ADME variants. But this coverage remains for all of the ADME genes.

There are also chips marketed with targeted pharmacogenetic content, such as the DMET panel (Enzymes and transporters metabolizing metabolics of Affymetrix) or the iPLEX PGx Pro panel from Agena, which provide targeted coverage of the most difficult pharmacogenetic variants. However, these panels have not been examined here since they do not offer genomic coverage, but they could possibly be used as complementary tests in addition to a genomic set in the context of pharmacogenomic studies [20].

Conclusion

Recent research has shown that Chloroquine could be a promising drug against coronavirus.

The work we undertook aimed to find the best platform that can cover the CYP2C8, CYP2D6 and CYP3A genes, the main isoforms affected or involved in the metabolism of Chloroquine.

To do this, we developed scripts in Python to calculate the coverage rate for each of these chips. Using the PLINK tool, we estimated the coverage rate obtained by the different chips by taking into consideration the SNPs markers which do not cover the ADME lists but which are in linkage disequilibrium (LD) with the variants of these lists.

The genotyping Axiom and enrichment SureSelect provided both genome-wide and pharmacogene coverage, which is crucial in the discovering of new variants responsible for drug adverse effects. This combination, which showed the best coverage of the core list, will help in the design of pharmacogenomic studies and will enable to find probably new therapeutic targets in the steps to fight against Coronavirus using Chloroquine treatment.

Declarations

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the first author or corresponding author on reasonable request.

Author contributions

NZ, LK, OB, YL and YZ contributed to conceptualization, data curation and formal analysis of the manuscript. NZ, IH, JT, KS and AC contributed to investigation, methodology and software. All authors contributed to validation, visualization participated in writing original draft and editing of the manuscript.

Data availability statement

All data generated or analyzed during this study are included in this published article.

Declaration of Competing Interest

Authors declare no conflict of interest.

Ethics approval and consent to participate

N/A.

Consent for publication

The contents and publication of the manuscript have been approved by all coauthors.

Funding

No funds received for this study.

References

1. Gao J, Tian Z, Yang X: **Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies.** *Biosci Trends* 2020, **14**(1):72-73.
2. Colson P, Rolain JM, Raoult D: **Chloroquine for the 2019 novel coronavirus SARS-CoV-2.** *Int J Antimicrob Agents* 2020, **55**(3):105923.
3. Savarino A, Boelaert JR, Cassone A, Majori G, Cauda R: **Effects of chloroquine on viral infections: an old drug against today's diseases?** *Lancet Infect Dis* 2003, **3**(11):722-727.
4. Vincent MJ, Bergeron E, Benjannet S, Erickson BR, Rollin PE, Ksiazek TG, Seidah NG, Nichol ST: **Chloroquine is a potent inhibitor of SARS coronavirus infection and spread.** *Virol J* 2005, **2**:69.
5. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE *et al.*: **Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial.** *Int J Antimicrob Agents* 2020:105949.
6. Vogenberg FR, Isaacson Barash C, Pursel M: **Personalized medicine: part 1: evolution and development into theranostics.** *P T* 2010, **35**(10):560-576.
7. Zaid N, Limami Y, Senhaji N, Errafiy N, Khalki L, Bakri Y, Zaid Y, Amzazi S: **Coverage rate of ADME genes from commercial sequencing arrays.** *Medicine (Baltimore)* 2019, **98**(3):e13975.
8. Lemieux Perreault LP, Zaid N, Cameron M, Mongrain I, Dube MP: **Pharmacogenetic content of commercial genome-wide genotyping arrays.** *Pharmacogenomics* 2018, **19**(15):1159-1167.
9. Ventola CL: **Pharmacogenomics in clinical practice: reality and expectations.** *P T* 2011, **36**(7):412-450.
10. Smith DA: **Evolution of ADME science: where else can modeling and simulation contribute?** *Mol Pharm* 2013, **10**(4):1162-1170.
11. Krejsa CM, Horvath D, Rogalski SL, Penzotti JE, Mao B, Barbosa F, Migeon JC: **Predicting ADME properties and side effects: the BioPrint approach.** *Curr Opin Drug Discov Devel* 2003, **6**(4):470-480.

12. Lubomirov R, di Iulio J, Fayet A, Colombo S, Martinez R, Marzolini C, Furrer H, Vernazza P, Calmy A, Cavassini M *et al*: **ADME pharmacogenetics: investigation of the pharmacokinetics of the antiretroviral agent lopinavir coformulated with ritonavir.** *Pharmacogenet Genomics* 2010, **20**(4):217-230.
13. Projean D, Baune B, Farinotti R, Flinois JP, Beaune P, Taburet AM, Ducharme J: **In vitro metabolism of chloroquine: identification of CYP2C8, CYP3A4, and CYP2D6 as the main isoforms catalyzing N-desethylchloroquine formation.** *Drug Metab Dispos* 2003, **31**(6):748-754.
14. Lu JT, Wang Y, Gibbs RA, Yu F: **Characterizing linkage disequilibrium and evaluating imputation power of human genomic insertion-deletion polymorphisms.** *Genome Biol* 2012, **13**(2):R15.
15. Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, Cawley S, Chung E, Connell S, Eshragh J *et al*: **Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array.** *Genomics* 2011, **98**(2):79-89.
16. **Abstracts from the 51(st) European Society of Human Genetics Conference: Posters.** *Eur J Hum Genet* 2019, **27**(Suppl 1):1-688.
17. Devlin B, Risch N: **A comparison of linkage disequilibrium measures for fine-scale mapping.** *Genomics* 1995, **29**(2):311-322.
18. Hovelson DH, Xue Z, Zawistowski M, Ehm MG, Harris EC, Stocker SL, Gross AS, Jang IJ, Ieiri I, Lee JE *et al*: **Characterization of ADME gene variation in 21 populations by exome sequencing.** *Pharmacogenet Genomics* 2017, **27**(3):89-100.
19. Gamazon ER, Skol AD, Perera MA: **The limits of genome-wide methods for pharmacogenomic testing.** *Pharmacogenet Genomics* 2012, **22**(4):261-272.
20. Agapito G, Settino M, Scionti F, Altomare E, Guzzi PH, Tassone P, Tagliaferri P, Cannataro M, Arbitrio M, Di Martino MT: **DMET(TM) Genotyping: Tools for Biomarkers Discovery in the Era of Precision Medicine.** *High Throughput* 2020, **9**(2).

Figures

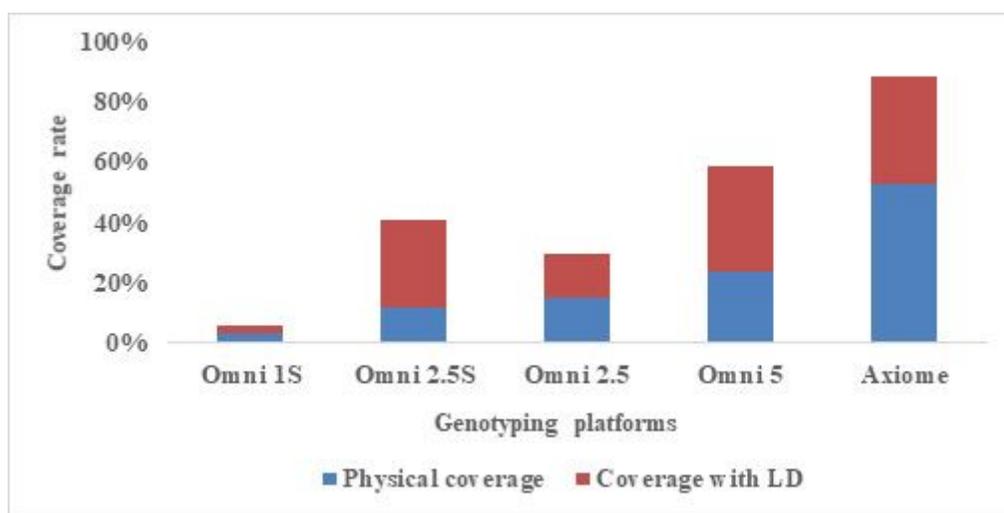


Figure 1

Comparison between physical coverage and coverage with LD

Supplementary Files

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

- [CovLetNZ.doc](#)