Host-pathogen-drug interactions in the context of antibiotic resistance:

How host xenobiotic metabolism can affect antibiotic efficacy in a Methicillin-Resistant *Staphylococcus aureus* infection

Volume 1 of 1.

Submitted by Cyrielle Floriane Tonneau

to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

in December 2018.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature ……

Thesis Abstract

Our arsenal of weapons to fight against bacterial infections is weakening: bacteria are gaining resistance to the common antibiotics, while industries are struggling to develop new effective ones. To avoid triggering *de-novo* antibiotic resistance, we need the right antibiotic for the specific bacteria, at a dose adapted to the patient genetics. Genes driving the degradation of antibiotics have indeed known genetic variants that can dramatically affect the kinetics of antibiotic metabolism from one patient to another. This could lead to treatment failure, excessive side effects or emergence of resistance.

I first investigated the clinical relevance of the vancomycin-rifampicin combination to treat Methicillin-Resistant *Staphylococcus aureus* infections (Chapter 3). I showed in various experimental settings that these two antibiotics may promote an environment prone for antibiotic resistance. Their interaction might be unstable *in vitro* because of environmental factors, one could wonder how the host environment might generate such instability.

I then explored how interactions between antibiotics and host xenobiotic genetics could influence antibiotic concentrations, potentially triggering increased treatment failure, side-effects and antibiotic resistance in patients carrying particular variants. *In silico*, I estimated the effects of genetic variants of the *Cytochrome P450 3A4* gene to its enzyme, and, as they are unequally distributed in the world, their global relevance (Chapter 4). *In vivo*, I focused on the *Carboxylesterase 2* gene and I found two of its variants, rs11075646 and rs8192925, capable of significantly altering the degradation of various drugs, including rifampicin and mycophenolate mofetil. A clinical study was designed, to explore possible correlations between genotype for these variants and treatment response in patients (Chapter 5).

Altogether, this body of work highlights the prescribing importance of considering not only the strain in bacterial infections, but also the genetics of the human host. This raises a need to make sure the right antibiotics are used in practices, at doses adapted to the patients. As part of personalised medicine, checking their genotype for these biomarkers could tailor their therapy, improving recovery while avoiding antibiotic resistance.

Acknowledgements

”*It does not do to dwell on dreams and forget to live*.” - J.K. Rowling.

In one of the numerous workshops I have attended during my PhD, the presenter stated that our PhD theses would probably be the very last writing pieces we would ever submit in our sole names, and that we should celebrate their ownership by using the first person (“I” and “my” - *oh my!*). In the classroom, this idea started a large debate, which grew to also cover the pros and cons of using the active or passive voices. I have to admit I did not participate much to this discussion, I was too busy contemplating where I would be without all of the persons who crossed my path. And the truth is, without any of them, I would not have been in this classroom. And I would not be finally submitting this present thesis neither.

As great writers have written way before me, we are the sum of our experiences and of the people we have ever met. I am no exception to this: from my birth to this very moment, all of my encounters made me who I am. There is no “I” or “my” without them. There is no “I” or “my” without you.

To be sure I would not forget anyone in these acknowledgements, I have briefly considered making an exhaustive list of all of these persons. But then my ecological commitments returned to my mind (*do you know that even numeric theses have a carbon impact*?) and I have made the decision to just focus on the most important encounters of my life.

(In order of appearance.) My parents. My nanny and her husband. My brother. My teachers. My childhood & high school & university friends. Everyone who helped me fall in love with the English language. My professors. My cat. My *Couchsurfing* hosts. My PhD supervisors. Everyone in GP322. Everyone in Team RNA. My fellow PhD students. My friends in Exeter. The *Pokémon Go* community in Exeter. My boyfriend. The *Pokémon Go* community in Toulouse. And you, who is currently reading this.

To all of you, a huge – no, a massive! - thank you for inspiring me, challenging me, helping me, motivating me. There is no “I” or “my” in most of this thesis, only because I was never truly alone.

- Cyrielle.

PS: Just like an Academy Awards winner, I am one of those who never stop listing everything that contributed to their success. And I am mostly thankful for Pumpkin Spice Lattes, cheerful Disney songs, ASMR videos (*Youtube* it!), wine and (a lot of) imported cheese!

Table of contents

List of figures

List of tables

Abbreviations

°	radius
°C	degrees Celcius
µL	microliter
2D	two-dimension
3D	three-dimension
A T C G	nucleobases adenine, thymine, cytosine, guanine
ADME	absorption distribution metabolism excretion
ALB	albumine
ALP	alkanine phosphatase
ALT	alanine transaminase
BIL	bilirubine
BMI	body mass index
bp	base pair
cDNA	complementary DNA
CDP	CCAAT displacement protein
cds	complete DNA sequence
CES2	carboxylesterase 2
CNS	central nervous system
Ct	cycle threshold
CYP (or p450)	cytochrome p450 monooxygenases
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DR	dose-response
E.coli	*Escherichia coli*
ECRF	Exeter clinical research facility
EDTA	ethylenediaminetetraacetic acid
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FICI	fractional inhibitory concentration index
gDNA	genomic DNA
GGTP	gamma-glutamyltransferase
GST	glutathione S-transferase
h	hour

HA	health-acquired
IC50	inhibitory concentration 50
i.e	*id est* (latin), in other words
indel	insertion-deletion
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
LB	Luria-Bertani
LBA	Luria-Bertani agar
LD	linkage disequilibrium
LFT	liver function tests
MAF	minor allele frequency
MH	Mueller-Hinton
MHA	Mueller-Hinton agar
MIC	minimum inhibitory concentration
Min	minute
miRNA	micro RNA
mL	milliliter
mm	millimetre
MMF	mycophenolate mofetil
mRNA	messenger ribonucleic acid
MRSA	methicillin-resistant *Staphylococcus aureus*
NA	not available
NAT	N-acetyltransferase
NC	non-communicated
ng	nanogram
NCBI	National Center for Biotechnology Information
NCLSI	National Clinical and Laboratory Standards Institutes
NIHR	National Institute for Health Research
Nm	nanometer
nt	nucleotide
OD	optical density
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PD	pharmacodynamics
PGx	pharmacogenomics

PK	pharmacokinetics
qRTPCR	quantitative real-time PCR
RD&E	Royal Devon & Exeter
RefSNP	reference single nucleotide polymorphism
RILD	Research Innovation Learning and Development
RNA	ribonucleic acid
rpm	revolutions-per-minute
SA	*Staphylococcus aureus*
sec	second
SNP	single nucleotide polymorphism
SR protein	serine/arginine rich protein
SSTI	skin and soft tissues infections
TF	transcription factor
Tm	melting temperature
UGT	uridine 5’-triphosphate glucuronosyltransferase
UK	United Kindgom
USA	United States of America
UTR	untranslated region
VIP	very important pharmacogene
WBC	white blood count
WHO	world health organisation

Author's declaration of contribution

All chapters were written by Cyrielle Tonneau under the supervision of Robert Beardmore and Lorna Harries. Unless otherwise stated, the results and data presented in this thesis were solely the work of Cyrielle Tonneau.

Chapter 3: The Hill function fitted to the planktonic data was written by Professor Robert Beardmore.

Chapter 5:

DNA samples for the genotyping of Exeter 10,000 cohort were extracted, prior to this study, by the Molecular Genetics laboratory in the Research Innovation Learning and Development (RILD) building (Exeter).

Sanger sequencing was performed by the Diagnostic laboratory in the RILD building (Exeter).

RNA samples for the allelic imbalance of Exeter 10,000 cohort were extracted, prior this study, by Ben Lee (Team RNA, RILD Building).

The clinical study is the result of an on-going collaboration between Cyrielle Tonneau, Professor Lorna W. Harries (Associate Professor in Molecular Genetics), Doctor Marina Morgan (Consultant Microbiologist), Doctor Rebecca Wollerton (Academic Clinical Fellow), Dr Bridget Knight (Nurse Manager at the Exeter Tissue Bank and NIHR Exeter Clinical Research Facility), Lidia Romanczuk (Experimental Medicine Research Officer at the NIHR Exeter Clinical Research Facility), Doctor Thomas Whitehead (Consultant Respiratory Physician), Ruth Unsworth (Respiratory / Tuberculosis Nurse Specialist), Doctor Lucy Smyth (Consultant Renal Physician) and Doctor Richard Haigh (Consultant Rheumatology Physician).

CHAPTER 1 - General introduction

1.1. Antibiotic resistance: causes and consequences

Bacteria are ubiquitous microscopic living organisms which have colonised most of Earth's habitats. Most are commensal to the organisms they are living with/in, including the human beings[1]. Yet pathogenic microorganisms can harm their hosts, causing infectious diseases. These transmissible infections are of various gravity, with a potential to become endemic or pandemic. Antibiotherapies are therefore prescribed to both treat the infected patients and avoid the dissemination of the bacteria in the community.

Contrarily to humans, bacteria can adapt quickly to their environment due to a high mutation rate (estimated to 10^{-6} mutations per base pair per year in bacteria and 10^{-9} in the average mammal)[2,3]. Given that this high genetic diversity occurs in large populations, most colonies comprise clones which will be equipped to survive to the next selective pressure. When infecting a human host being prescribed antibiotics, this bacterial capacity to adaptation allows them to select and propagate beneficial mutations resulting in their antibiotic resistance. Bacteria evolution into resistant strains occurs naturally in the environment[4,5], yet it has accelerated over the last decades due to a imprudent antibiotic use in both healthcare, agriculture and animal farming[6-9]. This strong selective pressure has allowed both the propagation of existing beneficial mutations and the introduction of *de novo* ones[10,11].

In the past, misuses, abuses and overuses of antibiotics used to drive physicians to prescribe another antibiotic from the pharmacy shelf[12], while nowadays a shortage of therapeutic options is predicted[13,14]. Strains are indeed getting more and more resistant to existing antibiotics (Figure 1.1A), whereas very few new antimicrobials are discovered, approved and introduced to clinical practices (Figure 1.1B)[15-18]. 700,000 deaths a year could be attributed to antimicrobial resistance currently, this number being predicted to reach 10,000,000 a year by 2050[19]. Antibiotic resistance is now described as a "slowly emerging disaster"[20], strategies need to be implemented to avoid such fate.

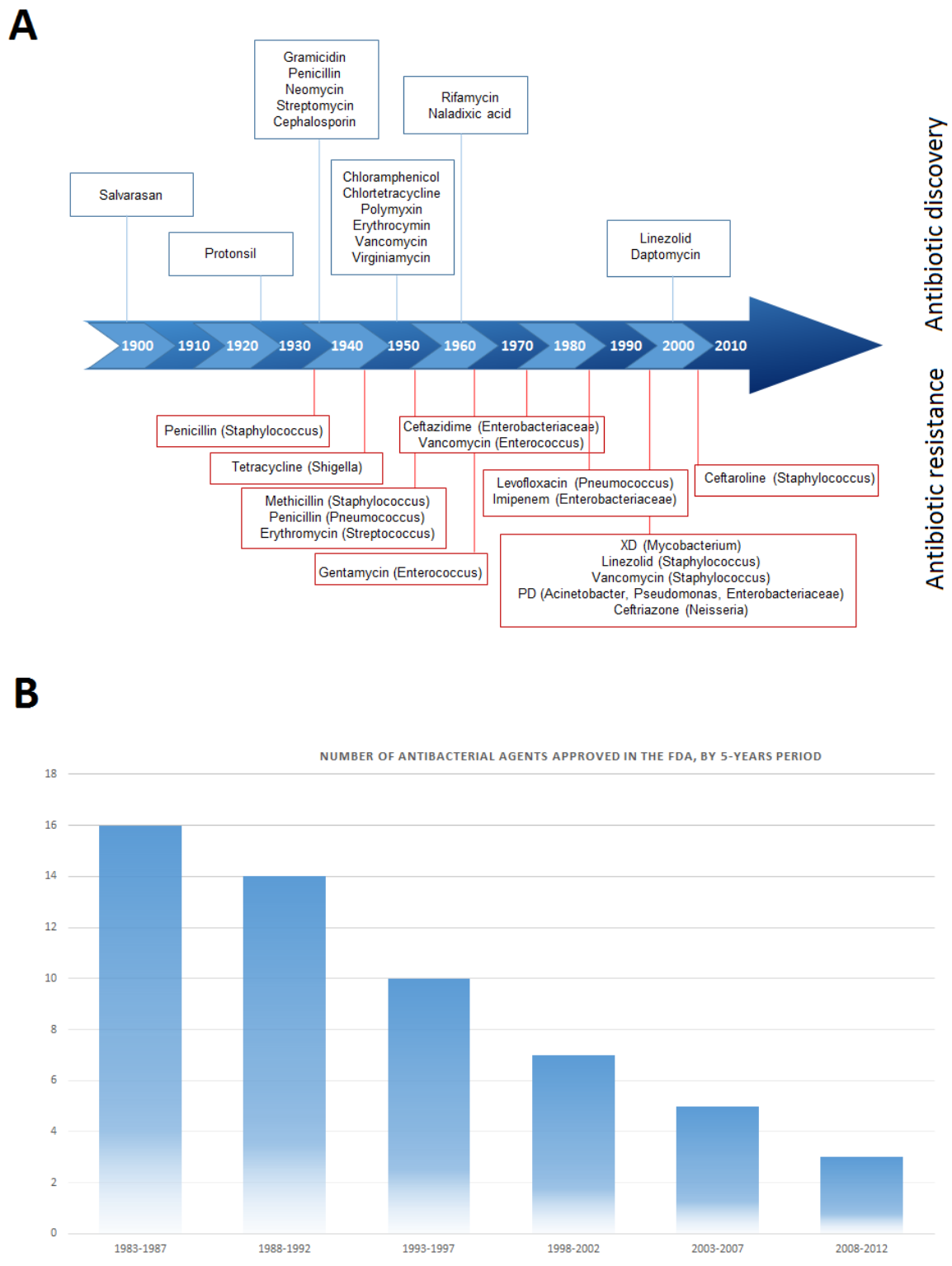


Figure 1.1

A lack of prescribable antibiotics. A – Antibiotic discovery and resistance history, by 10-years period, adapted from [16,18]. B - Recent antimicrobial approvals by the Food and Drug Administration, by 5-years period, adapted from [6].

1.2. Strategies against antibiotic resistance

1.2.1. Developing new antibiotics, and maximising the potency of existing antibiotics

In an attempt to slow down, stop or even reverse antibiotic resistance, either new antibiotics need to be developed, or the efficiency of the already-approved ones need to be optimised.

The high cost and low speed of drug development have led the pharmacology industries to decline most research and development investments in new antibiotics. Only 6 novel antibiotics have been approved since 2015 by the Food and Drug Administration[21], mostly based on already-known mechanisms of action[22,23]. Moreover, the World Health Organisation (WHO) estimated that, nowadays, the useful lifespan of a newly introduced antibiotic is of two years only, due to the rapid emergence and spread of resistances[24]. Research could focus first on discovering novel mechanisms, yet it would take a decade before they turn into distributed new antibiotics[25].

Maximising the efficiency of existing antibiotics is therefore crucial. By regulating their usage in agriculture and farming, this selective pressure could decrease[26]. Furthermore, antibiotic prescriptions to patients should also be dealt with an excessive care[12,17]. In practice this means, firstly, being certain the infection is due to a bacteria before proceeding to any antibiotic prescription and, secondly, prescribing wisely by selecting the appropriate antibiotic(s), at the appropriate dose and with the appropriate length of treatment[27]. Only this would allow the cure of the infection without triggering a *de novo* resistance that will impact on both the next infection outcome and the spread of resistance in the community.

1.2.2. Combination therapy

To minimise the probability of *de novo* resistance emerging when using our current arsenal of drugs, prescribing high doses of a single antibiotic (monotherapy) or a combination of two (or more) antibiotics from different classes targeting different bacterial mechanisms are the main strategies. Yet high doses apply a stronger selective pressure on the bacteria and are often associated to

stronger side-effects in patients[28-30]. Moreover, monotherapies have been reported to increase the risks of antibiotic resistance[31]. Combination therapy is therefore considered as the regimen of choice since the 1970s[32], but only if the combined spectrum of its antibiotics is broaden than their monotherapies' ones, and if the antibiotics are acting in synergy[33-35].

A key concept needs to be introduced when considering antibiotics interaction: the Minimum Inhibitory Concentration (MIC) of an antibiotic. MIC is linked to the antibiotic efficiency in monotherapy, has coexistent definitions and is often associated to concepts such as Mutant Prevention Concentration, Mutant Selection Window or Minimum Bactericidal Concentration[36-40]. Throughout this thesis, the MIC of one drug will be defined as the lowest concentration at which the visible bacterial growth is inhibited *in vitro*. Consistent with this definition, the study of the interaction of two antibiotics, used at equal proportions in a combination, will be performed at their Inhibitory Concentration 50 (IC50), i.e. the concentration at which 50% of the visible bacterial growth is inhibited in monotherapy[41].

For *in vitro* experiments, using both combined drugs at their respective IC50s could seem counter-intuitive compared to those prescribed in clinical settings, where prescribed concentrations are exceeding the set MICs. These clinical high doses are based on the reasoning that, if one bacterium spontaneously develops a beneficial mutation for one drug (due to high selective pressure), the same bacterium would unlikely simultaneously benefit from a resistance to the other drug, targeting a different mechanism. Yet, in patients, prescribed (high) concentrations are not the concentrations found at the site of infections, due to various processes of absorption, distribution, tissue penetration, etc. For example, 88% of the antibiotic rifampicin is bounded to plasma proteins after intravenous intake, making these bounded molecules ineffective[363].

The following definitions will describe the interactions of drugs in combination. A synergy between two antibiotics should result in a significantly better bactericidal effect than the use of either drug in monotherapy (decrease in both killing time and pathogen count). An antagonism between two antibiotics should result in a

significantly worse bactericidal effect than the use of either drug in monotherapy. Additivity and indifference are in-between those two distinct types of interaction, with a combination of additive effect showing the same result as the expected sum of both drugs used together (they act independently and do not interact), and indifference indicating usually that both drugs target the same metabolic pathway – the effect of the combination being the same as the one from the most potent drug used alone[42-45]. The underlying mechanisms of such drug interactions are still to be characterized, potentially using novel techniques of systems biology[46,47].

In the literature, the bacterial growth in a combination therapy with equal proportions of antibiotics at IC50 is usually compared to the bacterial growth in the most effective monotherapy at IC50[41], by using one-day checkerboard assays (in which two-fold dilutions of each drug are delivered in 96- or 384- well plates, one following the x-axis and one the y-axis)[48]. Yet, the MIC of an antibiotic in monotherapy has been reported to depend on environment parameters (such as the pH, or the concentrations in available metabolites or nutrients[24]), it would therefore be inaccurate to assume the bacterial growth in combination is independent from such factors. Moreover, studies have shown that drug interactions are dynamic in nature, observing that synergy could turn into antagonism overtime[24,41]. A synergistic combination is assumed to be beneficial for the patients as it increases antibacterial efficiency without increasing drug toxicity. It is also expected that, if a bacterial clone gains a resistance to one of the antibiotics, the other antibiotic will still active against it - and it could even take advantage of the cost associated to such resistance[24]. Yet, numerous synergistic therapies have been unsuccessful, and in the event those synergies were not artefact, one could wonder if such drug pair could actually apply a stronger selective pressure on the bacteria than their monotherapies[41,49-52]. The use of combinations of antagonist antibiotics have been questioned, yet they are associated to the prescription of high doses of each antibiotic, increasing drug toxicity in patients[53,54].

With such variations in MIC definitions and interpretations, it is not surprising that antibiotics interactions against every particular strain are not well defined in the literature[55,56]. Despite being prescribed for decades, evidences are still missing to fully support the use of several combination therapies used against clinically-relevant infections.

1.2.3. The case of vancomycin and rifampicin against Methicillin-Resistant Staphylococcus aureus infections

Methicillin-Resistant *Staphylococcus aureus* (MRSA) was first observed in 1961 and is currently indicative of *Staphylococcus aureus* (SA) with multiple resistances to antibiotics from the beta-lactam family[57] and cephalosporins[58]. MRSA is also commonly resistant to fluoroquinolones, kanamycin, tobramycin and clindamycin[33]. MRSA being an opportunist pathogen, its staphylococcal infections are of various gravity depending on the site of infections, the patient immunity and the efficiency of the prescribed therapy[59]. Pathogenic MRSA infections frequently spread in hospitals, they are named Healthcare-Acquired (HA) or nosocomial infections, as opposed to Community-Acquired infections[60]. Because of its ubiquity and its multi-resistances, MRSA is estimated to cause 10-20,000 deaths a year in the United States of America (USA)[61 183] - WHO even considers it to be one of the seven bacteria of "great concern"[62].

Numerous therapeutic regimens have been studied to treat (MR)SA infections successfully while trying to avoid the emergence of *de novo* resistances.

Vancomycin (Figure 1.2A) was selected as monotherapy of choice against SA infections in the 1960s[63] thanks to its interaction with the cell-wall cross-linking enzymes of the bacteria, inhibiting therefore a proper cell-wall synthesis[64]. Despite being bactericidal on SA, this "last resort" therapy is associated to several deficiencies (including nephrotoxicity and the emergence of resistance)[53,64-66], its use is therefore recommended in combination with another antibiotic.

Discovered in 1957, rifampicin ("rifampin" in the USA, Figure 1.2B) used to be prescribed as monotherapy to treat bacterial infections, as it can penetrate most

tissues, enter the bacterial cells and inhibit their RNA synthesis[45]. Due to the rapid emergence of resistance and to its hepatotoxicity / nephrotoxicity[45,67], rifampicin is now used in combination with other antibiotics to treat MRSA and *Mycobacterium tuberculosis* infections[45]. It is considered bactericidal towards MRSA rifampicin-susceptible strains, including in biofilms – an asset as this structured environment has been shown to be 100 to 1000 times more resistant to antibiotics than planktonic environments[33,45,68].

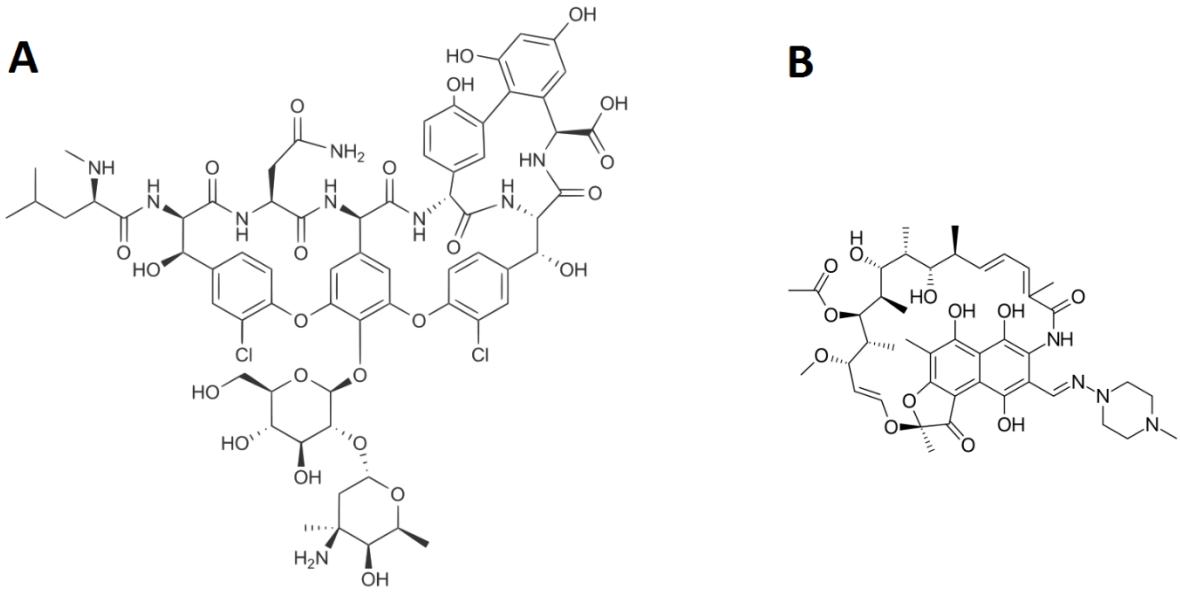


Figure 1.2

Chemical structures of vancomycin ($C_{66}H_{75}Cl_2N_9O_{24}$ **, A) and rifampicin (**$C_{43}H_{58}N_4O_{12}$ **, B).**

Vancomycin and rifampicin having two different bactericidal effects on MRSA (respectively inhibitions of the cell-wall and of the RNA synthesis), their synergistic effect was assumed, leading to the introduction of their combination in clinics in the 1970s[69,70]. In practice, this combination has been very controversial ever since: several *in vitro*, *in vivo* and clinical studies have been conducted, and they concluded on opposite types of interactions of these antibiotics against MRSA (synergy, antagonism, indifference, undetermined) (as seen in Table 1.1), rendering their potential use in combination either advantageous or clearly dangerous for the patients.

Date	Type of study	Observation	Reference
1984	*In vitro* (time-kill or checker-board)	Synergy or antagonism or undetermined (depending on method).	[71]
1985	*In vivo* (rabbit, endocarditis)	No evidence of antagonism.	[72]
1987	*In vitro* (minimal bactericidal concentration[45])	Synergy.	[73]
1991	Clinical (42 patients with endocarditis, prospective, randomised, some flaws[55])	Increased bacteraemia compared to vancomycin alone.	[74]
1993	Clinical (5 neonates with bacteraemia, prospective)	Favourable clinical outcome (but no control group).	[75]
1999	Clinical (14 patients with septicaemia and burns, prospective, some flaws[55])	Favourable clinical outcome (but no control group).	[76]
2007	Clinical (37 patients with osteomyelitis, retrospective)	Various outcomes and emergence of rifampicin resistance.	[77]
2008	Clinical (84 patients with endocarditis, retrospective, some flaws[55])	Increased hepatotoxicity, no improved outcome compared to vancomycin alone.	[78]
2009	*In vitro* (biofilm)	Synergy or indifference (depending on biofilm density).	[79]
2009	Clinical (4 patients, retrospective, some flaws[55])	No clinical success compared to linezolid alone.	[80]
2010	Clinical (83 patients with pneumonia, prospective, randomised, controlled)	Improved clinical outcome and mortality compared to vancomycin alone. Increased in hepatotoxicity and emergence of rifampicin resistance.	[81]

Table 1.1

Representative selection of the *in vitro*, *in vivo* and clinical studies published on the combination of vancomycin and rifampicin against Methicillin-Resistant *Staphylococcus aureus* infections.

The combination of vancomycin and rifampicin against MRSA being still questioned by the scientific community, physicians are reported to follow their local official guidelines, which contain published contradictory recommendations (Table 1.2) [82-84].

Disease or type of infection	USA recommendations[82]	UK recommendations[83]
Common SSTI or Severe SSTI treated outside the hospital	Not recommended.	"unable to make any recommendations on the use of combined therapy"
Severe SSTI treated in the hospital	Only recommended if recurrent and if no effect from changes.	"unable to make any recommendations on the use of combined therapy"
Infective endocarditis on prosthetic valve	Recommended on adults.	Not recommended.
Pneumonia	Unsure.	Not recommended.
Osteomyelitis	Recommended by "some experts".	"recommend the combination of a glycopeptide (vancomycin) alone or in combination with rifampicin" while waiting for new trials.
CNS	Only on adults.	Not recommended.

Table 1.2

Comparison between local guidelines regarding the combination of vancomycin and rifampicin against *Methicillin-Resistant Staphylococc*us aureus infections. Data presented from the guidelines issued for the United States of America (USA) and the United Kingdom (UK) have been chosen to highlight their differences and incertainties. *SSTI stands for Skin and Soft Tissues Infections, CNS for Central Nervous System.*

Deciphering the true underlying interaction of vancomycin and rifampicin in MRSA infections is critical, we need to elucidate why so many inconsistencies have been published. Various *in vitro* and *in vivo* studies seemed flawed (lack of control samples, low sample sizes, etc.), most were non-replicated and could show different results because of differences in strains, techniques, handling methods, etc. Moreover, the published clinical studies were mostly performed on small samples of patients, while sometimes lacking specific data, control groups and/or randomisation[55]. Their discrepancies could also be explained by the different nature of the staphylococcal infections and by differences between patients' immunity. It was notably pointed out that this regimen could not be optimal for all sites of infections, as both antibiotics are not able to penetrate easily the same tissues[33,45]. Furthermore, this combination was deemed more successful when the patients were suffering a "low organism burden"[45]. Inter-individual differences could also play a role in those inconsistencies, as they are currently not considered to adapt the dosages to the patients, and dosages are at the core of any antibiotherapy.

1.3. Considering antibiotic dosing

One key parameter to consider for any type of antibiotherapy is the prescribed dose of antibiotic(s). As previously discussed with the concept of MIC, a precise antibiotic dosage is crucial to inhibit all the bacteria, and yet most of the commonly-used dosages of antibiotics were defined during the antibiotic's golden era (1950-1969), only considering their pharmacokinetics (PK) and pharmacodynamics (PD) parameters[85-87]. At this time, two important paradigms about antibiotherapy were still followed: i) the need to "hit hard"[88] and ii) "one size fits all"[86,87].

Contrarily to what Paul Ehrlich claimed in 1913[88], using a high dosage of antibiotics is not necessarily the best option. As Alexander Fleming said in his Nobel Prize acceptance speech in 1945[89], "If you use penicillin, use enough" – "enough" being now considered as not too little, but not too much either. Not only a very high dosage is indeed usually associated to stronger side-effects in patients, but a lower, if not sublethal, dosage could be best to kill all bacteria and yet apply a weaker selective pressure for resistance[24]. Over the last decades, evidence have accumulated against the "one size fits all" paradigm[86,90]: the major drug dosages are only efficient for up to 60% of the population[91]. Empirical data supports now the need for physicians to slightly adapt their prescriptions to their patients baseline personal data (age [82,92-94], sex [45,95], Body Mass Index[96], concomitant diseases and treatments [97-100], etc.[101]).

Worryingly, antibiotherapy guidelines have not being revised since the antibiotic golden era: most of antibiotics are still delivered at the same non-optimised (mostly high) dosage, indifferently of the patients' characteristics[102]. Furthermore, recent studies reported on the critical effect host genetics (pharmacogenomics, PGx) could have on antibiotic dosage differences between patients.

1.4. The impact of polymorphism of xenobiotic metabolism genes

To treat efficiently patients while avoiding *de novo* antibiotic resistance, it is crucial to use the right antibiotic, at the right dose and for the right length of time[27,103]. Ensuring an appropriate dose of antibiotics is circulating the patients also means considering how their host genetics can alter their circulating concentrations by impacting on xenobiotic metabolism – yet the current prescriptions rarely take it into account.

1.4.1. The xenobiotic metabolism

A.D.M.E. describes the fate of drugs in patients: in the case of antibiotics, after their administration, they reach the blood circulation (Absorption) to be delivered

to the site of infection (Distribution). In the meantime, the blood is filtered in the liver and kidneys, allowing their detection as “xenobiotic” (foreign compounds) to be metabolised and eliminated (Metabolism, Excretion)[86].

Also called detoxification, this metabolism is a set of pathways that detects and modifies external substances, stimulating their excretion from the organism[104]. Enzymes from the xenobiotic metabolism catalyse the conversion of lipophilic compounds into polar / hydrophilic metabolites[105]. This affects the duration and the intensity of the pharmacological activity (bioavailability) of the antibiotics. Thus, it is linked to their PK parameters and impacts on their appropriate dosage[98,104].

This biotransformation happens in every biological tissue, yet it mainly occurs in hepatocytes (liver cells)[101]. It can be divided into three Phases.

1.4.1.1. Phase I

Phase I is associated with a first chemical modification of the foreign substances. Enzymes from this phase allow the solubilisation of lipophilic chemicals through various types of chemical reactions (hydroxylation, reduction, hydrolysis, etc.). These enzymes are of various natures (esterases, epoxide hydrolases, etc.), the main ones being the cytochrome P450 monooxygenases (CYPs, or p450s)[101]. Out of 57 isozymes, CYP1A1/2, CYP1B1, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and predominantly CYP3A4/5/7 are estimated to be responsible for 70-80% of the drug metabolism[106]. These haemoproteins are mainly performing a monooxygenase reaction, e.g. the hydroxylation (an oxidative process) of the compound, as follows:

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+ \ [107].$$

They have a relatively low-specificity and can consequently detoxify most of the existing xenobiotics. They are low-speed catalyser, but are inducible by various compounds (including medicine, and not necessarily those they metabolise). They are particularly concentrated where they will be directly exposed to lipophilic chemicals[101].

1.4.1.2. Phases II and III

Phase II is the conjugation of primary metabolites with polar groups, to complete their solubilisation and decrease their toxicity[104]. It precedes the active transport of the hydrophilic metabolites out of the cells (Phase III)[105].

The main Phase II enzymes are Uridine 5'-triphosphate GlucuronosylTransferases (UGTs), Glutathione S-Transferases (GSTs) and N-AcetylTransferases (NATs)[108]. As for Phase I enzymes, they exist into different isozymes, for example GSTM1, GSTT1 and GSTP1 are well-known amongst the GSTs[105].

1.4.1.3. Characteristics

Understanding of xenobiotic metabolism has facilitated the development of optimised medical treatments. The fate of most drugs has been characterised as being metabolised in the liver, their metabolites elimination occurring in the bile, in order to reach the intestinal lumen and the faeces. They can also be redirected to the kidneys via the bloodstream and be processed in the urines[105]. This has highlighted the need to ensure liver and kidneys functions in patients before treatment, and to monitor closely the patients affected by those organs' dysfunction. Moreover, some drugs have been developed as pro-drugs: they will be activated by the host xenobiotic metabolism, but will still be water-soluble and readily excreted[105,109], as opposed to the usually bioactive antibiotics losing their (antibacterial for antibiotics) effect after Phase I metabolism. It is also important to note that most medications are degraded by the xenobiotic metabolism, the involvement of zero, one, two or all three phases depending on the drug chemical structure / polarity[105]. A variety of different metabolites are produced for each drug, depending on the metabolic enzymes in charge - emphasizing on their broad spectrum[105].

As the xenobiotic metabolism can alter the bioavailability of the drugs (including antibiotics), considering it when prescribing a drug is vital.

1.4.2. Polymorphism at the xenobiotic metabolism genes

Most of the Phase I and Phase II enzymes are encoded by highly polymorphic genes: they present numerous sequence variations, with allele frequencies above 1%, leading to the expression of various enzymatic isoforms[98,101,106,110,111]. These genetic variations can be of various sizes, and the most frequent are single nucleotide substitutions, insertions and deletions - usually referred altogether as Single Nucleotide Polymorphisms (SNPs). Despite involving only one nucleotide, SNPs can affect the xenobiotic metabolism gene expression and/or enzyme activity. Depending on their localisation in the genome and the type of variation occurring, they can for example alter the protein sequence (leading to a dysfunctional enzyme), the gene transcription or the gene stability (leading to differences in the enzymatic concentrations)[112]. All these alterations could promote differences in drug metabolism. Polymorphisms in the xenobiotic metabolism genes are common occurrence of clinical relevance[101,108]. Illustrating this, the intronic *CYP1A2* SNP rs762551 is found in 35% of the world population and associated to an increased oral clearance for olanzapine. Similarly, 26% of the world population harbours *CYP2B6* rs2279343 (codon change), reporting an increased drug clearance for bupropion, efavirenz and cyclophosphamide. Likewise, the null allele resulting from a splicing defect due to *CYP2C19* rs4244285 was associated to a lack of drug clearance, particularly regarding antidepressants, antimalarials and antifungals - it is present in 19% of the world population.[101]

Considering antibiotics which are mostly delivered as active drugs, a lower (or slower) xenobiotic metabolism in “Poor Metaboliser” patients could induce an accumulation of the active antibiotic in their organism. This could increase their experience of side-effects and lead to drug overdose. In case their xenobiotic metabolism is higher (or quicker) than those from the average patient, a higher amount of antibiotic would be metabolised, lowering its bioavailability and impacting therefore on treatment efficacy in these “Extensive” or “Ultrarapid Metabolisers”. Median impacts have also been reported in patients considered as “Intermediate Metabolisers” [91,113]. Overall, by impacting on the antibiotic PK/PD parameters, polymorphisms at the xenobiotic metabolism genes could alter the exact concentration required to inhibit all bacteria while avoiding the emergence of antibiotic resistance[27].

The effect of genetic polymorphism on genes coding for Phase I CYPs is particularly well documented: 20-25% of patients are estimated to be clinically affected by genetic variations on their copies of the *CYP3A4/5, CYP2C9, CYP2C19* and *CYP2D6* genes[113]. These genes, known to metabolise more than 50 clinically-relevant drugs, have indeed been reported to present at least 40 different SNPs - each. For example, due to their *CYP2C9* genotype, Poor Metabolisers for the cardiovascular drug warfarin require a critical decrease in their drug dosage: up to 75% decrease in some patients[114].

Genes coding for Phase II enzymes have also being studied regarding the impact of their polymorphism. Responsible for up to 20% of clinically-relevant drug Phase II metabolism[106], the GST family contains common SNPs with critical effect: partial or entire gene deletion mutations are present in patients for the *GSTM1* and *GSTT1* genes, resulting in non-functional null alleles[115,116]. The null allele for *GSTT1* has for example being associated with an increased likelihood of toxicity in tuberculous patients treated with isoniazid, pyrazinamide and rifampicin[117].

These polymorphisms at the xenobiotic metabolism genes have mostly been studied in the context of cancer risk (the GSTT1 enzyme – for example – is involved in the elimination of carcinogenic compounds) or drug-drug interactions. Numerous evidences support indeed that one particular drug can stimulate (or inhibit) the expression of one particular enzymatic gene which is promoting therefore the quicker (or slower) degradation of another drug, reducing (or increasing) its bioavailability[98,101,106,110,118-120].

Ethnic differences have been reported for such polymorphisms. As documented with their continental Minor Allele Frequencies (MAF), some xenobiotic metabolism SNPs are indeed differently distributed around the world[110,118,121,122]. It is notably the case of many of the deleterious *CYP2D6* SNPs which have been found mostly in Caucasian populations: if 9% of the world population express the *CYP2D6*4* allele (harbouring rs3892097), resulting in the absence of detectable

protein in the liver, it was found that up to 25% of Caucasians are affected by such SNP[101,123]. Similarly, and to name a few, *CYP1A1* rs1048943 was found in 20-26% of the Asian populations but mostly absent in descents from Africa. *CYP1A2* rs2069514 is also dominantly expressed in African ethnic groups compared to other populations. 16-27% of the Asian populations express *CYP2A6* rs28399433, contrarily to African (4-12%) or European (4-5%) descents.[101]

Likewise, the complete *GSTM1* or *GSTT1* gene deletion has even been found in up to 50% of the patients in populations from European descent[115]. One study reported also a large and differential distribution of *NAT2* gene polymorphisms at global and micro-geographic scales in ethnic groups[124].

1.4.3. The case of vancomycin, rifampicin and Mycophenolate MoFetil metabolism

The impact of polymorphism at the xenobiotic metabolism genes may be well-characterized when it involves the CYP enzymes, yet not all clinically-relevant drugs are metabolised by these enzymes. Let us consider again the combination of vancomycin and rifampicin against MRSA infections.

Vancomycin is known to be directly poorly metabolised[125]. Readily reabsorbed, it is excreted from the blood through the kidneys[126]. Vancomycin clearance can still be influenced, not by polymorphisms at its metabolism by renal dysfunctions[76,126,127]. In opposition, rifampicin has been extensively studied because of its induction of the xenobiotic metabolism, involving it into drug-drug interactions with various compounds. After being administrated, it reaches the blood circulation[45,128] where 80% of the drug interacts with proteins - reducing its half-life. In the liver, rifampicin is degraded into its metabolites (mainly 25-desacetyl-rifampicin), which will then be excreted through the urine[45,129].

For decades, rifampicin has been described by the scientific community as an “auto-inducer of its own metabolism”, and it is believed to both i) induce the expression of the *CYP* genes and ii) be degraded by the CYPs[33,45]. The metabolism of rifampicin by the CYPs was actually never established: to the best

of our knowledge, no experimental proof has been published. Citations usually mention the work of Gianni Acocella who reported in 1972, simultaneously with other studies, that rifampicin induces its own metabolism (by reducing its own half-life)[130-132] – but the specific role of CYPs was never mentioned through his career[133-135]. Other studies describe actually that rifampicin is submitted to a deacetylation reaction (hydrolyse) in the liver[136] and that it increases its own rate of deacetylation[137]. This suggests rifampicin metabolism by the xenobiotic metabolism Phase I beta-esterases, and particularly the carboxylesterase 2 (CES2, former nomenclature: “hCE-2” for “human cocaine esterase 2”). CES1 and CES2 are indeed contributing to 80% of the hydrolytic activity in the liver[105], and evidences support both the induction of *CES2* by rifampicin and the metabolism of rifampicin by CES2[33,105,129,138-140].

With vancomycin being poorly metabolised in the organism and the metabolism of rifampicin being performed by the CES2 enzyme, polymorphisms at the *CES2* gene could impact on rifampicin concentrations and therefore on its interaction with vancomycin to treat MRSA infections. As previously mentioned, such polymorphisms in other xenobiotic metabolism genes have been associated to both clinical impact and ethnic differences. When it comes to *CES2*, its polymorphic nature was reported, and epidemiology studies have shown that the specific incidence of MRSA bacteraemia was associated with ethnicity[141] – one could wonder if those ethnic variations may be due to SNPs in *CES2*. Polymorphism at the *CES2* gene has been partly investigated in South-Korean tuberculosis patients, focussing first on 10 SNPs found in these patients. Thousands of SNPs have been reported on the *CES2* gene, including 70 being registered as “common”, i.e. found in at least 1% of the world population[111,142-148]. Studying those *CES2* SNPs could perhaps provide another explanation on the inconsistencies found in the literature about the combination of vancomycin and rifampicin towards MRSA infections.

Mycophenolate Mofetil (MMF) is yet another of the clinically-relevant drugs metabolised by CES2. This pro-immunosuppressor is prescribed notably to kidney transplant patients, to weaken their immunity and therefore reduce their rejection risks. Clinicians have witnessed unexplained different response to this

treatment between their patients[149-152]. One could wonder if their polymorphisms at the *CES2* gene could be associated to such differences, affecting the MMF concentrations in patients. It is particularly important to elucidate this due to the main adverse effect associated to MMF: a severe diarrhoea, causing a life-threatening dehydration. A rapid MMF withdrawal is then necessary, followed by the prescription of its primary metabolite (mycophenolic acid or salt mycophenolate sodium). In case these inter-patient differences are actually the result from polymorphisms in *CES2*, and given how critical transplant success is, developing a "predictive test" associating patients genotypes to the likelihood of MMF treatment success could be very useful in clinics.

1.5. Thesis aims

Given the decline in new antibiotic discovery and the rise of antibiotic resistance[6], it is crucial to use the right antibiotic, at the right dose and for the right length of time in order to treat efficiently patients while avoiding *de novo* antibiotic resistance[27,103]. Antibiotic prescription should be dealt wisely and consider the interactions between i) the bacteria and the antibiotic, ii) the antibiotic and the host, and iii) the bacteria and the host. This PhD thesis aims to expand our knowledge on the first two of these interactions, particularly regarding the clinically-relevant antibiotics prescribed to treat MRSA infections.

Firstly, Chapter 3 of this thesis will focus on the combination of vancomycin and rifampicin against MRSA infections. Given their importance and severity, making sure this regimen is effective, i.e. vancomycin and rifampicin synergise against MRSA, is key. Numerous contradictions have been published on this topic[55,82,83], and several hypotheses have been drawn to explain such discrepancies: for *in vitro* studies, differences in laboratory environment factors and techniques are questioned. Here, one operator will study their combination in various experimental settings, in the same laboratory and using the same materials. Chapter 3 will show that the interaction of vancomycin and rifampicin is antagonistic in planktonic and biofilm environments, and either indifferent or undetermined in more structured environments and using clinically-relevant

Etests. Overall, no data from these experiments was found to support the synergy of vancomycin and rifampicin on MRSA, and the impact of monotherapy MIC variations is discussed.

Secondly, Chapters 4-5 will investigate the impact host genetics can have on differential antibiotic bioavailability between patients.

Chapter 4 of this thesis will concentrate on the gene polymorphisms in *CYP3A4*[105]. Despite being associated to the metabolism of up to 30% of clinically-relevant drugs, rare are the common *CYP3A4* variants to have been both characterized and associated to an effect on the gene expression or enzyme activity. Using the available literature, open access data and *in silico* tools, the most common *CYP3A4* SNPs will be identified, and their potential importance on the *CYP3A4* gene expression or enzyme activity will be evaluated. Overall, this chapter will predict the importance of eight *CYP3A4* SNPs (rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116), only two of which being already reported in the literature. These eight SNPs will be classified according to their predicted potential impact score and their distribution in the world and in continental populations will be discussed.

The next chapter will focus on the overlooked rifampicin metabolism by CES2[138,139], as this enzyme is responsible for the metabolism of rifampicin and MMF – both drugs being associated to variable pharmacokinetics between patients. Firstly, using open access data and *in silico* tools, three of the most common - yet poorly characterised - *CES2* SNPs (rs11075646, rs8192925 and rs28382828) will be identified as potentially important for the *CES2* gene expression or enzyme activity. Their impact will then be investigated using an allelic balance assay, showing that the presence of the minor variant for rs11075646 and/or rs8192925 are associated to significant differences at the *CES2* mRNA levels. Finally, this chapter will assess these associations with a pilot clinical study. Patients treated with rifampicin or MMF will be genotyped for rs11075646 and rs8192925 to detect if there is an association between their genotype and their phenotypic response to their CES2-metabolised treatment. The preliminary results of this pilot study will be discussed.

Overall, in Chapters 4-5, a total of 10 poorly characterised xenobiotic metabolism SNPs is associated to predicted *CYP3A4* or *in vitro* significant *CES2* mRNA alterations. These *in silico* and *in vitro* predictions need to be validated, yet they are novel characterisations with the potential to explain, perhaps, inter-patients' variability in the kinetics of CYP3A4-metabolised cardiovascular simvastatin[153] and immunosuppressant tacrolimus[154], or in the kinetics of CES2-metabolised antibiotic rifampicin and pro-immunosuppressant MMF.

Overall, the scientific community really needs to investigate the xenobiotic metabolism of every clinically relevant drug, as well as the genetic polymorphisms affecting the enzymes responsible for the metabolism of these drugs. They could have a critical role in some patients, with potential ethnical divergences due to the differential distribution of those genetic backgrounds. With the rise of molecular biology and genetics, and particularly of the rapid and inexpensive genome sequencing, this improved knowledge could help develop personalized medicine[90,106,120], taking this concept to the next level: tailoring the treatment of diseases to the genotype of the patients. In the context of antibiotics, this could potentially explain some contradictions published in clinical studies, and this could most likely being useful to help us select the appropriate antibiotic doses to treat the patients effectively while avoiding the emergence of antibiotic resistance.

In each of the Chapters 3-5, a different research question will be addressed and presented as a manuscript containing an introduction, followed by the methods, the results and the discussion sections. A reflection on the research was added afterwards to explain how allowing more time to these projects could have helped solve research obstacles.

CHAPTER 2 - Materials and Methods

The following outlines generic concepts and techniques included in Chapters 3-5.

2.1. Antibiotics and bacteria

2.1.1. Definitions of regimens

Throughout this thesis, the following terms and definitions are being used.

Patients are being delivered a treatment by their physicians. In the case of antibiotics, the treatment is an antibiotherapy. If only one antibiotic is administrated during the treatment, we talk about “monotherapy”. Generally speaking, a combination therapy (or combined therapy) involves the presence of two (or more) antibiotics in the treatment, either used in mixture (simultaneously), or in sequence (for example a 2-weeks treatment consisting in Antibiotic A on the first week, and Antibiotic B alone for the second week) or in cycling (for example a treatment in which the two administrated antibiotics are switched every other day) [155,156].

In this body of work, combination therapy will refer only at a mixture of antibiotics used simultaneously. By extension, such terms and definitions will be applied to the laboratory infection control methods used in Chapter 3.

2.1.2. Concept of combination therapy

As monotherapies are often associated with the development of antibiotic resistance[31], the addition of at least one antibiotic to the prescription is recommended. It should also benefit even more the patient who will be dispensed two (or more) antibiotics at lower doses, decreasing therefore their risk in experienced side-effects. A combination of two antibiotics is deemed effective when they are in synergy against one bacterial strain - their concomitant use leads to a significantly better clearance of the bacteria than when used alone, as opposed to antagonism (significantly worse clearance), additivity or indifference (approximately the same clearance ratio)[41-45].

To assess the efficiency of one antibiotic monotherapy, the bacterial growth is evaluated in several increasing concentrations of said antibiotic. These dose-responses allow the estimation of the Minimum Inhibitory Concentration (MIC), the concentration at which 100% of the bacterial growth is inhibited. Guidelines from National Clinical and Laboratory Standards Institutes (NCLSI) recommend the use of two-fold dilution method, yet this tends to overestimate the MIC value, creating a stronger selective pressure on bacteria and/or stronger side-effects[24]. In this body of work, the increasing concentrations tested were therefore within the range of sub-inhibitory doses.

The estimation of the MIC allows the estimation of the Inhibitory Concentrations (IC) 50, 80, 90 and 99 (for example) which are the smallest concentrations that inhibit respectively 50%, 80%, 90% and 99% of the bacterial growth[41].

To visually assess the type of interaction between two antibiotics (A and B), we could plot at time T the bacterial growth (d) as a function of the drug combination delta factor or relative drug fraction (θ), i.e. the ratio of each antibiotic in the combination. Figure 2.1 shows an example where IC90A and IC90B are used as basal concentrations for each antibiotic. Θ represents then any combination along the equidosage line: each value of θ is therefore simultaneously associated to a drug A concentration of θ*IC90A and to a drug B concentration of (1 - θ)*IC90B. By fixing those basal concentrations of A and B, θ is the only variable parameter allowing the determination of the optimal antibiotherapy (the one associated to the lowest d). This ensures a fair comparison between regimens, prohibiting the use of arbitrarily high doses which could lead, over time, to antibiotic resistance and/or excessive side-effects[41].

As developed in the literature, the nature of the interaction between two antibiotics can be defined compared to a neutral interaction (additivity or indifference) using the Loewe drug interaction profile i(θ). i(θ) equals to d(T;θA,(1−θ)B) and, in the case of synergy - when the combination has more effect than the sum of the separate effects of each antibiotic: $i(\theta) < \theta * i(A) + (1-\theta) * i(B)$, with i(θ) being therefore a convex function of θ. For

antagonism, $i(\theta) > \theta * i(A) + (1-\theta) * i(B)$, with $i(\theta)$ being a concave function of θ. If $i(\theta)$ is a constant, it is independent of θ and the drug interaction is defined as additive – this would be visualised as a straight line on Figure 2.1[41]. In other words, on Figure 2.1, in the case of a synergy, more bacteria density is expected at the extremes of θ than in a 50%/50% combination, the plot is expected to show a smile-like feature (in blue on Figure 2.1). For antagonism, more bacterial growth is expected when both drugs are provided together than alone, creating a frown-like feature (in red on Figure 2.1). As previously reported, the smile configuration could lead to a frown over time due to greater selective pressure at 50%-50% of the drugs than in monotherapies, it is therefore important to assess the evolution of the combination of A and B over time[41].

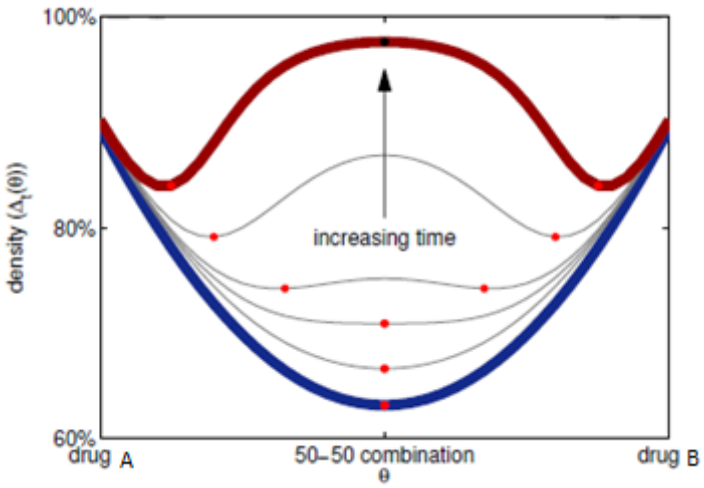


Figure 2.1

Illustration of a smile configuration when two drugs, A and B, synergise (blue bold line) and of a frown configuration when both drugs antagonise (red line). Interestingly, both conformations can be found over time in the same environment - the type of interaction between A and B is not necessarily stable. In this example, the drug combination delta factor or relative drug fraction θ is depicted as a function of the bacterial density ($\Delta t(\theta)$). *Adapted from* [41].

2.1.3. Optical density

For the planktonic experiments performed in *Chapter 3*, optical density (OD, with OD = Absorbance / optical pass Length) was read with a spectrophotometer at 600nm, as a proxy for biomass, or bacterial density (bacterial growth or inhibition, when compared to the negative and positive controls). This method is recognized as particularly effective as OD correlates well with live cell counts (colony forming units)[155,157]. OD may be biased by cell sizes and non-viable cells, yet this method was selected because it allows a minute-by-minute estimation of bacterial density for a large number of replicates and culture conditions, tested at the same time on the same microplate.

2.2. *In silico* predictions on the effect of a genetic variant on gene expression or enzyme activity

2.2.1. Genetic variants affecting gene expression regulation

Various mechanisms are known to regulate the gene expression, through transcriptional initiation, alternative splicing, gene stability, etc. In brief, *trans*-regulatory factors, such as transcription factors, can activate the gene transcription by binding to *cis*-regulatory elements (transcription binding sites). Intronic and exonic splicing enhancers and silencers (ISE, ISS, ESE, ESS) can modulate the alternative splicing through their binding on *cis*-acting RNA sequence elements in pre-messenger RNA (mRNA), depending on the 5'- and 3'- splice sites (donor and acceptor sites) and branch sites. MicroRNA (miRNA) can recognize motifs on the mRNA, stimulating their degradation (gene silencing)[158-161].

Variants expressed on one gene can affect the regulatory mechanisms of its expression. They can modify the DNA (and RNA) sequences, disrupting or creating motifs recognized by transcription factors (mostly in 5'-UnTranslated Regions (5'-UTRs)), ISS/ISE/ESE/ESS (unlikely in UTRs), miRNA (mostly in 3'-UTRs). When located in the vicinity of such DNA or RNA regions, by modifying the DNA or RNA conformation, they can impact indirectly on the binding on

regulatory elements. Variants can also directly modify the donor and acceptor splice sites, as well as the branch sites in alternative splicing[159-170].

Genetic variants can also have a functional impact on gene expression or enzyme activity by modifying translated exonic sequences (substitution leading to a missense or non-sense, or insertion/deletion of a few nucleotides potentially causing frameshifts). Synonymous substitutions have also been reported to impact on the protein secondary structure. Other types of impacting genetic variants include variations affecting epigenetic marks (methylation, histone marks)[168,171-174].

2.2.2. Linkage (dis)equilibrium

When studying a number of human genetic variants *in silico*, it is important to consider their linkage. Linkage equilibrium is the independent occurrence of the alleles of two different genetic variants: their (rare) concomitant inheritance is considered to be due to randomness. Contrarily, variants in linkage disequilibrium are considered linked, their alleles commonly occur together, non-randomly. This is mostly due to their physical proximity on gene sequence, which impacts on the frequency of recombination events[175,176].

To measure the linkage (dis)equilibrium of two variants, D' and R^2 are calculated, given the known distribution of haplotypes in a selected population. D' is considered as an indicator of allelic segregation, with value ranging from 0 to 1 (the closer to 1, the tighter the linkage). D' = 1 indicates that at least one expected haplotype combination is not observed. R^2 is more sensitive to allele frequency in populations. It measures the correlation of alleles, with value ranging from 0 to 1 (the closer to 0, the more independent the alleles).

2.2.3. Estimation of the distribution of genetic variants in different ethnic groups

To estimate the distribution of selected genetic variants in different ethnic groups, data from phase 3 of the 1000Genome Project was extracted[177,178]. This project is considered the largest catalogue of human variation data. It includes the genotype of such variants in 2,504 people from ancestry populations of known and detailed ethnic background: Africa (661 individuals), America (347), East Asia (504), South Asia (489) and Europe (503).

2.3. Molecular biology techniques

2.3.1. Polymerase Chain Reaction

Most of the molecular biology techniques in *Chapter 5* are based on the Polymerase Chain Reaction (PCR) technique that allows the amplification of a selected sequence of a DNA template in a thermal cycler. A typical PCR is carried out in a reaction buffer containing the DNA template to synthesize, deoxyribose Nucleoside TriPhosphates (dNTPs), a specific forward and reverse couple of oligonucleotides and the Taq DNA polymerase enzyme, which creates a complimentary copy of the template within the target region.

Synthesis accuracy is often linked to the nature of the Taq DNA polymerase itself, but also to the parameters of the reaction in the thermal cycler:

	Time	Temperature
Step 1	15 min	95°C
Step 2	30 sec	94°C (denaturation)
Step 3	1 min	60°C (annealing)
Step 4	1 min	72°C (synthesis)
Step 5	*Repeat steps 2-4 34 times*	
Step 6	Forever	4°C

Table 2.1

Example of the cycles of a conventional PCR. *Min and sec stand for minutes and seconds, respectively.*

By varying the concentrations inside the PCR mix, by adapting the cycling parameters (time, temperature, number of repeats, etc.) and by adding, when necessary, DiMethyl SulfOxide (DMSO) and/or Mg2+ (which makes the DNA more labile for heat denaturation) to the reaction mix, it is possible to optimise a PCR in order to improve synthesis accuracy and to obtain the required amount of PCR product.

Real-time PCR (RTPCR) is a commonly used technique that can be used quantitatively (qRTPCR). It consists in the monitoring, in real-time, of the amplification during the PCR, thanks to the use of sequence-specific DNA probes (oligonucleotides labelled with a fluorescent reporter activated by the probe hybridization with its complementary sequence). By using two sets of probes, each recognizing the major or the minor allele of a genetic variant, and each associated to a different fluorescent reporter, it is possible to quantify the PCR products associated to each allele, over time. A cycle threshold (Ct) value is attributed to each probe: it is the number of PCR cycles required for the probe

fluorescence to cross the threshold / background level. A high Ct is therefore correlated to a low amount of target nucleic acids in the sample.

2.3.2. Allelic imbalance assay

Allelic imbalance assay, using (separately) genomic and complementary DNA (gDNA and cDNA) samples from individuals heterozygous for one genetic variant, is a powerful tool to quantity the gene expression (mRNA), in each sample, associated to each allele (major or minor allele). It allows to distinguish differences in gene expression due to the presence of the major or the minor allele. Because both alleles are extracted in a single sample from one single individual, and because the measurements are performed independently for each type of DNA and for each individual, this method is considered insensitive to differences in environmental factors or extraction efficiency. No “control” genes are required[179].

Allelic imbalance assay can be performed using a qRTPCR and the Ct values are used to quantify the mRNA levels associated to each variant as following, for each sample x:

DCt(gDNA,x) = Ct (major,x) – Ct (minor,x) using the gDNA values

DCt(cDNA,x) = Ct (major,x) – Ct (minor,x) using the cDNA values

DDCt(cDNA,x) = DCt(gDNA,x) - DCt(cDNA,x)

DDCt(gDNA,x) = DCt(gDNA,x) - median(DCt(cDNA))

gDNA allelic ratio = $2^{0-DDCt(gDNA,x)}$ *to normalize the gDNA value to 1 (the gDNA values are expected to follow a 1:1 ratio)*

cDNA allelic ratio = $2^{0-DDCt(cDNA,x)}$ *to consider the logarithmic nature of PCR.*

gDNA allelic ratios and cDNA allelic ratios are pooled (separately) and compared: significant divergence in cDNA allelic ratio from 1:1 is correlated to an allelic imbalance[180].

2.4. Statistics

As a general rule and following the general usage in health science, throughout this thesis, a p-value of 0.05 was used as the cut-off for significance. Data were checked beforehand for normality and parametric state. Used tests includes t-test, Mann-Whitney test, one-sample variance comparison, two-samples t-test, Levene's test, Shapiro-Wilk test, Kruskal-Wallis, etc. A Hill equation was also used to fit the data to an expected distribution, and to deduce their correspondence (R-square). The selection of the appropriate statistical tests to determine significance is detailed throughout the chapters.

CHAPTER 3 - Interaction between vancomycin and rifampicin towards Methicillin-Resistant *Staphylococcus aureus*, in various experimental settings.

3.1. Overview

The purpose of this chapter is to understand the extent to which antibiotics vancomycin and rifampicin synergise - particularly in the context of the clinically-relevant Methicillin-Resistant *Staphylococcus aureus* (MRSA, especially here the MRSA252 strain) - in order to investigate if such interaction could be influenced, at some degree, by the environment (experimental methods or host genetics).

We first determined the dose-response of each drug in monotherapy in planktonic culture and we observed variability in the concentrations required to inhibit all the bacterial growth. We then sought the synergy of vancomycin and rifampicin with different experimental methods (most of them based on the planktonic Minimum Inhibitory Concentrations, MIC) and found no evidence of synergy. Vancomycin and rifampicin were antagonising or indifferent against MRSA.

Given the long-standing debate in the literature on whether this drug pair synergises - or not[55,71], we can only provide data in support of the former case. This could be influenced by the observed variability in MIC for both drugs, these concentrations being expected to fluctuate as well in patients given inter-individual differences in antibiotic diffusion and host genetic background.

3.2. Introduction

Given the rise of antibiotic resistance[20,181] and the lack of novel antibiotics being successfully developed and approved[6,182], the scientific community needs to ensure the current arsenal of antibiotics is used appropriately. Paul Ehrlich proposed in 1913[88] that we need to “hit fast and hit hard”, yet nowadays the consensus is that we need to use the right antibiotic, at the right dose and for the

right length of time[27] to inhibit bacteria as much as practicable while also avoiding antibiotic resistance.

Amongst the so-called “superbugs”, Methicillin-Resistant *Staphylococcus aureus* (MRSA) has been flagged as a particularly serious threat[181]. A report from the Centers for Disease Control and Prevention attributed more than 80,000 infections per year to MRSA in the United States of America (USA), leading to 10-20,000 deaths a year[61 183]. With such a high mortality rate, MRSA infections have been estimated to be 64% more likely to kill their human host than their non-resistant counterparts[184]. Moreover, they need to be dealt with due care given their predominance in hospitals[185], where the patients’ immune systems are often compromised.

To combat MRSA infections wisely, several regimens are proposed and their usage depending on the nature of the infection itself[82,83]. Among these is the combination of vancomycin and rifampicin – in use since the 1970s[69,70] – and yet it is still controversial, with guidelines varying greatly between countries. In the United Kingdom (UK), the last published recommendations proposed to restrict this combination to hospitalized patients either with cellulitis on surgical sites which are unresponsive for glycopeptide monotherapy, or with bone/joint infections (including prosthetic)[83]. Interestingly, not only did this report state that these guidelines were due to a lack of clear alternatives, but the American counterpart recommended this combination for very different types of infections, including infective endocarditis and several central nervous system infections[82].

This lack of consensus on the usage of vancomycin and rifampicin could be attributed to the variability in the outcomes of various *in vitro*, *in vivo* and clinical trials where this regimen was tested. Despite extensive studies on this combination, the nature of the interaction of vancomycin and rifampicin towards MRSA is, indeed, still in question: publications evidence both synergy and antagonism (i.e. beneficial or disadvantageous combination for the patients[42-45]), as well as indifference and additivity[55,71-81]. Variabilities affecting the data from *in vitro* experiments have been questioned[71,72,186-196], as well as differences in clinical settings[55]. For the latter, the current best practices to determine antibiotic

resistance and/or optimal drug dosages in patients are yet to be defined[45,82,83], most of them differing from laboratory practices[197].

Inconsistencies in clinical trial outcomes could also be caused by differences in MRSA strains, antibiotic dosages, study protocols (including sample sizes, or lack of control groups and randomisation)[55] and inter-individual differences[98]. The dosage of rifampicin is, indeed, only adapted to body weight according to current practice[82,83], with no consideration on potential differences in the patients' immune system, age, sex or even xenobiotic metabolism[92,95-97].

The purpose of the following study is to assess the effectiveness of the combination of vancomycin and rifampicin against MRSA in five different experimental settings: i) a planktonic experiment that produces a non-structured environment in a shaking-liquid media, ii) a biofilm experiment that produces a biofilm (a common biological structured environment for MRSA[33]), iii) and iv) two solid agar-plate experiments that produce structured environments, and finally v) the use of reagent strips of antibiotics on top of agar plates (Etests), commonly used in Clinical Diagnosis laboratories[56].

By inhibiting MRSA growth using vancomycin and rifampicin on culture media, the manifold variables from the patients are removed from the context, which should make fundamental pharmacological questions about MRSA treatments more tractable. Moreover, having just one single person performing the entire range of this wide variety of experiments, in the same laboratory environment, and with the same equipment and stock of supplies (MRSA strain, antibiotic stock and culture material), this study could help decipher if the published contradictions about the interaction of vancomycin and rifampicin on MRSA were be due to inter-laboratory/practices or inter-individual differences, or if they are the simple reflective of the unpredictable nature of MRSA under this specific combination.

Using the five different tests, we consistently rejected the hypothesis that vancomycin and rifampicin synergise in a MRSA infection, though some tests were better at answering this question than others. The only significant evidence we can provide is for a stable antagonism of vancomycin and rifampicin on MRSA252.

3.3. Materials and Methods

3.3.1. Bacteria strains

The MRSA strain MRSA252 was used in this study. Although vancomycin is not commonly thought to be active against *E. coli*[198], dose-response assays in liquid media exhibited the full range of inhibition (data not shown), allowing us to test the synergy of vancomycin and rifampicin on *E.coli* Wcl during preliminary experiments (only “square dish” experiments are presented here with the use of Wcl instead of MRSA, for health and safety purposes).

Bacteria strain	Description / Type	Source
E. coli Wcl	Assay strain Tet^S MC4 100-CFP / pCS-lambda	[52]
MRSA252	Hospital-acquired strain from the United Kingdom, representative of the epidemic EMRSA-16 clone.	[199]

Table 3.1

Description of the bacteria strains used in this study.

3.3.2. Bacteria culture media

E. coli Wcl was cultured in broth in the minimal growth media M9 (M9), prepared by mixing two concentrated salt solutions into distilled water (dH_2O), in order to reach the following concentrations: 40mM dipotassium phosphate (K_2HPO_4), 15mM monopotassium phosphate (KH_2HPO_4), 2.3mM trisodium citrate ($Na_3C_6H_5O_7$), 7.5mM ammonium sulfate ($(NH_4)_2SO_4$) and 1mM magnesium sulfate ($MgSO_4$) (all Fischer Scientific, UK or Sigma Aldrich, Germany). After being autoclaved, M9 was supplemented with filtered-sterilized 0.2% glucose ($C_6H_{12}O_6$, Fischer Scientific) and 0.1% casamino acids (casein hydrolysates, Duchefa, Netherlands). When required, M9 was also supplemented with agar (Sigma Aldrich), at 5g/L of agar for soft agar and 20g/L for hard agar plates in 150mm round Petri dishes (Greiner, UK).

MRSA252 was cultured in the rich Luria Bertani growth media (LB), composed of dH_2O with 1% (w/v) tryptone, 0.5% (w/v) bacto-yeast extract and 0.5% (w/v) sodium chloride (NaCl) (as an already-made mix from Sigma Aldrich). When

appropriate, it was supplemented with agar, at 5% (w/v) for soft agar plates or 20% (w/v) for hard agar in 150mm round Petri dishes.

For the Etest experiments, the MRSA252 strain was also cultured on Mueller-Hinton Agar (MHA) plates according to the Etests' manufacturer protocol. MHA powder (Fisher Scientific) was diluted in dH_2O to provide a solid media containing 17g/L of agar, 17.5g/L of acid hydrolysate of casein, 2g/L of beef extract and 1.5g/L of starch.

The aim for all experiments was to inhibit the growth of MRSA in/on a fresh media.

3.3.3. Antibiotics

When required, antibiotics were added to the sterile culture media, prior to use. Vancomycin and rifampicin powders (Sigma Aldrich) were dissolved to provide stock solutions at 50mg/mL in dH_2O (adjusted at pH7.3 with the addition of a few drops of hydrogen chloride for rifampicin), before being stored at -20°C in the dark.

Etests (bioMerieux, France) were used for antibiotic susceptibility testing on vancomycin and rifampicin on LB-agar and MHA plates. The VA 256 Etest was diffusing vancomycin from 0.016 to 256µg/mL, and the RI 32 Etest was diffusing rifampicin from 0.002 to 32µg/mL (Figure A12). These concentrations are defined by bioMerieux with the use of standard agar dishes (in volume and softness) as per manufacturer's protocol.

3.3.4. General bacteria culture

All bacteria strains were stored frozen at -80°C in 30% glycerol before use. They were all manipulated in sterile conditions, using autoclaved media and tools. Sterility was checked during the experiment with the systematic use of negative controls whereby media were not inoculated by a bacterium but still underwent every other process of the experiment. Results were discarded completely if the negative control appeared to be contaminated: this was checked by Optical Density (OD, with OD = Absorbance / optical pass Length) at 600nm as a proxy for cell population density (see *Chapter 2*), or visually – depending on the type of

experiment. Positive controls were also systematically provided with the bacterial inoculation of/on the growth media in the absence of antibiotics. Other controls were undertaken when deemed relevant.

Bacteria were grown in various conditions – depending on the experiment. They were always used after an overnight culture of 14-16 hours (h) at 37 for *E.coli* or 42°C for MRSA in an aerated incubator with a 180-revolutions-per-minute (rpm) shaking. The overnight culture consisted of directly thawing a *µL* sample from their frozen aliquot to 10mL of media broth[155].

3.3.5. Bacterial culture in a planktonic environment (broth dilutions)

Using a 96-pins replicator (Sigma-Aldrich), the equivalent of 1.5µL of the overnight culture of *E.coli* Wcl (in the preliminary steps of this experiment) or MRSA252 was transferred into the wells of a crystal-clear flat-bottom 96-wells microplate (Greiner) filled with 150µL of growth media (with or without the presence of antibiotics)[155]. This microplate was sealed with a transparent adhesive film and each well was aerated by carefully drilling a hole in the film. Bacteria growth was recorded automatically every 20 minutes (min) via the OD at 600nm by eLx807 UV/ABS (BioTek, USA) or Infinite 200Pro (Tecan, Life Sciences, Switzerland) plate readers, at 37°C (*E. coli*) or 42°C (MRSA). To keep the planktonic nature of the environment and avoid the formation of a biofilm of bacteria, plates were shaken in the plate readers. To keep the OD measurement consistent in-between wells, the hole in the transparent film was located at the same position for each well. Control growth was established in the absence of any antibiotic in the growth media (negative control).

Unless otherwise specified, each experimental condition was tested in three wells from the same overnight culture, and its OD is the average of the OD in these triplicates minus the OD from the negative control wells.

Antibiotic monotherapies were assessed first with a planktonic dose-response (DR) experiment allowing the plot of a DR curve, as previously described [41]. The

DR assay of one particular antibiotic towards one particular bacterium was performed in the wells of a microplate containing row-wise increasing concentrations of said antibiotic (Figure 3.1A). After bacteria inoculation, the microplate was incubated in the plate reader to measure the OD every 20min. The DR curve was plotted with the OD value at 24h as a function of the antibiotic concentration (Figure 3.1B). This allowed the estimation of the Inhibitory Concentrations (IC) 50, 80, 95 and 99 which are the smallest concentrations that inhibit respectively 50%, 80%, 95% and 99% of the bacterial growth[41]. The Minimum Inhibitory Concentration (MIC) is the minimal drug concentration at which no visible growth of a bacterium is observed after *in vitro* overnight culture, it is usually associated to IC95 or IC99[155].

Important note: To fairly compare the efficacy of a drug combination to their monotherapies, each basal drug concentration needed to be normalised to achieve equal inhibitory effect at time T. When deemed possible by the type of experiment (as it is the case in planktonic environment and biofilms), the IC50 of each antibiotic were chosen as basal drug concentrations, based on the fact that clinicians may prescribed higher than MIC concentrations to patients, yet these concentrations are often not reached at the site of infections – they are highly diluted by processes of absorption, plasma protein binding, tissue penetrations, etc.[363]. For example, in bone and joints - tissues often infected by MRSA, rifampicin and vancomycin penetrations are estimated to be particularly efficient, allowing those antibiotics concentrations to reach IC90 at these sites[364]. Yet, in the MRSA infections-susceptible Central Nervous System, rifampicin concentrations can exceed the MIC while vancomycin concentrations are lower[365]. With 88% of rifampicin being protein-bounded after administration[363], a direct association between dose-response curves *in vitro* and in patients is therefore not straightforward.

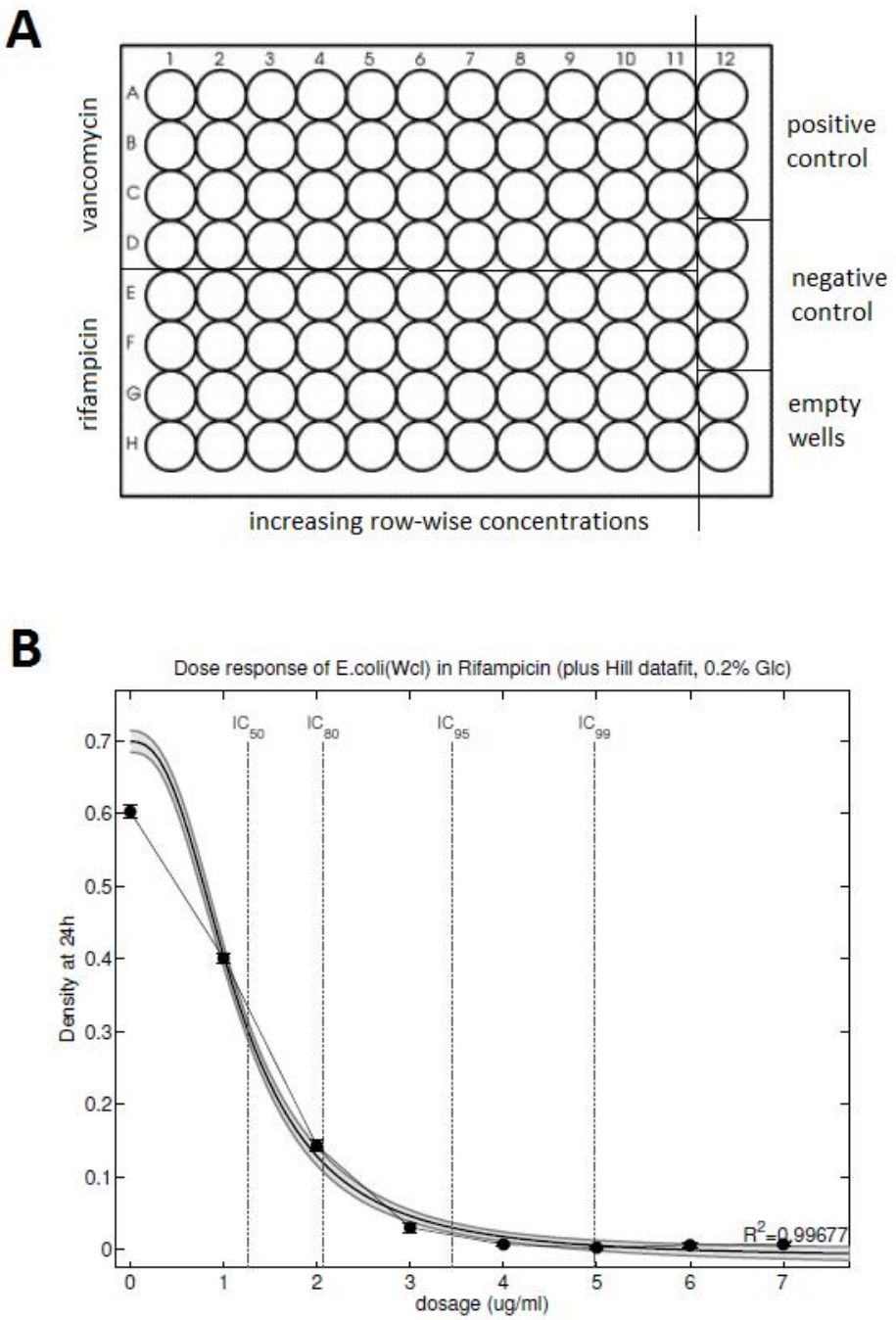


Figure 3.1

Testing antibiotic monotherapies in planktonic and biofilm experiments. A- Example of a microplate set-up producing an antibiotic monotherapy for both vancomycin and rifampicin. The negative control consisted in the non-inoculated non-antibiotic-supplemented growth media, the positive control contained non-antibiotic-supplemented growth media and has been inoculated by the bacteria. B- Example of the dose-response of *Escherichia coli (E.coli)* Wcl strain growing on rifampicin at 0.2% glucose, with the indication of the Inhibitory Concentrations (IC) IC50, IC80, IC95 and IC99. This example being imperfect due to the data point in absence of rifampicin not reaching the optical density of the overnight culture (positive control). *On B, the full points represent the average of the*

observed data, and the vertical lines the standard deviation between the observed data. The data follows the expected biochemical interaction of the bacteria and the antibiotic (Hill equation fitted to the data, with 95% confidence intervals around a predicted mean in grey).

Here, the extrapolated IC50 of each drug were therefore used to produce seven combinations of vancomycin and rifampicin, ranging from 100% of one drug at IC50 to 100% of the other one at IC50 (drug combination delta factor or relative drug fraction θ, Table 3.2)[41]. Each of these combinations was tested in four wells from the same overnight culture, and, on the same plate (with the same overnight culture and the same fresh media), monotherapies of vancomycin and rifampicin were re-assessed as dose-responses - in quadruplicates as well - to control if variations of the extrapolated IC50s occurred, and to observe the evolution of each dose-response over time, during 3 days[41] (Figure 3.2).

Combinations	**Content in vancomycin at IC50 (%)**	**Content in rifampicin at IC50 (%)**
Combination #1	0	100
Combination #2	20	80
Combination #3	35	65
Combination #4	50	50
Combination #5	65	35
Combination #6	80	20
Combination #7	100	0

Table 3.2

Composition of the seven tested combinations of vancomycin and rifampicin. These combinations were used in planktonic and biofilm experiments, with their respective experimental Inhibitory Concentrations (IC) IC50 concentrations.

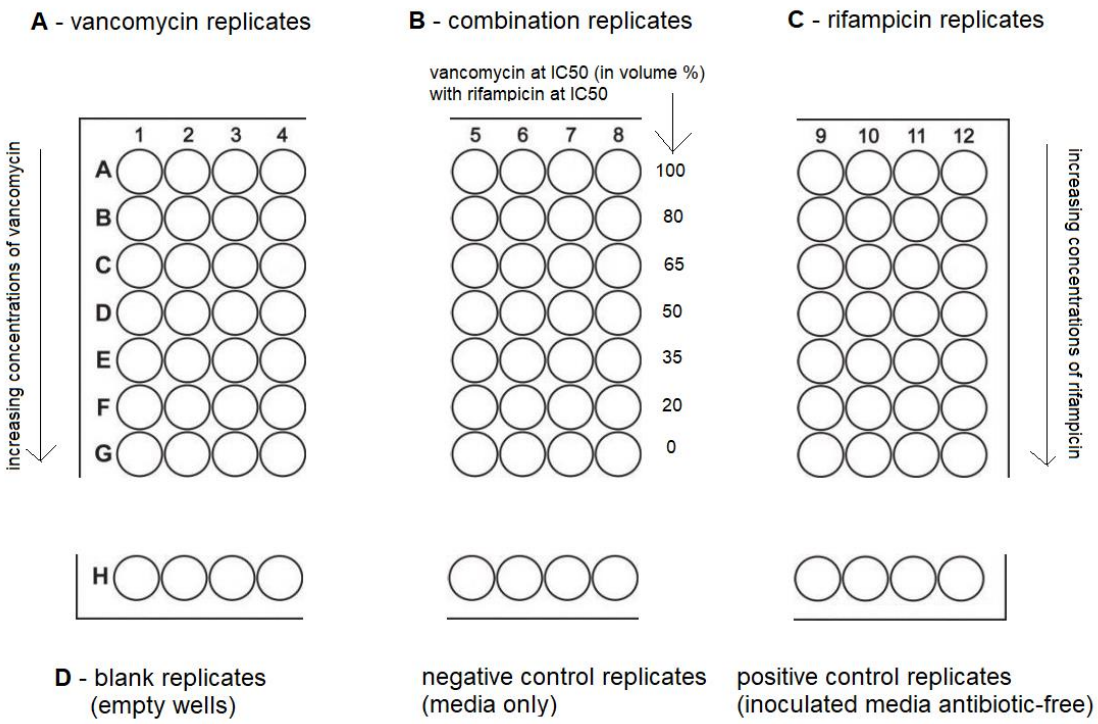


Figure 3.2

Example of a microplate set-up producing both monotherapies and the seven drug-drug combinations, used in the planktonic and biofilm experiments. A- Area of the plate for the vancomycin monotherapy. B- Area of the plate for the combination of vancomycin and rifampicin, with the volume of vancomycin at IC50, leading to the drug combination delta factor or relative drug fraction θ (not shown). C- Area of the plate for the rifampicin monotherapy. D- Area of the plate with the controls: empty wells, negative control containing only the growth media, positive control containing the inoculated media without any antibiotics. *IC50 is the antibiotic concentration inhibiting 50% of the bacterial growth (Inhibiting Concentration).*

On Day 1, 3 microplates (D1-3) were filled with these combination of antibiotics (Table 3.2 and Figure 3.2). The 96-pins replicator was used to inoculate the first plate (called D1) before incubation in the plate reader (where the plate was shaken, and the OD was read every 20min), the other plates (D2 and D3) were stored at 4°C in the dark. Every 24h, one plate (Dx) was retrieved from the fridge and approximately 1% of the inoculum of each well of the previous plate (Dx-1) was transferred into the new plate (Dx) using the 96-pins replicator, to refresh the carbon source in the media. The Dx plate was then incubated in the plate reader,

while the Dx-1 plate was saved in the freezer at -80°C. This serial transfer allowed to assess the evolution of the dose-responses on 3 days.

By plotting the bacterial density as a function of the drug combination delta factor, a smile or a frown was expected to show synergy or antagonism, respectively (as fully developed in *Chapter 2*). All presented results have passed the quality control: the negative control was not contaminated over the three days of the experiment, and the extrapolated IC50s for vancomycin and rifampicin were approximatively reached on the D1 microplate. The experiment was successfully replicated, providing the same outcome.

3.3.6. Bacterial culture in a biofilm environment

To create biofilms into the wells of a crystal-clear flat-bottom 96-wells microplate, 200µL of LB (with or without antibiotics) were inoculated with 2µL of the overnight culture of the MRSA252 strain. The plate was sealed and placed into a 37°C incubator without shaking, for 48h. After incubation, the OD was first read at 600nm by the eLx807 UV/ABS (BioTek) or the Infinite 200Pro (Tecan) plate readers, to determine the wells in which bacteria growth happened. Changing tips at every well from now on, media was pipetted out from the wells, without disrupting the bottom of the wells which contained the biofilm (*this special care applies to all of the remaining steps of this protocol*). Wells were washed with 200µL of 1x Phosphate Buffered Saline (PBS, Fischer Scientific) before being filled with 200µL of 0.1% crystal violet stain (Fisher Scientific) for 15min. This classic dye binds to negatively charged molecules, allowing the quantification of total biomass including bacteria and the extracellular polymeric substances of the biofilm. It was then removed from the wells, which were then washed five times with PBS. The microplate was let to dry, and 200µL of 1x ethanol (EtOH, Fisher Scientific) was poured into the wells to dissolve the biofilm-bound crystal violet. After an incubation at room temperature for 10min, the content of each well was

homogenized by pipetting, and OD was measured again to quantify the biofilm formation[200,201].

Unless otherwise specified, each experimental condition was tested in four wells from the same overnight culture and its OD is the average of the OD in these quadruplicates, minus the OD from the negative control wells (LB without bacterial inoculation, in triplicates). Control growth, tested in triplicates, was established in the absence of any antibiotic in LB.

The MIC of each drug was tested first in the experimental condition of biofilm, by growing MRSA252 on increasing row-wise concentrations of vancomycin or rifampicin for 24-48h without shaking (as seen for the planktonic experiment in Figure 3.1A). This allowed the creation of a DR curve for each antibiotic and therefore the estimation of the IC50 for each drug in this condition.

As in *Bacterial culture in a planktonic environment*, the extrapolated IC50 of each drug were used to produce seven combinations of vancomycin and rifampicin, ranging from 100% of one drug to 100% of the other one (drug combination delta factor or relative drug fraction θ - as seen for the planktonic experiment in Table 3.2). Both monotherapies were re-assessed on the same microplate than the combination therapy to control if variations of their expected IC50 occurred (as seen for the planktonic experiment in Figure 3.2).

By plotting the bacterial density as a function of the drug combination delta factor, a smile or a frown was expected to show synergy or antagonism, respectively (as fully developed in *Chapter 2*). All presented results have passed the quality control: the negative control was not contaminated during the experiment, and the extrapolated IC50s for vancomycin and rifampicin were approximatively reached on the microplate. The experiment was successfully replicated, providing the same outcome.

3.3.7. Bacterial culture inside a "square dish" structure

3.3.7.1. Description and protocol

For this structured environment, 150mm square Petri dishes (Greiner) were used. This structured method was designed to keep an entirely homogenous dish in terms of bacteria inoculation, nutrients and drugs, avoiding therefore an unequal diffusion of drugs in the dish[24]. 50mL of soft M9-agar, at 0.2% glucose, inoculated with 1 or 5% of *E. coli* Wcl (in the preliminary experiment, instead of MRSA252) was poured into the dish, except of a circular area in the middle which was kept empty by using a 60mm round Petri dish (Greiner) (Figure 3.3A-B). When the agar was solidified, the small dish was removed, and a solid plastic band, whose length and width were similar to the diameter and the thickness of the small dish (respectively), was placed in the centre of dish to divide it in two : half of the centre of the dish was then filled with non-inoculated vancomycin-supplemented soft M9-agar. After solidification, the solid plastic band was removed to allow the pouring of soft M9-agar containing rifampicin in the other half. The tested doses of antibiotics to use in combined therapies were of 5, 10 or 15 times the daily MIC tested in planktonic conditions (up to 6µL/mL for vancomycin and to 70ng/mL for rifampicin) due to Wcl being a facultative anaerobe - as MRSA [202,203].

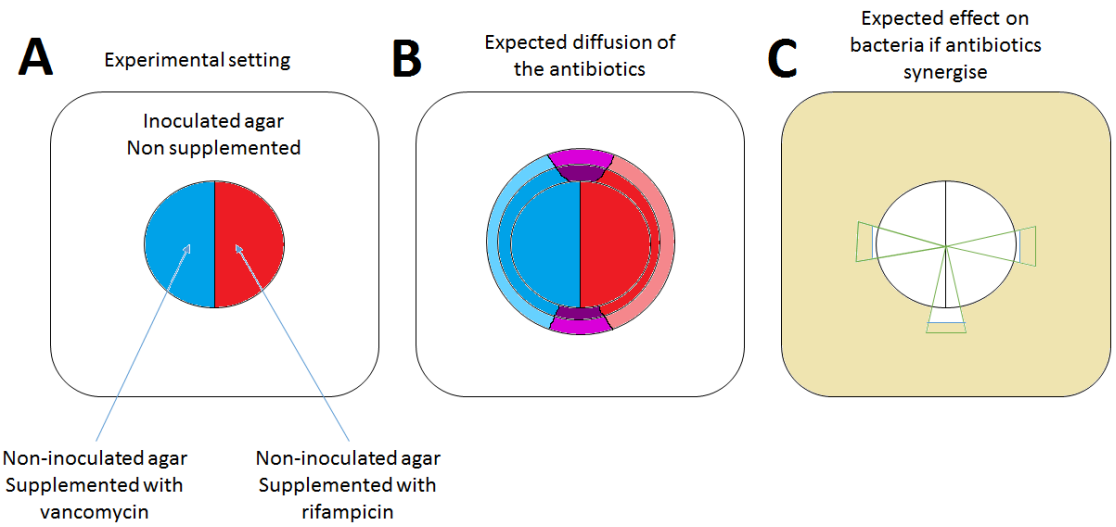


Figure 3.3

Illustration of the "square dish" methodology to produce a gradient of drug combinations in a structured environment. A - The inner circle of the square dish was divided in two and filled with non-inoculated agar supplemented with vancomycin (in blue) on one side and with rifampicin (in red) on the other side. The rest of the dish was filled with media-agar inoculated with bacteria. B - The antibiotics were expected to diffuse from the inner centre of the dish, with lower concentrations localised away from the inner centre (visualised in lighter shades of blue or red). There should be a monotherapy of each drug (in shades of blue or red) as well as a combination of them on the dish, this latter being approximatively localised where both half-circles meet (visualised in shades of purple). C – After incubation, if the antibiotics synergise, there should be a bigger clearance of bacterial growth (visualised as clear agar) where the bacteria were growing in the combination of both drugs, compared to where they are subjected to any of the monotherapies.

Dishes were incubated at 37°C for 24-48h in the laboratory BioBox, which is an automated photography chamber using the EOS 1100D digital SLR camera (Canon, UK) with white light. Square dishes were used in this experiment as they were fitting perfectly to the square platform of this station, allowing the settings to be consistent between experiments. Camera was set to take a picture and record it every 20min. Images were analysed with the ImageJ software (National Institutes of Health, USA) to correct the perspective and glare.

Image analysis of the final picture undertaken on ImageJ after subtracting, on each pixel, the background (picture taken at 0h) (Process > Image Calculator). Using the Radial Profile Angle plugin extension (https://imagej.nih.gov/ij/plugins/radial-profile-ext.html) (Plugins > Radial Profile Angle), the centre of the small dish was localised and its coordinates were recorded (Figure A1 left). This was performed three times, independently. With the same plugin, radial angles of 3° and 700 pixels of radius – starting at the previously located centre of the small dish – were created and their pixel intensities were extracted, the resulting data being the normalized integrated intensities around concentric circles as a function of distance from the centre of the small dish. This was performed for the three conditions (vancomycin monotherapy, rifampicin monotherapy and combination therapy, Figure A1 right – bottom for vancomycin monotherapy), with three independent readings per conditions.

The final pixel intensity (as a proxy for cell density or bacterial growth) was plotted as their average for each, as a function of the average distance from the centre of the dish. Differences in the thickness of the halo of inhibition were expected to be visualised, between the three conditions, as differences in the lengths between the two peaks in pixel density (as illustrated on the drawing in Figure 3.4). Lengths between the peaks were calculated from raw data.

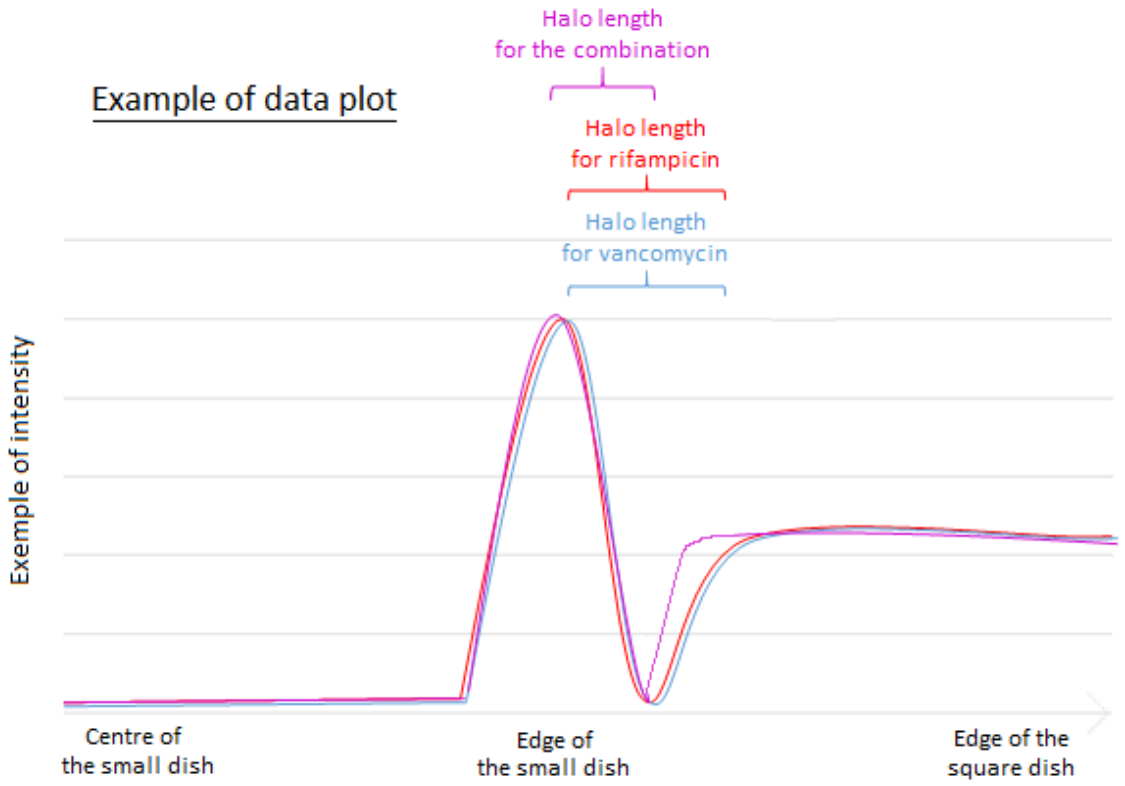


Figure 3.4

Illustration of the methodology used to visualise the halo of inhibition of bacterial growth for every condition in the "square dish" structure. The distance between the highest peak (edge of the small dish) and the second peak (the resumption of the bacterial growth) were calculated for each condition from raw data. In this example of antagonism, the halo length for the combination (in pink) would be shorter than the ones for vancomycin (in blue) or rifampicin (in red).

The control dishes consisted of using only one antibiotic or no antibiotic at all in the centre of the dish. A negative control consisted of a dish not being inoculated at all with the bacteria. This experiment has been completely replicated three times successfully (validation of the control dishes). Each experiment consisted of duplicated dishes per each setting condition (monotherapy controls, negative control…), and the analysis was performed for each dish with triplicate readings of 3° radial angles. The presented results and analysis are those associated to the experiment produced at 15 times the daily MIC (tested in planktonic

conditions), they represent all observed outcomes from all experiment, regardless of the concentrations used.

3.3.7.2. Expectations

As the “square dish” structure was produced using an unfamiliar protocol, the following details the expectations from performing this experiment.

The “square dish” method was designed to show, if any, a difference in the sizes of halo of bacterial growth inhibition between where bacteria were only in contact with one antibiotic (vancomycin or rifampicin) and where they were in contact with both antibiotics. The size of the halos would be visualised on the plots as the lengths between the edge of the small dish (where the first peak is visible) and where the bacteria growth is recovered (next pixel intensity peak).

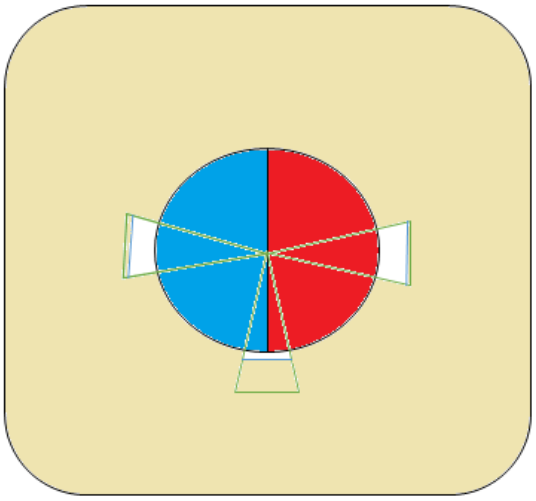


Figure 3.5

Illustration of an antagonistic result from the “square dish” methodology. After incubation, if the antibiotics antagonise, there should be a bigger clearance of bacterial growth (visualised as clear agar) where the bacteria were growing in monotherapies (of vancomycin in blue, of rifampicin in red), compared to where they are subjected to the combination therapy (bottom triangle).

On the same dish, in case of synergy, it was expected to observe a smaller halo of agar without bacteria growing in it where the bacteria are only in contact with monotherapies than where they are treated with combined therapy (Figure 3.3C).

By plotting the pixel intensity as a function of the distance from the centre of the dish, for the three therapies, the halo length would be significantly shorter for the monotherapies than for the combination therapy (Figure 3.6A).

Conversely, in case of antagonism, a bigger halo of agar without bacteria growing in it was expected where the bacteria are only in contact with monotherapies than where they are treated with combined therapy (Figure 3.5). The pixel intensity plot would therefore show a significantly longer halo length for the monotherapies than for the combination therapy (Figure 3.6B).

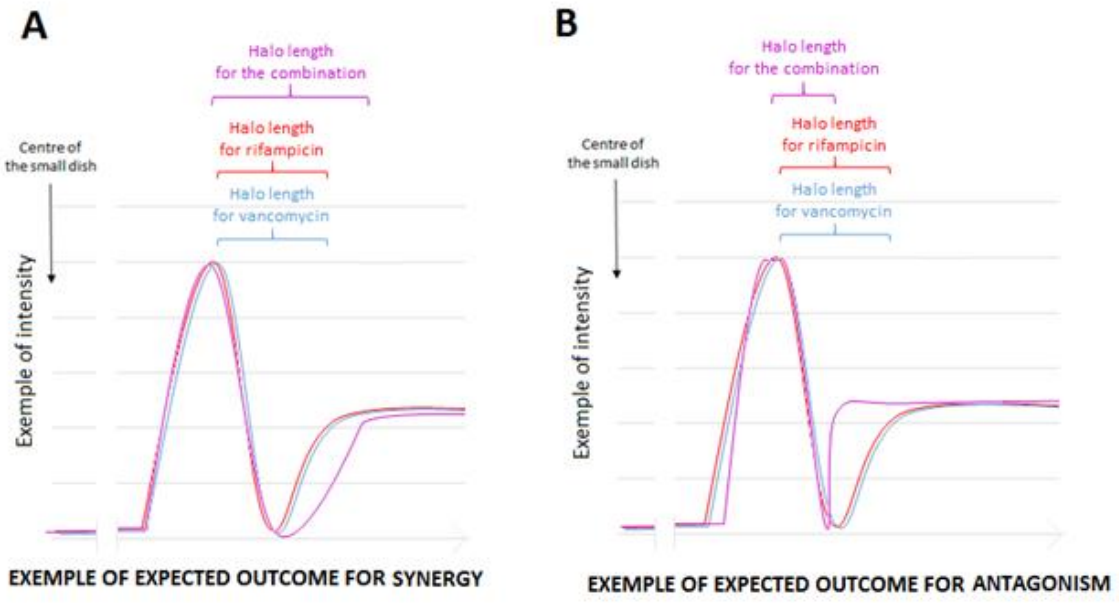


Figure 3.6

Illustration of the expected plots for synergy or antagonism interactions of vancomycin and rifampicin from the "square dish" structure. The distance between the highest peak (edge of the small dish) and the second peak (the resumption of the bacterial growth) could be calculated for each condition from raw data. Left is an example of synergy, in which the length associated to a combination is (significantly) longer than those from the monotherapies. Right is an example of antagonism, in which the length associated to a combination is (significantly) shorter than those from the monotherapies. *Vanc would be in blue, rif in red and the combination of both in pink.*

3.3.8. Bacterial culture on a "layer dish" structure

3.3.8.1. Description and protocol

The "layer dish" structured environment is named after the two unbalanced layers of soft LB-agar placed inside 150mm round Petri dishes (Figure 3.7). This structure, also known as "gradient plate" in the literature, was manufactured by pouring a first layer of soft agar on a non-levelled Petri dish, and by waiting for it to dry before levelling up the dish and pouring a second layer of soft agar, of equal volume[204-206]. Each layer was supplemented with a different antibiotic (Figure 3.7A) whose homogenous diffusion in the soft agar was assumed (here, as both the stock solutions of vancomycin and rifampicin are made of polar and water-soluble compounds, their hydrophilic nature were expected to favour their diffusion through the agar[207-209]). This created a gradient of antibiotic inside the dish, from the pole A containing 100% of antibiotic A and 0% of antibiotic B, to the pole B with 0% of antibiotic A and 100% antibiotic B (Figure 3.7B). When the second layer was dry, the MRSA252 overnight culture was swabbed onto its surface in two linear streaks: one from pole A to pole B, and one - a few millimetres away - from pole B to pole A. Dishes were incubated at 42°C for 24-48h and bacterial growth was checked visually and recorded with a camera. As for the "square dish" experiment, the tested doses of antibiotics to use in combined therapies were of 1, 5 or 15 times the daily MIC tested in planktonic conditions (up to 9µL/mL for vancomycin and to 100ng/mL for rifampicin) due to MRSA being facultative anaerobe [202,203].

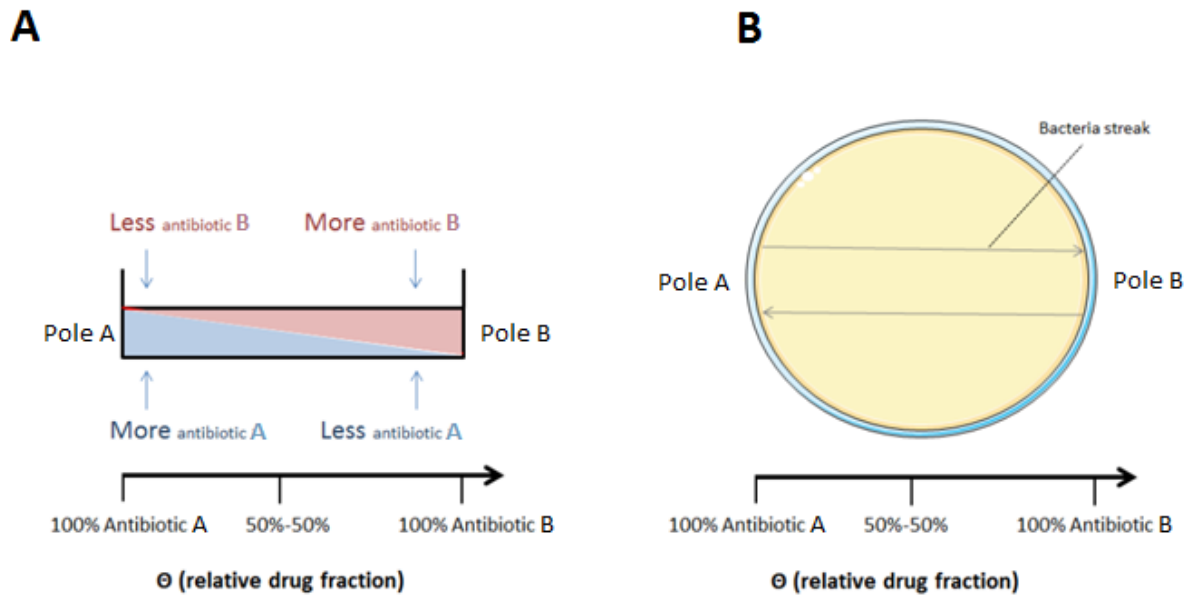


Figure 3.7

Illustration of the "layer dish" methodology to produce a gradient of drug combinations in a structured environment, with the antibiotics A and B. A- Content of the Petri dish, in a cross-section view, showing the two unbalanced layers of soft agar. B- Surface of the Petri dish where the bacteria were streaked in one single line going from Pole A to Pole B, and another streak going from Pole B to Pole A. *"Pole A" marks the pole of the dish containing 100% of antibiotic A and 0% of antibiotic B, "Pole B" marks the opposite pole with 0% of antibiotic A and 100% of antibiotic B. Both layers should allow the drug combination delta factor or relative drug fraction θ along the streaks.*

Every experiment consisted in producing this "layer dish" with all possible arrangements of layers consisting of either vancomycin, or rifampicin, or no antibiotic (Table 3.3). For the purpose of clarity, the nomenclature of the layer dishes is "A1B2", with A the initial of the antibiotic used in the first (bottom) layer and B the initial of the antibiotic used in the second (top) layer (using V for vancomycin, R for rifampicin and C for the control without antibiotic). Therefore, as an example on a C1R2 dish, the Pole A of the dish contained 100% of non-supplemented agar and the Pole B contained 100% of rifampicin-supplemented agar.

Conditions	First layer	Second layer	Streaks	Comments
Negative control (C1C2*)	No antibiotic	No antibiotic	Non-inoculated liquid media	No bacterial growth was expected.
Positive control (C1C2)	No antibiotic	No antibiotic	Inoculated liquid media	Bacterial growth was expected along the streaks.
Vancomycin control 1 (V1C2)	Vancomycin	No antibiotic	Inoculated liquid media	Should be equivalent to C1V2.
Vancomycin control 2 (C1V2)	No antibiotic	Vancomycin	Inoculated liquid media	Should be equivalent to V1C2.
Rifampicin control 1 (R1C2)	Rifampicin	No antibiotic	Inoculated liquid media	Should be equivalent to C1R2.
Rifampicin control 2 (C1R2)	No antibiotic	Rifampicin	Inoculated liquid media	Should be equivalent to R1C2.
Combination vancomycin / rifampicin (V1R2)	Vancomycin	Rifampicin	Inoculated liquid media	To be analysed if both V1C2 and C1R2 are validated.
Combination rifampicin / vancomycin (R1V2)	Rifampicin	Vancomycin	Inoculated liquid media	To be analysed if both R1C2 and C1V2 are validated.

Table 3.3

Description of the seven different conditions tested in the "layer dish" experiment. The nomenclature of the layer dishes is "A1B2", with A the initial of the antibiotic used in the first (bottom) layer and B the initial of the antibiotic used in the second (top) layer - using V for vancomycin, R for rifampicin and C for the control without antibiotic.

On each dish, after recording the bacterial growth with a camera before and after incubation, image analysis was performed on ImageJ (National Institutes of Health) by subtracting the background of the dish (as developed in the “square dish” experiment), by plotting a straight line from pole A to pole B, on both bacteria streaks, and by extracting the pixel intensity (profile) along these lines (Figure A2). This was replicated three times per streak, and the resulting profile (as a proxy for cell density or bacterial growth) was plotted as the average of the three readings per bacteria streak, as a function of the distance to the pole A of the dish. The final bacterial growth (expressed in percentage along the streaks, from 0% at one pole to 100% at the opposite pole) was plotted. The exact antibiotic concentration inhibiting the bacterial growth in these structured environments were calculated using the antibiotic concentration inside the antibiotic layer, and the localisation of the growth inhibition – which is related to the relative drug fraction distribution. For example, with a bacterial growth along the streak at 65% for 3-times MIC in the antibiotic layer, the bacteria were inhibited using a minimal concentration of 3x0.65-times MIC.

Up to four replicates were produced with the same dish composition per experiment. As a rule, every dish was validated only when its two streaks showed the same pattern, and every dish composition was validated only if all replicates were showing the same results. Negative and positive control dishes were also produced every time this experiment was performed, and all presented results have passed the quality control by i) showing the expected outcome for both these controls (see *Results*), and ii) not presenting any sign of bacterial contamination or of dried agar. All presented results have passed the quality control: the negative control was not contaminated during the experiment. The presented results and analysis are representative of all observations.

3.3.8.2. Expectations

As the “layer dish” structure was produced using an unfamiliar protocol, the following details the expectations from performing this experiment.

To be able to visualise a potential synergy or antagonism between vancomycin and rifampicin, the “layer dish” environment had to be produced both at low doses and at high doses of each antibiotic (Figure 3.8). Synergy could indeed only be visualised at low doses, i.e. concentrations at which the antibiotic monotherapy control dishes allowed the bacteria to grow homogenously from pole A to pole B – for both rifampicin and vancomycin. In this context, if vancomycin and rifampicin synergise, their combination dishes would show an inhibition of the bacterial growth in the centre of both A-B streaks, while bacteria would be growing at the poles (Figure 3.8A). Similarly, antagonism could only be visualised at high doses, i.e. concentrations at which the antibiotic monotherapy control dishes allowed no bacteria to grow homogenously from pole A to pole B – for both rifampicin and vancomycin. In this case, if vancomycin and rifampicin antagonise, their combination dishes would display bacterial growth only in the centre of both A-B streaks (Figure 3.8B). No interpretation could be made from other growth patterns on the dishes.

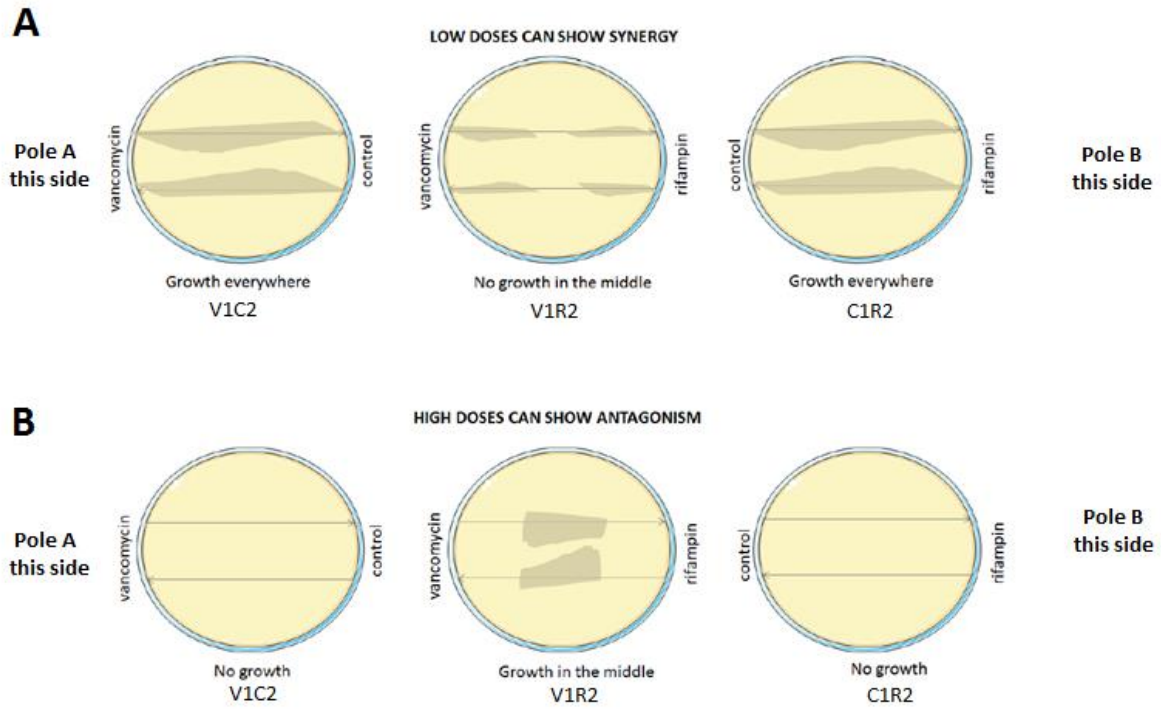


Figure 3.8

Illustration of the methodology to observe, in the "layer dish" experiment, the interaction between both antibiotics depending on the doses of antibiotics used. A – If any, low doses of antibiotics could show a synergy between vancomycin and rifampicin: no bacterial growth is visible in the middle of the streaks on the V1R2 plate, compared to growth on the plates V1C2 and C1R2. B- If any, high doses of antibiotics could show an antagonism between vancomycin and rifampicin: bacterial growth is visible in the middle of the streaks on the V1R2 plate, compared to an absence of bacterial growth on the plates V1C2 and C1R2. *C stands for control layer, V for a layer with vancomycin and R for a layer with rifampicin.*

To provide a clear and definitive interpretation of the interaction between vancomycin and rifampicin, this experiment was produced with both low doses and high doses dishes. For each condition, the observations had to be similar between the two combination dishes (V1R2 and R1V2), which were compared respectively to the bacterial growth on V1C2 and C1R2, and R1C2 and C1V2 (Table 3.3 for the full nomenclature and composition of the conditions).

For the antibiotic control dishes, and only if the bacterial growths along the streaks were linear as expected (as seen on the example of control dishes V1C2 or C1V2 in Figure 3.9A, and as opposed to the growth pattern in both V1R2

dishes in Figure 3.8), the bacterial growth was recorded in percentage along the streak, representing the relative distance from the pole with 0% antibiotic to where the bacterial growth is fully inhibited (example in Figure 3.9A with 100% of bacterial growth on the V1C2 dish and approximatively 60% of bacterial growth on C1V2). On those monotherapy “layer dishes”, a curve following a Hill equation pattern was expected by plotting the average bacterial growth along the streaks as a function of the antibiotic concentration in the antibiotic layer (Figure 3.9B). The bacteria were indeed predicted to fully grow along the streak (100%) in very low antibiotic concentrations, and to follow an exponentially decreasing dose-response with increasing antibiotic concentrations in the antibiotic layer.

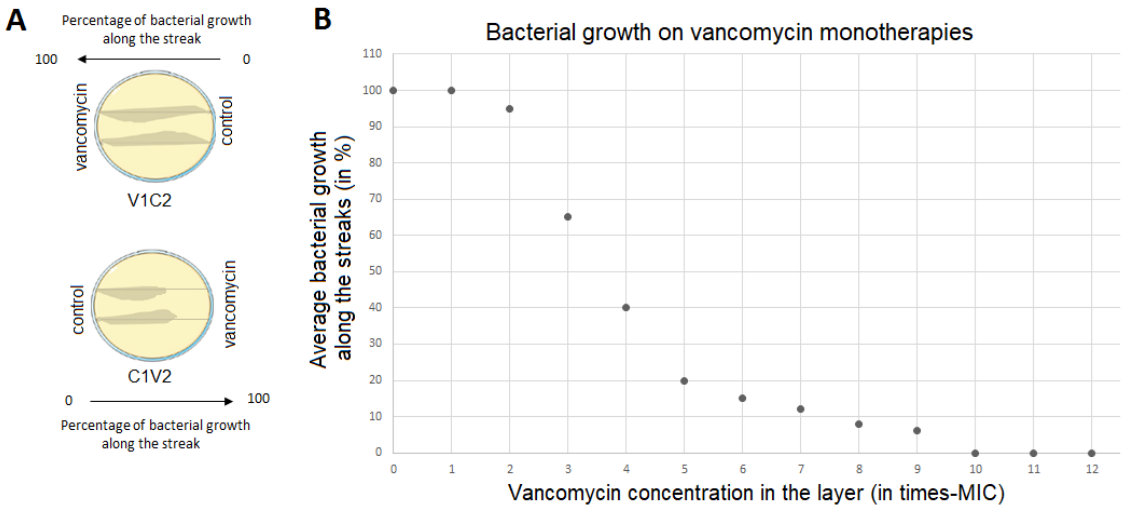


Figure 3.9

Illustration of the analysis for antibiotic control plates in the “layer dish” experiment, in cases of linear growth along the streaks. A – In cases of linear growth along the streak, the average bacterial growth was recorded in percentage of the streak on which the bacteria were able to grow, from the control pole to the antibiotic pole. In this example, the streaks on V1C2 and C1V2 would be recorded as, respectively, 100% and approximatively 60% of average bacterial growth. B- The average bacterial growth along the streak was then plotted as a function of the antibiotic concentration in the antibiotic layer. The curve is expected to follow a Hill function with limited antibiotic concentrations allowing bacterial growth along the streaks and high antibiotic concentrations inhibiting the bacterial growth. *C stands for control layer, V for a layer with vancomycin, R for a layer with rifampicin and MIC is Minimum Inhibitory Concentration. Notes: in A, the same pattern are expected between V1C2 and C1V2 to allow for a plot following the example exhibited in B. The examples in A and B are therefore unrelated.*

3.3.9. Bacterial culture on agar plates using Etests

After pouring LB-agar or MH-agar in 150mm round Petri dishes, MRSA252 was streaked into the dry dishes using a disposable L-shaped spreader (Sigma Aldrich) and RI 32 and/or VA 256 Etests were carefully positioned onto the dishes with forceps – sterilely and in one single movement, as per manufacturer's instructions. Dishes were then incubated for 18-24h at 42°C.

MIC was determined by applying only one antibiotic Etest on the surface of the inoculated LB-agar or MHA dish. The way to determine the MIC was dependant on the microorganism and on the drug - for rifampicin and vancomycin, and on MRSA, the manufacturer protocol advised a reading after incubation as following: rifampicin MIC was read where the edge of the inhibition ellipse intersects with the side of the strip (or at the next higher value if it sits in-between markings) and, vancomycin being considered a bactericidal agent with a high molecular weight, its MIC is reached where the growth halo intersects with the side of the strip as well (Figure 3.10A). With vancomycin, special care had to be undertaken to avoid false results due to trailing growth (hazes, micro- and macrocolonies) – only the higher MIC value had to be considered.

When used in combination, the two Etests were placed perpendicularly to each other (with a 90° angle) according to their MIC readings from the previous day[186-189,210]. In the case of the combination of antibiotic A and antibiotic B, both possible configurations were tested: on one dish, Etest(A) was placed first and Etest(B) was positioned on top of it with a 90° angle (Figure 3.10B), and on another dish, Etest(B) was placed first and Etest(B) was positioned on top of it with a 90° angle (Figure 3.10C). The combination MIC was read in the area where both Etests intersect, where the edge of the inhibition ellipse intersects with the side of both strips.

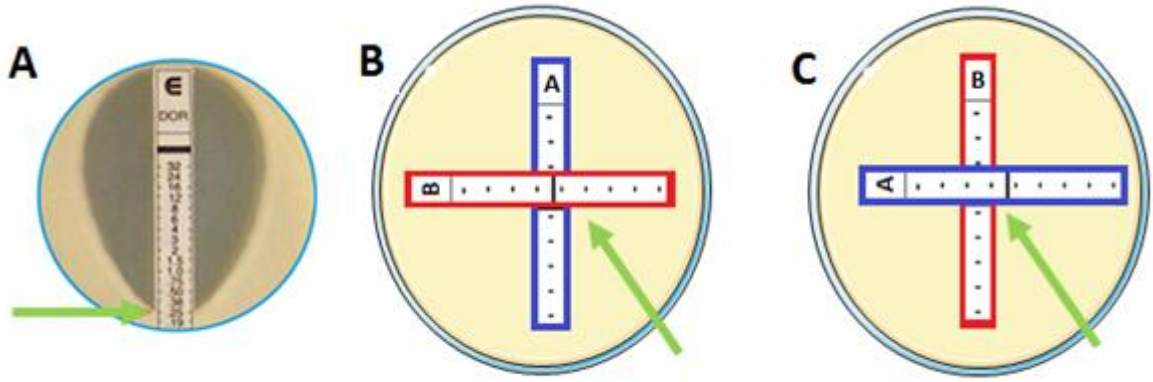


Figure 3.10

Illustration of the Etest methodology to produce a gradient of drug combinations in a structured environment, with the antibiotics A and B. A- The Etest was placed on an agar plate surface which was inoculated with bacteria prior to use. Using one Etest (monotherapy), the Minimum Inhibitory Concentration (MIC) was read at where the edge of the inhibition ellipse intersects with the side of the strip (green arrow). B and C- For a combination assay, the Etests of both antibiotics were placed perpendicularly to each other, at their respective MIC breakpoints. B shows the configuration in which the Etest of antibiotic A was placed first, C shows the configuration with the Etest of antibiotic B being first on the agar plate. In both cases, the MIC in combination was read in the area pointed by the green arrow.

Visual observations and photograph recordings were taken at 24h. Control dishes included the re-determination of the MIC of monotherapies on the day of the combination, along with both: i) the combination of two vancomycin Etests and ii) the combination of two rifampicin Etests. One control dish was also simply streaked with the bacteria (positive control) and another one was incubated without bacteria nor Etest (negative control). Both control dishes were produced for each experiment. All presented results have passed the quality control by not presenting any sign of bacterial contamination or of dried agar. The application of the Etests on the agar plates was particularly troublesome in the case of Etests combination, as it was difficult to place the Etests perpendicularly to each other at their exact MIC, in one single movement and using forceps. Only the most precise applications were incubated, and therefore analysed.

After preliminary experimentation, the whole experiment was replicated four times. Up to two exact replicates were produced per control and tested conditions, per experiment. The presented results and analysis are representative of all observations.

3.3.10. Plotting and statistics

Differences in bacteria growth were followed through the Optical Density as a function of drug concentration (monotherapies), or drug combination delta factor or relative drug fraction θ, at one time point (2D graph) and over time (3D graph) (combined therapies) using MatLab 2018 (MathWorks, USA) for the planktonic experiment, and SPSS Statistic 24 (IMB, USA) for the biofilm experiment. As detailed in *Chapter 2*, this allowed the visualisation of possible synergies as smile-features and of antagonism as frown-features[41]. For the planktonic experiments, statistical differences were assessed by MatLab 2018 using t-test (with $p < 0.05$ considered significant) and Hill equation was used to fit the data points to show standard deviation and standard error (*Supplementary Data*). For the biofilm experiment, error bars represent the standard deviation of the four replicates and statistical differences were assessed using XLSTAT (Addinsoft, France) and Stata/SE 15 (Statacorp, USA) by a Mann-Whitney test (*Supplementary Data*).

For the “square dish” experiment, the three halo lengths produced in each experiment were plotted as clustered bar charts using Excel 2016 (Office 365, Microsoft, US) for visualisation purpose and two-samples t-tests were performed using Stata/SE 15 (*Supplementary Data*).

For the “layer dish” experiment, the average bacterial growth along the streaks was plotted as a function of the antibiotic concentration, using a scatterplot on Excel 2016 to visualise the disparities in results - with error bars representing the standard deviations along streaks on similar dishes (same composition, experiment performed on the same day).

For the Etest experiments, the distribution of the MICs was plotted using X, Y scatter plots on Excel 2016. For each combination, the Fractional Inhibitory Concentration Index (FICI) was calculated as the sum of ratios of the MIC of the combination of Antibiotics A and B divided by the MIC of the antibiotic A (or B) monotherapy (Equation 1). The FICI data was interpreted using the following criteria: antibiotics synergize if FICI ≤ 0.5; they are indifferent if 0.5 < FICI < 4.0; and they antagonise if FICI ≥ 4.0 [186,187,189,190,210-212].

Equation 1: $\mathrm{FICI} = \frac{MIC\ A\ (in\ combination\ with\ B)}{MIC\ A\ (in\ monotherapy)} + \frac{MIC\ B\ (in\ combination\ with\ B)}{MIC\ B\ (in\ monotherapy)}$

3.4. Results

The aim of this study was to assess the effectiveness of the combination of vancomycin and rifampicin against MRSA infection in five different experimental settings, to see if the inconsistencies between previously reported clinical studies could be due to inter-laboratory/practices or to inter-individual differences, or if they are potentially the result of a non-optimal regimen.

3.4.1. Bacterial culture in a planktonic environment (broth dilutions)

Research question 1: what is the interaction of vancomycin and rifampicin towards MRSA in a planktonic culture - an environment typically used in clinical laboratories[45]?

A planktonic experiment was performed to produce a non-structured environment in which MRSA252 could grow at different concentrations of vancomycin and/or rifampicin. Usually referred as “broth (micro) dilution”[213], this method of antibiotic susceptibility testing, commonly used in research laboratories and clinics[191,193,209,213,214], was adapted here to be able to follow, on one microplate, the dose-responses of each monotherapy, and to compare them to the combination therapy, both over time (3 days).

Over 24h, the bacterial growth of MRSA252 was inhibited, in monotherapies, when reaching the MIC of approximately 300ng/mL for vancomycin and of approximately 3.5ng/mL for rifampicin (as visualised in Figures 3.11A and 3.12A). Their respective IC50s were of 60.7 and 1.25 ng/mL. Interestingly, the dose-response of MRSA in a rifampicin monotherapy was not stable over time: a dose completely inhibiting the bacterial growth at 24h was found to be ineffective after 48h (see 3ng/mL in Figure 3.12B, for example).

By repeating the monotherapy experiments, it was witnessed that the MICs read at 24h, and therefore their resulting determined IC50s, tend to fluctuate from one day to another (example of a 2ng/mL MIC for rifampicin in Figure A3) – despite great handling care to follow precisely the protocol. It was therefore challenging to proceed to the combination part of the experiment, as it required this particular dosage to be reached for both vancomycin and rifampicin. This was nevertheless successfully achieved, when - on the same microplate - the observed IC50s of the monotherapies were in accordance with the predicted ones, concentrations which were used in the combination therapy (Figures 3.13-14).

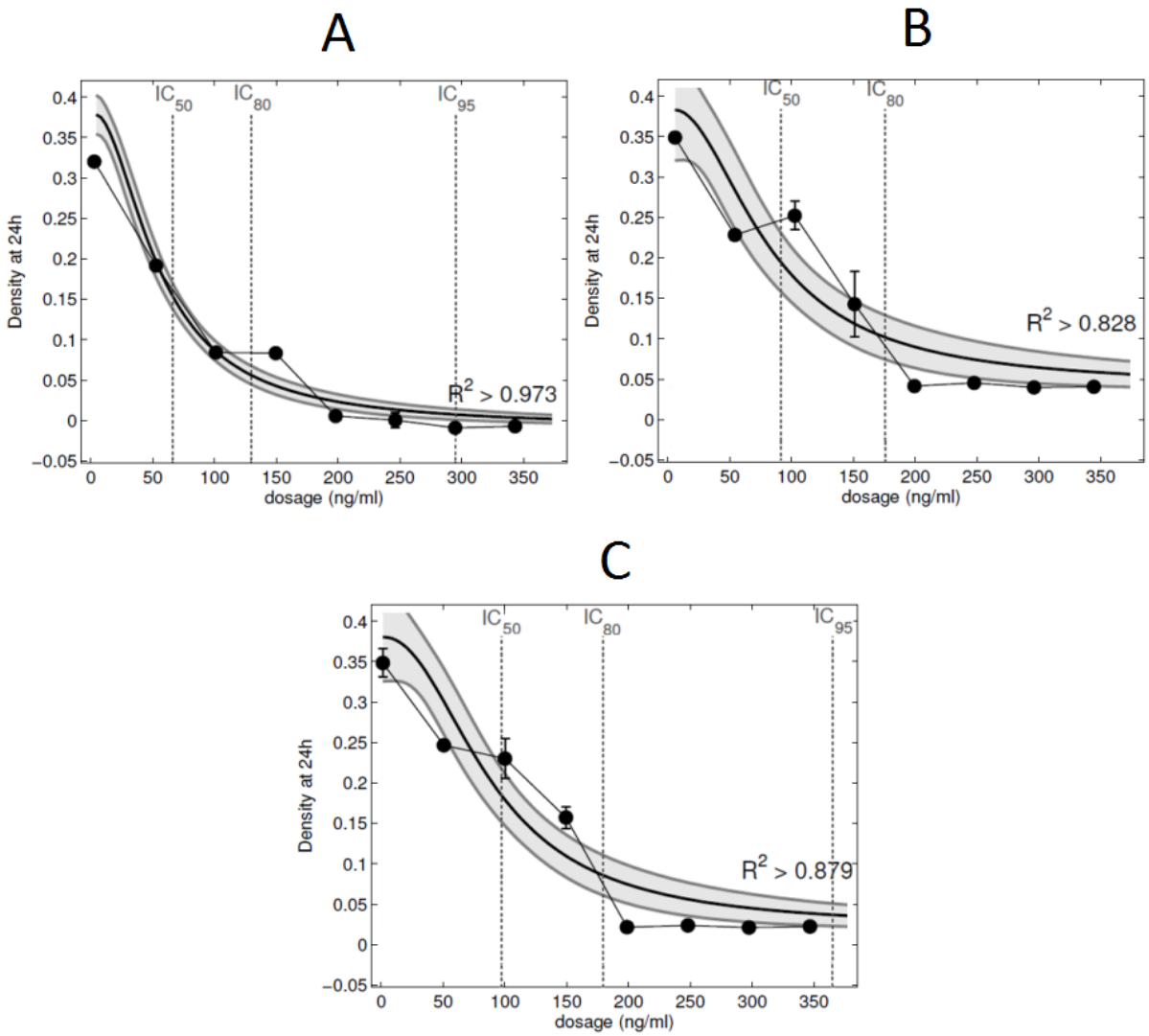


Figure 3.11

24h-dose-response curves for MRSA252 in monotherapy of vancomycin, followed over time as the optical density (OD) at 600nm at increasing vancomycin dosages in broth liquid. A- Dose-response at 24h on day 1, *allowing the estimation of the Minimum Inhibitory Concentration at approximatively 300ng/mL.* B- Dose-response at 24h on day 2, after one serial transfer. C- Dose-response at 24h on day 3, after a second serial transfer. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data. A Hill equation is fitted to the data, with 95% confidence intervals around predicted mean, associated to the expected biochemical interaction of the bacteria and the antibiotic, in the grey area along the black curve (superimpos ed Hill curve). The data follows the Hill curve, with the correlation coefficient R^2 indicated on the graph. The Inhibitory Concentrations 50 (IC50), 80 (IC50) and 95 (IC95) are indicated.*

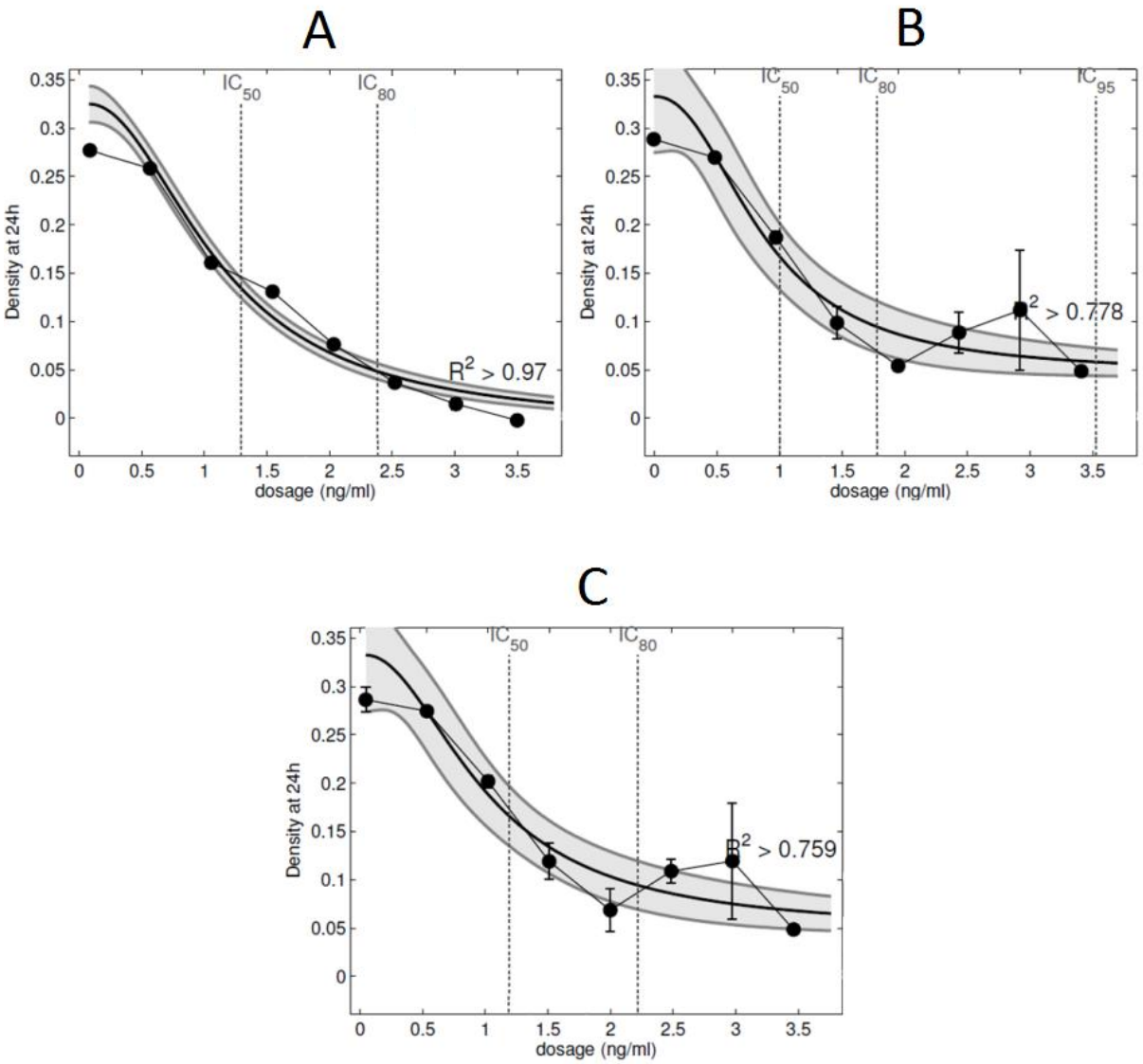


Figure 3.12

24h-dose-response of MRSA252 in monotherapy of rifampicin, followed over time as the optical density (OD) at 600nm at increasing rifampicin dosages in broth liquid. A- Dose-response at 24h on day 1. B- Dose-response at 24h on day 2, after one serial transfer. No Minimum Inhibitory Concentration was estimated from this growth curve, the bacteria never being completely inhibited. C- Dose-response at 24h on day 3, after a second serial transfer. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data. A Hill equation is fitted to the data, with 95% confidence intervals around predicted mean, associated to the expected biochemical interaction of the bacteria and the antibiotic, in the grey area along the black curve (superimposed Hill curve). The data follows the Hill curve, with the correlation coefficient R^2 indicated on the graph. The Inhibitory Concentrations 50 (IC50), 80 (IC50) and 95 (IC95) are indicated.*

Interestingly, following the monotherapies over time allowed us to observe in Figures 3.11-12 that, as expected, the definition of Minimum Inhibitory Concentration may be biased. It involves the visual observation of no bacterial growth at a given concentration, and does not consider the limitations of such visual observations: no visible bacterial growth ($\leq 10^5$ bacteria) could actually hide the presence of low-density populations, as persister cells or slow growing phenotypes for example[24]. If the optical density of 0 were reached in both monotherapies on day 1 (Figures 3.11A and 3.12A), these OD zero values were not found in days 2 or 3, for both vancomycin and rifampicin.

Focussing now on the combination therapy tested on the same microplate (presented in Figures 3.13-14), the OD at 600nm was plotted as a function of the drug combination delta factor or relative drug fraction θ, and is expected to exhibit a smile-feature when the antibiotics synergise, or a frown-feature when they antagonise[41]. On the first day of the combination (Figure 3.13A), at 16.7 hours, a frown was visible: there was significantly more bacterial growth at θ ½ (where an equal amount of vancomycin and rifampicin is present in the broth) than in either of the monotherapies (θ = 0/1 or 1/1). This significant frown was stable over time, up to 3 full days of experiment (Figure 3.13B-C, B being at 13.3h on the second day of the experiment, and C at 13.3h on the third day of the experiment). As illustrated with the 3D visualisation in Figure 3.14 (and in Figure A4), the frown was stable during the 24 hours of each day of the experiment. Vancomycin and rifampicin were therefore significantly antagonising against MRSA252 in this broth dilution environment.

Important note: Figures 3.11-14 show the outcome of the same experiment (one single dataset). The optical densities of each monotherapy can therefore be read as following: on day 1 at 16.7h on Figure 3.13A, on day 1 at 24h on Figures 3.11A and 12A, on day 2 at 13.3h on Figure 3.13B, on day 2 at 24h on Figure 3.11B and 12B, on day 3 at 13.3h on Figure 3.13C and on day 3 at 24h on Figure 3.13C.

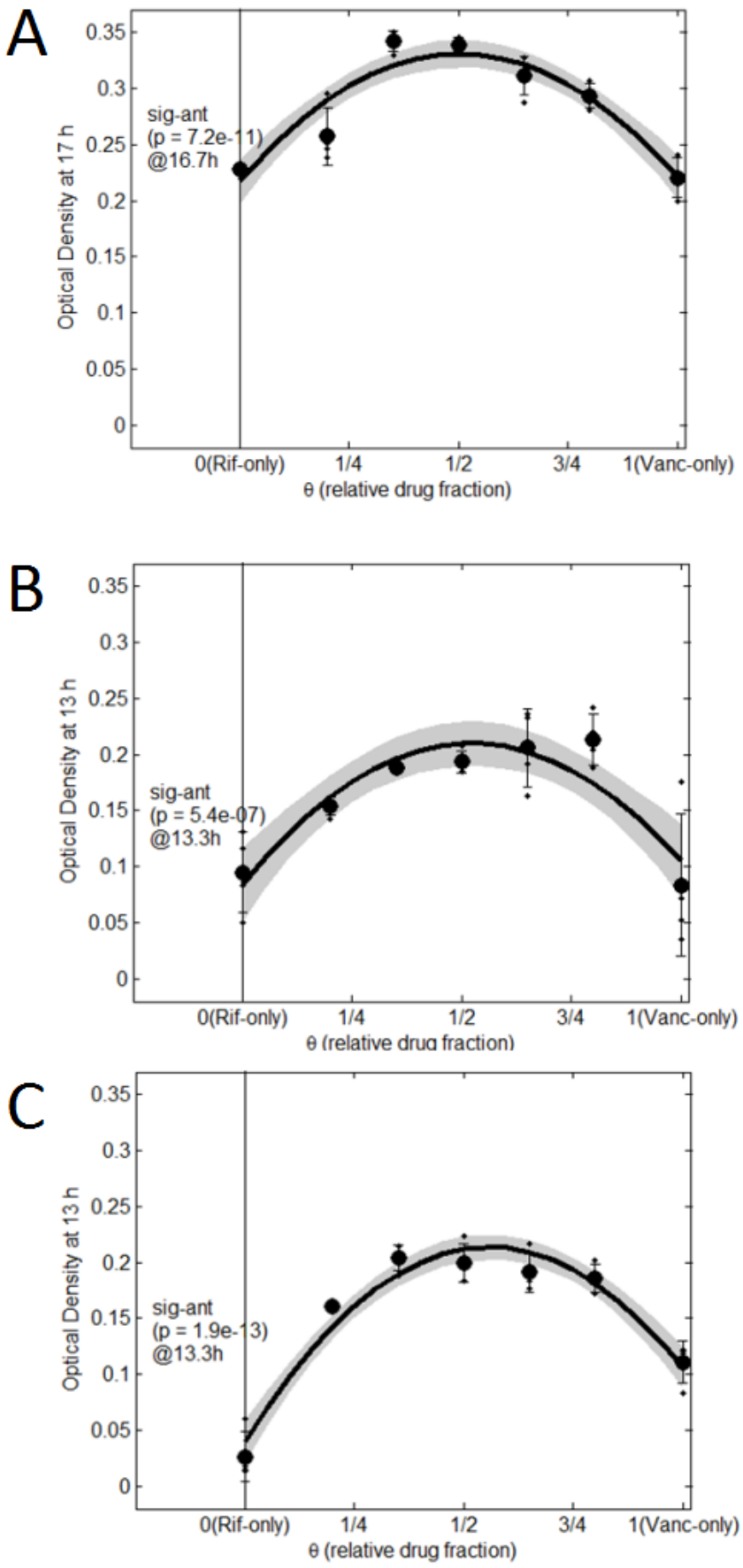
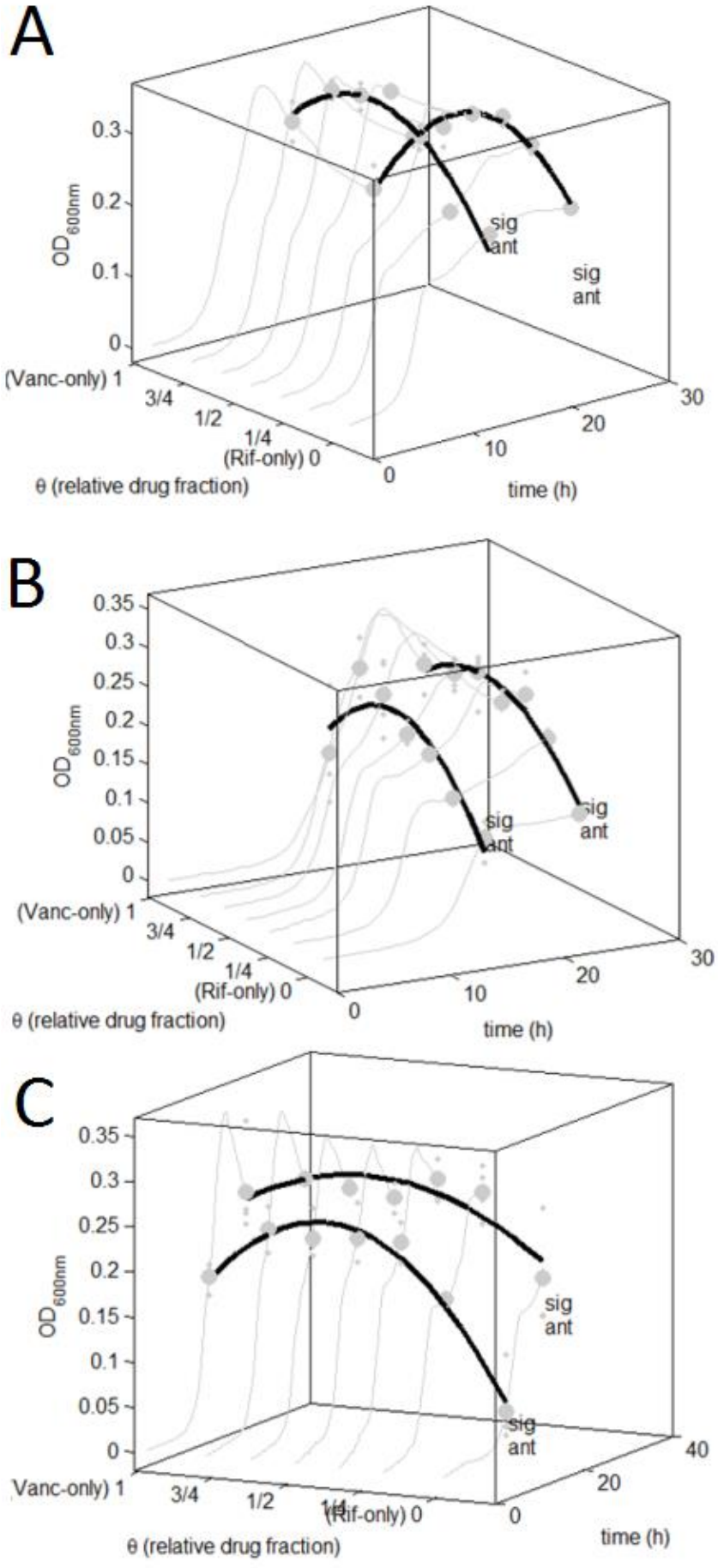


Figure 3.13 (described next page)

Vancomycin and rifampicin significantly antagonise against MRSA252 in broth liquid, over three days (2D visualisation). The combination of vancomycin and rifampicin at Inhibitory Concentration 50 (IC50) does not inhibit the bacterial growth of MRSA252 in broth liquid, over 3 days. A frown is observed here in 2D visualisation of the combination therapy as a function of the drug concentration delta factor or relative drug fraction θ. The antagonism of vancomycin and rifampicin is stable and significant ("sig.ant"), associated to a p-value < 10^{-7}, over three days (A – day 1 read at 16.7h, B- day 2 read at 13.3h, C- day 3 read at 13.3h). *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data. A Hill equation is fitted to the data, with 95% confidence intervals around predicted mean, associated to the expected biochemical interaction of the bacteria and the antibiotic, in the grey area along the black curve (superimposed Hill curve). The data follows the Hill curve, with the correlation coefficient* R^2 *indicated on the graph. The Inhibitory Concentrations 50 (IC50), 80 (IC50) and 95 (IC95) are indicated. Vanc stands for vancomycin and rif for rifampicin, and drug concentration delta factor or relative drug fraction θ is expressed in vancomycin ratio in the combination (from 0 to 1).*

Figure 3.14 (next page)

Vancomycin and rifampicin significantly antagonise against MRSA252 in broth liquid, over three days (3D visualisation). The combination of vancomycin and rifampicin at Inhibitory Concentration 50 (IC50) does not inhibit the bacterial growth of MRSA252 in broth liquid, over 3 days (A- day 1, B- day 2 after serial transfer, C- day 3 after serial transfer). A frown is observed in 3D visualisation of the combination therapy as a function of the drug concentration delta factor or relative drug fraction θ and time. The antagonism of vancomycin and rifampicin is stable and significant over the course of the three days ("sig ant"). *The black curves follow the average of the observed data along delta factor, the light grey curves follow OD over time. Vanc stands for vancomycin and rif for rifampicin, and the drug concentration delta factor or relative drug fraction θ is expressed in vancomycin ratio in the combination (from 0 to 1).*


A
0.3
0.2
0.1
0
OD600nm
sig
ant
sig
ant
(Vanc-only) 1
3/4
1/2
1/4
(Rif-only) 0
θ (relative drug fraction)
0
10
20
30
time (h)
B
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
OD600nm
sig
ant
sig
ant
(Vanc-only) 1
3/4
1/2
1/4
(Rif-only) 0
θ (relative drug fraction)
0
10
20
30
time (h)
C
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
OD600nm
sig
ant
sig
ant
Vanc-only) 1
3/4
1/2
(Rif-only) 0
θ (relative drug fraction)
0
20
40
time (h)

Conclusion 1: the interaction of vancomycin and rifampicin towards MRSA is significantly antagonistic, and stable over time, in a planktonic culture. Both vancomycin and rifampicin exhibit both variability in their Minimum Inhibitory Concentrations.

3.4.2. Bacterial culture in biofilms

Research question 2: what is the interaction of vancomycin and rifampicin towards MRSA in biofilm - a structure typically formed by MRSA[33,45]?

Colonies of *Staphylococcus aureus* are known to form biofilms[33,45]. This highly-structured environment is beneficial for MRSA as it is less permeable to foreign compounds such as antibiotics[33,45,215]. Since rifampicin has been described as potentially able to penetrate such biofilms[45], the combination of vancomycin and rifampicin is favoured to eradicate device-related staphylococcal infections, particularly during prosthetic bone/joint replacements or insertions of intravenous catheter[33,45]. Yet this specific usage is still questioned[82,83].

To drive MRSA252 into forming biofilms, the bacteria were cultured in liquid media inside the wells of a microplate, for 48h without being shaken – in the presence of monotherapies or combinations of vancomycin and rifampicin. The Optical Density was read at 600nm before and after a staining with crystal violet, this technique allowing the quantification of the biofilm formation, which proportional to the bacterial load[79,197,200,201,215].

The dose-response of each antibiotic alone was monitored using monotherapies. The MIC was reached at approximatively 6.65ng/mL for rifampicin, and not reached at all for vancomycin (Figure A7). The IC50 for rifampicin was calculated to be approximately of 2ng/mL. These concentrations were slightly higher than those observed in planktonic cultures, which was expected as the formation of

biofilm is beneficial for MRSA to survive antibiotic therapy. However, it was not anticipated that the rifampicin MIC in biofilm would be double than in broth dilution. This was particularly surprising as rifampicin was known for its anti-biofilm properties[33,45,68].

When vancomycin was associated to rifampicin, at the IC50 of each antibiotic, a significantly more robust biofilm of bacteria was detected in the presence of both drugs combined compared to either monotherapies: a frown-feature was visible, before and - more importantly - after crystal violet staining, when plotting the OD (as a proxy for, respectively, the bacterial growth and the biofilm formation) as a function of θ (Figures 3.15-16). Vancomycin and rifampicin were therefore significantly antagonising against MRSA252 in this biofilm environment.

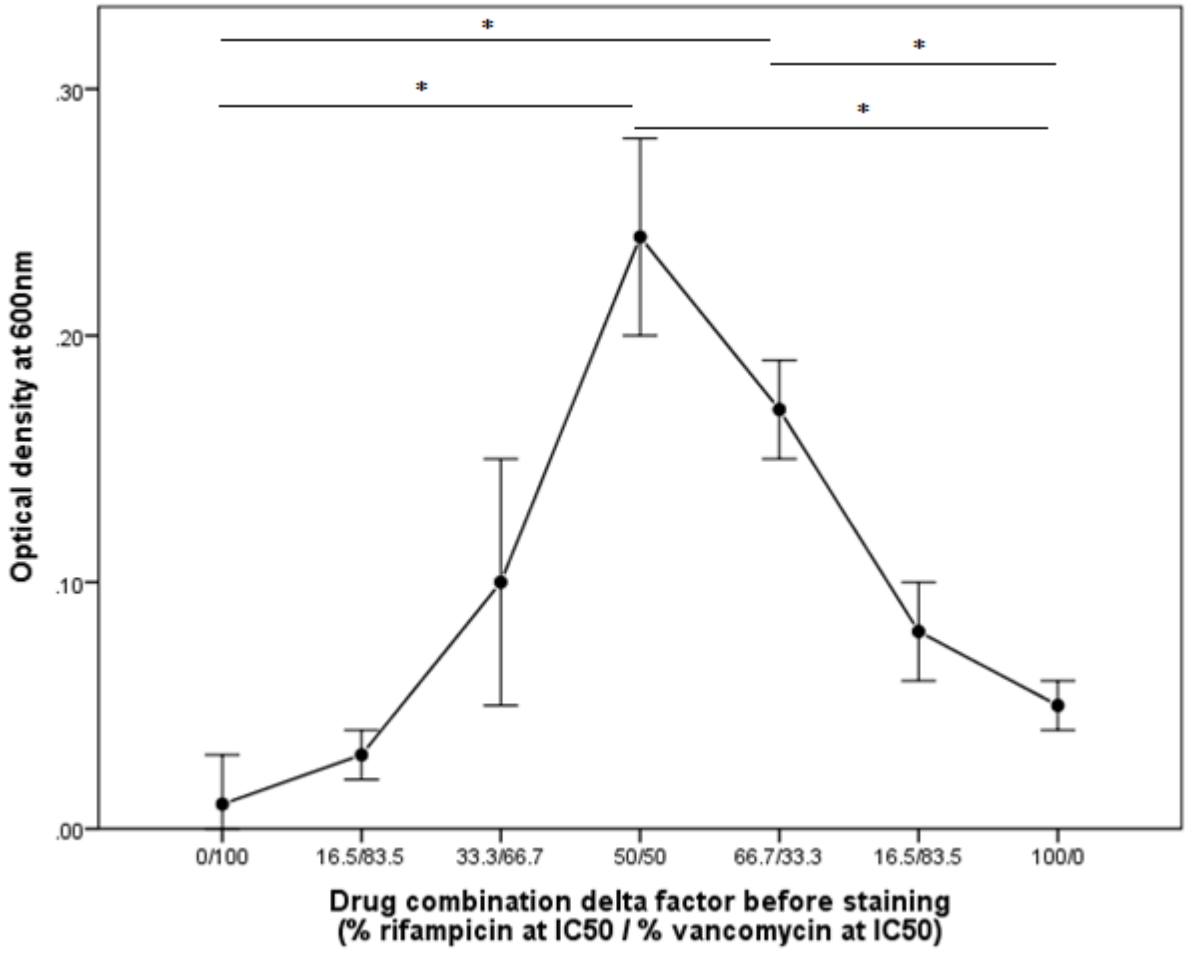


Figure 3.15

MRSA252 bacterial growth is significantly improved in the presence of both vancomycin and rifampicin than with their monotherapies, in a biofilm environment. Data shown represent the optical density at 600nm, as a proxy for MRSA252 bacterial growth, as a function of the drug concentration delta factor or relative drug fraction θ – before staining the biofilm with crystal violet. Vancomycin and rifampicin were used at their Inhibitory Concentrations 50 and combined at different percentages (from 0% rifampicin + 100% vancomycin to 100% rifampicin + 0% vancomycin). *The error bars represent the standard deviation of the four replicates. * indicates $p < 0.05$ from a Mann-Whitney test (Supplementary Data).*

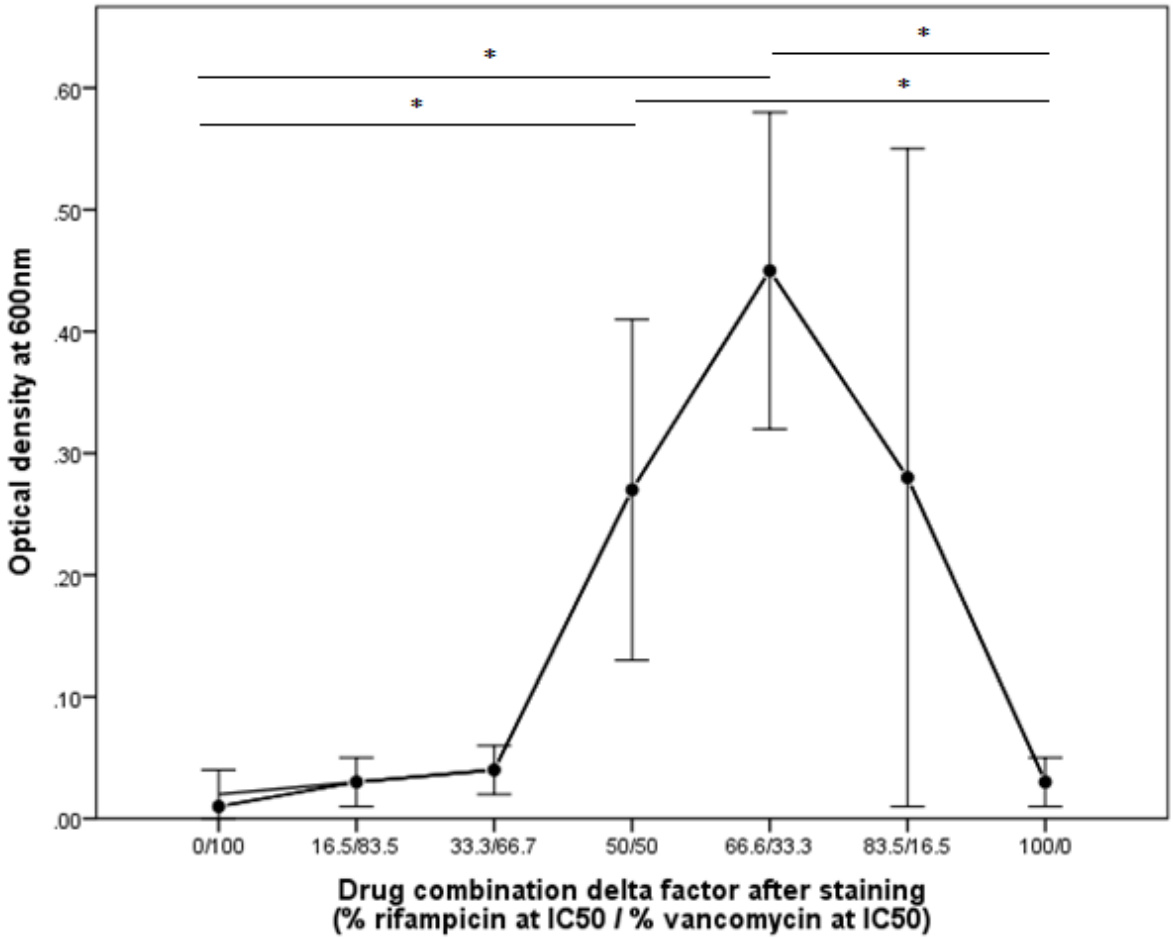


Figure3.16

MRSA252 produces significantly more biofilm in presence of both vancomycin and rifampicin, than with their monotherapies. Optical density, as a proxy for biofilm formation by MRSA252, as a function of the drug concentration delta factor or relative drug fraction θ – after staining with crystal violet. Vancomycin and rifampicin were used at their IC50 and combined at different percentages (0% rifampicin + 100% vancomycin to 100% rifampicin + 0% vancomycin). *The error bars represent the standard deviation of the four replicates. * indicates p < 0.05 from a* Mann-Whitney test *(Supplementary Data).*

Conclusion 2: the interaction of vancomycin and rifampicin towards MRSA is significantly antagonistic, in a biofilm environment.

3.4.3. Bacterial culture inside a "square dish" structure

Research question 3: what is the interaction of vancomycin and rifampicin towards MRSA (or Wcl) in a "square dish" structure?

In this pilot study using *E.coli* Wcl instead of MRSA252 (for sterility purpose), the bacteria were cultured inside a media supplemented with a small dose of agar, in presence of a gradient of vancomycin alone, of rifampicin alone and of a combination of both antibiotics (used here at 15 times the daily MIC measured in planktonic culture). This methodology aimed to provide a homogenous bacterial inoculation in the agar, along with a homogenous diffusion of both drugs, alone and in combination[24].

In the first step of the analysis of the produced "square dishes" at 15-times-MIC in 0.2% glucose, as described in *Chapter 2*, the data capture was processed to eliminate the background from the "square dishes" pictures taken at the end of the experiment. Resulting data was extracted as pixel intensity (as a proxy for bacterial density), using radial profile angles in the area of the plate associated to i) rifampicin monotherapy, ii) vancomycin monotherapy or iii) combination therapy.

For visualisation purpose, the pixel intensity in each condition was plotted as a function of the distance from the centre of the plate, allowing - if any - the direct visualisation of differences in the halos of inhibition (distance between the two peaks of pixel intensity) between the three conditions. Figure 3.16 illustrates the visualisation associated to a "square dish" representative of all performed experiments. No clear difference in the distance between the two peaks of pixel intensity were detected from this. The raw data was therefore analysed statistically.

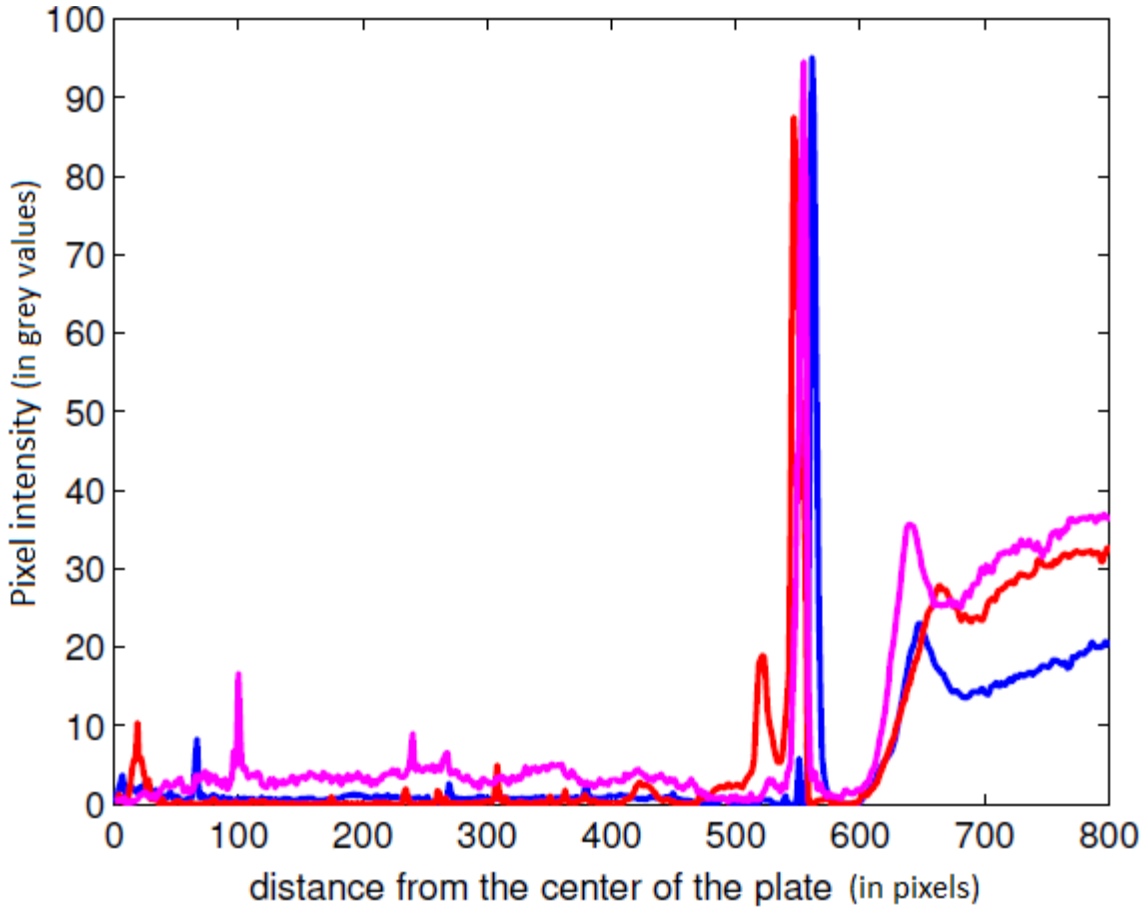


Figure 3.16

Distribution of the "pixel density" in different radial angles (localised in monotherapies or combination of vancomycin and rifampicin), as a function of the distance from the centre of the dish, using the "square dish" methodology. The pixel intensity was a proxy for *Escherichia coli* Wcl growth in 0.2% of glucose with a monotherapy of vancomycin (blue line), a monotherapy of rifampicin (red line), or in the presence of both antibiotics (pink line).

Using the radial profile raw data, the localisation of each peak of pixel intensity was measured, for each condition. By subtracting the localisation of the first peak (associated to the edge of the small dish) to the localisation of the second peak (associated to the end of the halo of inhibition), halo lengths could be measured for each condition (Table 3.4). Data was analysed using two-samples t-tests to compare the halo lengths associated to vancomycin monotherapy, rifampicin monotherapy and combination therapy (Figure 3.17). An inconclusive outcome was reported: the halo length of the combination therapy may be (significantly) shorter than the vancomycin one, yet it is longer than the rifampicin one. No interaction between these antibiotics could be interpreted from such data.

Table 3.4 (next page)

Resulting halo lengths (in pixels) between the areas where *Escherichia coli* Wcl were introduced in monotherapies of vancomycin or rifampicin, and where they were growing in the combined therapy – both with 0.2% of glucose in the media, in a "square dish". ** indicates $p < 0.05$ from two-samples t-tests (Supplementary Data).*

Experimental condition	Halo length (in pixels), representative for the halo of inhibition)		
Vancomycin monotherapy	116.6		
Rifampicin monotherapy	84.75		*
Combination therapy	85.66		

Table 3.4

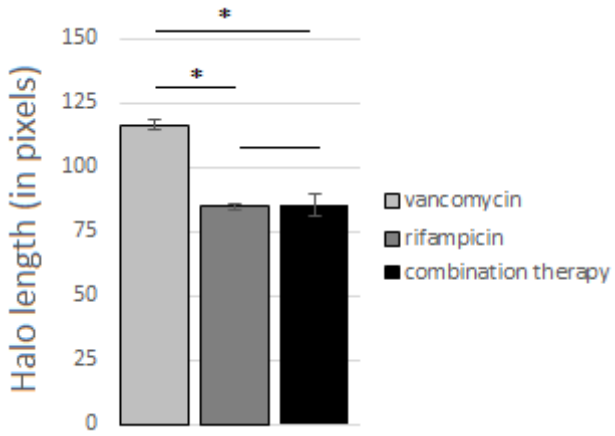


Figure 3.18

Conflicting outcomes from the halo length comparisons between monotherapies or combinations of vancomycin and rifampicin, using the "square dish" methodology. When comparing the length of the halos of inhibition of *Escherichia coli* Wcl strain in 0.2% glucose, in presence of antibiotic monotherapies or of their combination (at 15-times-Minimum Inhibitory Concentrations), the average halo lengths for the combination therapy is significantly shorter than the one from the vancomycin monotherapy – but no significance is found when comparing the combination therapy to rifampicin. D*ata represents the average halo lengths of three radial profiles in each therapy, the error bars indicate the standard deviation. Monotherapies are in shades of grey (lighter for vancomycin) and the combination is in black. * indicates p < 0.05 from two-samples t-tests (Supplementary Data).*

The presented results are representative of all observed outcomes from all experiment, regardless of the concentrations used. This could mean the experiment did not work (the observations do not meet the expectations, with a lack of significance and opposite interpretations depending on the monotherapy the combination therapy was compared to) or is not suitable to analyse the interaction of vancomycin and rifampicin towards Wcl, especially given the size of the effect that, if there was any, could be potentially not large enough to be clearly visualised with this methodology. The BioBox parameters were indeed not optimised yet at the time of this pilot study: for example, the surface of the dish was not perfectly parallel to the objective of the camera, and for sterility reasons, the incubation and the photographic captures were performed with the lid of the dish, creating reflections on the photographs. This could have decreased the sensitivity of this test, and therefore the precision of the measurement.

Another explanation for such inconclusive outcome could be associated to the reported variability in MIC for vancomycin and rifampicin in planktonic culture. These concentrations were used (at 15-times-MIC each) in this study and could have increased 15-fold times daily variations in monotherapy MIC, dys-stabilising a potential interaction between vancomycin and rifampicin.

It is important to note that the “square dish” experiment was performed using *E.coli* Wcl because of safety reasons (the BioBox was not insuring sterility and safety for other laboratory users and the environment). Given the infectious state of MRSA252 and the presented results from this preliminary study on “square dishes”, another structured environment was developed: the “layer dish” structure.

Conclusion 3: the “square dish” methodology was relatively flawed to answer the research question on the interaction of vancomycin and rifampicin towards MRSA or E. coli. This could be due to the observed variability in the Minimum Inhibitory Concentrations of both antibiotics.

3.4.4. Bacterial culture on a "layer dish" structure

Research question 4: what is the interaction of vancomycin and rifampicin towards MRSA in a "layer dish" structure?

The "layer dish" method produces a structured environment for the bacteria to grow in contact with gradients of antibiotics using unbalanced layers of soft agar. Using low or high doses of each antibiotic in separated layers inside the Petri dish could allow the visualisation of potential interaction between vancomycin and rifampicin.

Low and high doses of vancomycin or rifampicin were initially chosen according to the daily planktonic MICs of each antibiotic, starting at 2-times MIC for each drug in the pilot experiments and adjusting the concentrations according to the actual bacterial growth on the antibiotic control dishes (V1C2, C1V2, R1C2 and C1R2). At "low enough" doses, the bacteria were indeed expected to grow along the streaks on the antibiotic control dishes, enabling the possible visualisation of a synergy on the V1R2 and R1V2 dishes. Similarly, at "high enough" doses, a complete inhibition of the bacterial growth along the streaks of the antibiotic control dishes was expected, allowing the possible visualisation of a complete inhibition of the growth on the V1R2 and R1V2 dishes.

The experiment was repeated 22 times in total and did not provide any evidence of the nature of the interaction between vancomycin and rifampicin.

Firstly, despite great handling care, there were often disparities between both streaks from the same dish (Figure A9), and/or between replicated dishes produced on the same day (Figure A10), and/or between the two different control dishes produced on the same day (for example R1C2 and C1R2, which were supposed to provide a similar structure and therefore a similar bacterial growth – Figure 3.20). This variability in the bacterial growths along those streaks is

visualised as standard deviations with error bars on Figure 3.19, focussing on the rifampicin control dishes.

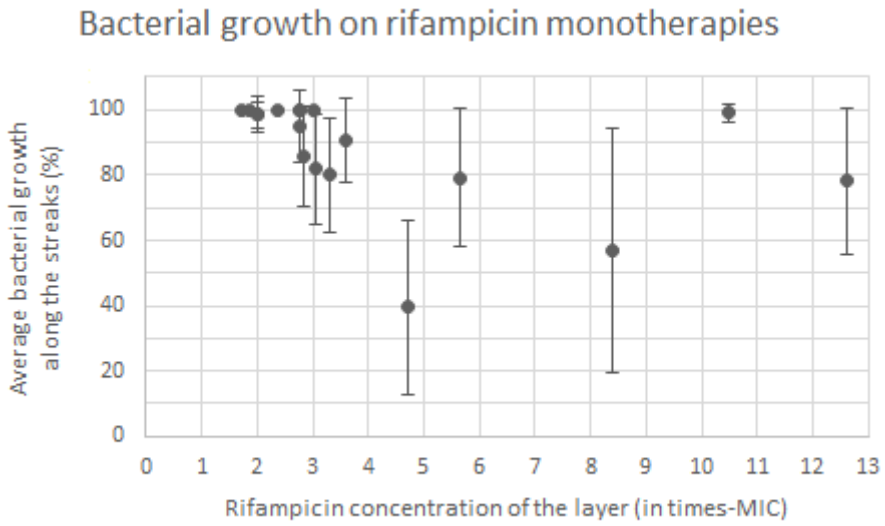


Figure 3.19

In-between replicates and in-between experiments variations in bacterial growth on rifampicin monotherapies. Rifampicin concentrations are expressed in times-Minimum Inhibitory Concentrations (MIC) inside the rifampicin layer of the "layer dish". *The average bacterial growth is expressed in percentage along the streaks, with 0% being a complete inhibition of growth, 100% being bacteria growing everywhere on the streaks and 33% being the bacterial growth on the first third of the streak – starting at the absence of antibiotic. Each data point represents the average of all streaks performed on the same day on the control plates C1R2 and/or R1C2, with two streaks per dish and up to six replicated dishes - standard deviations are shown using error bars.*

No expected Hill curve could be fitted to these data, a sign of randomness in their values and distribution as a function of the rifampicin concentration.

Secondly, it was challenging to produce either of the two full sets of conditions, by reaching either i) the highest low doses for both antibiotics simultaneously, allowing the growth along the streaks on both the antibiotic control dishes, or ii) the lowest high doses for both antibiotics simultaneously, inhibiting completely the growth along the streaks on both control dishes. As for the planktonic structure experiment (Figures 3.12A and A3), there were indeed variations in the inhibition patterns for similar concentrations in monotherapies – for example, a sufficient rifampicin concentration on one day could be either too strong or too weak on the next day, using the same protocol and the same stock solutions. Figure 3.20 depicts this for rifampicin, with for example 3-times MIC which did not inhibit the bacterial growth on one day (top dot with 100% bacterial growth and standard deviation = 0) and 3.04-times MIC which inhibited the bacterial growth at 29% on the same week (with a standard deviation of approximatively 17% for 8 analysed streaks). A representative illustration of this effect is provided in Figure A8.

Despite great care to produce these dishes in the same conditions every time (with the same stock of antibiotics and media), and as homogenous diffusion of these antibiotics inside the soft agar was assessed in the literature[204,205,207-209], these two types of variations (in-between samples and in-between experiments) affected the distribution of the bacterial growth along the streaks as a function of the antibiotic concentration in the layer: the expected Hill curve fitting the data was not retrieved (Figure 3.19). One could argue that the antibiotic concentrations were not appropriate to allow a complete inhibition of the bacterial growth (literature reporting 10-50µg/mL for rifampicin in agar media[216,217], compared to up to 100ng/mL here), yet the concentrations used here were consistent with those used with doxycycline on *E.coli* in our laboratory (up to 20-times difference between planktonic and “square dish” method[24]), moreover the common concentrations are known to follow the “hit hard and fast” paradigm[88] and were not designed for this specific type of dishes and bacteria strain. Furthermore, low doses of antibiotics were necessary to visualise a possible synergy and those concentrations were able to inhibit partially the bacterial growth.

Hypothesis could be drawn to explain these inconsistencies. A first one involves the extreme precision needed to handle these antibiotics. As seen in the

planktonic experiments, with much smaller volumes, the MIC of both antibiotics were variable. Here, bigger concentrations could impact this inherent variability, despite using bigger volumes (generally associated to less errors). A second hypothesis could be based on a possible instability of the interaction between each antibiotic and MRSA252 in this type of structured environment. It is also possible that the diffusion of each antibiotic inside the agar is not homogenous, as expected. Yet studies have being published about vancomycin and rifampicin used in agar media[218,219]. Finally, studies have reported the possible effect of circadian rhythm on laboratory experiments using bacteria, this could partially explain the lack of replicability of these experiments[366].

Altogether, and because it was difficult to reproduce the same outcomes every time the experiment was performed in full, it was challenging to validate the observations of the control conditions, and therefore to be able to observe, analyse and interpret the results provided by the combination dishes.

However, at rare occasions, both antibiotic control dishes (C1R2 and V1C2, or C1V2 and R1C2) were validated and it was then possible to observe the resulting combination dish (V1R2, or R1V2) (Figure 3.20). It was not possible to interpret such observations, with not only a lack of replicability of the experiment (the main experimental outcome in A was not retrieved on the next day, Figure 3.20B), and the two supposedly similar conditions V1R2 and R1V2 tested on the same day not providing the same outcome (Figure 3.20A versus C). The validity of this experimental methodology to answer our research question was therefore even more questioned.

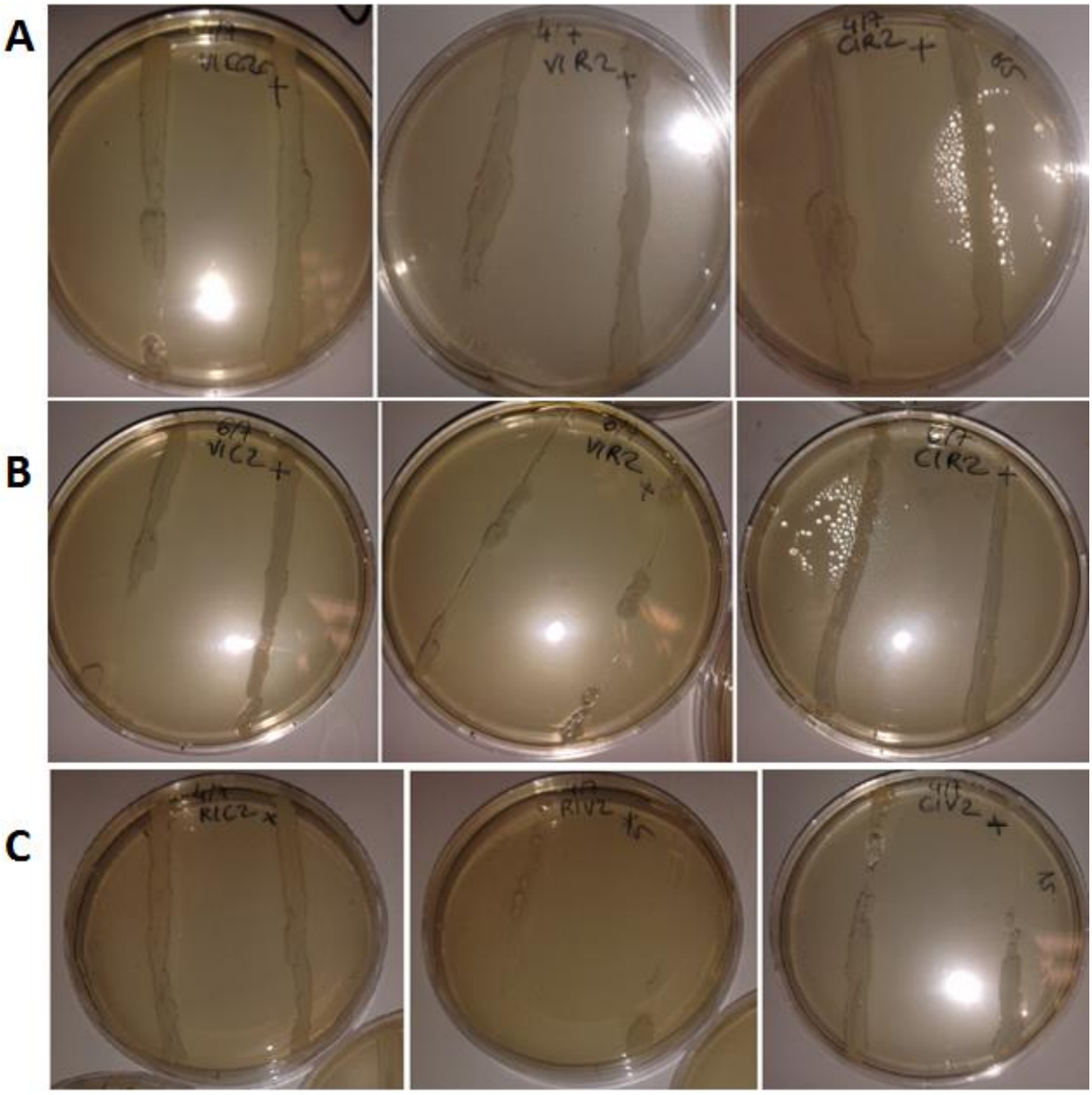


Figure 3.20

Illustration of the resulting combination dishes V1R2 and R1V2, after validation of their respective antibiotic control dishes (V1C2, C1R2; R1C2, C1V2), using low doses of antibiotics (4.6MIC for vancomycin and 2.76MIC for rifampicin). A and B - On two different days (A and B), when bacterial growth was observed on the V1C2 and C1R2 control dishes produced at high doses, different observations were made on the combination V1R2 dish. Taken alone, A could be interpreted as an absence of synergy; yet B was not interpretable. A and C – The R1C2, R1V2 and C1V2 plates in C were produced on the same day than the V1C2, V1R2 and C1R2 plates in A, and are supposed to show a similar bacterial growth along the streaks. They do not, no interaction could be interpreted from the plates in C, discrediting even more the validity of this experimental methodology to answer our research question.

Despite a great care to follow a precise protocol, this “layer dish” experiment was found incapable of providing any evidence about the type of interaction of vancomycin and rifampicin against MRSA252 in such a structured environment. This could be due to the methodology itself that was proven challenging, possibly worsened by the extreme variability of the MICs for both vancomycin and rifampicin (as seen in the planktonic experiments and in Figure 3.20). Another hypothesis could be a possible instability of the nature of the interaction of both antibiotics towards MRSA. This could be tested by improving the experiment using a third layer of agar[220] or by assessing the expected diffusions of the antibiotics in the agar[206].

Conclusion 4: the “layer dish” structure was relatively flawed to answer the research question on the interaction of vancomycin and rifampicin towards MRSA. This could be due to the observed variability in the Minimum Inhibitory Concentrations of both antibiotics.

3.4.5. Bacterial culture using Etests

Research question 5: what is the interaction of vancomycin and rifampicin towards MRSA using commercialized Etests - commonly used in clinics?

Etests (bioMérieux) are commercialized strips containing known gradients of antibiotic. Commonly used in Clinical Diagnosis laboratories for antimicrobial susceptibility testing, this quantitative technique allows for the determination of the MIC of the antimicrobial agent on agar media against the tested microorganism, after an overnight incubation. The use of Etests is not a referenced method, yet it has been validated for Gram-positive aerobic bacteria, including *Staphylococcus aureus*, by both the American Society for Microbiology and the European Society for Clinical Microbiology and Infectious Diseases[191].

Repeated monotherapies of vancomycin or rifampicin, against MRSA252 cultured on LB- or MH- agar plates, showed major variations in MIC (Figure 3.21 and Figure A12). The MIC of rifampicin on LB- or MH-agar plates was fluctuating from 0.003µg/mL to 0.012µg/mL and from 0.004µg/mL to 0.016µg/mL respectively. More importantly, the MIC of vancomycin was particularly instable, showing bacteria eradication from 0.125µg/mL and 4µg/mL (with one extra reading at 24µg/mL) for LB-agar plates and from 0.38µg/mL to 4µg/mL on MH-agar plates. These differences could not entirely be attributed to differences in the laboratory environment (or the media / agar plates handling techniques) or to the subjectivity of this methodology (two different operators were determining the MIC, separately, while following the supplied protocol and guidelines for test reading). It is interesting to note that, in most Clinical Diagnosis laboratories, commercialized agar plates and an automatic MIC reader are often provided, to avoid such human errors[56,221].

	Mean	Median	Standard Deviation
Vancomycin and LB-agar	3.14	0.75	6.67
Vancomycin and MH-agar	2.35	3	1.85
Rifampicin and LB-agar	0.0067	0.006	0.003
Rifampicin and MH-agar	0.0092	0.008	0.005

Table 3.5

Descriptive statistics regarding the distribution of the Minimum Inhibitory Concentrations for vancomycin and rifampicin towards MRSA252 using Etest on two different media. As detailed in Figure 3.21.

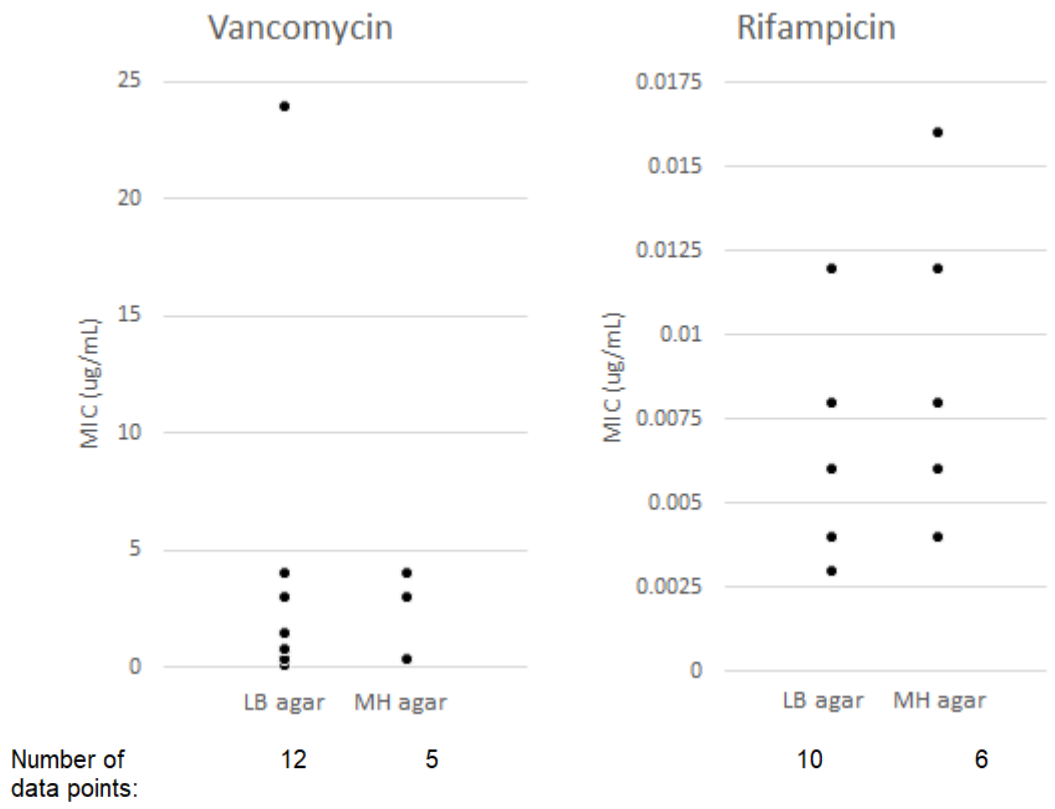


Figure 3.21

Distribution of the Minimum Inhibitory Concentrations for vancomycin and rifampicin towards MRSA252 using Etests on different media. Antibiotic susceptibility testing using Etests for vancomycin (left panel) and rifampicin (right panel) monotherapies on both Luria Bertani (LB) and Mueller-Hinton (MH) agar plates, previously streaked with MRSA252, showed an heterogenous distribution of the Minimum Inhibitory Concentrations (MICs) in most cases. *The total of data points per category is the following: (vancomycin) 12 for LB, 5 for MH, (rifampicin) 10 for LB, 6 for MH. These experiments were conducted during the same couple of weeks using the same stock material to prepare the media. Descriptive statistics are detailed in Table 3.5.*

Each day, after determining the daily MIC for vancomycin and rifampicin from the previous day, combination therapies of vancomycin and rifampicin were performed using two different Etests crossed together at their daily MIC – both combination conformations were tested (one with the vancomycin Etest applied

first on the agar plate, and the other one with the rifampicin Etest being applied first). The MIC for the combination of both drugs was determined on the next day, for both conformational plates (Table 3.6 and Figures A13-14). No bacterial inhibition was visible on any of the LB-agar plates, which could explain why their usage is not particularly recommended in Clinical Diagnosis laboratories. Considering the MHA plates (that are recommended in most Clinical Diagnosis laboratories[221]), three experiments out of the four applying vancomycin first on the plate did not inhibit the growth of MRSA252. Bacterial inhibition of MRSA252 was visible on the other five configurations, the values of the MICs in combination were varying, for vancomycin between 0.19µg/mL and 1µg/mL, and for rifampicin between 0.002µg/mL and 0.008µg/mL. With such daily variations at the MIC for monotherapies, it was not surprising to obtain such heterogeneous observations, yet the variations between the primarily application of either vancomycin or rifampicin were questioning the validity of this whole antimicrobial susceptibility testing methodology

Experiments		Observed combinated MICs for each experiment (µg/mL)			
		A	**B**	**C**	**D**
On LB agar plates	Vancomycin On rifampicin	No bacterial inhibition for these four experiments.			
	Rifampicin On vancomycin				
On MH-agar plates	Vancomycin On rifampicin	Vancomycin: 0.5, Rifampicin: 0.006	No bacterial inhibition for these three experiments.		
	Rifampicin On vancomycin	Vancomycin: 0.75, Rifampicin: 0.008	Vancomycin: 1, Rifampicin: 0.003	Vancomycin: 0.5, Rifampicin: 0.002	Vancomycin: 0.19, Rifampicin: 0.003

Table 3.6

Example of several experiment outcomes of antimicrobial susceptibility testing of combination of vancomycin and rifampicin. The combinated Minimum Inhibitory Concentrations (MICs) of Etests, on Luria Bertani (LB) and Mueller-Hinton (MH) agar plates streaked with MRSA252, are expressed when bacterial inhibition was reached in the experiment. *These experiments A-D were conducted during the same couple of weeks using the same stock material to prepare the media*.

When used in Clinical Diagnosis laboratories, the FICI is calculated to determine if the combination of two drugs is synergistic or antagonistic, or if it is indifferent to the bacteria to be subjected to either monotherapies or combination therapy. The FICI data of the experiments on LB and MH-agar plates, for both conformations of combination of vancomycin and rifampicin, is presented in Table 3.7. Here, the combination of vancomycin and rifampicin appeared to not be beneficial to the eradication of MRSA252, as it is associated to indifference. Interestingly, indifference is sometimes limited with FICI > 2.0 in the literature, yet the presented data is still associated with indifference according to this definition.

Experiments		**FICI result for each experiment**			
		A	**B**	**C**	**D**
On LB-agar plates	Vancomycin on rifampicin	NA			
	Rifampicin on vancomycin				
On MH-agar plates	Vancomycin on rifampicin	0.83 (I)	NA		
	Rifampicin on vancomycin	1.16 (I)	1.25 (I)	0.58 (I)	1.26 (I)

Table 3.7

Example of several Fractional Inhibitory Concentration Index (FICI) of antimicrobial susceptibility testing of vancomycin and rifampicin in Etests combinations, on Luria Bertani (LB) and Mueller-Hinton (MH) agar plates streaked with MRSA252. FICI ≤ 0.5 indicates a synergy (S) of both drugs against the microorganism, 0.5 > FICI > 4.0 to an indifference (I) and FICI ≥ 4.0 to an antagonism (A).
NA indicates an absence of validated MIC data for the condition (as seen in Table 3.6).

As the FICI calculation considers the MIC of the antibiotic used alone, the FICI was also calculated for these experiments with the MIC of the monotherapies obtained on the same day of the combination (as developed in *Supplementary Data*). It did modify the FICI prediction of indifference between vancomycin and rifampicin in the case of rifampicin on vancomycin on a MH-agar plate: synergy was detected (Table A1).

It seemed interesting to observe what would happen in terms of bacterial growth if LB- or MH-agar plates, previously streaked with MRSA252, were subjected to a combination of the same drug, at its MIC in monotherapy (with Etest(A) and Etest(B) being both vancomycin, for example). This would allow to test the validity of the use of Etests and/or the FICI measure themselves, as the combination of the same dose of the same drug should be equal to the double of the dose. As presented in Table 3.8, all experiments concluded with an indifference of the use of these “combinations”. The FICI was also calculated for these experiments with the MIC of the monotherapies obtained on the same day of the combination (as developed in *Supplementary data*). It did modify the FICI prediction of indifference for two double vancomycin combination, the one on MH-agar was associated to antagonism and the LB-agar one to synergy (Table A1). This data should be considered if Clinical Diagnosis laboratories are only using the Etest methodology to determine the antibiotic regimen to prescribe to their patients. FICI may not be the best value to interpret antibiotic interactions on Etest. The Weighted Cohen’s kappa statistic may be of better use, yet its interpretation and use are not straightforward due to reported prevalence and bias effects[188,222].

Experiments		Observed combinated MICs (µg/mL) and FICI result for each experiment							
		A		B		C		D	
		MIC	FICI	MIC	FICI	MIC	FICI	MIC	FICI
On LB-agar plates	Double vancomycin combination	**0.38 and 0.5**	**1.17 (I)**	MRSA growing		**0.38 and 0.25**	**0.84 (I)**	MRSA growing	
	Double rifampicin combination	MRSA growing		**0.003 and 0.003**	**2 (I)**	NA		NA	
On MH-agar plates	Double vancomycin combination	NA		NA		**1 and 0.75**	**1.16 (I)**	**0.25 and 0.25**	**1.55 (I)**
	Double rifampicin combination	**0.003 and 0.004**	**0.58 (I)**	NA		NA		**0.003 and 0.004**	**1.75 (I)**

Table 3.8

Example of several experiment outcomes (in Minimum Inhibitory Concentrations (MIC) and Fractional Inhibitory Concentration Index (FICI)) of antimicrobial susceptibility testing of Etest combinations of the same antibiotic, on Luria Bertani (LB) and Mueller-Hinton (MH) agar plates streaked with MRSA252. These experiments were conducted during the same couple of weeks using the same stock material to prepare the media. FICI ≤ 0.5 is correlated to a synergy (S) of both drugs against the microorganism, 0.5 > FICI > 4.0 to an indifference (I) and FICI ≥ 4.0 to an antagonism (A).
NA indicates an absence of validated experimental condition; no MIC was retrieved and therefore no FICI could be calculated.

Conclusion 5: No synergy was detected between vancomycin and rifampicin towards MRSA using Etests. Yet, the validity of such method is questioned (combinations of the same drug, FICI data depending on the variable MIC).

3.5. Discussion

The combination of vancomycin and rifampicin is commonly used in practices to treat MRSA infections[82,83]. Nevertheless, this regimen has been studied over the last decades in a variety of clinical studies, and reports lead to contradicting results: some demonstrated a successful combination, in which the antibiotics synergise effectively, and the others a failing therapy which could trigger the emergence of antibiotic resistance[55]. These inconsistencies in the literature could be due i) to the antibiotics themselves (their interaction may not be synergistic, or could be inherently unstable[44,46,48,49,51]), or ii) to inter-individual differences between the patients[45,82,83,92,95-98], or iii) to differences between the parameters of the studies (quality of the antibiotics, dosages, type of infections, etc. [55,71,72,186-196]). By studying the inhibitory effect of the combination of vancomycin and rifampicin on MRSA252 in fresh cultures in laboratory, inter-individual differences could be removed from the equation, and by doing so in five different experimental settings, in the same laboratory environment with the same operator using the same equipment and stock of supplies, it could be possible to assess the effectiveness and stability of the combination of vancomycin and rifampicin. Overall, we consistently rejected the synergy of vancomycin and rifampicin towards MRSA252, and we also observed that some experimental tests were more suitable to answer our research question than others.

The planktonic experiment in a shaking-liquid media is very similar to the standardised serum bactericidal test[45]. It produced a non-structured environment which did not favour combination therapy over monotherapies: the bacteria were significantly growing less in presence of only vancomycin or rifampicin at IC50 than with their mixed combination. The biofilm experiment, producing a more biologically relevant structured environment[33,45,197], delivered the same outcome of significant antagonism between vancomycin and rifampicin, quite unsurprisingly given the literature[68,197,223]. There was no clear result from the two solid agar-plate experiments that produced structured environments (the “square dish” and the “layer dish” ones), possibly due to variabilities in the antibiotic concentrations – as they were particularly sensitive to minor handling differences from one day to another – but also to the difficulties to follow their protocol precisely. Finally, the use of reagent strips (Etests) of antibiotics on top of agar plates showed no sign of synergy between those drugs – indifference was observed. Yet, the

validity of this commonly used methodology was questioned using supposedly indifferent combination of the same monotherapy – in support of the lack of evidences on this test and on the computing of FICI data[56,224]. Overall, it is therefore possible to conclude from these five types of experiments that no data supports a synergy of vancomycin and rifampicin in such laboratory settings. As laboratory experiments provide much simpler infection models than clinical infections and are easier to monitor closely[225,226], one could wonder the impact inter-individual differences (patients' immune system, xenobiotic metabolism, etc.) could have on the nature of the interaction of combination of vancomycin and rifampicin against MRSA. These differences have indeed been flagged as capable of modifying the pharmacokinetics/pharmacodynamics of drugs[98,101,119,173,227-234], and this metabolism has been reported to degrade rifampicin[117,127,133,135,137,138]. By modifying the circulating concentrations of rifampicin, inter-individual differences could therefore alter an already very sensitive interaction between vancomycin and rifampicin. However, correlations between *in vitro* and clinical settings have been long questioned[45].

Vancomycin and rifampicin are two antibiotics which seemed particularly difficult to use at the right dose in the laboratory, therefore we wonder if the reported synergies[71,73,75-77,79,81] could have been due to the variabilities in their MIC, leading to changes in their subtle and sensitive concentrations equilibrium. Not only MIC is known to not be a fixed measure - as it is sensitive to its environment (temperature, pH, concentrations of available metabolites, etc.[24,191]) – but any use of rifampicin should be cautious as it is deemed unstable[235,236]. Indeed, rifampicin is readily oxidised in alkaline media, and hydrolysed in highly acidic media[367]. Very sensitive to moisture, rifampicin is also affected by light, air and oxygen. Its half-life in patients is estimated between 1.5 and 5 hours[133].

Furthermore, differences in MIC readings and analysis, for the same strain, have been reported in the past, as well as the association of vancomycin to increased MIC for some common clones[56,237-239] – the variability in MIC of vancomycin and rifampicin should not be surprising here. In patients, drug gradients have been reported[240], they can also destabilise the required balanced therapy, while acting as a stronger pressure selecting for resistance. Given how difficult it was to inhibit the growth of MRSA252 in

those laboratory settings when using vancomycin and rifampicin in combination, and given the possible effects that inter-individual differences in patients could have on the pharmacokinetics/pharmacodynamics of rifampicin (and therefore on this very sensitive combination), the use of alternative regimen should probably be advised to treat MRSA infections while avoiding antibiotic resistance. Indeed, synergistic combinations are more than needed when it comes to the use of vancomycin and rifampicin: counterintuitively, antagonist combinations may have been of interest[49,52,54] yet, they require high dosages, which will be nephrotoxic for vancomycin, and hepatotoxic as well for rifampicin[45,53,67].

Further studies should be carried out to conclude on the combination of vancomycin and rifampicin on all five experiment settings to bring along a clearer result. The most challenging part of the presented experiments was without any doubt the variations in vancomycin and rifampicin MICs in monotherapies, in planktonic cultures. Even with the best possible care, differences were observed at the MIC levels between the replicated experiments. These MICs being used to determine the IC50 to use in the combination of vancomycin and rifampicin, it was not surprising it was challenging as well to produce, on the same microplate, a perfectly balanced combination therapy. Yet, it could be interesting to reproduce the experiment with a final subculture in antibiotic-free media, to observe the growth of the content of the wells after antibiotherapy, or to re-inoculate the D3 microplate with fresh MRSA at the end of the experiment, as a control for drug degradation[24]. The amount of living bacterial cells could have been quantified in both the planktonic and the biofilm structures[241]. A more precise data capture method could have been used to analyse the “square dish” halos. The “layer dish” experiment seemed very difficult to handle as well. Higher concentrations could have been tested in order to observe an antagonism – if any. Anyway, the antibiotic concentrations of these two structured environments were based on those, particularly unstable, retrieved from the planktonic experiment, therefore difficulties to even replicate an observed outcome should have been expected. Given Etests are commercialised gradients of antibiotics, the difficulty in replicating their results from one day to another was particularly worrisome, especially as these strips are often used in clinics to test the antimicrobial susceptibility of the patients’ strains[56]. The precise application of each strip required precise handling skills

and their use in combination was more than delicate. However, it was striking to realize that MRSA252 was mostly indifferent to the use of any combination, even those of two strips of the same antibiotics (where additivity should be expected). With more material, another combination technique involving two Etests placed on top of each other[190,211,212], still challenging to handle, could have been tested to help figuring out if Etests and their FICI data are good indicators of the effectiveness of antibiotic combinations[188,222]. It could have been interesting to re-evaluate the diffusion of the antibiotics in the agar, especially given the natural instability of rifampicin[235,236]. For all five experiments, it could maybe have been beneficial to pre-culture a single colony of MRSA252 before a shorter overnight culture, to avoid it evolving into different sub-populations during this overnight culture. For all experiments but the Etest protocol, dosages higher than IC50 (but lower than MIC) for both antibiotics could have been tested. It could have been also interesting to assess the combination of vancomycin and rifampicin in a single-cell experiment [242-244], or to carry out an agar dilution experiment (using soft agar inside the wells of a microplate[188,194,213]) to study the susceptibility of MRSA252. Another MRSA strain could also have been tested.

No evidence of a synergy between vancomycin and rifampicin, in the context of a MRSA infection, was found in this body of work using five different antibiotic susceptibility testing methods. It was established that these antibiotics are very instable: there are critical variations in their MIC, which may have led to some contradicting results in the past. As some of these clinically-relevant methods were proved to be particularly difficult to handle, and to provide a clear and definitive answer to the type of interaction vancomycin and rifampicin have, it is crucial to either recommend the usage of a less controversial regimen[82,83], or conduct a comprehensive clinical studies on this mixed combination therapy - considering important patient parameters such as their immune system and xenobiotic metabolism, both of them playing a role to fight against their MRSA infection[98]. The clock is ticking, antibiotic resistance has been predicted to cause 10 million deaths per year by 2050[19], and MRSA is one of the main actors of this terrible projection.

3.6. Reflection on the research

In this chapter, five types of experiments were performed in order to answer our research question about the interaction of vancomycin and rifampicin towards MRSA.

Two of them were successful to deliver an answer (significant “antagonism” for both the planktonic environment and biofilms), but the two experiments involving a structured environment using agar (“layer dish” and “square dish” structures) failed to answer the research question, while bacterial cultures using Etests provided more questions on its validity than answers about our research theme.

The "square dish" environment was supposed work inside the laboratory BioBox. The parameters of this laboratory-made automated photography chamber were unfortunately not perfectly set at the time of the experiments. Given more time, the BioBox could have been impeccably fitted to this experiment, allowing us perhaps to conclude on the type of interaction between those two antibiotics. During the BioBox refit, another structured environment could even have been tested: for example, we could have inoculated, with the same amount of bacteria, a large series of agar dishes - each of them containing a different combination of vancomycin and rifampicin (in terms of concentrations and proportions) - in order to spot the combinations inhibiting the growth of MRSA.

The “layer dish” environment was supposed to provide an improved approach compared to the "square dish" structure, yet it was limited to the use of low antibiotic doses to see synergy or high doses to spot antagonism, and it was very sensitive to fluctuations in the planktonic MICs used to set the layer concentrations. In case allowing more time on the “square dish” method did not provide a better outcome, repeating this “layer dish” experiment would perhaps not help, but it could have been good to validate its methodology by dosing the antibiotic concentrations along the streaks.

Finally, the bacterial culture using Etests gave rise to a plethora of questions about the validity of such antibiotic susceptibility testing method. Without any restriction on our budget (Etests are very expensive), we would have investigated even more on these commercialized strips. Do they diffuse well the antibiotics inside the agar? What about the antibiotic degradation? What is the best combination technique? Why does a combination of two strips from the same batch (same antibiotic) seem to be linked with auto-indifference of the antibiotic? Is the FICI data reliable? What about the Weighted Cohen’s kappa coefficient? All of these open questions could definitely be conducted in a new research project on its own.

CHAPTER 4 - *In silico* investigation of the potential impact of the most common genetic variants on the *CYP3A4* gene expression or enzyme activity.

4.1. Introduction

Xenobiotic metabolism is the process in which enzymes, mostly hepatically expressed, degrade and excrete compounds that are detected as foreign from the organism – including medicines. This biotransformation has been extensively studied for decades, with the main objective of avoiding drug-drug interactions: one drug stimulating (or inhibiting) the degradation of another, leading to a low treatment efficiency (or to the emergence of side-effects)[108,245]. In the case of antibiotics, this metabolism can affect the circulating concentrations, impacting therefore on the critical dose required to eliminate all bacteria while avoiding antibiotic resistance[27,91,113].

Amongst the xenobiotic metabolism enzymes, the most common are the cytochrome p450 (CYP) enzymes[101,108]. This superfamily of 57 mono-oxygenases takes part in Phase I of the xenobiotic metabolism, in which CYP-1A2, -2B6, -2C9, 2C19, -2D6 and -3A4/5 are the most popular members[101]. They are highly expressed in the liver and account for the metabolism of around 75% of the clinically relevant drugs – including antibiotics[101]. Differences in various drugs metabolism have been reported between patients[112], many have been attributed to polymorphisms at genes coding for xenobiotic metabolism enzymes[98,101,106,110]. 20-25% of patients are estimated to be possibly clinically affected by genetic variations (mainly Single Nucleotide Polymorphisms, or SNPs) on the *CYP3A4/5, CYP2C9, CYP2C19* and *CYP2D6* genes[113]. For example, the presence of the minor variant for either rs1799853 and rs1057910 on *CYP2C9* has been shown to dramatically affect the metabolism of warfarin: patients poorly metabolising this cardiovascular drug require a 75% decreased dosage[114]. Similarly, Phase II xenobiotic metabolism *Glutathione S-Transferase Mu 1* (*GSTM1)* and *Theta 1* (*GSTT1)* genes are expressed as non-functional null alleles in patients[115,116], increasing for example for *GSTT1* in tuberculous patients their likelihood of experiencing treatment toxicity to isoniazid, pyrazinamide and rifampicin[117].

Both *CYP3A4* and *CYP3A5* genes code for two distinct enzymes metabolising about 30% of the clinically relevant drugs altogether[246], classify them as Very Important Pharmacogenes (VIP)[245,247]. Their enzymatic broad substrate specificities often overlap, rendering their separate study challenging[248,249]. Most clinically relevant drugs (including antibiotics) are metabolised in the liver[108,250], where *CYP3A4* expression is four-fold higher than *CYP3A5* expression[251]. *CYP3A4* is therefore of great interest, and yet its genetic variants have been less characterized than those associated to *CYP3A5* - there is three-times less pharmacogenomics-based prescribing information available in the literature for CYP3A4 than CYP3A5[252].

CYP3A4 presents thousands of SNPs[253], although their impact on gene expression or enzyme activity have not being indisputably established[251]. Some studies reported an association between *CYP3A4* SNPs and differences in the metabolism of several drugs (including the immunosuppressant cyclosporin and the oestrogen receptor modulator tamoxifen)[228,234,254-257]. Notably, the minor variants for rs35599367 and rs2242480 were associated to, respectively, a decreased statin clearance (a cardiovascular drug) and an increased R-warfarin clearance (an anticoagulant). But contradictory results have being published, for example the same rs2242480 has been shown to not influence the pharmacokinetics and pharmacodynamics of ticagrelor (an platelet aggregation inhibitor)[231,258-261]. This could potentially be explained by a compensation of the enzymatic activity by another enzyme - possibly CYP3A5[248,262-265]. Another hypothesis could be drawn based on both i) a high linkage equilibrium between the *CYP3A4* SNPs and ii) the very low Minor Allele Frequency (MAF) of most of them[234,253]. The *CYP3A4* SNPs are indeed sometimes found grouped in the same haplotype and studied together[266], perhaps making a direct correlation between one SNP and its specific impact of the *CYP3A4* gene expression or enzyme activity difficult to evidence. Understandably, for one variant, the lower its MAF, the less likely it will be found in a particular studied population, and consequently its studied sample size will be low and associated to a limited significance. Ethnic differences have been reported for pharmacogenetics polymorphisms: studies have shown differential distribution for some xenobiotic metabolism SNPs around the world[110,118,121,122]. This was notably evidenced in Caucasian populations expressing many of the deleterious *CYP2D6* SNPs[123], or the complete *GSTM1* or *GSTT1* gene deletion found in up to 50% of the patients in populations from European descent[116].

The aim of this study was therefore to identify *CYP3A4* SNPs harboured by a large proportion of the world population, to evaluate *in silic*o their potential impact on the *CYP3A4* gene expression or enzyme activity, to finally classify them with a score of potential impact (derived from the type of evidence, and both the nature and the localisation of the variation). It was also aimed to consider their distribution amongst different ethnicities with a view to explaining geographical differences in drug response. If the predictions presented in this study are validated by *in vitro* and clinical studies, this could help establish a list of relevant *CYP3A4* variants to genotype in patients, to be able to tailor their drug regimen (choice of drugs, and dosage of said drugs) according to their genetic background. This would allow an improved treatment efficiency while avoiding the emergence of side-effects. CYP3A4 being known to metabolise several antibiotics (such as the following antibiotics: clarithromycin, clindamycin, erythromycin[101]), adapting their regimen to the patients' genetics could also help fighting against the rise of antibiotic resistance.

4.2. Materials and Methods

4.2.1. Sequences

The human *CYP3A4 gene* sequences (as complete coding DNA sequence, cds) and messenger RNA (mRNA) sequences (as transcript variants sequences) were extracted from NCBI's Nucleotide or the UCSC Genome Browser (https://www.genome.ucsc.edu), in FASTA format. The *CYP3A4* gene sequence was found under the reference NG_008421.1 (GenBank) or at chr7:99756960-99784188 (UCSC Genomic Sequence). *CYP3A4* transcript variants 1 and 2 (mRNA) are associated respectively to NM_017460.5 and NM_001202855.2 (NCBI Reference Sequence). Detailed information is available in Table B1.

4.2.2. Identification of the CYP3A4 variants of interest

The *CYP3A4* gene is located on the minus chromosomal strand, yet its genetic variants have been reported on either strand (i.e. with “A > C” equivalent to “G > T” for the same SNP, depending on the use, or not, of the reference strand). In this body of work, for standardisation purpose, the SNPs are expressed as referenced on dbSNP Short Genetic Variations browser (https://www.ncbi.nlm.nih.gov/projects/SNP).

The UCSC Genome Browser (on the most recent human assembly and with the most recent Short Genetic Variations database – respectively GRCh38/hg38 and dbSNP17) was used to identify all genetic variants in the gene, and to localise them on it (the selected options are detailed in *Supplementary Data*). The selection of *CYP3A4* variants of interest was undertaken using an *in silico* pipeline (Figure 4.1), to select the most “common” variants which have the potential to be relevant to a large proportion of the world population. On UCSC Genome Browser, “common” variants were defined as having a global MAF above the 1% threshold (as documented on the UCSC Genome Browser, *Supplementary Data*). They also had to be documented in the 1000Genome Project (http://www.internationalgenome.org/): this project being considered the largest catalogue of human variation data, variants not referenced in it are unlikely to be found in large proportion of the world population[178].

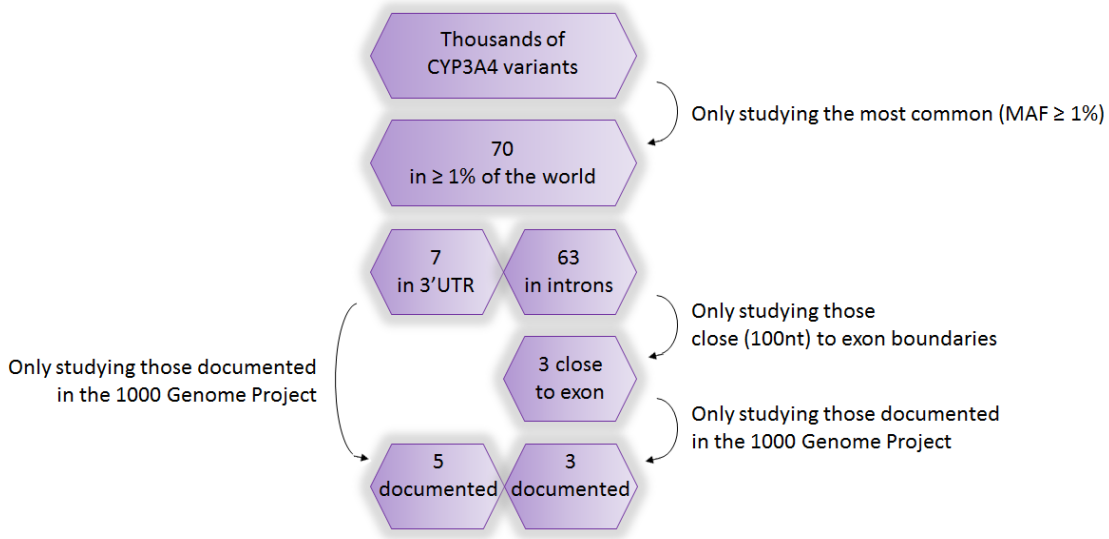


Figure 4.1

Overview of the methodology to select the *CYP3A4* variants of interest. Out of the 3,100 variants on the *CYP3A4* gene, 70 are found in at least 1% of the world population: 7 are located in the 3'-UnTranslated Region (UTR) and 63 are in introns. Amongst the 7 exonic variants, 5 are documented in the 1000Genome Project. Between the 63 intronic variants, 3 are situated at a maximum of 100 nucleotides (nt) from the exon boundaries (as detailed *in-text*) and were selected accordingly - due to a potential predicted impact on the gene expression or the enzyme activity (*see text*). All 3 selected intronic variants are documented in the 1000Genome Project. This study focuses on *CYP3A4* "common" variations in order to predict, if any, an impact of the variant on gene expression or enzyme activity which has the potential to be relevant to a large proportion of the world population. "Common" variations were therefore defined as those associated both to a world Minor Allele Frequency (MAF) of 1% at a minimum, and those documented in the 1000Genome Project (a project considered the largest catalogue of human variation data). Variations with below-threshold MAF or not documented in the 1000Genome Project were assumed to be infrequently distributed in the world and therefore not relevant to a large proportion of the world population.

As developed in *Chapter 2,* the SNPs were pre-selected when they were localised either in introns in vicinity from the closest exon (at a maximum of 100 nucleotides from the exon boundary), on in exons. In intron, it is assumed that the closer to the exon boundary, the more likely the sequence contains regulatory elements[159,165,167,170,267-271]. When variants are localised in such vicinity of the

boundaries, they can alter the expression of a gene or the activity of its protein, by affecting the binding of regulatory Serine-aRginine rich (SR) proteins[158] on intronic splicing enhancers or silencers (modulating alternative splicing in both directions)[164,169]. They can also impact on 5'- or 3'- splice sites, and branch points, necessary for alternative splicing. When in the 3'-UnTranslated Region (UTR), genetic variations could lead to amino-acid changes in translated proteins[272], alterations to microRNA (miRNA) binding sites (gene stability)[160] and, less likely, exonic splicing enhancers or silencers. Variations altering the protein sequence could lead to a non- or dys-functional protein, impacting on its activity[272]. Variations altering the *cis*-regulatory elements could lead to alternative splicing: the gene could be transcribed into different mRNAs leading to their translation into different protein isoforms which could have different enzymatic activities[164]. Finally, variations altering the miRNA binding sites could trigger differences in gene silencing, affecting its expression and therefore the level of mRNA being transcribed[273-276], and variations altering the DNA or RNA conformations could impact indirectly on the binding of regulatory elements. Overall, the localisation of the SNPs (in introns close to exon boundaries or in exons) was assessed using Alternative Splicing Graph from the Swiss Institute of Bioinformatics[159,165-167,170,267-271] (Figure B1).

4.2.3. In silico searches to predict the potential effect of CYP3A4 variants

Untranslated and intronic variants have been reported to be potentially able of affecting the gene expression or enzyme activity by altering transcription, alternative splicing or silencing mechanisms[158,160,272], or DNA/RNA conformations. Several open access databases were used to predict such potential effect of the *CYP3A4* variants. The searches on the following integrated predicting tools focussed on localising the *CYP3A4* sites on which transcription factors, *trans*-regulatory proteins, miRNA or splicing factors could bind. Several databases were perceived as redundant, yet their concomitant use is established[277,278], possibly because of their differing default thresholds which have both being recommended[277,279-283]. Unless otherwise specified and argued, the pre-set parameters were therefore used (as detailed in *Supplementary Data*).

RegRNA 2.0 (http://regrna2.mbc.nctu.edu.tw/) is an integrated website predicting functional RNA motifs from inputted RNA sequences[279,284]. This database was searched to find potential alterations by the variants of interest on transcription factor binding sites, polyadenylation sites, ribosome binding sites and RNA-RNA interaction regions (including miRNAs). The scan function was performed twice, to allow for the comparison of the outputs from *CYP3A4* RNA sequences with or without the presence of the minor variants of interest[285].

Poly-miRTS on microRNA binders (http://compbio.uthsc.edu/miRSNP/) was also accessed to locate the miRNA target sites on the *CYP3A4* gene[286-288] . They are classified into four classes, depending on their likelihood to alter the repression control of the gene. By inputting the gene name into the search option, the outputs associate, when possible, miRNA(s) to each *CYP3A4* exonic SNPs, alongside a function class predicting the functional impact of the variants[289].

The potential impact of both intronic and exonic SNPs of interest on alternative splicing was studied using the Human Splice Finder 3.1 (http://www.umd.be/HSF3/), Fruit Fly (http://www.fruitfly.org/seq_tools/splice.html), ESE Finder 3.0 (http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home) and SpliceAid 2 (http://193.206.120.249/splicing_tissue.html). Human Splice Finder 3.1[282] was used with the "performing a quick mutation" tool to mutate the sequence at the SNP locus from major to minor variants. Fruit Fly[290] was accessed by using the whole sequence of the gene (with or without the expression of the minor allele for the variants of interest) and comparing the two outputs (list of acceptor or donor sites). Likewise, overlapping 4900nt gene sequences, with or without the expression of the minor alleles, were inputted into ESE Finder 3.0[283,291]. Even smaller overlapping gene sequences were used in Splice Aid 2[292] to compare the output between the presence and the absence of the minor variants of interest in said sequences.

As CYP3A4 metabolism is predominant in the liver[108], the hepatic expression of all flagged miRNA and splicing proteins was checked using GeneCard (https://www.genecards.org/) and GTex (https://gtexportal.org/home/).

Genetic variants can impact on the gene expression or enzyme activity by affecting the folding conformation of the DNA, RNA or protein sequences (for example, DNA sequences of different conformations are differentially recognized by regulatory proteins, depending on the physical availability of the motif they bind to)[293-295]. Differences in folding conformations on both DNA and RNA sequences by the expression of the minor allele for each selected locus were assessed using the mfold Web Server (http://unafold.rna.albany.edu/?q=mfold) with the pre-set parameters of temperature and concentrations in sodium and magnesium cations (as detailed in *Supplementary data*). A series of structures, differing in "initial free energies", were provided for each sequence (type of genetic material, with or without the major variant being expressed). The presented results are associated to those ranked as most predictive with their initial dG and are being representative of them all.

The scientific literature regarding the SNPs of interest was accessed with their Reference SNP (RefSNP) reports on the dbSNP Short Genetic Variations browser (https://www.ncbi.nlm.nih.gov/projects/SNP) - which is linked to both PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and LitVar (https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/LitVar/#). SNPedia (https://www.snpedia.com/) and PharmGKB (https://www.pharmgkb.org/) were also searched for this purpose.

This literature search was performed in a systematic manner, using directly their current Reference SNP identification number (rs#) in the search bar: all publications referencing the selected SNPs in these databases, available in open access through Pubmed and written in English were considered for characterisation of the effect of the *CYP3A4* variant on gene expression or enzyme activity. These evidences were classified depending on their type (predictions, or laboratory study, or clinical study) and their power – considering various study parameters such as: was the study directly associating the SNP to the effect? Was the study including homozygous individuals for the minor variant of the SNP? Has the study a large sample size? Was the study controlled, randomised and/or including a placebo group? Was the study ethnically heterogenous?

4.2.4. Distribution of the genotypes in the world populations and in different ethnic groups

As developed in *Chapter 2,* and because *CYP3A4* variants are often inherited together due to their linkage disequilibrium (LD)[176,297], the D' and the R^2 of those variants of interest, assessed in pairs, were calculated using the world frequencies option on the LDLink matrix and pairwise tools (https://ldlink.nci.nih.gov/)[175].

The MAFs of the variants of interest were directly assessed using the open access 1000Genome Project data (http://phase3browser.1000genomes.org/index.html) which was based on the GRCh37 reference assembly then updated to GRCh38/hg38. It displays the world MAF, and MAFs associated to different ethnic groups (amongst the 2,504 individuals from this study, 661 were originating from Africa, 347 from Americas, 504 from East Asia, 489 from South Asia and 503 from Europe). Full information regarding the 2,504 individuals from these ethnic groups is available in a previous study and in *Chapter 2*[178].

By using the options "individual genotypes", it was possible to build the genotype profile of all individuals for all of the studied SNPs. For clarity purpose, terms such as "Europeans" will be used through this chapter for "individuals from European ancestry populations".

4.3. Results

4.3.1. Identification of the most common and potentially altering CYP3A4 variants

All 3,100 *CYP3A4* variants were considered in this study, they were counted on the gene directly with db150[253]. Using the UCSC genome browser with the Human GRCh38/hg38 Assembly from December 2013, their distribution was assessed: it was uniform (Figure B1). The majority of *CYP3A4* variants appeared to be SNPs (variations affecting only one nucleotide) which were associated to a "reference SNP ID number" (namely, an accession number for referenced SNPs, such as "rs123456789"). Most of these SNPs were intronic, the other variants being either located in UnTranslated Regions (UTRs) or in one of the 13 *CYP3A4* translated exons (75% of them coding for non-synonymous codons).

This study focuses on variations associated to MAF above 1% of the world population, in order to predict, if any, an impact of the variation on gene expression or enzyme activity which can be potentially relevant to a large proportion of the world population. On the genome browser, only 70 “common” variants were displayed (Figure 4.2): 7 of them were SNPs in *CYP3A4* 3’-UTR and the remaining 63 were SNPs located in introns. No “common” variant was found in the 5’-UTR of *CYP3A4.* No “common” variant was found in coding regions neither, where insertions-deletions (indels) are prone to impact on the gene or its enzyme, by causing frameshifts.

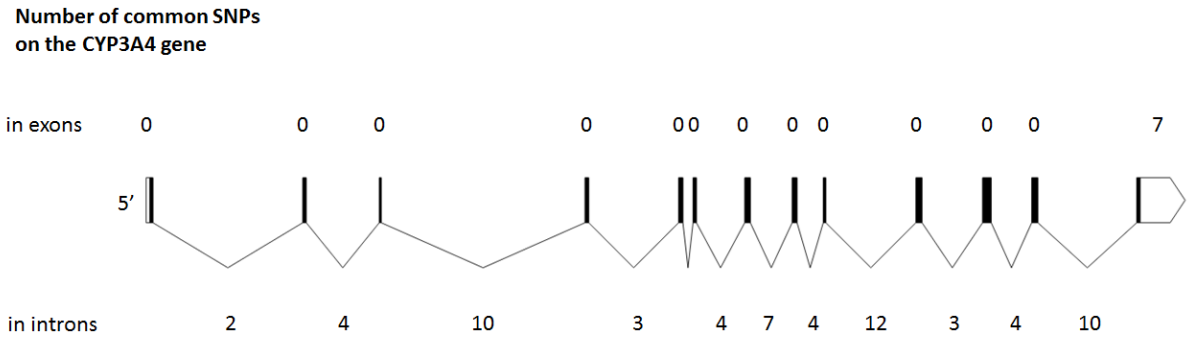


Figure 4.2

Distribution of the 70 “common” Single Nucleotide Polymorphisms (SNPs) along the *CYP3A4* gene. The 70 “common” *CYP3A4* SNPs are visualised from the 5’-UnTranslated Region (UTR) to the 3’-UTR (both in white) – i.e. between chr7:99,756,722 and chr7:99,784,327. The number of exonic SNPs are provided above the gene drawing, the position of these numbers indicating on which exon they are located. Similarly, the number of intronic SNPs are given below the gene drawing, the position of the number indicating on which intron they are located. *“Common” SNPs are defined here as associated to a world Minor Allele Frequency of 1% at a minimum. Black vertical boxes represent the exons, black lines the introns.*

Amongst the pre-selected 63 intronic SNPs, only 3 were located in the regions more likely to contain regulatory elements[159,165,167,170,267-271]. Amongst the remaining 3 intronic and 10 exonic SNPs, 2 exonic SNPs were discarded due to their lack of documentation in the largest catalogue of human variation data, 1000Genome Project[178]. Ultimately, 8 SNPs were identified as “of interest” due to their loci and MAF: the 3 intronic SNPs (rs2687116 at 33 base pair (bp) downstream of exon 7/13,

rs2242480 at 11bp downstream of exon 10/13, and rs12721620 at 10bp upstream of exon 12/13), and 5 SNPs located in the *CYP3A4* 3'-UTR (rs28988603, rs28988604, rs28969391, rs28371763 and rs28988606) (Figure 4.2). Interestingly, rs28969391 was the only "common" indel, yet it is located in the 3'-UTR of the gene, where it can not cause frameshift.

The distributions of the genotypes for the 8 SNPs identified as "of interest" (separately) in the world and amongst ethnic groups are presented in Table 4.1 as Minor Allele Frequencies.

SNPs		Type of variation (from 1000Genome Project)	Minor Allele Frequencies (%)					
			World	Africa	America	East Asia	South Asia	Europe
Located in the 3' -UTR	rs28988603	A > C	1% C	4% C	1% C	0% C	0% C	0% C
	rs28988604	G > A	3% A	6% A	2% A	0% A	1% A	2% A
	rs28969391	A > -	35% -	63% -	20% -	32% -	37% -	10% -
	rs28371763	T > A	1% A	0% A	1% A	0% A	3% A	2% A
	rs28988606	G > C or T	1% C 0% A	4% C	1% C	0% C	0% C	0% C 0% A
Located in introns	rs12721620	A > G	9% A	33% A	3% A	0% A	2% A	0% A
	rs2242480	T > C	42% T	85% T	39% T	27% T	37% T	8% T
	rs2687116	C > A	22% C	72% C	10% C	1% C	5% C	3% C

Table 4.1

Minor Allele Frequencies of the 8 identified Single Nucleotide Polymorphisms on the *CYP3A4* gene, in the world and among ethnic groups. This gene is on the minus chromosomal strand. For each Single Nucleotide Polymorphism (SNP), located in the 3'-UnTranslated Region (UTR) or in intron, the variation is expressed as referenced in db150, with the ancestral nucleobase followed by the mutant nucleobase (and a ">" sign placed in-between). The Minor Allele Frequency (MAF) of each SNP is shown for the world population and different ethnic groups (populations from African, American, East-Asian, South-Asian and European ancestry). Interestingly, the minor variant is not necessarily the one associated to the mutant nucleotide for the intronic SNPs. For each SNP, differential MAF distributions amongst the ethnic groups, are written in red, with an arbitrary threshold of 10-percentage point between the ethnic groups with the lower and the highest MAF. *A stands for adenine, C for cytosine, G for guanine and T for thymine. "–" indicates a nucleotide deletion. Data from the 1000Genome Project.*

4.3.2. In silico prediction of the effect of the eight CYP3A4 variants of interest

In silico data was extracted from various databases to predict the possible effect, if any, of these 8 SNPs of interest. When a possible effect is predicted by these databases, they associate particular protein or miRNA binding sites, or splice sites, to the SNP locus on the *CYP3A4* sequence[164,276,279,282-284,286-288,292,298]. As the liver is the main organ for the CYP3A4 xenobiotic metabolism[108], the hepatic expression of the flagged proteins and miRNAs was validated using GeneCard and GTex. Only the relevant, namely hepatically-expressed, proteins and miRNAs are presented (Table 4.2).

Table 4.2 (next page)

Associations of the minor variant of the eight identified *CYP3A4* Single Nucleotide Polymorphisms to potential changes in the silencing or the splicing of the gene. For each Single Nucleotide Polymorphism (SNP), the potential changes in gene silencing have been searched on RegRNA2.0 and Poly-miRTS (column A) and the potential changes in gene alternative splicing have been searched on Human Splice Finder, Fruit Fly, ESE Finder 3.0 and SpliceAid 2 (column B). *"–" denotes a potential loss of the miRNA or splicing element binding, or a potential shortening of the splice/branch site, due to the presence of the minor variant. "+" indicates a potential gain of the miRNA or splicing element binding, or a potential creation of a splicing site, due to the presence of the minor variant. "/" denotes an absence of flagged changes for the SNP.*

**rs28988606 minor variant can be G or T*

SNPs		A – Silencing elements potentially affected by the presence of the minor variant	B- Splicing elements potentially affected by the presence of the minor variant
Located in the 3' -UTR	rs28988603	hsa-miR-5584-3p (--)	/
	rs28988604	none	SRSF1 (-)
	rs28969391	hsa-miR-197-3p (+) hsa-miR-1305 (+) hsa-miR-5096 (+)	splicing acceptor site (+) human 5' splice site (-)
	rs28371763	none	hnRNPU (-) hnRNPK (-) hnRNPA1 (-) FMR1 (-) RBM5 (-)
	rs28988606	hsa-miR-3679-5p () for both*	SRSF5 (-) for both* SRSF6 (-) for both* SRSF3 (-) for both* YBX1 (-) for G* MBNL1 (+) for G*
Located in introns	rs12721620	/	CELF2 (+) SRSF6 (+)
	rs2242480	/	branch site (-)
	rs2687116	/	MBNL1 (-) hnRNP1 (+) TIA1 (+) TIAL1 (+)

Table 4.2

Hepatically-expressed miRNA binding sites were localized on the 3'-UTR of the *CYP3A4* gene using RegRNA2.0 and Poly-miRTS. Out of the five *CYP3A4* SNPs located in 3'-UTR, only rs28988603, rs28969391 and rs28988606 were shown to be potentially associated to differences in the binding of miRNAs between minor and major alleles (Table 4.2 – column A), these differences potentially impacting the silencing of the gene. The case of rs28988603 was particularly highlighted by these searches as it is associated to a "functional impact score D", meaning that "the derived

allele disrupts a conserved miRNA site (ancestral allele with support >= 2)"[287]. All the flagged miRNAs are poorly characterised in the literature.

Databases regarding alternative splicing (Human Splice Finder, Fruit Fly, ESE Finder 3.0, SpliceAid 2) were searched to predict the impact of the presence of the minor variant of each SNP on splicing, by associating the SNP to at least one alternative splicing regulatory elements: donor or acceptor splice sites (respectively located at the 5'- or 3'- end of introns), or branch sites, or serine/arginine rich proteins (SR proteins)[164,165,167,169,273,274]. Nothing was detected with the Human Splice Finder, but the other databases associated all 8 SNPs with differences between minor and major variants (Table 4.2 – column B) - these SNPs could therefore potentially impact alternative splicing. Most of the flagged hepatically-expressed SR proteins have been characterized as modulator of the selection of alternative splice sites, by binding to splicing enhancers or silencers - one could wonder how the gain or loss of their binding could impact on *CYP3A4* gene regulation. Depending on the direction of the modulation, they may stimulate or inhibit alternative splicing. Alternative splicing also requires two splicing sites (one donor and one acceptor) and a branch point, therefore the creation of such motif on the gene could stimulate alternative splicing – if other sites and point are available. Similarly, the break or shortening of one splice site (or one branch point) could inhibit *CYP3A4* alternative splicing[158,159,161,163-167,169-171,273,274,299]. Overall, by targeting alternative splicing, such variants could alter the concentrations in functional CYP3A4 isoforms in the organism, leading to an increased or a decreased drug metabolism.

As genetic variants can impact on the gene expression or enzyme activity by affecting the folding conformation of the DNA, RNA or protein sequences[293-295,298], the mfold Web Server was used to examine changes in the conformation of the DNA and RNA sequences for all of the 8 SNPs of interest. Major DNA and RNA conformations changes were interpreted as such when the observed conformations were visibly different between those associated to the sequence with the major or the minor variant for each SNP. This was the case for four of the SNPs localized in *CYP3A4* 3'-UTR: rs28371763 (Figure 4.3), rs28988603, rs28988606 and rs28988604 (Figure B1) - the latter being only predicted to impact on the RNA conformation. By impacting on the DNA or RNA conformation, these variants can render *cis*-regulatory elements

unavailable for binding, leading to differences in the regulation of the gene, and therefore possible differences in CYP3A4 concentrations and differences in CYP3A4-mediated drug metabolism.

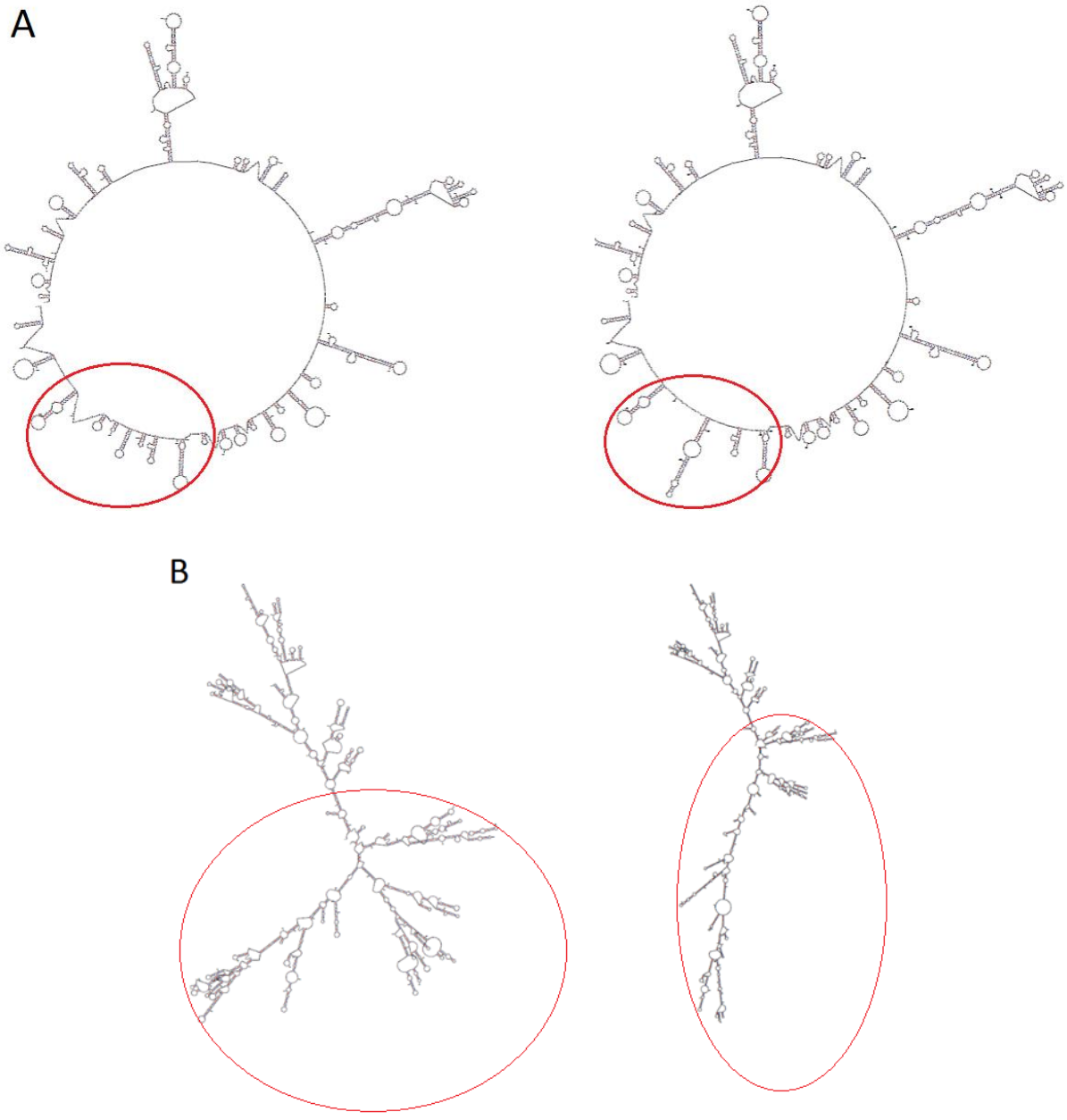


Figure 4.3

Examples of conformation changes between the major and the minor variants of the identified *CYP3A4* Single Nucleotide Polymorphisms, at 37°C. A- DNA folding conformation change between the major (left) and the minor variant (right) of rs28371763. B- RNA folding conformation change between the major (left) and the minor variant (right) of rs28371763. *Conformations retrieved from mfold web server.*

Finally, the scientific literature was systematically searched for references to rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 in LitVar, Pubmed, SNPedia and PharmGKB. The aim was to find evidences of (potential) effect of such SNPs on the *CYP3A4* gene expression or enzyme activity.

Most of these SNPs have not been of particular interest for the scientific community in the past decades (Table 4.3). In these renown databases, rs28988606 was not mentioned in any publication, both rs28371763 and rs28988603 were only cited in 1 publication each, both rs28969391[234] and rs12721620[300] in a total of 2 publications each, and rs28988604 in a total of 3 publications[234,277]. However, rs2687116 was mentioned in a total of 11 distinct articles and rs2242480 (as part of the specific haplotype *CYP3A4*1G*) in a total of 104 articles.

SNPs	**Number of articles mentioning the SNP in the database** total (open access only)				**Total of distinct articles mentioning the SNP** total (open access)
	LitVar	**Pubmed**	**SNPedia**	**PharmGKB**	
rs28988603	0 (0)	0 (0)	0 (0)	1 (0)	1 (0)
rs28988604	2 (2)	0 (0)	0 (0)	1 (0)	3 (0)
rs28969391	1 (1)	0 (0)	0 (0)	1 (1)	2 (2)
rs28371763	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
rs28988606	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
rs12721620	1 (1)	0 (0)	0 (0)	1 (0)	2 (0)
rs2242480	99 (80)	34 (16)	12 (8)	0 (0)	104 (83)
rs2687116	6 (5)	3 (2)	1 (0)	3 (0)	11 (7)

Table 4.3

Number of articles mentioning the eight identified *CYP3A4* Single Nucleotide Polymorphisms in four main databases: LitVar, Pubmed, SNPedia and PharmGKB. For each Single Nucleotide Polymorphism (SNP), the total number of articles is provided, with the total of open access articles in brackets.

To study the evidences published on the effect, or lack thereof, of those 8 SNPs on the *CYP3A4* gene expression or enzyme activity, the search was restricted to articles available in open access through Pubmed links: a grand total of 88 articles were retrieved. All of them were written in English, therefore they were all assessed. Tables 4.4 and 4.5 summarize the findings from these systematic reviews for, respectively, rs28988603, rs28988604, rs28969391, rs28371763, rs28988606 and rs12721620, and rs2242480 and rs2687116.

Table 4.4 (next page)

Summary of the scientific evidence regarding the effect on *CYP3A4* gene expression or enzyme activity of six of the identified Single Nucleotide Polymorphisms: rs28988603, rs28988604, rs28969391, rs28371763, rs28988606 and rs12721620. *Data retrieved from all open access articles, about the selected Single Nucleotide Polymorphism (SNP), found on LitVar, Pubmed, SNPedia and PharmGKB.*

SNPs	Summary of the scientific evidences regarding *CYP3A4* gene expression or enzyme activity	Type of study and reference
rs28988603	No study in open access	
rs28988604	rs28988604 is predicted to influence *CYP3A4.*	*In silico,* predictions regarding microRNA binding sites.[277]
	No significant association between genotype and mRNA levels.	*In vitro*, using allelic balance assay and cloning techniques.[234]
rs28969391	No significant association between genotype and mRNA levels.	*In vitro*, using allelic imbalance assay and cloning techniques.[234]
	One article was not related to *CYP3A4* nor the SNP.	
rs28371763	No significant association between genotype and mRNA levels.	*In vitro*, using allelic balance assay and cloning techniques.[234]
rs28988606	No published study	
rs12721620	No functional data	

Table 4.4

SNPs	Summary of the scientific evidences regarding *CYP3A4* gene expression or enzyme activity	Type of study	Comments and reference
rs2242480	No significant association between genotype and mRNA levels.	*In vitro*, using allelic balance assay and cloning techniques.[234]	
	Significant association between genotype and transcriptional activity. Significant association between genotype and coronary heart disease.	*In vitro* (using reporter gene) and clinical.	Controlled study - 322 cases with coronary heart disease and 306 controls. Study considered diabetes, hypertension, and smoking or drinking patterns. Not a direct clinical study, but also *in vitro* findings. Study with individuals from the same ethnic group (Chinese). Gender effect on results.[372]
	No significant association between genotype and colorectal cancer.	Clinical	Large controlled study - 2,575 cases with colorectal cancer (carcinogen metabolism) and 2,707 controls. Study with individuals from the same ethnic group (British Caucasian). No mention of concomitant medication, smoking patterns, etc. Not a direct study – possible enzyme compensation for carcinogen metabolism. Cannot exclude a small effect in cancer risk associated with the SNP. Did not study the gene-environment effect on cancer risk.[368]
	Significant association between genotype and ticagrelor metabolite concentration. No significant association between genotype and platelet aggregation.	Clinical	Only 14 individuals, all healthy males with normal range laboratory values. Numerous exclusion criteria including frequent smokers and grapefruit consumers. Kinetic of ticagrelor metabolite concentrations was followed on 15 timepoints, over 48h, while platelet aggregation was assayed at 7 timepoints. Study with individuals from the same ethnic group (Chinese). Possible mechanism of enzyme compensation to metabolise ticagrelor.[231]
	No significant association between genotype and ticagrelor plasma concentration, or platelet aggregation.	Clinical	Only 18 individuals, all healthy and non-smokers males. Numerous exclusion criteria including caffeine, alcohol and juices consumption. Kinetic of ticagrelor metabolite concentrations was followed on 15 timepoints over 48h, while platelet aggregation was assayed at 5 timepoints over 24h. Study with individuals from the same ethnic group (Chinese).[258]

No significant association between genotype and clopidogrel (and its metabolites) concentration, or platelet aggregation.	Clinical	Only 13 cases and 69 controls, all having coronary disease. Study with individuals from the same ethnic group (Caucasian). No mention of concomitant medication, smoking patterns, etc. No homozygous for the minor variant. The pharmacokinetics of clopidogrel was followed in only 44 individuals, at 8 timepoints over 24h of the last day of treatment. Platelet aggregation was assayed at only 1 timepoint. Issues with low concentrations, below the quantitation limit of the methodology.[371]
Significant association between genotype and clopidogrel-driven platelet aggregation.	Clinical	94 cases and 97 controls, all having coronary disease. Considered the concomitant medications. No significant differences in the distribution of age, sex, smoking, alcohol, etc. between responders and non-responders. Study with individuals from the same ethnic group (Chinese). Platelet aggregation was assayed at only 1 timepoint a day, over 5 days. More hypertension in the responders[254].
Significant association between genotype and steady-state finasteride concentration.	Clinical	597 cases and 676 controls, all from the same demographic and from various ethnic groups in the USA. Placebo-controlled study. Minorities were oversampled, yet ethnic groups along with age, sex, alcohol consumption, etc. were studied. Concentrations were tested at a very specific timepoint (one day, between 1- and 7-years post baseline data), may not be the same depending on duration of use. Serum concentrations may not be similar to tissue concentrations. Self-reported doses, low or undetectable levels could be due to missed doses because of low half-life of finasteride.[309]
Significant association between genotype and breast cancer.	Clinical	Large controlled study - 1,508 cases with breast cancer (carcinogen metabolism) and 1,556 controls. Dietary intake was assessed (self-report). Physical activity, smoking status, etc. were studied. Study with individuals from the same ethnic group (white from Long Island, USA). Not a direct study – possible mechanism of enzyme compensation for carcinogen metabolism.[370]

rs2687116	No significant association between genotype and colorectal cancer.	Clinical	Large controlled study - 2,575 cases with colorectal cancer (carcinogen metabolism) and 2,707 controls. Study with individuals from the same ethnic group (British Caucasian). No mention of concomitant medication, smoking patterns, etc. Not a direct study – possible mechanism of enzyme compensation for carcinogen metabolism. Cannot exclude a small effect in cancer risk associated with the SNP. Did not study the gene-environment effect on cancer risk.[368]
	No significant association between genotype and neviparine "apparent" clearance.	Clinical	Only 129 patients with HIV-1, all with good tolerance to the drug (bias against potential genotypes of high effect). Study with individuals from the same ethnic group (Cambodian). No exclusion criteria, nor study on concomitant parameters (alcohol, smoking, etc). Self-report of medication by the patients: issues with compliance and missing doses. Clearance was tested at only 2 timepoints: one morning after 1.5 and 3 years of treatment. Individual "apparent" clearance, due to a lack of pharmacokinetic model.[259]
	No significant difference in Lopinavir therapeutic drug monitoring between the alleles. No significant association of genotype to failure of treatment.	Clinical	Pubmed entry for a poster presentation – lack of data. Only 39 patients with HIV-1. The mean in therapeutic drug monitoring was yet lower for one allele.[369]

Table 4.5

Summary of the scientific evidence regarding the effect on *CYP3A4* gene expression or enzyme activity of two of the identified Single Nucleotide Polymorphisms: rs2242480 and rs2687116. *Data from a representative selection of open access articles, about the selected Single Nucleotide Polymorphism (SNP), found on LitVar, Pubmed, SNPedia and PharmGKB. The articles not investigating functional characterisation of these two identified SNPs were removed from this summary. HIV-1 stands for Human Immunodeficiency Virus 1, UK for United Kingdom and USA for United States of America.*

No evidence of any kind about the potential effect of rs28988603, rs28988606 and rs12721620 on *CYP3A4* were retrieved from these systematic literature searches. The *in vitro* study on rs28988604 advocated for a lack of association between genotype and *CYP3A4* gene expression or enzyme activity[234], contrarily to a *in silico* study prediction using microRNA binding sites[277]. No clinical study was available for this SNP. For rs28969391 or rs28371763, no association was found between their genotypes and mRNA levels in the only open access article mentioning those SNPs[234] (Table 4.4). No significant association between rs2687116 genotype and *CYP3A4* gene expression or enzyme activity was published in any of the 3 articles in open access[259,367,368].

Generally, large and controlled clinical studies regarding rs2242480 were not assessing directly the effect on the *CYP3A4* gene expression or enzyme activity – their focus was on disease risks. There was no real consensus between the large clinical studies: significant associations and non-significant ones were both published[368,370]. Clinical studies with smaller sample size acknowledged several limitations including a lack of controls, sampling data points or ethnical heterogeneity[231,254,258,309,371]. Overall, their results were contradictory – even when considering the metabolism of the same medication in-between ethnic groups[254,371] or the same medication in the same ethnic group[231,258]. Even *in vitro* experimental results on rs2242480 were inconsistent[234,372] (Table 4.5). Still, there was more publications on a potential impact of rs2242480 to *CYP3A4* gene expression or enzyme activity, including articles associating it to tacrolimus or sunitinib treatment[304,305], in methadone overdoses[232], in imatinib mesylate plasma levels[306], in beta-lactam allergies[307], in vitamin D concentrations[308], etc. In particular, the variant allele for rs2242480 was associated to the fluoxetine response status, a drug involved in the treatment of depression[310]. This association remained significant after a general linear regression to control for the impact of age, gender and depression score. It was also reported that the presence of the minor variant for rs2242480 was correlated to an increase in either the *CYP3A4* gene expression or the CYP3A4 protein function[311].

In general, out of the 8 variants of interest, the open access literature could only flag rs2242480 as “potentially altering *CYP3A4* gene expression or enzyme activity”. Its effect is not undisputed, yet it has the merit of being way more referenced on this particular thematic than any of the other SNPs of interest.

4.3.3. Introducing a classification for the potential importance of the eight CYP3A4 variants of interest

Given the data obtained *in silico*, the importance of the 8 SNPs of interest could be scored according to the predictions of their impact on the gene expression or enzyme activity. As developed below, the higher the score, the most likely the SNP could have an effect.

Despite controversies, the systematic literature search on the open access articles regarding the 8 variants of interest could only flag rs2242480 as “potentially altering *CYP3A4* gene expression or enzyme activity”, impacting potentially on drug concentrations, drug efficacies and even disease risks. None of the remaining 7 SNP of interest was significantly associated to any published laboratory or clinical effects. This could perhaps be explained by their lower minor allele frequencies in the world (between 1 and 35%, compared to 42% for rs2242480). Interestingly, the present *in silico* search did not predict an effect of the presence of the minor variant for rs2242480 on miRNA or splicing element bindings, or on DNA or RNA conformations, contrarily to the other SNPs of interest. It could then be assumed that this intronic SNP could impact the gene expression or enzyme activity through another route, such as the binding of transcription enhancers[312], and that the 7 remaining SNPs of interest really need to be characterised – and in open access articles.

To score the potential impact of the SNPs on *CYP3A4* gene expression or enzyme activity, as a rule, published literature evidence documenting the potential impact of a SNP through empirical laboratory or clinical evidences was scored higher than any *in*

silico predictions, published or presented in this thesis. rs2242480 impact score was therefore inherently the highest.

To rank the 7 remaining SNPs not associated to open access literature evidence of effect (rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, and rs2687116), every type of *in silico* predictions was weighted equally to the other ones (predictions about miRNA bindings = predictions about splicing elements = predictions regarding the DNA/RNA conformations = prediction from non-validated references in the literature). Equal weight was indeed applied on these different types of change as they all have the potentially to entirely inhibit the gene expression or the enzyme activity (respectively through complete miRNA silencing[276], or alternative splicing leading to a non-functional enzyme[164], or dysfunctional protein-RNA interactions[298]). For each of the 7 remaining SNPs of interest, it was then possible to calculate their total score as the number of different types of predictions they were associated to (Table 4.6) and to classify them accordingly - from the one which are predicted to have a lower impact on the *CYP3A4* gene expression or enzyme activity to the ones predicted to have a higher impact (Figure 4.4).

***CYP3A4* SNPs of interest (with world MAF)**	**Predicted changes involving miRNA binding sites**	**Predicted changes involving splicing elements binding sites**	**Predicted changes on the DNA/RNA conformations**	**Total score**
rs12721620 (9%)	NO	YES	NO	1
rs28988603 (1%)	YES	NO	YES (DNA and RNA)	2
rs28988604 (3%)	NO	YES	YES (RNA)	2
rs28969391 (35%)	YES	YES	NO	2
rs28371763 (1%)	NO	YES	YES (DNA, RNA)	2
rs2687116 (22%)	NO	YES	NO	1
rs28988606 (1%)	YES	YES	YES (DNA, RNA)	3

Table 4.6

Summary of the predicted changes on gene expression or enzyme activity when the minor variant of seven of the *CYP3A4* variants of interest is expressed. Presented data only recovers potential changes in hepatically-expressed miRNAs and splicing elements, as well as DNA/RNA folding conformations, associated to 7 of the Single Nucleotide Polymorphisms (SNPs) of interest*, as presented in Table 4.2, Figure 4.3 and Figure B1 following predictions on RegRNA2.0, Poly-miRTS, Human Splice Finder, Fruit Fly, ESE Finder 3.0, SpliceAid 2 and mfold. Minor Allele Frequency (MAF) in the world are shown for each SNP.

All of the *in silico* predictions about these 7 SNPs of interest were weighted equally: a "YES" in any category was associated to a SNP when any change in binding site or conformation was associated to the presence of the minor variant. Each "YES" was related to a score of 1. A total sum of scores was calculated to each SNP by cumulating the number of "YES". The total score of 0 was associated to a lack of predictions of any type of impact on *CYP3A4* gene expression or enzyme activity, while the total score of 3 was associated to a subjective higher likelihood for the SNP to impact on *CYP3A4* gene expression or enzyme activity.

*rs2242480 was excluded from this summary as, contrarily to the other SNPs, it is associated to a high number of literature evidences about its potential impact on *CYP3A4* gene expression or enzyme activity. These evidences are contradictory and with several limitations, there is a real need to investigate further on this particular SNP given it is found as minor variant in 42% of the world population.

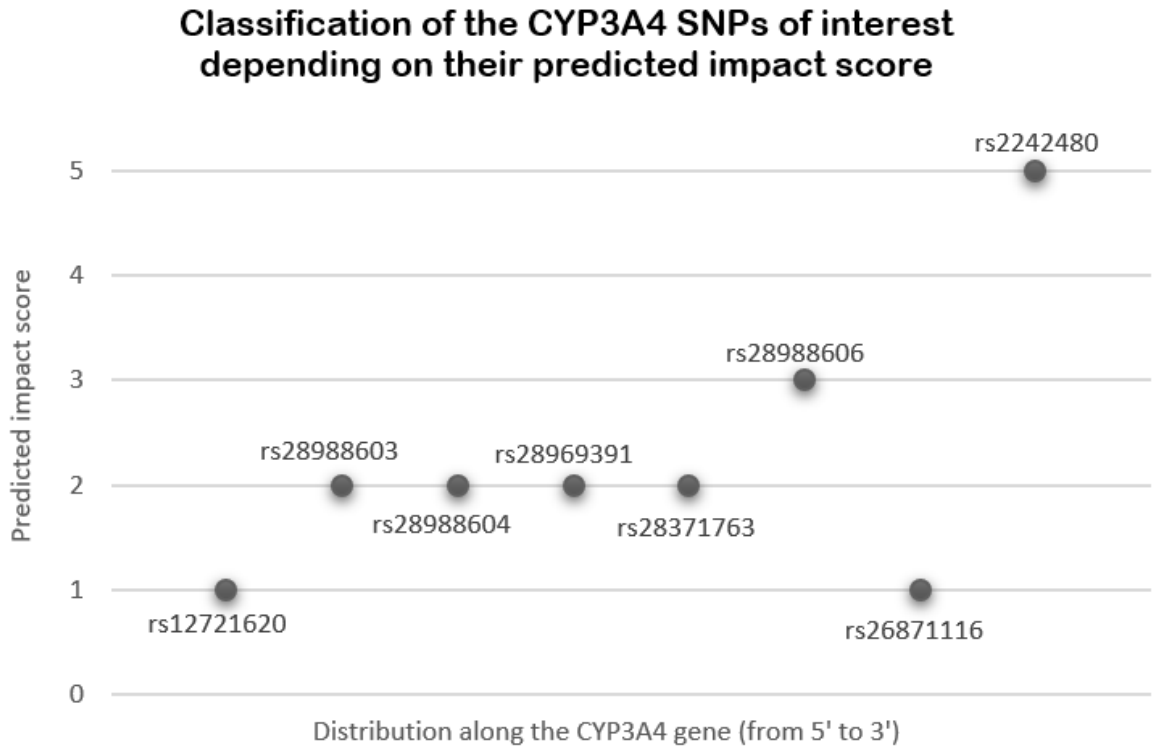


Figure 4.4

Classification of the eight *CYP3A4* variants of interest depending on their predicted score. Single Nucleotide Polymorphisms (SNPs) are displayed on the x-axis according to their 5'-3'- loci and, on the y-axis, their predicted impact score is shown. The higher the score, the most likely this SNP could have an impact on the gene expression or enzyme activity. A score of 5 was arbitrarily chosen for rs2242480 as it is studied in numerous literature publications, as discussed in-text.

4.3.4. Distribution of the genotypes for the eight CYP3A4 variants of interest in the 1000Genome Project

rs2242480 being predicted to be the one more likely to affect the *CYP3A4* gene expression or enzyme activity amongst the 8 SNPs of interest, their linkage

disequilibriums (LDs) were investigated in case the minor allele for rs2242480 is non-randomly associated to the minor allele for one of the other SNPs of interest, in the world population. In population genetics, two loci are in LD when the frequency associated to their association is different from the expected random association frequency (see *Chapter 2*)[175,176]. To study the association of the 8 SNPs of interest, the two common measures associated to LD (D' and R^2) were accessed using the LDLink matrix tool, which uses dbSNP 142 and therefore does not include the most recently characterized rs28969391, rs28371763 and rs28988606 SNPs.

Amongst the 5 SNPs of interest present in dbSNP 142, R^2 values were unsurprisingly low (Figure 4.5), because of differences in allele frequencies (Table 12), and D' values were assumed mostly high (Figure 4.6). The following pairs were flagged as in linkage disequilibrium: rs12721620 and rs28988603, rs12721620 and rs2687116, rs12721620 and rs2242480, rs28988603 and rs28988604, and rs2687116 and rs2242480. This meant that the inheritance of the minor variant for rs2242480 was strongly associated to the presence of the minor variant for both rs12721620 and rs2687116 (respective R^2 were of 0.142 and 0.269, with respective D' of 1.0 and 0.834).

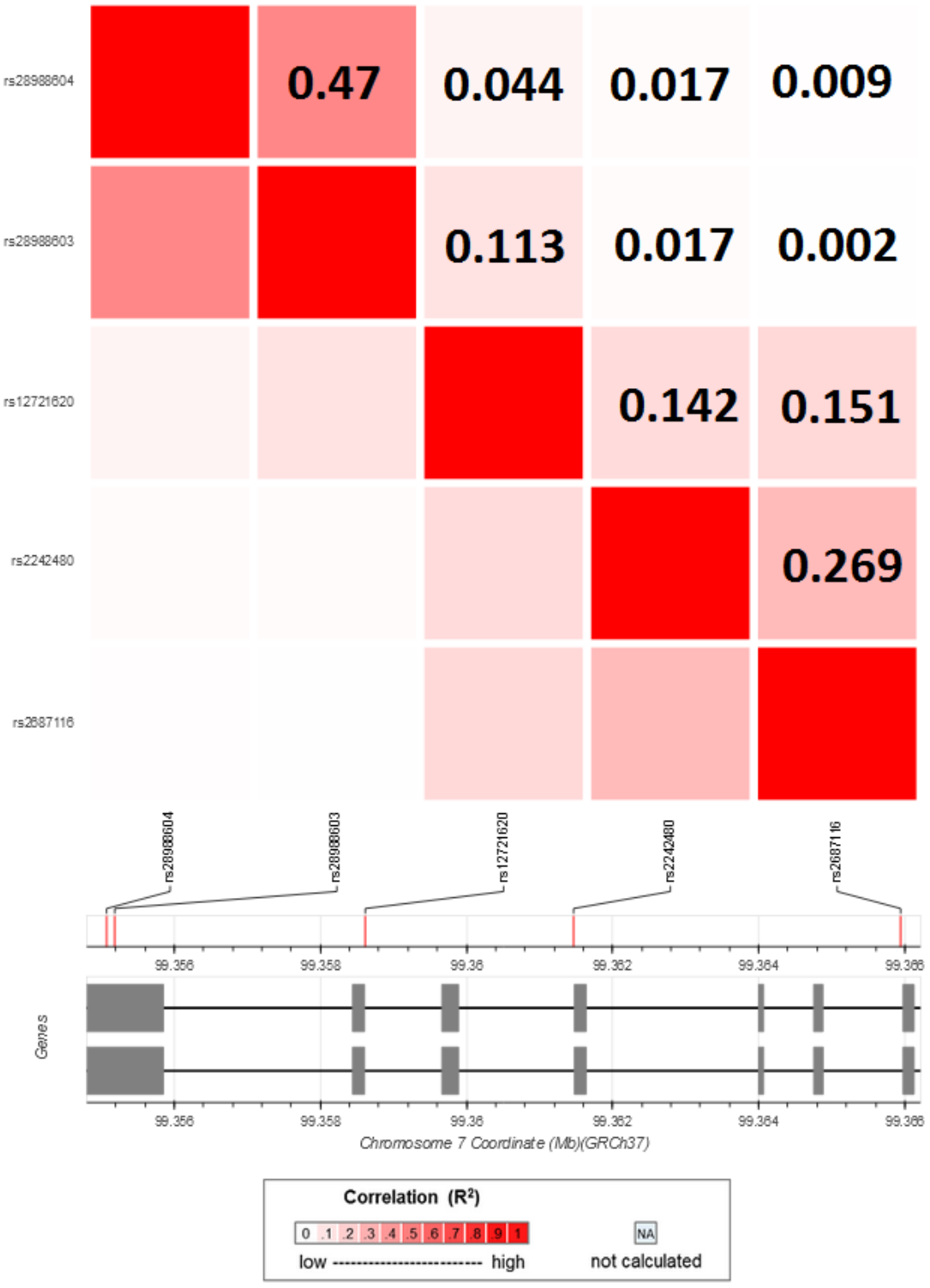


Figure 4.5

Linkage (dis)equilibrium matrixes of R^2 correlation value for *CYP3A4* variants rs28988603, rs28988604, rs12721620, rs2242480 and rs2687116. The correlation of two Single Nucleotide Polymorphisms are studied pairwise and estimated "high" when in red. *Data from LDLink, using the world population data.*

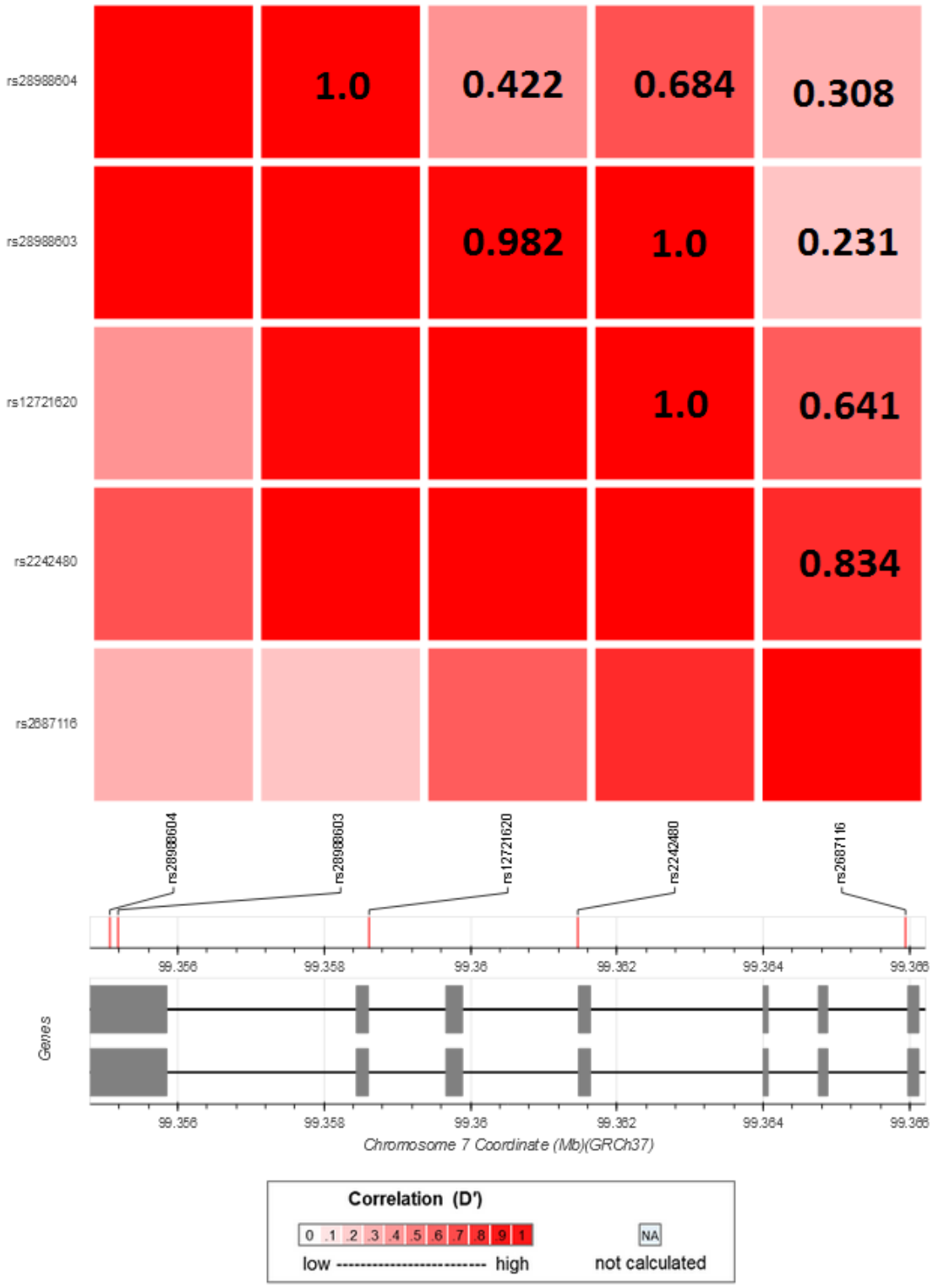


Figure 4.6

Linkage (dis)equilibrium matrixes of D' correlation value for *CYP3A4* variants rs28988603, rs28988604, rs12721620, rs2242480 and rs2687116. The correlation of two Single Nucleotide Polymorphisms are studied pairwise and estimated "high" when in red. *Data from LDLink, using the world population data.*

rs2242480 being associated to the highest impact score (Figure 4.4) and non-randomly associated with other SNPs of interest (Figures 4.5-6) using the world MAF data, the distribution of the different possible genotypes involving those 8 SNPs of interest was estimated, while focusing on rs2242480 - with or without rs28988603 and rs12721620 which are in LD with it. Indeed, it was important to estimate the proportion of the population being potentially affected by the predicted phenotypical differences due to their rs2242480 genotype on the *CYP3A4* gene expression or enzyme activity. They were expected to metabolise a different amount of clinically-relevant drugs, compared to patients carrying the major variants for this SNP.

Individual data from the participants of the 1000Genome Project was accessed to obtain the estimated distribution of the different possible genotypes involving rs28969391, rs28371763, rs28988604, rs28988606, rs28988603, rs12721620, rs2242480 and rs2687116, in the world and in ethnic groups. Globally, 70 different genotypes were found (as listed in Table B3). The most frequent genotype in the world was established when each of the 8 SNPs of interest are present as homozygous for the major variants (36.7% of the population, this genotype being named “homozygous for the 8 major variants”), and a total of 53 different genotypes were present in less than 1% of the world population, each (Figure 4.6A and Table B3).

Focussing on the distribution in ethnic groups (Figure 4.7B-F), 76.1% of the European population is homozygous for the 8 major variants, the rest of the population being distributed amongst 17 other genotypes (Figure 4.7B). In South Asia and East Asia, 22 and 15 different genotypes were accounted for, respectively with 35.8% and 44.2% of the populations being homozygous for the 8 major variants (Figure 4.7C-D). America was demonstrating a pretty heterogeneous population separated into 36 different genotypes, the most common being homozygous for the 8 major variants as well (37.5% of the population, Figure 4.7E). Finally, Africa was accounting for a total of 53 different genotypes, the most common - homozygous for the 8 major variants - being found in only 10.1% in this ethnic group (Figure 4.7F).

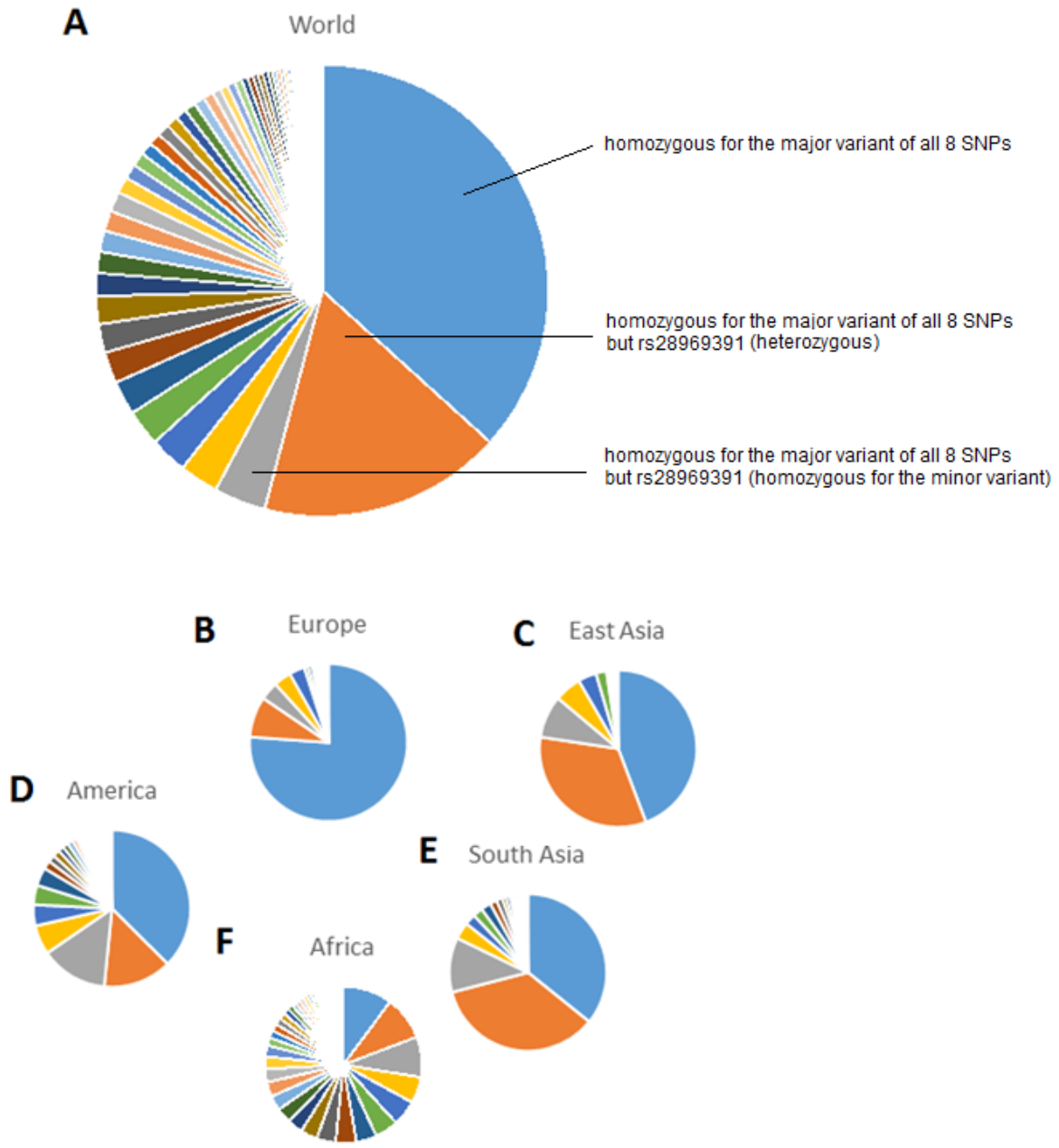


Figure 4.7

Distribution of the different genotypes for the eight *CYP3A4* variants of interest amongst the world population and between ethnic groups. Calculated from the 1000Genome Project phase 3 data, these distributions of the different genotypes for rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 are shown for the world population (A) as well as different ethnic groups (Europeans (B), East Asians (C), Americans (D), East Asians (E) and Africans (F)). Each colour represents one particular genotype (out of 70 possible in total, as described in Table B3), the light blue being the one associated to homozygous for the 8 major variants. Faded parts represent the least frequent genotypes.

As the validated literature only reported rs2242480 as potentially impacting the *CYP3A4* gene expression or enzyme activity[254,304,306,308,310,311], the rest of this population genetics study focuses on the distribution of the minor variant of rs2242480 in the world and in populations from different ethnic groups (Figure 4.8). It was found that 26.2% of the world population is homozygous for the minor variant of rs2242480, but massive disparities were found between ethnic groups, as 1.0% of the Europeans, 14.9% of the South Asians, 6.9% of the East Asians, 18.2% of the Americans and, most importantly, 72.4% of the Africans present such genotype. These major differences between ethnic groups in the distribution of the minor variant for rs2242480 evidence that, if this SNP is indeed impacting on the *CYP3A4* gene expression or enzyme activity, it could possibly have a major effect on drug concentrations in patients from different genetic backgrounds. It would be critical to investigate on the direction of the potential effect associated to the minor allele for rs2242480: are people homozygous for this minor allele Poor Metabolisers or Ultra-Rapid ones? [91,113] As shown by the critical frequency difference for the genotype homozygous for the minor variant of rs2242480 between Europeans versus Africans, one could wonder how to adapt a potential CYP3A4-metabolised drug dosage between these populations: which of these two populations is most likely to require a lower drug dosage (to decrease the experience of side-effects) or a higher drug dosage (to increase the treatment efficacy)?[101,106,108]

It may be important to consider the ethnicity of the patients before prescribing a regimen metabolised by CYP3A4. Yet even one of the 8 SNPs of interest expressed as a heterozygous allele could possibly have a critical impact on the CYP3A4 enzyme concentration or activity, and therefore on the drug metabolism. A systematic genotyping of particular SNPs on xenobiotic metabolism enzymes could therefore be proposed as part of a personalised medicine[90,106,120].

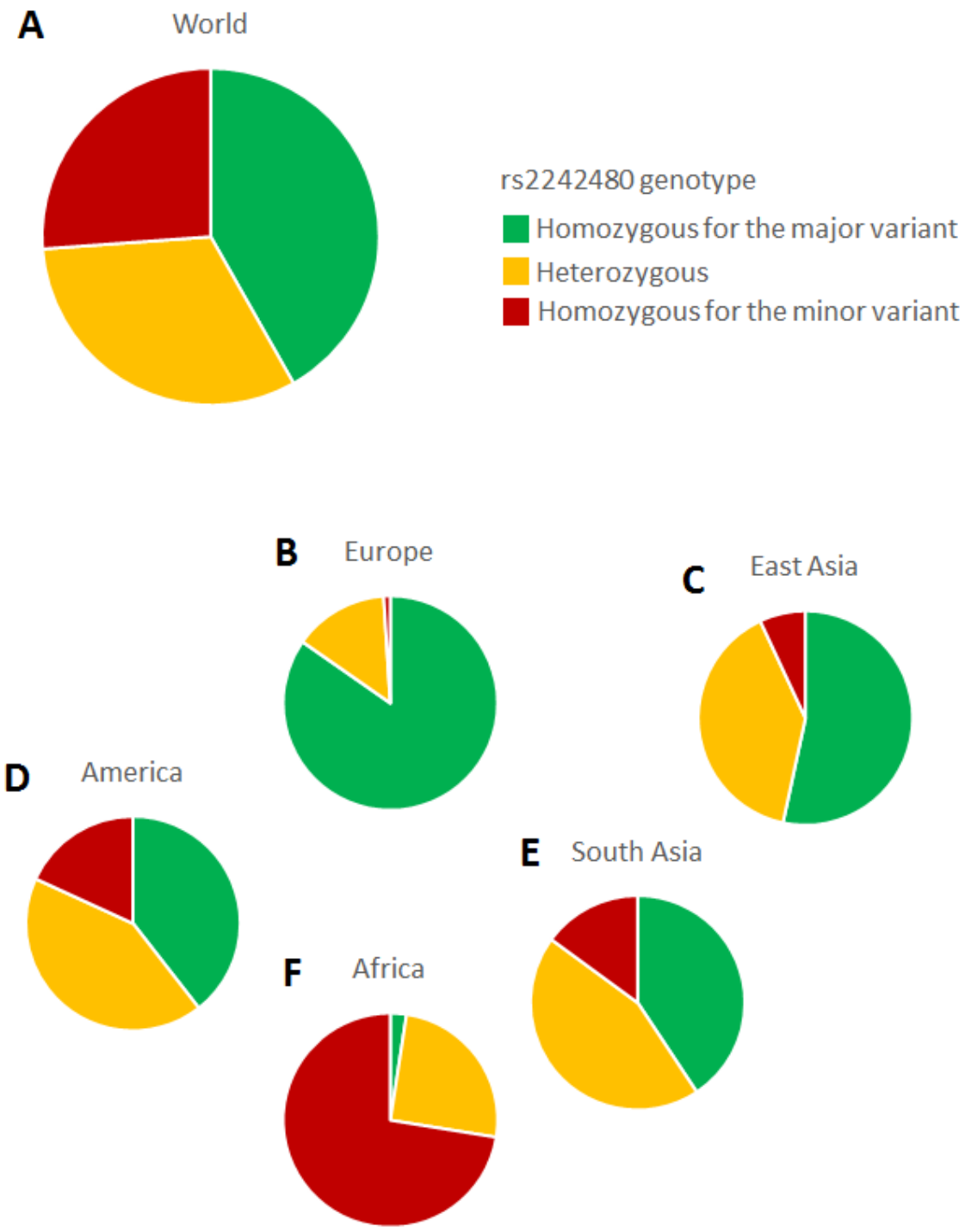


Figure 4.8

Distribution of the different genotypes for *CYP3A4* rs2242480 amongst the world population (A) and between Europeans (B), East Asians (C), Americans (D), South Asians (E) and Africans (F). Data from the 1000Genome Project – phase 3, with green depicting homozygous for the major variant of rs2242480, orange heterozygous and red homozygous for the minor variant.

4.4. Discussion

This study aimed to identify the most common and potentially important variants on the *CYP3A4* gene using publicly available bioinformatics tools. Common variants were defined with a MAF above 1% in the world population. *In silico* tools were used to predict the importance of the selected SNPs: searches were performed to associate, if any, splicing sites or the binding sites on various regulatory elements (transcription factors, miRNAs, SR proteins) to the selected SNPs. Overall, this study acknowledged 8 SNPs (rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116), they were all predicted to have a potential impact on the *CYP3A4* gene expression or enzyme activity, at different scales. A classification of those SNPs was established, and rs2242480 was scored as the most likely to affect *CYP3A4,* or at least the most important to investigate further on. The distribution of all possible genotypes for the 8 SNPs of interest was considered, with a focus on rs2242480, in the world and among ethnic groups. It showed major differences in the distribution of the genotypes between ethnic populations. Altogether, the present study emphasises on the importance of studying xenobiotic metabolism variants. It may be critical to generalise the use of personalised medicine, a tailoring of drug dosages according to the patient genetic background[90,106,120].

This study being performed using *in silico* tools, its predictions would need to be confirmed *in vitro*, if not *in vivo* or in clinical studies. Firstly, subjective, although informed, thresholds were used to select the 8 SNPs of interest. For example, the positional effect of SNPs on gene function was assumed[161,162,165,169,268,313], excluding therefore intronic SNPs not present in the vicinity of the exon boundaries. Indeed, a threshold of 100 nucleotides from the boundaries was suggested in previous papers[159,163-165,167,169,170,267-271], although several other studies proposed to extend it to 200 nucleotides[159,267] or even expected a distribution of intronic regulatory elements at variable distances from the exons[165]. Secondly, this study did not consider the extensive list on how variants can affect the gene expression or enzyme activity. We focused on the most common ones, excluding for example genetic variants affecting *CYP3A4* epigenetics marks[172], despite reports of the differential expressions of *CYP3A4* due to epigenetic factors[251]. Thirdly, strong linkage disequilibriums have been found to correlate the alleles of many of the xenobiotic metabolism genes, *CYP3A4*

included,[230,297,314-316] therefore a broader approach could have been performed to consider all 3,100 *CYP3A4* variants, instead of 8.

Fourthly, and more notably, the predictions retrieved from the *in silico* searches were computed together using subjective scores, discarding a potential asymmetry in the distribution of actual regulatory elements on the gene[165,267]. We also assumed that associations to *CYP3A4* impact published in clinical studies weighted more than predictions about possible alterations of the miRNA silencing or of alternative splicing or of DNA/RNA conformation (those predictions were weighted equally in the scoring system).

Fourthly, it is important to note that the systematic review of all open access publications on the 8 SNPs of interest only flagged rs2242480 as potentially impacting on *CYP3A4*, yet the search was limited to open access articles. Notably, this systematic literature review on open access articles showed that most studies on rs2242480 had several limitations: clinical study with a large sample size were not assessing directly the effect of the SNP but were focussing on disease risks[368,], while much smaller sample-sized clinical studies were lacking controls, sampling timepoints, etc.[231,254,258,259,309,370,371]. It is therefore not surprising that most of these published evidences were contradictory - for example two studies on the same medication, performed on the same ethnic group, had distinct conclusions on the potential impact of the SNP[231,258]. However, rs2242480 was still flagged as having the "potential to effect on *CYP3A4* gene expression or enzyme activity" given that studies focussing on similar xenobiotic metabolism enzymes have evidenced at the same time i) differences in the distributions of the variants amongst ethnic groups, and ii) an association between those variants and differential phenotypic response (drug response)[91,98,106,113-117,317,318]. It would be unlikely that CYP3A4 is the "only" xenobiotic metabolism enzyme not subjected to such alterations from its gene polymorphisms. The studies focussing on other xenobiotic metabolism enzymes have already advocated for an adaptation of the treatment to the patient genetics. This present body of work points out to 8 more SNPs to consider when administrating CYP3A4-metabolised drugs. These 8 SNPs are of clinical importance as this enzyme metabolism is predominant for clinically-relevant drugs[108,247,251].

In the literature, CYP3A4 and CYP3A5 were often studied together due to their

overlapping substrate specificity for many of the clinically-relevant drugs[246,248,249]. Yet, differences in their tissue expression and catalytic efficacy[251] were reported, alongside the metabolism of a major antibiotic agent, erythromycin, by CYP3A4 and not CYP3A5[251]. It is therefore particularly important to consider the SNPs being able to potentially impact on the *CYP3A4* gene expression or enzyme activity. Especially now that differences in their distribution amongst particular ethnic groups have been shown (in the present study), such as differences between populations from European or African ancestry. Previous studies actually evidenced that, in Caucasian populations, the non-functional CYP3A5*3 allele is highly represented and therefore the xenobiotic metabolism of drugs metabolised by CYP3A4/5 is only undertaken by CYP3A4[246]. Even though "only" 1.0% of the Europeans were estimated to be homozygous for the minor variant of rs2242480 (that could potentially be responsible for an abundance or lack of *CYP3A4* gene expression or enzyme activity[101,108]), this represents 7.4 million potential patients being at risk of experiencing low treatment efficiency and/or increased side-effects due to their genetic background, and the possible lack of enzymatic redundancy between CYP3A4 and CYP3A5[251].

In summary, there are several implications of the present findings. This *in silico* study shows that common SNPs on the *CYP3A4* gene could potentially have an impact on its expression or enzyme activity. This needs to be confirmed *in vitro* or *in vivo*, as this could impact on the concentrations on the many drugs metabolized by CYP3A4, leading to ineffective treatment, over-doses and side-effects[106]. As CYP3A4 is involved in the metabolism of many clinically-relevant antibiotics[101], those SNPs could affect their circulating concentrations, destabilising the exact dosage required to assure both treatment efficacy and the avoidance of antibiotic resistance[27]. More importantly, this study evidences differences in the distribution of various "common" *CYP3A4* SNPs between ethnic groups, and previous reports on other xenobiotic metabolism enzymes have associated such differences to alterations in drug response[110,118,121,122]. It was notably the case for Caucasian populations expressing many of the deleterious *CYP2D6* SNPs[123] or for the complete *GSTM1* or *GSTT1* gene deletion found in up to 50% of the patients in certain populations[116]. The clinical effect of such polymorphisms was reported, with examples of i) up to 25% of the world population being affected by genetic variations on the main *CYP* genes (including

CYP3A4/5)[113], ii) a specific CYP2C9 SNP dramatically affecting warfarin metabolism[114], or iii) the common *GSTT1* gene deletion increasing the toxicity to isoniazid, pyrazinamide and rifampicin for tuberculous patients[117].

If one or more of the 8 SNPs highlighted in the present study are indeed proven to impact the *CYP3A4* gene expression or enzyme activity, it would be interesting to develop personalized medicine for the xenobiotic metabolism genes[90,106,120]. Patients would then be genotyped for targeted SNPs before being delivered any regimen, their prescription being tailored later on, according to their genetic background. In areas in the world where systematic genotyping is still an out-of-reach technique of molecular biology, knowledge about the distribution of the most common *CYP3A4* SNPs in different ethnic groups should be spread out to make sure clinicians adapt treatments and treatment dosages to the average patients living in the area.

In this *in silico* study, rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 were found to be potentially impacting the CYP3A4 enzyme activity. Further research is needed to address whether these variants are indeed impacting CYP3A4 and whether they can influence, alone or together, drug efficacy and/or toxicity.

4.5. Reflection on the research

This *in silico* study acknowledged 8 *CYP3A4* SNPs, all predicted to have a potential impact on the *CYP3A4* gene expression or enzyme activity, but at different scales – a classification of predicted importance was shared.

Given more time, it could have been interesting to seek access to the closed-access publications on these SNPs in the literature, as part of their classification is based on the scientific evidences published on their potential effect on CYP3A4. It could also have been interesting to translate this *in silic*o research into an experimental one, using an allelic balance assay. We could have quantified the gene expression associated to each allele (major or minor), using genomic and complementary DNA samples from individuals being heterozygous for only one of the selected genetic variants. It could have helped us distinguish – if any - differences in gene expression

due to the presence of the major or the minor allele of each of these 8 SNPs. Considering the genetic linkage found between those 8 SNPs, finding and recruiting such individuals would be time consuming though.

A more realistic approach of extending this research would be to consider all *CYP3A4* SNPs, instead of focusing only those in 3'-UTR and those in vicinity of exon boundaries. In total, 60 intronic SNPs were discarded due to their loci: they were localised in regions "less likely to contain regulatory elements", yet there is still a chance these SNPs could have effect on *CYP3A4* gene expression or enzyme activity. Finally, still using *in silico* methods, given more time, we could have investigated *CYP3A5* as well, as CYP3A4 and CYP3A5 are closely linked – one could wonder if they compensate each other when one presents an altering SNP.

CHAPTER 5 - Investigating the impact of *Carboxylesterase 2* genetic variants on host xenobiotic metabolism, treatment efficacy and adverse effects.

5.1. Introduction

With the global threat of antibiotic resistance, the right antibiotics need to be delivered to the patients, at the appropriate dose and for the appropriate length of treatment, to be able to treat them efficiently whist reducing the probability of antibiotic resistance emergence[27]. To improve the treatment against Methicillin-Resistant *Staphylococcus aureus* (MRSA) infections, many clinical trials have focussed over the last decades on the combination of vancomycin and rifampicin[55]. Due to serious published inconsistencies about the type of their interaction towards MRSA (both synergies, antagonisms and indifferences have been reported)[71-81], physicians have to base their prescriptions on highly contradictory treatment guidelines: this combination is for example recommended for infective endocarditis on prosthetic valve in the United States of America, but discouraged in the United Kingdom, without clear explanation[82,83]. Several hypotheses have been formulated to explain the discrepancies between the *in vitro*, *in vivo* and clinical studies on this interaction[33,45,55]. During clinical trials, inter-individual differences between patients (such as age, sex, body weight, concomitant diseases and treatments, and genetic background) might have influenced these antibiotics xenobiotic metabolism during clinical trials[91,101,113].

In the host organism, xenobiotic metabolism detects and processes foreign compounds, including medicine, to facilitate their rapid excretion[108]. This biotransformation is divided into three enzymatic phases, involving enzymes coded by highly polymorphic genes[98,101,106,108,110] – i.e. their genes are known to contain various genetic variants, present in at least 1% of the world population, that could alter their gene expression or influence their enzyme activity. Amongst xenobiotic metabolism Phase I enzymes, the well characterised cytochrome P450 monooxygenases (CYPs or p450) are estimated to be involved in 70-80% of the clinically relevant drug metabolism[101,106,108]. Some of their common genetic variants have been reported to interfere with drug metabolism, treatment efficacy and adverse effects[113]. It is notably

the case for patients treated with warfarin, a major cardiovascular drug: carriers of either minor allele for rs1799853 and rs1057910 on *CYP2C9* may require up to 75% decrease in warfarin dosage to reduce their experienced side-effects[114]. Similarly, the intronic *CYP1A2* variant rs762551 was associated to an increased oral clearance for the mental health medication olanzapine, and *CYP2B6* rs2279343 to increased clearance of the HIV-1 drugs efavirenz and cyclophosphamide, to name a few relevant Phase I SNPs[101]. Polymorphisms in the Phase II genes have also been associated to altered gene expression or enzyme activity. In particular, the non-functional null alleles of the *Glutathione S-Transferase T1 (GSTT1)* or the *GSTM1* gene were reported, separately, to increase the likelihood of toxicity from antituberculous treatment[115-117,319]. Likewise, the slow acetylator variants for N-Acetyltransferase 2 were associated with increased hepatotoxicity from the antituberculous isoniazid[124]. A large proportion of differences in treatment efficacy between patients could therefore be associated to their genetic background, amongst other causes (lack of drug monitoring, influence of immune system, sex, age, etc.[45,82,92,95,97,98]). This could be of critical importance as genetic variation shows strong effects of ethnicity, i.e. a heterogenous distribution of such variants between the different ethnic groups[108,110,116,118,121,122]. For example, many of the deleterious *CYP2D6* genetic variants are expressed mostly in Caucasian populations[123], and up to 50% of individuals of European descent exhibit the deletion of *GSTM1* or *GSTT1* genes[115].

Vancomycin is known to be poorly metabolised[125] and, until recently, rifampicin xenobiotic metabolism was wrongfully attributed to the CYPs[33,45]. Although rifampicin is indeed a potent inducer of the CYP enzymes[101], it was recently reported that it is primarily metabolised by the Carboxylesterase 2 (CES2) enzyme[33,108,129,136-140]. CES2 has been overlooked[320] despite being responsible for the metabolism of several clinically-relevant drugs, including the pro-immunosuppressor Mycophenolate MoFetil (MMF) “dampening down” the immune system after organ transplant[149,150]. CES2 is encoded by the *CES2* gene, which harbours a number of genetic variants with potential effects on gene expression or enzyme activity[144-146,148,321]. Most of these polymorphisms have been poorly characterised, despite their potential interference with drug metabolism, treatment efficacy and adverse effects. Rare *CES2* genetic variants have been recently associated with differences in rifampicin metabolism in Korean tuberculosis patients[138], yet little is known about the most common ones or

their distribution amongst ethnic groups[142,144,145,147,148]. This could potentially explain why rifampicin, currently prescribed to MRSA and tuberculosis patients at a dose which only considers their body weight[82,83], has been frequently associated to hepatotoxicity or unresponsiveness in patients [322,323]. Likewise, MMF fixed dose[133] has long been questioned, with different groups of responders being reported (including unresponsive or with increased experiences of side-effects). Critically-ill patients can develop severe diarrhoea, leading to severe dehydration and MMF withdrawal[151,152,324]. Their prescribed doses are not considering the potential impact of the host genetics or baseline parameters (age, concomitant diseases, etc.). Regarding antibiotic prescription, such inter-individual differences could alter the circulating antibiotic concentrations, modifying therefore the effective dose required to treat them efficiently whilst eradicating all bacteria and reducing the probability of emergence of antibiotic resistance. In other words, inter-individual differences such as genetic variations could convert a successful combination of vancomycin and rifampicin into an ineffective (if not hazardous) antibiotic regimen in patients infected with MRSA[27].

To characterise the impact of *CES2* genetic variants on the host xenobiotic metabolism, this body of work aimed first to associate “common” *CES2* genetic variants to *in silico* predictions of gene expression or enzyme activity alterations. Using an allelic imbalance assay, the effect of potentially-altering genetic variants on *CES2* mRNA levels was established: rs11075646 and rs8192925 were identified as associated with distinct effects on allele-specific mRNA output - the minor variant for rs11075646 is associated with significant increased level of *CES2* transcripts and the minor variant for rs8192925 with significantly reduced *CES2* mRNA. As differences at *CES2* mRNA levels might be correlated with differences at the enzyme level[325,326] – with potential inter-patient differences in drug efficacy and adverse treatment effects, a pilot clinical study was then set up to evaluate a potential association between the response of patients to rifampicin or MMF, and their genotype for *CES2* rs11075646 and rs8192925. The preliminary data of this clinical study is developed in this chapter. If these *CES2* SNPs are shown to interfere with drug metabolism and treatment response, their presence as minor alleles in the genotype of patients treated with rifampicin could affect their recovery, whilst potentially triggering *de-novo* antibiotic resistance[27]. Thus, it could be useful for both the patients and the community to

proceed to pre-operative screenings before antibiotic prescription, allowing the tailoring of medication dose based upon genotype. This would be another step towards personalized medicine[90,106,120].

5.2. Materials and Methods

5.2.1. Bioinformatics

5.2.1.1. Identification of gene sequences and genetic variation

CES2 gene sequences (as complete DNA sequence, cds) and messenger RNA (mRNA) transcript sequences were extracted from NCBI's Nucleotide or the UCSC Genome Browser (https://www.genome.ucsc.edu), in FASTA format. Sequence references and lengths (in base pair, bp) are provided (Table 5.1).

Gene / Transcript variant	Sequence reference	Sequence length (bp)	Link
CES2 gene	AY851164.1 (GenBank) Or Genomic Sequence (chr16:66,934,444-66,945,069) (UCSC)	14,667	FASTA Or FASTA
CES2 transcript variant 1 (mRNA)	NM_003869.5 (NCBI Reference Sequence)	3,955	FASTA
CES2 transcript variant 2 (mRNA)	NM_198061.2 (NCBI Reference Sequence)	3,907	FASTA
CES2 transcript variant 3 (non-coding RNA)	NR_036684.1 (NCBI Reference Sequence)	4,193	FASTA

Table 5.1

Information about the genes and transcript variants sequences for *Carboxylesterase 2 (CES2)*. *mRNA stands for messenger RNA, and bp for base pair.*

CES2 genetic variants were identified using the UCSC Genome Browser, on the most recent human assembly (GRCh38/hg38) with the most recent Short Genetic Variations database (dbSNP17). The search was limited to those carried by at least 1% of the world population, as they are those defined as "common" in dbSNP (*Supplementary Data*). The commonality of the selected genetic variants was important to observe, if there is any, an impact on gene expression or enzyme activity which has the potential to be relevant to a large proportion of the world population. The Minor Allele Frequencies (MAFs) of the variants, and their alleles frequencies in the world and in different ethnic groups (populations with descents from Africa, Americas, East Asia, South Asia, Europe), were accessed using the 1000Genome Project data (http://phase3browser.1000genomes.org/index.html), which was based on the GRCh37 reference assembly then updated to GRCh38/hg38. Full information regarding the 2,504 individuals from these ethnic groups is available in *Chapter 2* and in a previous study[178]. For clarity purpose, terms such as "Europeans" will be used through this chapter for "individuals from European ancestry populations". *CES2* genetic variants were classified according to their world MAF, the 3 more frequent exonic *CES2* genetic variants were preselected for our study given their MAF (*Supplementary Data*). The rationale to select genetic variants localised in exons was that studies have reported variant positional effects on function[161,162,169,268,313], and exonic regulatory elements have been more characterised than their intronic counterparts[168].

As SNPs are sometimes non-randomly associated and could therefore biased the each other' reported potential effect, the Linkage Disequilibrium (LD) was estimated between the 3 SNPs of interest on the LDLink database (https://ldlink.nci.nih.gov), using the LD Pair and LD Matrix functions in the world population, for determination of the D' and the R^2 of the pairs of variants of interest[175] (see *Chapter 2*).

5.2.1.2. In silico assessment of potential effects on RNA transcription or processing

Following identification and classification of the *CES2* variants, 3 SNPs were included in this study: rs11075646, rs8192925 and rs28382828, due to their commonality and

their exonic loci. They are found on the chromosome 16, at their respective locus of chr16:66935273-66935273 for rs11075646 (C>G), chr16:66944094-66944094 for rs8192925 (A>G) and chr16:66944326-66944325 for rs28382828 (- >AGAA or else). They are cited in their 5'-3'- order through this chapter. As detailed below, *in silico* studies were performed to assess their localisation in regards of regulatory elements (sequences regulating the gene), in order to validate and specify the general prediction that their position could affect *CES2* gene expression or enzyme activity[161,162,169,268,313].

rs11075646 being localised in *CES2* 5'-UnTranslated Region (UTR), it could most likely alter gene transcription, by disrupting the binding of a transcription factor, or - less likely - alternative splicing by disrupting the binding of an Exonic Splicing Enhancer or Silencer (ESE/ESS)[161,166,169,170,291,327]. rs8192925 and rs28382828 being localised in the *CES2* 3'-UTR, they could most likely affect mRNA stability by influencing microRNA (miRNA) binding[162] or - less likely - alternative splicing by disrupting ESE/ESS (more details on gene regulation mechanisms in *Chapter 2)*. *In silico* searches - detailed below - were therefore performed to evaluate the localisation of the 3 SNPs of interest regarding known or predicted transcription factor binding sites (for rs11075646), splicing sites (for all), noncoding RNA binding sites (for rs8192925 and rs28382828), etc. When available on the *in silico* tool, the search was made with or without the minor variant, to observe if differences are highlighted.

The co-localisation of the SNPs with regulatory elements was primarily assessed using the human assembly build GRCh37/hg19 on UCSC Genome Browser. When necessary given the SNP localisation on the *CES2* gene, potential effects on transcription were evaluated using "ENCODE TFBS ChIP-seq". Effects on miRNA or other non-coding RNA regulation were assessed using "Regulation", when needed.

Predictions for transcription factor binding sites were also checked using the Regulatory Analysis of Variations in ENhancers database (http://www.cisreg.ca/cgi-bin/RAVEN/a)[327] using all of the proposed reference sequences to access the list of transcription factor binding sites affected by *CES2* SNPs. RegRNA 2.0 database (http://regrna2.mbc.nctu.edu.tw/)[279,284] was also used for this purpose and to assess

several other regulatory elements: polyadenylation sites, ribosome binding sites and RNA-RNA interaction regions (including miRNA).

miRNA binding sites were also validated through the Poly-miRTS database[286-288] (http://compbio.uthsc.edu/miRSNP). They are classified into four classes, depending on their likelihood to alter the repression control of the gene. By inputting the gene name into the search option, the outputs associate, when possible, miRNA binding site(s) to each *CES2* exonic SNPs, alongside a function class predicting the functional impact of the variants[289]. Likewise, Exiqon (mirseach), Target Scan[328], MirSNP[329], miRSearch[330,331] and miRBase[332] were also accessed to study the miRNA target sites on *CES2* gene (*Supplementary Data*).

DNA and RNA conformation changes due to genetic variants being reported to have a potential impact on the gene expression or protein activity[293-295], the predicted effect of conformation changes at the DNA or RNA level because of the presence of the minor variant(s) were analysed using the mfold Web Server[281] (http://unafold.rna.albany.edu/?q=mfold).

Unless otherwise specified, all of these databases were accessed using the default options and thresholds (*Supplementary Data*), as they were rationalised and recommended in the literature[279-283], consistently with this study's general approach. When necessary, the motifs recognized on regulatory elements were accessed in Jaspar (jaspar.genereg.net).

5.2.1.3. In silico assessment of expression patterns of factors potentially binding to CES2 regulatory elements

The CES2 metabolism of clinically relevant drugs being localised mainly in the liver[320], hepatic expression patterns for transcription factors predicted to bind to sites disrupted by rs11075646 were assessed using the Expression Atlas (https://www.ebi.ac.uk/gxa/home), the dataBase for Gene Expression Evolution (https://bgee.org/), the Jaspar CORE (http://jaspar.genereg.net/) and the Tissue-specific Gene Expression and Regulation (TiGER, http://bioinfo.wilmer.jhu.edu/tiger/) databases. Jaspar CORE was also accessed to associate the factors to their specific

motifs. Where miRNA binding was predicted to be disrupted by rs8192925 or rs28382828, their likely presence in the appropriate tissue was assessed using the Genotype-Tissue Expression portal (https://www.gtexportal.org/home/), the miRmine database (http://guanlab.ccmb.med.umich.edu/mirmine/) and data from the Expression tab on the miRNA Genecards (https://www.genecards.org).

5.2.2. Measurement of relative allelic expression

5.2.2.1. Samples

300 peripheral blood DNA and RNA samples from anonymised participants in the Exeter 10,000 study were collected and used here with the ethical permission from Exeter National Institute for Health Research (NIHR) Clinical Facility (REC 09/H0106/75). Exeter 10,000 is a cross sectional population study consisting of samples collected from volunteer individuals living in the South West of England and recruited since 2010.

5.2.2.2. Determination of genotypes

DNA samples were genotyped for the presence of rs11075646 and rs8192925 using Polymerase Chain Reaction (PCR) technique with TaqMan chemistries on the ABI 7900HT Fast RealTime PCR System or StepOnePlus Real-time PCR System as appropriate (ThermoFisher, Warrington, UK). 22ng of genomic DNA was amplified in a total volume of 20 µL containing 200nM of the specific primers to one SNP (the VIC-tagged primer recognizes the major variant, whereas the FAM-tagged primer attaches to the minor one, Table 5.2) and 10µL Taqman Genotyping MasterMix (ThermoFisher), as per manufacturer's instruction. Cycling conditions were an initial cycle of 60°C for 30 seconds (sec), and 95°C for 10 minutes (min), then 40 cycles of 95°C for 15 sec and 60°C for 1 min, and a final read at 60°C for 30 sec. Clustering of genotypes was achieved using the post-read function.

		rs11075646 (custom)	rs8192925 (assay ID: shelf C___2847570_10)
Primers	Forward	*ACGTGCACATCCTCAGAGAAG*	*Proprietary*
	Reverse	*GTCCACAGGCCCGATGAG*	
Reporter sequence	VIC	*CC CTCCTATCGAT**C**CCCCAG*	*Proprietary*
	FAM	*CTCCTATCGAT**G**CCCCAG*	
Reporter Data		1 (C) is VIC 2 (G) is FAM	1 (A) is VIC 2 (G) is FAM
Context sequence		*CC CTC CTA TCG AT* *[C/G]* *CCC CAG*	*GGGTCAGCCTGCTGTGCCCACACAC* *[A/G]* *CCCACTAAGGAGAAAGAAGTTGATT*
Final concentration of the assay mix		Used at 1x (containing an equal concentrations of each primer, and equal concentrations of each reporter).	Used at 1x (containing an equal concentrations of each primer, and equal concentrations of each reporter).

Table 5.2
Description of the Taqman Genotyping primers for rs11075646 and rs8192925. The primers for rs11075646 were selected using the Pyromark Assay Design Software (Qiagen). The rs8192925 assay is inventoried under the assay ID "C___2847570_10", which is proprietary. For both Single Nucleotide Polymorphisms, the major variant (1) is associated to the VIC-tag, while the minor variant (2) is recognized by the FAM-tag, allowing for their distinction.

Despite using the Pyromark Assay Design Software (Qiagen, Limburg, Netherlands) to design the primers for rs28382828 genotyping, this Custom Taqman SNP Assay was failing – potentially because of its DNA structural conformation (Figure C1). In the presence of the minor variant for rs28382828, it requires a melting temperature (Tm) of 54.9°C to be denatured, yet the designed probe Tm was of 49.9°C, causing low

levels of cross-hybridisation. Genotyping for rs28382828 was therefore performed using an alternative approach: conventional PCR followed by Sanger sequencing. Specific primers containing a M13-cap were selected using Pyromark Assay Design (Qiagen) (Tables 5.3 and C1). They were checked for off-target pairing with the UCSC In-Silico PCR tool (http://genome.ucsc.edu/cgi-bin/hgPcr) and the BLAST-like alignment tool (https://genome.ucsc.edu/cgi-bin/hgBlat/). A large region of 503 bp - containing both rs8192925 and rs28382828 - was amplified in a thermocycler (ThermoFisher) and Sanger sequencing was ordered from the RILD Diagnostic laboratory. Each amplification consisted in 25µg of genomic DNA content in a total volume of 10µL containing 5µL of MegaMix-Royal (Gel Company, San Francisco, USA) with the primers at 200nM. This was performed with the following conditions: 95°C for 12 min, 30 cycles of 95°C for 30 sec and 63°C for 1 min and 72°C for 1 min, and a final step at 72°C for 10 min. The PCR products were then Sanger sequenced to obtain genotyping data for both rs8192925 and rs28382828.

Description		Sequences (5'- to 3'-)
Complete primers	Forward Primer	*TGTAAAACGACGGCCAGTAGCCGAGGAGCCTGAAGA*
	Reverse Primer	*CAGGAACACGCTATGACCGGATTAGGGGCATGAGCCAC*

Table 5.3
Sequences of the primers amplifying by Polymerase Chain Reaction rs8192925 and rs28382828. The primers were selected using Pyromark Assay Design (Qiagen) and a M13 tag was added to the 5'- end of the primers for sequencing ease.

5.2.2.3. Allelic imbalance assay

In order to determine potential effects of the presence of each SNP on mRNA expression or stability, allelic imbalance assays were performed for the 3 SNPs of interest, separately. Using genomic DNA and RNA samples from Exeter 10,000 individuals heterozygous for the studied SNP, this technique allowed for the distinct quantification of the mRNA output from the major and the minor variant. As fully developed in *Chapter 2*, these two measurements were compared after being normalized to the expected 1:1 ratio obtained from genomic DNA. Significant deviation from the expected 1:1 allelic ratio was associated to an unequal expression from each allele.

RNA samples were previously extracted using PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) and the PAXgene Blood miRNA Kit (Qiagen) on the QIAcube automated extraction platform (Qiagen), according to the manufacturer's instructions. 200 ng of each RNA sample was DNAse-treated using the Turbo DNA-free kit (ThermoFisher), according to the manufacturer's instructions, and subsequently reverse transcribed using the VILO cDNA synthesis kit (ThermoFisher) in a total volume of 20µL with the following conditions: 25°C for 10 min, 42°C for 60 min and 85°C for 5 min. cDNA samples were stored at-20°C for a next-day use or at -80°C for long term storage.

For rs11074656 or rs8192925 (separately), allele-specific mRNA levels were determined using the Custom TaqMan SNP Genotyping Assays, on their respective heterozygous gDNA or cDNA samples. Quantitative Real-Time PCR (qRTPCR) was performed as previously described (see *Determination of genotypes*). For each type of sample, the cycle thresholds (Ct) associated to each allele were pooled, and the dCt (being Ct(minor)-Ct(major)) and the ddCt were computed as previously described[180] to obtain an allelic ratio of RNA data normalised to the allelic ratio associated to genomic DNA.

For rs28382828, as the Custom TaqMan SNP Genotyping Assay failed to amplify the sequences (see *Determination of genotypes*), another approach was followed – the

pyrosequencing allele quantification method[333]. DNA pyrosequencing for allele quantitation was performed following published protocols[179,334-337], after conventional amplification by PCR of a 290 bp fragment - spanning rs28382828 - in a total volume of 20µL, using specific forward and reverse primers (Table 5.4, the forward primer displaying a biotin cap). T hese primers were checked for off-target pairing with the UCSC In-Silico PCR tool (http://genome.ucsc.edu/cgi-bin/hgPcr) and the BLAST-like alignment tool (https://genome.ucsc.edu/cgi-bin/hgBlat/). Using the Pyrosequencing Vacuum Prep Tool on a Pyromark workstation, the PCR products were first denatured, and single-stranded biotinylated fragments were captured on streptavidin sepharose beads. After annealing to the sequencing S1 primer (Table 5.4), products were loaded on the Pyromark Q24 plates with the Pyromark Gold Q24 Reagents. The plates were then run on the Pyromark Pyrosequencing system, following manufacturer's instructions to detect the following sequence: *CTTCTCAAAAAAAAAAAAAAAA[AGAA]AGA*. Pyrogram outputs were analysed by the PyroMark Q24 software (all Qiagen) to determine the proportion, in each individual, of product associated to a positive or negative detection of the expected sequence, according to relative peak height. Data was pooled to analyse, separately, samples heterozygous for both rs11074656 and rs28382828, and samples heterozygous for both rs8192925 and rs28382828. The full protocol is provided as *Supplementary Data*.

Primers	**Sequence**
Forward	[btn]*TCAGGCATGATGGCCCATA*
Reverse	*CCACAGTGCCTGGCCTATTAT*
S1	*TAGCTTCTAATCACACACTC*

Table 5.4

Sequences of the primers amplifying by Polymerase Chain Reaction rs28382828, before pyrosequencing. The primers were designed using PyroMark Assay Design (Qiagen), and a biotin-cap (btn) was added to 5'- end of the forward primer.

For rs11075646, rs8192925 and rs28382828, the mean average measurement values from respectively 39, 19 and 52 samples are presented. Stata (Statacorp, USA) and SPSS Statistic 24 (IBM, USA) were used to analyse the data and express the

significance of the results. The Levene's test for equality of variance was used, as well as the Independent sample t-test in SPSS (*Supplementary Data*). A p-value of 0.05 was used as the cut-off for significance.

5.2.3. Clinical study

5.2.3.1. CES2 genetic variations rs11075646 and rs8192925

Both *CES2* SNPs rs11075646 and rs8192925 can be expressed either as major variant (represented as +) or minor variant (-). For each, it is possible to be either homozygous for the major variant (+/+), or heterozygous (+/-) or homozygous for the minor variant (-/-). Overall, nine different genotypes are possible (Table 5.5, with rs11075646 being cited first throughout this chapter).

Genotypes for (alleles presented as "Parent 1/ Parent 2")			Terminologies used throughout this chapter
rs11075646	**rs8192925**	**rs11075646 and rs8192925***	
+/+	+/+	++/++	"double major homozygous"
+/+	+/-	++/+- (or +-/++)	"heterozygous for rs8192925"
+/+	-/-	+-/+-	"minor homozygous for rs8192925"
+/-	+/+	++/-+	"heterozygous for rs11075646"
+/-	+/-	++/-- (or +-/-+)	"double heterozygous"
+/-	-/-	+-/--	"heterozygous for rs11075646 and minor homozygous for rs8192925"
-/-	+/+	-+/-+	"minor homozygous for rs11075646"
-/-	+/-	-+/-- (or --/-+)	"minor homozygous for rs11075646 and heterozygous for rs8192925"
-/-	-/-	--/--	"double minor homozygous"

Table 5.5

Nine different genotypes are possible, involving rs11075646 and rs8192925. For each Single Nucleotide Polymorphism, + indicates the presence of the major variant and - the minor variant. Alleles are presented as "Parent 1 copy/Parent 2 copy", with rs11075646 in first position on each copy. *It is important to note that the genotyping method used in this chapter does not allow chromosomal discrimination, it was therefore impossible to know, for example in the case of a double heterozygous, if both minor alleles were present on the same copy of the chromosome 16 or not, making "++/--" equivalent to "+-/-+" in this chapter*.

5.2.3.2. Patient selection / recruitment

Over 40 patients from the Royal Devon & Exeter (RD&E) hospital were targeted for recruitment through the RD&E Tissue Bank (ETB), the respiratory medicine department and the renal transplant unit. Inclusion and exclusion criteria are provided in Table 5.6.

Inclusion criteria	**Exclusion criteria**
Adults (18 years and above).	Children and youth (up to 18 years).
Patient have capacity to give informed consent to participate to the study.	Patient does not have capacity to give informed consent to participate to the study.
Patient has tuberculosis or has received a kidney transplant.	Patient does not have tuberculosis or has not received a kidney transplant.
Patient has been treated with rifampicin (tuberculosis) for a minimum of 6 weeks or with MMF (kidney transplantee), suggesting drug tolerance.	Patient has not been treated with rifampicin (tuberculosis) or with MMF (kidney transplantee), for a minimum of 6 weeks.
Patient is currently treated with rifampicin (tuberculosis) or with MMF (kidney transplantee).	Patient is not currently treated with rifampicin (tuberculosis) or with MMF (kidney transplantee).
Patient is regularly followed by clinicians from the RD&E hospital.	Patient is not regularly followed by clinicians from the RD&E hospital.
Patient accepts to donate an extra blood sample as part of their routine care.	Patient does not accept to donate an extra blood sample as part of their routine care.
Patient accepts to be genotyped for rs11075646 and rs8192925.	Patient does not accept to be genotyped for rs11075646 and rs8192925.
Patient accepts to have their baseline data (see *Sample donation*) collected and anonymously analysed.	Patient does not accept to have their baseline data (see *Sample donation*) collected and anonymously analysed.

Table 5.6

Inclusion and exclusion criteria to participate to the study on rifampicin or Mycophenolate MoFetil (MMF) metabolism at the Royal Devon & Exeter (RD&E) hospital.

Recruited patients were separated into two different cohorts, a respiratory cohort and a renal one. The respiratory cohort was focused on tuberculosis patients treated with rifampicin, while the renal cohort consisted in kidney transplant patients treated with MMF. 20 patients were expected to join each cohort, the sample size of this pilot study being developed in the results section. For both sub-studies, patients were identified beforehand by their clinicians either as responding adequately to the treatment (control group) or as not responding well to the treatment (lack of treatment efficacy or experience of side-effects).

Tuberculosis patients on rifampicin regimen were selected for this study after 6 weeks of treatment, as clinicians use this breakpoint to estimate patients' response to treatment [338]. Critically non-responding patients, or patients experiencing severe side-effects, were at this time already withdrawn from rifampicin, and could not be included in the study. It was "hoped" that the remaining patients would show various degrees of efficacy or side-effects (abnormal liver function tests results indicating hepatotoxicity), to allow a potential correlation between this and their genotypes for rs11075646 and rs8192925. Amongst the patients on MMF, clinicians were asked to tell apart groups of patients depending on the side-effects they were experiencing: i) no side-effect, ii) diarrhoea only, iii) low white blood count only, iv) both diarrhoea and low white blood count, and v) other side-effects. It is important to note that critically-responding patients and their (expected critical) genotypes were ruled out from the study because these patients, experiencing organ rejection or severe diarrhoea (leading to life-threatening dehydration), were rapidly withdrawn from MMF. *Changes have been made in the recruitment process a posteriori to include them.*

5.2.3.3. Sample donation and baseline data collection

For each patient, a blood sample of 7.5mL was obtained in addition to those taken as part of their routine care. Samples were collected in EDTA blood tubes and stored at -20°C by the NIHR Exeter Clinical Research Facility (NIHR ECRF).

Samples were anonymised using each patient unique ETB study identifying (ID), under which important baseline data known to affect xenobiotic polymorphism was recorded (Table 5.7). As knowledge about the role of CES2 in xenobiotic metabolism is still relatively scarce[108,146,339], personal anthropometric data including sex, age and

Body Mass Index (BMI, *Supplementary Data)* was collected as they could feasibly affect xenobiotic metabolism driven by cytochrome P450 enzymes. It has been indeed known for decades that sex and age could influence drug metabolism[29-30], and that other xenobiotics (medications, tobacco) could stimulate or inhibit metabolic enzymes[101]. Interestingly, there could indeed be a difference in CES2 activity between males and females, as it was shown with cocaine metabolism in rats[340] – and CES2 is the main metabolic enzyme for cocaine[229]. Moreover, if the BMI is already taken into account for rifampicin dosing[322,323], it is not the case for MMF[151].

Baseline data	Complementary information
Sex	As evidenced in [229,340]
Age	As evidenced in [341]
Ethnicity	As *CES2* SNPs are differently distributed in ethnic groups[147].
Body Mass Index (or height and weight)	As evidenced in [151,322,323]
Procedure(s) undertaken	
Concomitant treatments and diseases	As evidenced in [101]
Smoking: history and current status	As evidenced in [101]
Baseline liver function test results (LFT) (*Supplementary Data*, Table C3)	ALanine Transaminase (ALT), Gamma-GlutamylTransferase (GGTP), bilirubin (BIL), ALkanine Phosphatase (ALP), albumin (ALB). Baseline LFT results being used to estimate the activity of the patient liver (hepatotoxicity of rifampicin)[342].
Recent blood test results	Full Blood Count, White Blood Count (WBC), Urea & Electrolites, C-Reactive Protein. A low WBC being a possible side-effect from MMF treatment[343] and blood test results in general indicating if the immune system has been successfully "dampen down".
Results of sputum culture and chest X-ray reports.	For patients under rifampicin, to estimate their treatment efficacy.

Table 5.7

Baseline data parameters known to affect xenobiotic metabolism. They were collected during the study as they can potentially impact on the Carboxylesterase 2 (CES2) metabolism of both rifampicin and Mycophenolate MoFetil (MMF). *SNP stands for Single Nucleotide Polymorphism, all other acronyms being explained in-table.*

Data collection, recording, storage and archiving were undertaken as per standard RD&E Research & Development protocols and procedures to protect patient confidentiality. This involved standardised ETB study specific data collection form, ETB study specific database, password-protected computers and locked filing cabinets accessible only by the ETB research team.

Baseline data not reported in the results section has not being retrieved yet, at the time of submission, as it is the case notably for ethnicity, recent blood test results, results of sputum culture and chest X-ray reports.

5.2.3.4. Sample analysis

DNA was extracted from 200µL of frozen blood sample, using PureLink Genomic DNA kits (ThermoFisher) as per manufacturer's instructions. This required the sequential use of Proteinase K, RNAse A, the Purelink Genomic Lysis/Binding Buffer and 100% ethanol in a first stage, and then Purelink columns were used to purify the DNA per centrifugation – using PureLink Wash Buffers and Elution Buffer as medium. Successful DNA extraction was checked by quantitating the DNA content at 260nm, in replicate reads, on the NanoDrop 8000 Spectrophotometer (ThermoFischer) after a blank of PureLink Elution Buffer (ThermoFisher). After DNA extraction, the remaining frozen blood samples in EDTA blood tubes were stored in -80°C until plasma extraction for (future) rifampicin level analysis for the rifampicin cohort. DNA samples were stored at -20°C for a next-day use, or at -80°C for long term storage.

Genotyping for rs11075646 and rs8192925 were performed on DNA samples using Custom Taqman Genotyping assays (ThermoFisher) (*as described earlier*). Genotype for rs11075646 and rs8192925 was assessed in blinded fashion without knowing the degree of treatment response of patients.

Differences between the expected and observed numbers of people of each genotype in the responder and non-responder/over-responder groups were expected to be assessed by Chi-squared analysis, when the required numbers of analysed samples were met.

5.3. Results

5.3.1. In silico *predictions*

5.3.1.1. Selecting bioinformatically the CES2 genetic variants of interest

This body of work hypothesizes that genetic variation in the *CES2* gene could potentially impact the transcription or stability of *CES2* mRNA, affecting its enzymatic efficacy to metabolise drugs and therefore impacting on treatment efficiency. Using the most recent Short Genetic Variations database (dbSNP17) on UCSC Genome Browser, 1,977 genetic variants could be accounted for on *CES2*. It was important to study relatively "common" variants, to predict and observe - if there is any - an impact on gene expression or enzyme activity which has the potential to be relevant to a large proportion of the world population. The most "common" variants were therefore displayed using the "Variations" tab on UCSC browser: it revealed 18 SNPs carried out by at least 1% of the world population (Figure 5.1 and Figure C2). None of these SNPs were insertions-deletions (indels). Using the 1000Genome Project data entry associated to each SNP, these 18 "common" SNPs were ranked by Minor Allele Frequency in the world (Table C4). Their allelic distribution in different ethnic groups was also accessed, showing allelic divergences between ethnicities, that could possibly explain differences in drug efficacy between populations[338].

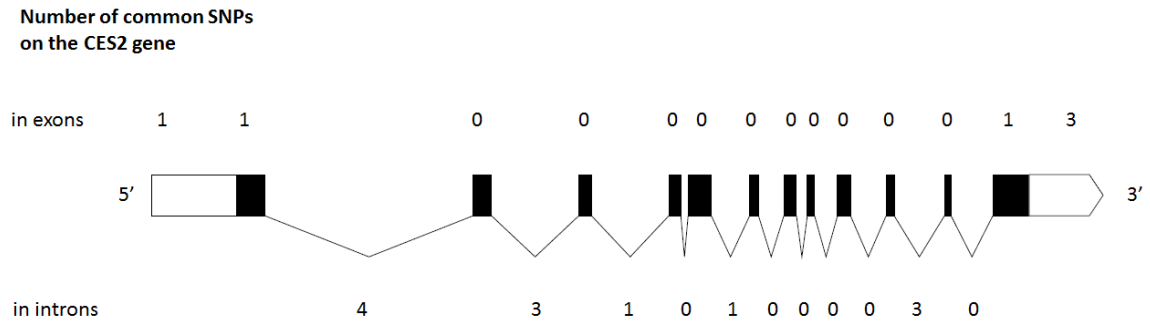


Figure 5.1

Distribution of the 18 "common" Single Nucleotide Polymorphisms (SNPs) along the *Carboxylesterase 2 (CES2)* gene. The 18 "common" *CES2* SNPs are visualised from the 5'-Untranslated Region to the 3'-UTR (both in white). The number of exonic SNPs are provided above the gene drawing, the position of the number indicating on which exon they are located. Similarly, the number of intronic SNPs are given below the gene drawing, the position of the number indicating on which intron they are located. *"Common" SNPs are defined here as associated to a world Minor Allele Frequency of 1% at a minimum. Black vertical boxes represent the exons, black lines the introns*.

As the positional effect of genetic variants on function is well characterized[161,162,179,268,283,291,313,327], the 18 "common" SNPs loci were evaluated (Figure 5.1). As fully developed in *Chapter 2*, it was historically assumed that exonic SNPs are more likely to affect protein concentrations than intronic SNPs, especially when exonic SNPs are associated to a non-synonymous sequence variation (although more recent studies have associated introns from the so-called "junk drawer" to functions[168]). Among the 18 "common" *CES2* SNPs, 6 were exonic: one (rs11075646) was located in *CES2* 5'-UTR, one (rs11863141) was reported in the transcribed exon 1, one (rs28382827) was located in the transcribed exon 12 and three (rs8192925, rs28382828 and rs28382829) were in *CES2* 3'-UTR (Figure 5.1). The exon 1 rs11863141 and exon 12 rs28382827 were discarded as they were both associated with a synonymous amino acid change (for each, a leucine was encoded indifferently of major and minor variants – Table C4). 4 "common" SNPs remained to be studied, and the top 3 in term of world allelic frequency were selected. Thus, the 3 *CES2* SNPs of interest to investigate here were rs28382828, rs11075646 and rs8192925,

associated respectively to MAF of 0.25, 0.13 and 0.09 (in comparison to rs28382829 with a MAF of 0.03).

The 3 SNPs of interest are all located in *CES2* UnTranslated Regions (in 5'-UTR for rs11075646 and in 3'-UTR for rs8192925 and rs28382828). Their potential effect on the gene expression or enzyme activity could be regulatory, by affecting the transcription (for rs11075646 only), gene stability (for rs8182825 and rs28382828), DNA or RNA conformations or – less likely – alternative splicing[161,162,166,167,169,170,267,268,270,281-283,291,292,313,327]. Importantly, rs28382828 is not a SNP *per se*: this is not a "single" nucleotide which is changed when the minor variant is expressed, but an insertion (indel) of either +AA or +GAAA or (most commonly) +AGAA, at chr16:66944326-66944325.

5.3.1.2. Investigating the potential impact of rs11075646, rs8192925 and rs28382828 with bioinformatics

Bioinformatics databases were accessed to assess possible effects of rs11075646, rs8192925 and rs28382828 on *CES2* gene expression or enzyme activity, through alterations in transcription, gene stability and DNA/RNA conformations[161,162,166,167,169,170,267,268,270,281-283,291,292,313,327]. All factors binding to such regulatory elements are reported here only if they were expressed in the liver, the main organ in which CES2 metabolism occurs[138,139,146,147,229,320,321,339].

CES2 5'-UTR variant rs11075646 was evaluated first using ENCODE transcription factor ChIp-Seq on the UCSC genome browser. Its locus was positioned near sequence motifs recognized by two transcription factors (TFs): Taf1 and Hey1. Despite being expressed in the liver, Taf1 and Hey1 were discarded as potential targets for impactful alterations. Indeed, there was no TATA-box (a consensus sequence of T and A base pairs repeats) that could be recognised by Taf1 in rs11075646 vicinity, and Hey1 is reported to recognize 5'-CACGTG-3', a sequence found only in reverse (3'-CACGTG-5') near rs11075646. rs11075646 was therefore deemed unlikely to affect *CES2* transcription through these TFs. The RegRNA 2.0 database[284] was then

accessed using TRANSFAC TF search on *CES2* gene. It associated rs11075646 locus to the binding of a transcription regulator called CCAAT Displacement Protein (CDP), which has hepatic expression. Also known as CDP/Cux or CUX1 or the human protein Cut, CDP is a homeodomain protein that may regulate gene expression[344,345]. It contains 3 regions (Cut Repeat x or CUTx, with x being 1 to 3) that bind to DNA sequences to prevent the binding of other TFs (amongst co-transcription factors such as the hepatic HNF-1 and NRF-1). Both repeats CDP_CR1 and CDP_CR3 HD recognize the motif ATCG(orA)AT, located one base pair upstream of rs11074656 on the *CES2* gene (Figures C1-C2)[346]. The minor allele of rs11075646 was therefore predicted to potentially alter *CES2* transcription through changes in CDP binding.

Through bioinformatic searches on the PolymiRTS, TargetScan, MirSNP and MirSearch databases[286-288,328-331], rs8192925 and rs28382828 were confirmed as to localise within predicted miRNA binding sites (Tables 5.8-9). For both SNPs, the miRBase tool[332] did not show such difference between miRNA binding sites for both alleles. Most of the reported miRNAs were poorly conserved and poorly characterized and no evidence for expression was found *in silico*. If any of the miRNAs reported in Tables 5.8-9 were to be expressed in the liver, the presence of minor variants for rs8192925 or rs28382828 could potentially impact these miRNAs binding on *CES2* mRNA, influencing therefore gene stability and leading to differences in CES2 enzyme concentrations.

miRNA associated to:	miRNA potentially not expressed in the liver	miRNA with no tissue expression
Major variant (A)	hsa-miR-609 (potential break)	hsa-miR-3650 (potential break)
	hsa-miR-4450	hsa-miR-4455 * (potential break)
		hsa-miR-6748-5p (potential break)
		hsa-miR-6772-5p (potential break)
Minor variant (G)	hsa-miR-585-3p	hsa-miR-1268-a (potential creation)
		hsa-miR-1268-b (potential creation)

Table 5.8

MicroRNA (miRNA) binding predictions associated to the expression of rs8192925 alleles. As *Carboxylesterase 2* is mainly expressed in the liver, these potentially disrupted miRNA binding sites are classified regarding the miRNA liver expression using GTEx. *Data retrieved from the PolymiRTS, TargetScan, MirSNP, miRSearch (validation indicated by *) and miRbase databases. Indication of "potential break / creation" is associated to PolymiRTS prediction.*

miRNA associated to:	miRNA potentially expressed in the liver	miRNA potentially not expressed in the liver	miRNA with no tissue expression data
Major variant (-)			hsa-miR-6817-3p
			hsa-miR-6873-3p
			hsa-miR-7110-3p
Minor variant (AGAA)	hsa-miR-130b-5p	hsa-miR-4768-5p (create)	hsa-miR-6809-3p
			hsa-miR-6833-3p

Table 5.9
MicroRNA (miRNA) binding predictions associated to the expression of rs28382828 alleles. As Carboxylerase 2 is mainly expressed in the liver, the potentially disrupted miRNA binding sites are classified regarding the miRNA liver expression using GTEx. *Data retrieved from the TargetScan, MirSNP, miRSearch and miRbase databases. Indication of "create" is associated to MirSNP prediction.*

Finally, genetic variants being also able to impact on the enzyme transcription because of conformation changes[281], *CES2* DNA and RNA structures, with and without each minor variant for rs11075646 and/or rs8192925 and/or rs28382828, were analysed with the mfold Web Server (Figure 5.2 and Figure C1). Visible conformation d)ifferences were established at DNA level when rs8192925 (or rs28382828, data not shown) was present as minor variant (Figure 5.2A). There was also a visible effect in the RNA conformation of the presence of the minor variant for rs11075646 (Figure 5.2B) and an effect in the RNA conformation of the presence of the minor variant for rs8192925. Overall, all 3 of the SNPs of interest are predicted to potentially affect *CES2* in its DNA and/or RNA conformation. By impacting on the DNA or RNA conformations, these variants can render *cis*-regulatory elements unavailable for binding, leading to differences in the regulation of the gene, and therefore possible differences in CES2 concentrations and differences in CES2-mediated drug metabolism.

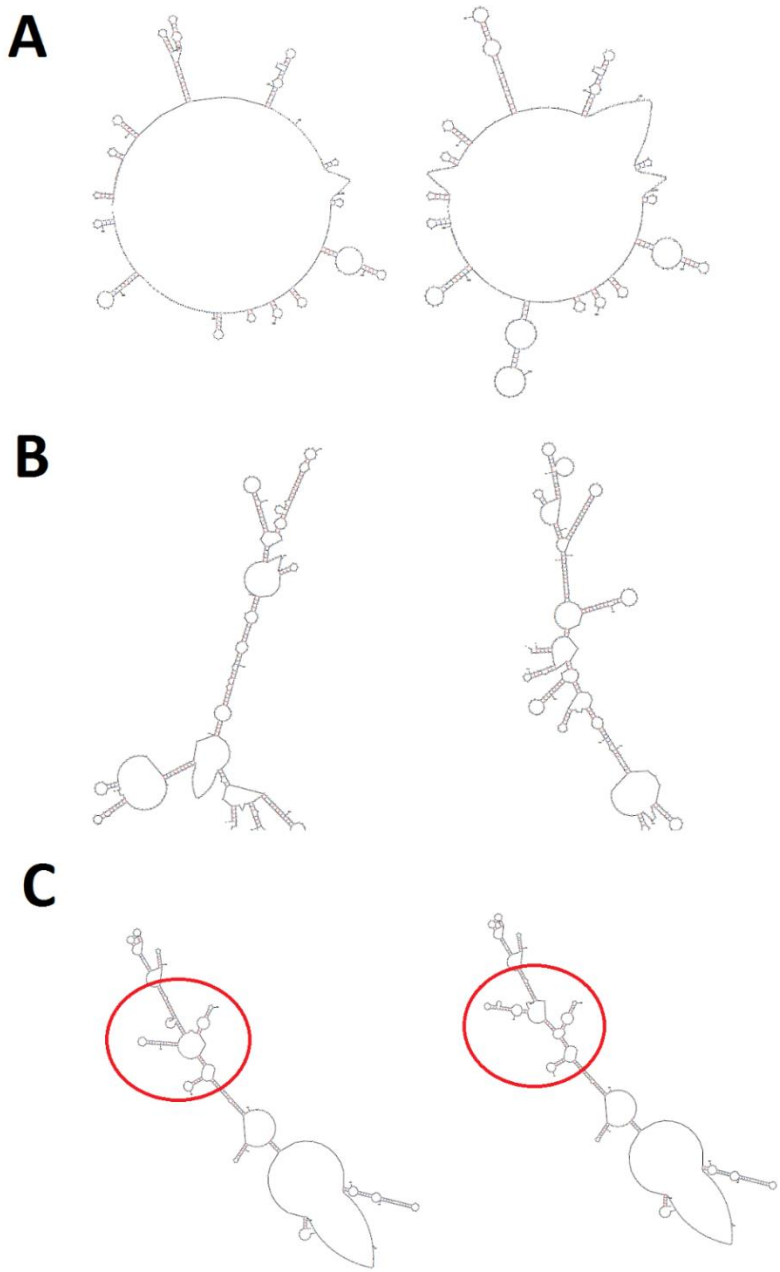


Figure 5.2

Predictions of *Carboxylesterase 2* DNA or RNA structure changes associated to the presence of the minor variants for rs11075646 and rs8192925. A - DNA conformation in the presence (left) or absence (right) of the major variant for rs18192925. B - RNA conformation in the presence (left) or absence (right) of the major variant for rs11075646. C- RNA conformation in the presence of the minor variant for rs8192925. *Conformations retrieved from mfold.*

To summarize, the “common” rs11075646, rs8192925 and rs28382828 were all predicted to be potentially affecting *CES2* (Table 5.10), through its transcription (for rs11075646), *CES2* stability regulation by miRNAs (for rs8192925 and rs28382828) and *CES2* DNA/RNA conformations (for all). Literature was therefore searched, to evaluate if these SNPs have been flagged for such alterations in the past. Searches were made using their current names (rs11075646, rs8192925 and rs28382828), but also their former dbSNP- and HGSV-based names (Table C5). Very little publications were found mentioning those polymorphisms, and most of the retrieved articles failed to study their potential impact on *CES2*. Only the article published by Song about rs8192925 (among other rare genetic variants) could be retrieved[138]. Studying those SNPs was therefore particularly important; thus their relative allelic expression was investigated.

	Regulatory factors predicted to be potentially affected by…		
	rs11075646	**rs8192925**	**rs28382828**
Transcription factor binding sites	CDP		
miRNA binding sites		hsa-miR-3650	hsa-miR-130b-5p
		hsa-miR-4455	hsa-miR-6817-3p
		hsa-miR-6748-5p	hsa-miR-6873-3p
		hsa-miR-6772-5p	hsa-miR-7110-3p
		hsa-miR-1268-a	hsa-miR-6809-3p
		hsa-miR-1268-b	hsa-miR-6833-3p
DNA/RNA conformations	RNA only	DNA and RNA	RNA mostly

Table 5.10
Summary of the predicted types of alterations associated to the presence of the three *CES2* Single Nucleotides Polymorphisms (SNPs) of interest. All listed regulatory factors are hepatically expressed. The CCAAT Displacement Protein (CDP) binding site was co-localised with rs11075646. Several micro RNA (miRNA) binding sites were associated to rs8192925 and rs28382828. rs11075646 and rs28382828 were associated to changes in RNA conformation, rs8192925 to changes in DNA conformations.

5.3.2. Measurement of relative allelic expression for rs11075646, rs8192925 and rs28382828

5.3.2.1. Finding heterozygous samples for the 3 variants of interest in the Exeter 10,000 cohort

In order to study the effect of rs11075646, rs8192925 and rs28382828 on *CES2* gene expression with an allelic imbalance assay, samples of DNA and RNA samples from individuals being heterozygous for at least one of these SNPs needed to be retrieved from the Exeter 10,000 cohort. Ideally, these samples should be heterozygous for only one of the studied SNP, but rs8192925 and rs28382828 were found to be closely associated through linkage disequilibrium search (developed in *Chapter 2*). It was reported that their major variants were frequently inherited together, and their minor variants were also correlated (D' = 0.894 and R^2 = 0.245). There was a linkage equilibrium between the other pairs (rs11075646 and rs8192925, rs11075646 and rs28382828), i.e. they were reported to be randomly inherited together, without any particular correlation (Figure 5.3 for correlation with R^2 – *Supplementary Data* for D', as well as both R^2 and D' data for all 18 "common" *CES2* SNPs in Figures C5-C7).

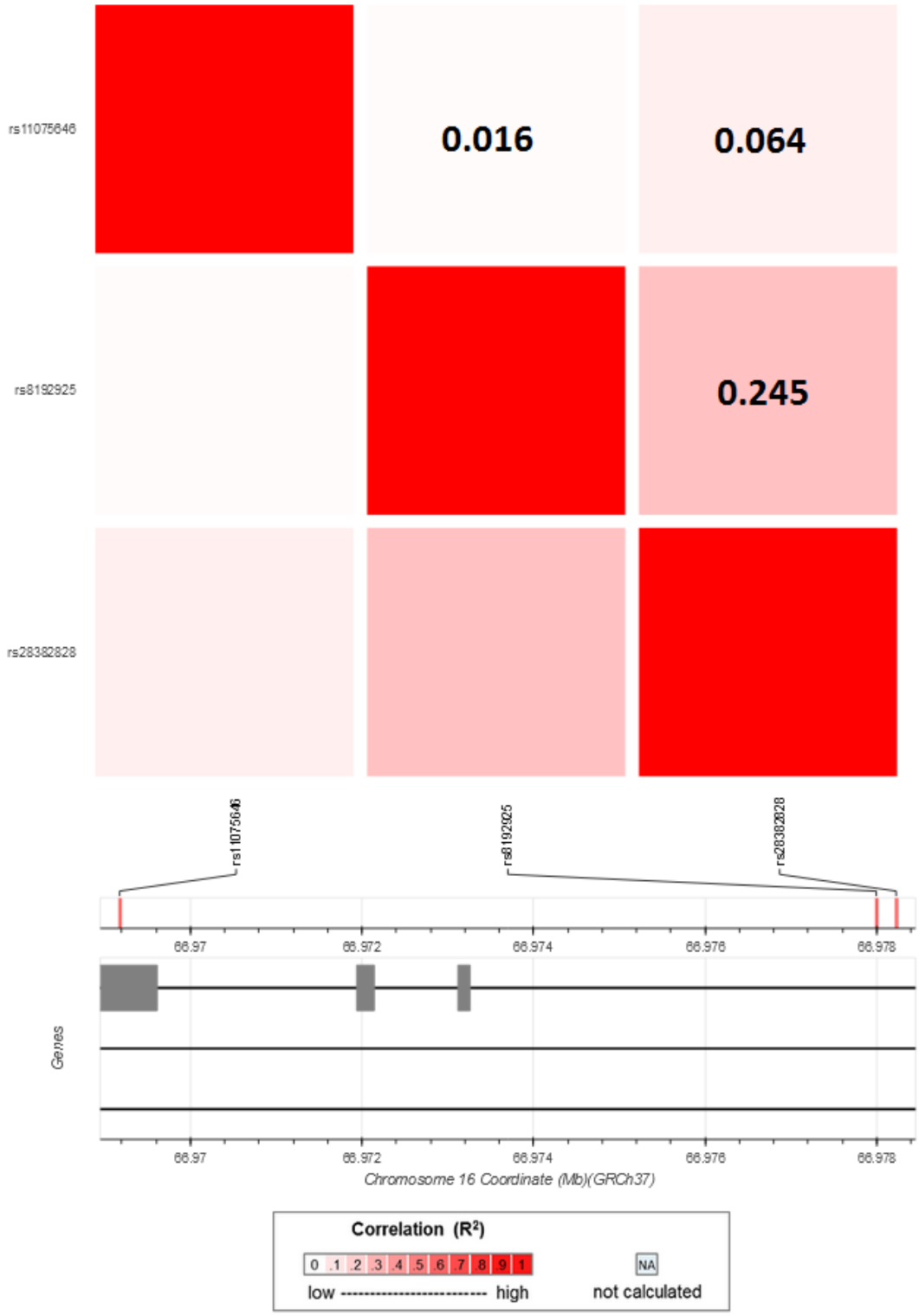


Figure 5.3

Matrix of pairwise linkage disequilibrium statistics in the world population between rs11075646, rs8192925 and rs28382828. *Red shows the correlation in $R^{2,}$ the value is provided. Matrix retrieved from the LDLink tool, based on the 1000Genome Project database.*

To estimate the sample size needed to retrieve heterozygous individuals for rs11075646, rs8192925 and/or rs28382828 from this cohort of people living in the South West of England, ethnicity allelic distribution data from the 1000Genome Project was accessed and the different genotype frequencies for these 3 SNPs were analysed (Table 5.11). It is important to note that, according to the 2011 census, people living in the South West of England are expected to be at 94.9% from a "White British ethnic background"[347], yet their genetic traits have shown particularities in another study[348]: they are expected to follow the genetic distribution trends of both Europeans and world populations. Interestingly, in the 1000Genome Project world population, no people being heterozygous for only rs11075646 or for only rs8192925 (and homozygous for the major variants of the other two SNPs) could be found. In Europe, only 0.4% of the population was heterozygous for only rs28382828 (and homozygous for the major variants of the other two SNPs), while this genotype was found in 7% of the world population.

Important divergences in rs11075646, rs8192925 and rs28382828 genotype frequencies exist between ethnic populations: for example, the most represented genotype in South Asia was the triple homozygous for the major variants (79% of the population) while this genotype is found in only 11% of the African population. If these genotypes are indeed correlated with differences in treatment efficacy, this could explain clinical studies inconsistencies found in the literature regarding CES2 metabolised drugs (rifampicin for example).

Table 5.11 (next page)

Frequencies of the different genotypes for rs11075646, rs8192925 and rs283828, in various populations (world, Africa, East Asia, South Asia, America and Europe). *Values are in red if range between 5-25%, in orange if between 26-46% and in green if above 47%. Data retrieved from the 1000Genome Project.*

Genotype for rs11075646 (C major)	Genotype for rs8192925 (A major)	Genotype for rs28382828 (- major)	Frequencies in…					
			the world	Africa	South Asia	East Asia	America	Europe
C|C	A|A	-|-	51.0%	10.9%	78.5%	52.8%	54.7%	72.4%
C|C	A|A	-|AGAA or AGAA|-	6.9%	23.1%	0%	0%	5.2%	0.4%
C|C	A|A	AGAA|AGAA	3.1%	11.8%	0%	0%	0%	0%
C|C	A|G or G|A	-|-	1.0%	0.2%	0%	1.6%	1.7%	1.8%
C|C	A|G or G|A	-|AGAA or AGAA|-	13.0%	1.1%	7.8%	32.1%	23.3%	8.1%
C|C	A|G or G|A	AGAA|AGAA	0.4%	1.13%	0%	0%	0.6%	0%
C|C	G|G	-|-	0%	0%	0%	0%	0%	0%
C|C	G|G	-|AGAA or AGAA|-	0.3%	0%	0%	1.0%	0.3%	0.2%
C|C	G|G	AGAA|AGAA	1.2%	0%	0.2%	4.6%	2.0%	0%
C|G or G|C	A|A	-|-	6.7%	9.2%	10.0%	5.4%	3.7%	3.6%
C|G or G|C	A|A	-|AGAA or AGAA|-	8.3%	17.2%	3.3%	0%	6.1%	11.1%
C|G or G|C	A|A	AGAA|AGAA	3.5%	13.2%	0%	0%	0.3%	0%
C|G or G|C	A|G or G|A	-|-	0.08%	0%	0.2%	0.2%	0%	0%
C|G or G|C	A|G or G|A	-|AGAA or AGAA|-	0.9%	0.6%	0.4%	2.4%	0.9%	0.4%
C|G or G|C	A|G or G|A	AGAA|AGAA	0.3%	6.4%	0%	0%	0.3%	0.4%
C|G or G|C	G|G	-|-	0%	0%	0%	0%	0%	0%
C|G or G|C	G|G	-|AGAA or AGAA|-	0%	0%	0%	0%	0%	0%
C|G or G|C	G|G	AGAA|AGAA	0%	0%	0%	0%	0%	0%
G|G	A|A	-|-	0.6%	1.8%	0.4%	0%	0%	0%
G|G	A|A	-|AGAA or AGAA|-	2.0%	6.7%	0%	0%	0.6%	0.6%
G|G	A|A	AGAA|AGAA	0.9%	2.6%	0%	0%	0.3%	1.0%
G|G	A|G or G|A	-|-	0%	0%	0%	0%	0%	0%
G|G	A|G or G|A	-|AGAA or AGAA|-	0%	0%	0%	0%	0%	0%
G|G	A|G or G|A	AGAA|AGAA	0%	0%	0%	0%	0%	0%
G|G	G|G	-|-	0%	0%	0%	0%	0%	0%
G|G	G|G	-|AGAA or AGAA|-	0%	0%	0%	0%	0%	0%
G|G	G|G	AGAA|AGAA	0%	0%	0%	0%	0%	0%

Table 5.11

As it was not possible to predict which of the global or the European trends in genotype distribution was followed our cohort, we extrapolated the genotype distribution for a sample size of 300 individuals from the Exeter 10,000 cohort in both cases (Table 5.12). Given the linkage disequilibrium between rs8192925 and rs28382828, it was not surprising that it was predicted be near to impossible to retrieve samples from people being heterozygous for only one SNP (and homozygous major variant for the other SNPs) with such a sample size, especially if the cohort distribution of genotype was following the European one. Yet only 291 DNA and RNA samples were available for this study to use, therefore the sample size was validated, with the hope that it ensured the representation of different genotypes by following the global trend (as it was estimated that a total of 43.8 single heterozygous could be found in such distribution, contrarily to 17.4 for a more European cohort).

It was also hoped that the allelic imbalance assay would only show an imbalance for one or two SNPs for this study to be able to distinguish the effects of the SNPs in the highly probable double heterozygous samples. This was not sample size dependant.

Genotype for rs11075646 (C major)	Genotype for rs8192925 (A major)	Genotype for rs28382828 (- major)	World frequencies	Global estimation for 300 samples	European frequencies	European estimation for 300 samples
C|C	A|A	-|AGAA or AGAA|-	6.9%	20.7	0.4%	1.2
C|C	A|G or G|A	-|-	1.0%	3	1.8%	5.4
C|C	A|G or G|A	-|AGAA or AGAA|-	13.0%	39	8.1%	24.3
C|C	A|G or G|A	AGAA|AGAA	0.4%	1.2	0%	0
C|C	G|G	-|AGAA or AGAA|-	0.3%	0.9	0.2%	0.6
C|G or G|C	A|A	-|-	6.7%	20.1	3.6%	10.8
C|G or G|C	A|A	-|AGAA or AGAA|-	8.3%	24.9	11.1%	33.3
C|G or G|C	A|A	AGAA|AGAA	3.5%	10.5	0%	0
C|G or G|C	A|G or G|A	-|-	0.08%	0.24	0%	0
C|G or G|C	A|G or G|A	-|AGAA or AGAA|-	0.9%	2.7	0.4%	1.2
C|G or G|C	A|G or G|A	AGAA|AGAA	0.3%	0.9	0.4%	1.2
C|G or G|C	G|G	-|-	0%	0	0%	0
C|G or G|C	G|G	-|AGAA or AGAA|-	0%	0	0%	0
C|G or G|C	G|G	AGAA|AGAA	0%	0	0%	0
G|G	A|A	-|AGAA or AGAA|-	2.0%	6	0.6%	
G|G	A|G or G|A	-|-	0%	0	0%	0
G|G	A|G or G|A	-|AGAA or AGAA|-	0%	0	0%	0
G|G	A|G or G|A	AGAA|AGAA	0%	0	0%	0
G|G	G|G	-|AGAA or AGAA|-	0%	0	0%	0

Table 5.12

Estimation of the frequencies of people being heterozygous for at least one Single Nucleotide Polymorphism in a group of 300 individuals from the Exeter 10,000 cohort. *Values are in red if range between 5-25%, in orange if between 26-46% and in green if above 47%. Rs11075646, rs8192925 and rs28382828 data was retrieved from the 1000Genome Project.*

291 DNA samples from the Exeter 10,000 cohort were genotyped for rs11075646, rs8192925 and rs28382828 (Table 5.13). Seven different genotypes were found, and the distribution of their frequencies amongst this cohort seemed to follow the expected distribution for the European data (Table 5.11). As estimated, the triple homozygous for the major variants was the most represented genotype (77.3% in the cohort, estimated 72.4% in Europe) and the two most frequent genotypes amongst those harbouring at least one heterozygous variant were, on one hand, the double heterozygous for rs11075646 and rs28382828 (homozygous for the major variant of rs8192925 - 13.7% of the cohort, estimated 11.1% in Europe) and, on the other hand, the double heterozygous for rs8192925 and rs28382828 (homozygous for the major variant of rs11075646 - 6.5% of the cohort, estimated 8.1% in Europe).

Genotype for rs11075646 (C major)	Genotype for rs8192925 (A major)	Genotype for rs28382828 (- major)	Distribution amongst the 291 samples
C|C	A|A	-|-	225 (77.3%)
C|C	A|A	-|AGAA or AGAA|-	0
C|C	A|A	AGAA|AGAA	0
C|C	A|G or G|A	-|-	0
C|C	A|G or G|A	-|AGAA or AGAA|-	19 (6.5%)
C|C	A|G or G|A	AGAA|AGAA	1 (0.3%)
C|C	G|G	-|-	0
C|C	G|G	-|AGAA or AGAA|-	0
C|C	G|G	AGAA|AGAA	0
C|G or G|C	A|A	-|-	1 (0.3%)
C|G or G|C	A|A	-|AGAA or AGAA|-	40 (13.7 %)
C|G or G|C	A|A	AGAA|AGAA	3 (1.0%)
C|G or G|C	A|G or G|A	-|-	0
C|G or G|C	A|G or G|A	-|AGAA or AGAA|-	0
C|G or G|C	A|G or G|A	AGAA|AGAA	2 (0.7%)
C|G or G|C	G|G	-|-	0
C|G or G|C	G|G	-|AGAA or AGAA|-	0
C|G or G|C	G|G	AGAA|AGAA	0
G|G	A|A	-|-	0
G|G	A|A	-|AGAA or AGAA|-	0
G|G	A|A	AGAA|AGAA	0
G|G	A|G or G|A	-|-	0
G|G	A|G or G|A	-|AGAA or AGAA|-	0
G|G	A|G or G|A	AGAA|AGAA	0
G|G	G|G	-|-	0
G|G	G|G	-|AGAA or AGAA|-	0
G|G	G|G	AGAA|AGAA	0

Table 5.13

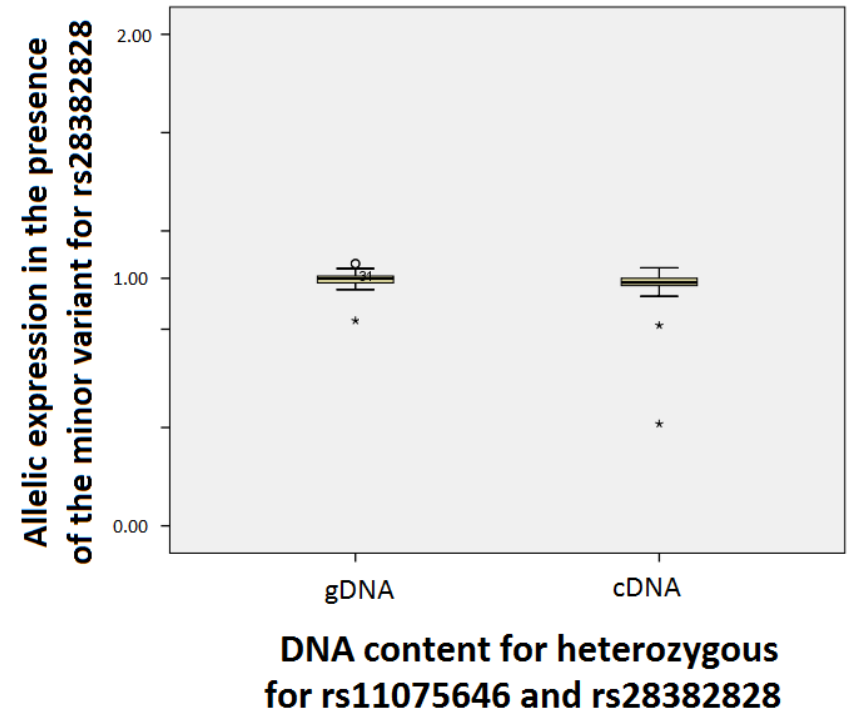
Distribution of the genotypes for rs11075646, rs8192925 and rs28382828 amongst the 291 samples of the Exeter 10,000 cohort.

As expected by the European distribution, this sample size did not allow for a clear distinction between the 3 SNPs of interest: no heterozygous for only rs8192925 or only rs28382828 were retrieved in the cohort, and only one heterozygous for only rs11075646 (being homozygous for the other two major variants) was found. The allelic expression was therefore assessed in samples being heterozygous for two SNPs - separately, with the 19 samples from individuals being double heterozygous for rs8192925 and rs28382828, and the 40 double heterozygous for rs11075646 and rs28382828.

5.3.2.2. Investigating the effect of rs11075646, rs8192925 and rs28382828 on allele-specific CES2 mRNA expression

Having "selected" the heterozygous individuals for rs11075646, rs8192925 and rs28382828 from the Exeter 10,000 cohort, their allelic expression was assessed by comparing their gDNA and cDNA levels with an allelic imbalance assay, for each SNP separately – as developed in *Chapter 2*. By normalising gDNA levels to 1 and comparing to those the cDNA levels, it allows the quantification of expression of mRNA deriving from the allele carrying the minor allele relative to that carrying the major allele.

The genotype of 59 individuals being heterozygous for rs28382828 were assessed for their allelic balance on this SNP by pyrosequencing. 7 samples failed to be successfully pyrosequenced, leaving 52 samples: 35 individuals being heterozygous for both rs11075646 and rs28382828, and 17 being heterozygous for both rs8192925 and rs28382828. Pyrosequencing data was analysed separately for these two sub-cohorts (Figure 5.4): it showed for each a non-significant allelic imbalance between their levels of gDNA and cDNA, meaning the expression of mRNA deriving from each allele was equal - rs28382828 has potentially no effect on *CES2* gene expression.



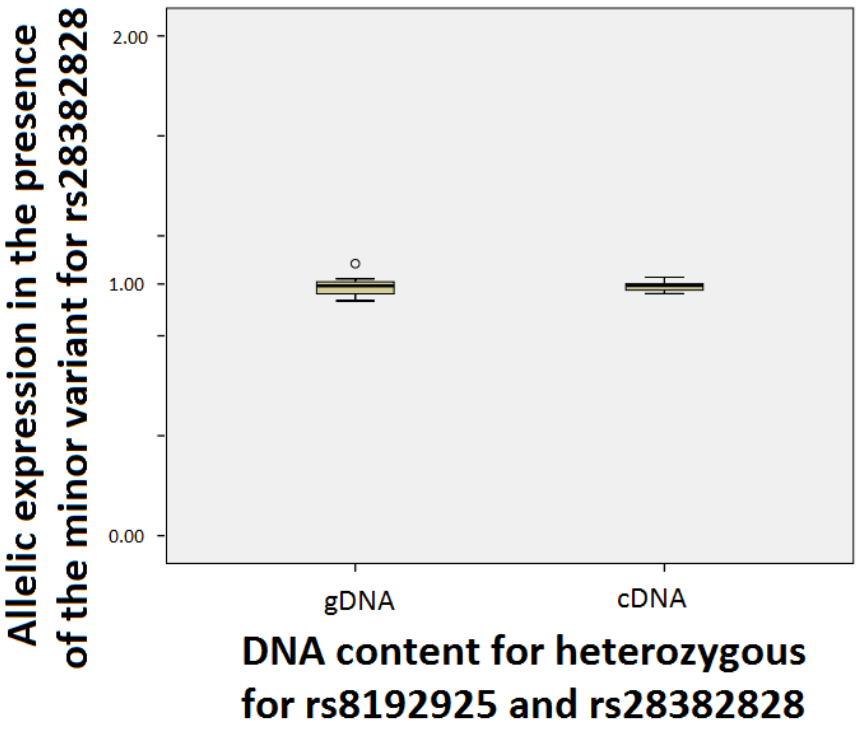


Figure 5.4

Allelic balance for rs28382828 between genomic and complementary DNA amounts in heterozygous individuals for rs28382828, and either rs11075646 or rs8192925. Top – Expression of mRNA deriving from the allele carrying the minor

allele relative to that carrying the major allele, in individuals being heterozygous for both rs11075646 and rs28382828 (and homozygous for the major variant of rs8192925). Bottom - Expression of mRNA deriving from the allele carrying the minor allele relative to that carrying the major allele, in individuals being heterozygous for both rs8192925 and rs28382828 (and homozygous for the major variant of rs11075646). Both - There is no significant difference between the means of genomic DNA (gDNA) and complementary DNA (cDNA) samples for this allelic expression of rs28382828 (respective p-values of 0.101 for top and 0.601 for bottom – Figure C8). *Data are represented as box plots, with the boxes bounding to the interquartile ranges divided by the median, and Tukey-style whiskers extending to a maximum of 1.5-times the interquartile range beyond the box. Open circles and stars are sample data points outside those values.*

The 40 individuals being heterozygous for rs11075646 and rs28382828 (and homozygous for the major variant of rs8192925) were assessed for their allelic expression for rs11075646 by qRTPCR. One sample was discarded due to a lack of qPCR effectiveness - the allelic imbalance assay outcome is presented for the 39 remaining individuals (Figure 5.5). gDNA levels were normalised to 1 and, compared to those, cDNA levels are in significant imbalance. cDNA levels are indeed significantly 3-fold higher than the gDNA levels in the presence of the minor variant for rs11075646. rs28382828 not been previously associated to an allelic imbalance (Figure 5.4), rs11075646 has a significant effect on the *CES2* gene expression, potentially by increasing its transcription by 3-fold when the minor variant is expressed.

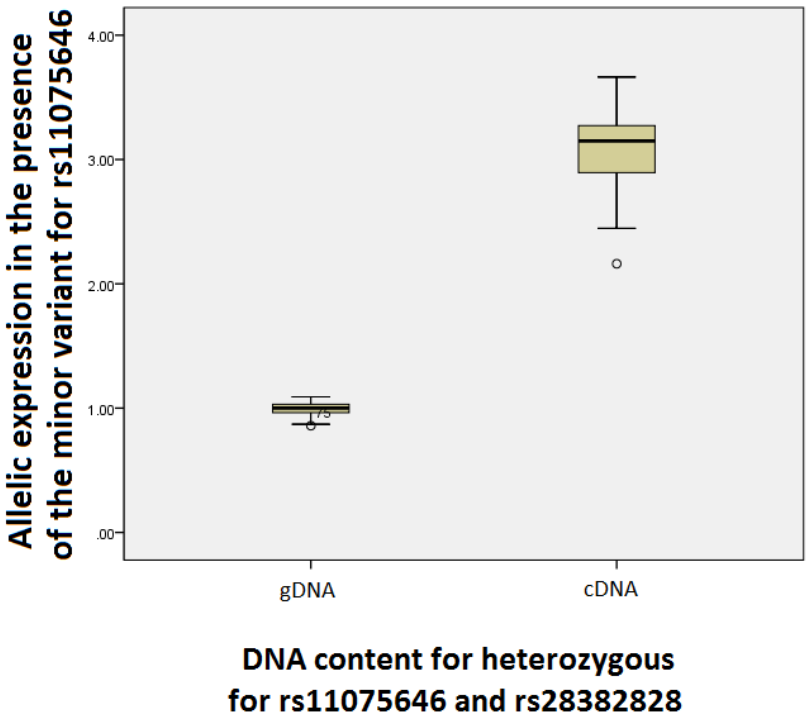


Figure 5.5

Allelic imbalance between the genomic and complementary DNA amounts in heterozygous individuals for rs11075646 (being also heterozygous for rs28382828 and homozygous for the major variant of rs8192925). Expression of mRNA deriving from the allele carrying the minor allele relative to that carrying the major allele, in individuals being heterozygous for both rs11075646 and rs28382828 (and homozygous for the major variant of rs8192925). There is a significant 3-fold difference between the means of genomic DNA (gDNA) and complementary DNA (cDNA) samples for this allelic expression of rs11075646 (p-value of 0.001 – Figure C9).

Data are represented as box plots, with the box bounding to the interquartile ranges divided by the median, and Tukey-style whiskers extending to a maximum of 1.5-times the interquartile range beyond the box. Open circles are sample data points outside those values.

The 19 individuals being heterozygous for rs8192925 and rs28382828 (and homozygous for the major variant of rs11075646) were assessed for their allelic expression for rs8192925 by qRTPCR. The allelic imbalance assay outcome is presented for the 19 individuals (Figure 5.6), it showed cDNA levels are in imbalance with their normalized gDNA levels for rs8192925. The cDNA levels are indeed significantly 1.13-fold lower than the gDNA levels in the presence of the minor variant for rs8192925. rs28382828 not been associated to an allelic imbalance, rs8192925 has a significant effect on *CES2* expression, potentially by increasing its silencing by 1.13-fold when the minor variant is present.

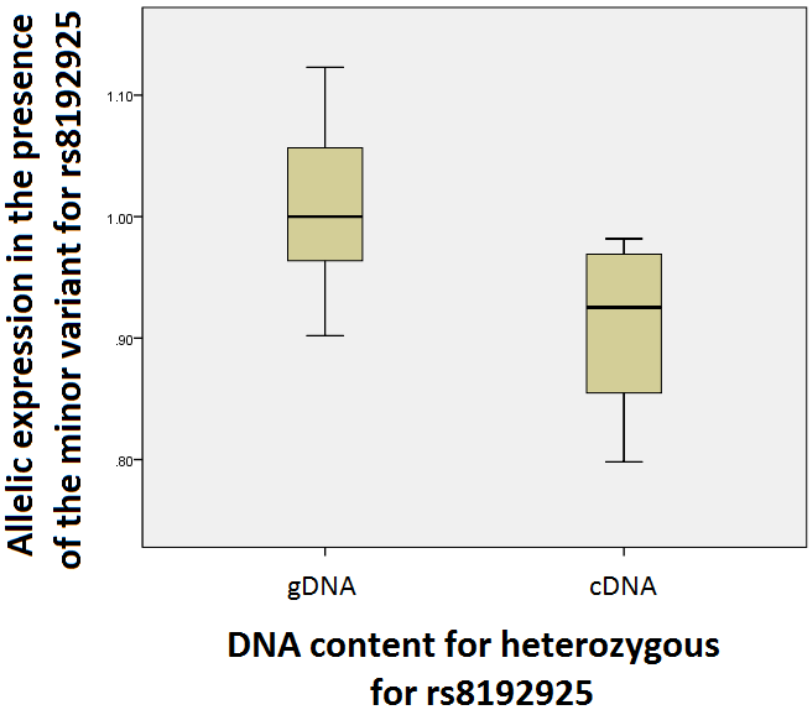


Figure 5.6

Allelic imbalance between the genomic and complementary DNA amounts in heterozygous individuals for rs8192925 (being also heterozygous for rs28382828 and homozygous for the major variant of rs11075646). Expression of mRNA deriving from the allele carrying the minor allele relative to that carrying the major allele, in individuals being heterozygous for both rs8192925 and rs28382828 (and homozygous for the major variant of rs11075646). There is a significant 1.13-fold difference between the means of genomic DNA (gnDNA) and complementary DNA (cDNA) samples for this allelic expression of rs8192925 (p-values of 0.001 – Figure C10).

Data are represented as box plots, with the box bounding to the interquartile ranges divided by the median, and Tukey-style whiskers extending to a maximum of 1.5-times the interquartile range beyond the box.

Overall, the minor allele for rs11075646 and rs8192925, but not rs28382828, were shown to be associated to significant differences in *CES2* mRNA levels, possibly by altering its transcription (rs11075646) or stability (rs8192925). As differences at the *CES2* mRNA levels have been correlated to differences at the CES2 protein levels for this enzyme[326], a clinical study was designed to explore these association further.

5.3.3. Clinical study of rs11075646 and rs8192925

5.3.3.1. Setting-up the clinical study: the sample size

This clinical study was planned in collaboration with clinicians from the RD&E Hospital and personnel from the Exeter Tissue Bank and from the NIHR Exeter Clinical Research Facility. Approximately 15-20 patients per annum are treated with rifampicin in the RD&E hospital for an active or latent *Mycobacterium tuberculosis* infection or another non-tuberculous mycobacterial infection (personal communication). Each year, around 250 renal transplant patients are treated with MMF at the RD&E hospital (personal communication). Due to budget constraints, 20 patients were expected to be selected in each cohort, if possible. In the 9 months of this pilot study, 6 patients were recruited in the respiratory cohort and 12 in the renal one. It was expected that the collected patients would yield the expected genotype distributions for both rs11075646 and rs8192925, predicted by the Hardy-Weinburg equilibrium [349] and the world data in the 1000Genome Project (Table 5.11). The choice of using the world distribution of these genotypes found in the 1000Genome Project is discussed in *Supplementary data*, where European and British distributions are also developed. The expectations were therefore that, amongst the 20 potential patients in each sub-study, 7 (39%) would carry a genotype that differs from double major homozygous (++/++), the genotype one would predict most likely to be found in patients responding optimally to their treatment. Undoubtedly, data was missing to make clear predictions with such small sample size, yet this is an ongoing collection.

rs11075646 genotype	rs8192925 genotype	World distribution (%)
+/+	+/+	60.98
+/-	+/+	18.49
+/+	+/-	14.30
-/-	+/+	3.43
+/+	-/-	1.52
+/-	+/-	1.28
+/-	-/-	0
-/-	+/-	0
-/-	-/-	0

Table 5.14
Expected distribution of each possible genotype for rs11075646 and rs8192925 in the world, based on the 1000Genome Project data (2,504 participants). For each Single Nucleotide Polymorphism, + indicates the presence of the major variant, - the minor variant.

However, a previous study showed a correlation between alteration of the *CES2* mRNA levels and a difference at the enzyme levels [325]. As both rs11075646 and rs8192925 minor variants were found to show very robust significant differences at the mRNA level (the minor variant of rs11075646 conferring a 3-fold increase and the minor variant of rs8192925 a 1.13-fold decrease of transcription), we predicted that were these variants to be present at the expected frequencies in our cohort, considerable divergences in phenotype may be observed in terms of treatment response, given a previous study in which less considerable effects were observed for similar sample size[350]. Yet, if no significant observation was made, the potential effects of such SNP was not expected to be ruled out.

5.3.3.2. Preliminary genotyping data

At the time of writing, only 6 samples from patients enrolled in the respiratory cohort and 12 from the renal one were available. DNA was extracted from the blood sample from these patients, and genotyping was performed to assess their genotype towards rs11075646 and rs8192925 (Table 5.15).

Cohort	Patient number	rs11075646 genotype	rs8192925 genotype	Double major homozygosis
RIFAMPICIN	R001	+/+	+/+	YES
	R002	+/+	+/+	YES
	R003	+/+	+/+	YES
	R004	+/+	+/+	YES
	R005	+/+	+/+	YES
	R006	+/-	+/+	NO (heterozygous for rs11075646)
MMF	R007	+/-	+/-	NO (double heterozygous)
	R008	+/+	+/+	YES
	R009	+/+	+/+	YES
	R010	+/+	+/+	YES
	R011	+/+	+/+	YES
	R012	+/+	+/+	YES
	R013	+/+	+/+	YES
	R014	+/+	+/-	NO (heterozygous for rs8192925)
	R015	+/+	+/+	YES
	R016	+/-	+/+	NO (heterozygous for rs11075646)
	R017	+/-	+/+	NO (heterozygous for rs11075646)
	R018	+/-	+/+	NO (heterozygous for rs11075646)

Table 5.15

Preliminary data - genotypes for the 6 patients from the rifampicin cohort, and the 12 patients from the Mycophenolate MoFetil (MMF) cohort. For each Single Nucleotide Polymorphism, + indicates the presence of the major variant, - the minor variant.

Out of the 6 rifampicin patients, 5 (83%) were major homozygous for both SNPs (++/++) with a predicted basal metabolism, and 1 (17%) was heterozygous for rs11075646 (++/+-). This fits with the general distribution of these alleles in the world and in Europe, these two genotypes being the two most common, and the major homozygous for both SNPs being dominant in proportion.

Amongst the 12 MMF patients, 7 (58%) were double major homozygous (++/++), 3 (25%) were of ++/-+ genotype (heterozygous for rs11075646), 1 (8%) was double heterozygous (++/--) and 1 (8%) was of ++/+- genotype (heterozygous for rs8192925). The repartition fits with the general distribution of those genotypes in the world and in Europe, ++/++ being indeed the most common and ++/-- and ++/+- being the rarest.

It is important to point out that this data is only preliminary with very limited sample sizes and precludes any correlation with phenotypes presently, as numbers are too small. Given the expression differences and the variability in expression, it was estimated that 20 patients per cohort would be a better sample size to analyse. However, it would be interesting to know, compared to the basal metabolisers in their cohorts, i) if the patient of ++/+- genotype on rifampicin has been identified as a fast metaboliser by their clinicians (through a less effective treatment or a lack of hepatotoxicity at high doses), ii) if the ++/+- patient in the MMF cohort was flagged as less responsive to treatment and iii) if the 3 MMF-treated patients of ++/-- or ++/+- are super-responsive to the treatment, with or without an increase of experienced side-effects – all of these predictions being made with the assumptions found in *Supplementary data*.

5.3.3.3. Preliminary clinical data

At the time of submitting this thesis, most of the important baseline data for the 18 genotyped patients was not retrieved yet. Preliminary clinical data is therefore shared (Tables 5.16-18 and Tables C6-7) and discussed here. It is important to note that the degree of response to treatment and the genotyping analysis were studied separately to avoid involuntary bias.

The following table (presented on the next page) presents preliminary clinical data from the 6 patients in the rifampicin cohort.

(Table 5.16 next page) Preliminary clinical data from the 6 patients in the rifampicin cohort. Preliminary data consists for each patient in: age (in years), sex, Body Mass Index (BMI), rifampicin dosage (in mg/day), selection of other concomitant diseases and treatments and liver function test results (LFT; ALanine Transaminase ALT, Gamma-GlutamylTransferase GGTP, bilirubin BIL, ALkanine Phosphatase ALP, albumin ALB – all being in red if there are outside the normal ranges, as developed in Table C6). The full list of concomitant diseases and treatments is available in Tables C6-7.

		Patient R001	Patient R002	Patient R003	Patient R004	Patient R005	Patient R006
Age (years)		33	73	41	41	32	61
Sex		Female	Male	Female	Male	Female	Female
BMI		23.4	24.4	32.1	23.8	31.0	18.5
Smoking history and status		never	Not currently (for 40 years)	Not currently (for 20 years)	Not currently (for 20 years)	never	never
Rifampicin dosage (mg/day)		600	600	900	300	120	900
Selection of concomitant	**diseases**	None	Vasculitis, infarct	Asthma, XXX chromosome syndrome	Pseudomonas infection, superimposed lower respiratory tract infection	Lymphade-nopathy	Bronchiectasis
	treatments	isoniazid pyridoxine	azithromycin cotrimoxalone ethambutol prednisolone rituximab	amikacin azithromycin ethambutol	carbocisteine colomycin isoniazid tazocin salbutamol	ethambutol isoniazid pyranamide pyrisoxine	azithromycin ethambutol clofazimine
LFT results	**ALT**	13	17	12	40	18	22
	GGTP	20	50	29	31	25	16
	BIL	5	10	5	12	2	14
	ALP	81	98	66	64	99	62
	ALB	47	44	44	45	47	46

Table 5.16

As described in Table 5.16, out of the 6 patients on rifampicin, which are from both sex, there was a wide variety of ages (32-73 years old) and BMI categories (developed in *Supplementary data*). Three of the patients were smoking heavily until recently, none are diabetic but 5 out of the 6 patients have several concomitant diseases and infections, not necessarily related to their tuberculosis infection. They all followed other courses of medications, including some that are usually influencing xenobiotic metabolism[101].

Contrarily to the current guidelines (which have been reported to be followed by physicians[84]), rifampicin dosage does not seem to be directly related to the BMI: the clinically-obese patient being prescribed the lowest amount of drug in this cohort. It may be related to the degree / stage of tuberculosis or the presence of morbidity risk factors.

The degree of treatment response was not retrieved yet for these patients on rifampicin, but the main side-effect of this treatment being hepatotoxicity, we could estimate if they were experiencing liver damage, using their LFT results (Tables 5.16 and C3). Out of the five LFT results per patient, no value were found to be outside of the normal range. Patients seemed therefore to tolerate well their treatment, but the low preliminary sample size does not allow to conclude on the projected effect of their genotype.

The following table (presented on two pages) presents preliminary clinical data from the 12 patients in the Mycophenolate MoFetil (MMF) cohort.

(Table 5.17 – next page) Preliminary clinical data from the 12 patients in the Mycophenolate MoFetil (MMF) cohort. Preliminary data consists in: age (in years), sex, Body Mass Index (BMI), MMF dosage (in mg/day), selection of other concomitant diseases, selection of other concomitant treatment and liver function tests results (LFT; ALanine Transaminase ALT, bilirubin BIL, ALkanine Phosphatase ALP, albumin ALB – all being in red if there are below or above the norm data developed in Table C3). Patients R008 and R009 were on MMF for less than 6 weeks. NC stands for Non-Communicated (at the time of thesis submission). The full list of concomitant diseases and treatments is available in Tables C6-7.

Patients		R007	R008	R009	R010	R011	R012
Age (years)		NC	56	31	NC	31	54
Sex		Male	Male	Male	Female	Male	Male
BMI		23.6	36	NC	Expected healthy	Expected overweight	30
Smoking status		NC	Currently	NC	N	Yes	NC
MMF dosage (mg/day)		1000	2000	2000	2000	2000	1000
Selection of concomitant	diseases	nephrotectomy (2013).	type 2 diabetes hypertension reiter's disease	hypertension.	nc	neuropathy.	type 2 diabetes.
	treatments	adoport aspirin atorvastatin prednisolone	adoport aspirin cotrimoxalone prednisolone	adoport aspirin paracetamol prednisolone	adoport aspirin atorvastatin	prograf prednisolone ramitidine	aspirin prednisolone nebivolol
LFT results	ALT	37	20	36	7	18	47
	BIL	7	5	**24**	3	6	17
	ALP	116	81	74	**171**	87	53
	ALB	46	40	47	37	46	47

Patients		R013	R014	R015	R016	R017	R018
Age (years)		36	49	20	53	64	50
Sex		Female	Male	Male	Male	Male	Female
BMI		NC	Expected obese	Expected healthy	Expected overweight	22.9	30
Smoking status		NC	N	N	Not currently	N	N
MMF dosage (mg/day)		500	1000	500	1000	1000	1000
Selection of concomitant	diseases	Type 1 diabetes cmv neutropenia	IgA nephropathy peritoneal dialysis	hypertension	deep vein thrombosis.	uretic stone.	multiple sclerosis Type 1 diabetes asthma
	treatments	NC	folic acid rampiril atrovastatin	prednisolone adoport atorvastatin	prednosolone lausoprasole adoport aprixaban	doxozosin prednisolone adoport	insulin adoport omeprazole
LFT results	ALT	71	9	17	23	8	17
	BIL	7	9	16	4	5	5
	ALP	221	60	130	48	49	66
	ALB	44	48	51	NC	42	43

Table 5.17

As seen in Table 5.17, most of the 12 patients treated with MMF following kidney transplant are men, and ages vary greatly in the cohort (from 20 to 64 years old). BMI and smoking status were not all indicated yet, some of the patients have Type 1 or Type 2 diabetes – amongst other concomitant diseases such as hypertension, asthma, multiple sclerosis, etc.

Ranges of MMF dosing went from 500mg to 2000mg per day – these do not seem to be proportionate to the BMI (or reported body weight), as the higher dose was attributed indifferently to expected healthy and obese patients. Clinicians clarified the dosing strategy in the unit: MMF is usually delivered at 1000mg per day, and doses are increased (or decreased) as a function of the response (increase in case of low efficacy, decrease in case of side-effects). Doses can also be “spaced out” to limit the severity of side-effects (personal communication).

As part of their kidney transplant and other concomitant diseases (full list available in Table C6), these patients are heavily medicated and other treatments include adoport, aspirin / paracetamol, predmisolone, rampiril, folic acids, omeprazole, etc., whose effects on the CES2 enzyme are not documented[108]. Contrarily to rifampicin patients, a third of the MMF patients present are potentially suffering from liver damage, as they present at least one LFT result as nonstandard. A statistical correlation between these LFT results and – separately - the responsive status reported by clinicians, and patients’ genotype for rs11075646 and rs8192925 could be attempted when the sample and data collection will be finalised.

Precise response to treatment was not retrieved yet for the 18 samples patients but preliminary data was shared about the side-effects experienced by the MMF-treated patients (Table 5.18). Out of the 12 patients, 3 did not experience any side-effects, 2 experienced only diarrhoea, 2 experiences only a low white blood count, 3 experienced both (diarrhoea and low white blood count), 1 experienced anaemia (but analysis were planned to make sure it was not the sign of an autoimmune disease) and 1 was transferred from another hospital and therefore side-effect history was not available.

Clinicians provided the insight that patients experiencing severe diarrhoea, given the high risks of dehydration and therefore withdrawal, were administrated “spread out” MMF dosages: instead of delivering 1000mg once a day, they were administrated 500mg twice a day. This was included in the dataset (Table 5.18), to estimate the degree of the reported side-effects.

Even though the sample size was not high enough to draw any conclusion yet, the repartition of the different genotypes amongst these groups of patients experiencing similar side-effects seemed heterogenous, with for example patients not being double major homozygous for rs11075645 and rs8192925 experiencing either no side-effect, or only diarrhoea, or only low white blood count, or both of these main side-effects, or even the possible anaemia. It is important to note two possible biases in the study: i) side-effects can be due to a concomitant treatment, and ii) these patients have been treated with MMF for various lengths of time (6 weeks up to 8 years), therefore the longer-treated patients (patients R012 and R017) had more time to experience and report any sign of side-effects than the shorter-treated patients (patients R009, R011 and R016).

MMF cohort Patient number	**Double major homozygosis**	**Type of experienced side-effects**	**Space out of daily MMF dosage?** (compared to 1g once a day)	**Date of MMF first dose**
R007	NO (double heterozygous)	Low WBC	YES	March 2018
R008	YES	Diarrhoea	NO	April 2018
R009	YES	None	NO	April 2018
R010	YES	Diarrhoea and low WBC	NO	May 2018
R011	YES	None	NO	March 2018
R012	YES	Diarrhoea	YES	2010
R013	YES	Diarrhoea and low WBC	YES	March 2017
R014	NO (heterozygous for rs8192925)	NC	YES	April 2018
R015	YES	Low WBC	YES	September 2017
R016	NO (heterozygous for rs11075646)	None	YES	April 2018
R017	NO (heterozygous for rs11075646)	Diarrhoea and low WBC	YES	July 2013
R018	NO (heterozygous for rs11075646)	Anaemia	YES	October 2016

Table 5.18

Preliminary clinical data about the side-effects experienced by the 12 patients in the Mycophenolate MoFetil (MMF) cohort. Reported side-effects include diarrhoea, low White Blood Count (WBC) and/or anaemia. They are reported along with the date of their introduction to MMF, and their genotype for rs11075646 and rs8192925.

Overall, this very preliminary data investigating the possible association of genotypes for rs11075646 and rs8192925 to the degree of response of patients treated with rifampicin or MMF was not able to provide any evidence yet to support or oppose this potential association - at the present time, but the study is ongoing and may provide answers in the future.

5.4. Discussion

Polymorphisms in xenobiotic metabolism genes have been shown to have a clinical impact on treatment efficacy and adverse effects experience in patients[91,101,106,113-115]. To elucidate if this is the case of CES2 as well, responsible for the metabolism of clinically relevant drugs such as rifampicin and MMF[33,129,136,137,140,233], *CES2* genetic variants were selected and allelic imbalance assays were performed, followed by the design of a clinical study. Amongst 18 “common” *CES2* SNPs, 3 were predicted, using *in silico* tools to assess the expression patterns of factors potentially binding to *CES2* regulatory elements, to have the potential to affect the transcription or stability of *CES2* mRNA: rs11075646, rs8192925 and rs28382828. Using allelic imbalance assays, only the predicted potential effects of rs11075646 and rs8192925 were demonstrated, associated with a significant 3-fold increase (for 39 individuals heterozygous for rs11075646) and a significant 1.13-fold decrease (for 19 individuals heterozygous for rs11075646) of *CES2* mRNA levels. An on-going clinical study was designed to investigate a possible association between patients genotype for rs11075646 and rs8192925 and their treatment response (in efficiency and experience of side-effects) to their CES2-metabolised treatment of rifampicin or MMF. The preliminary data was shared and discussed.

Performing *in silico* predictions regarding the impact of genetic variants on gene expression or enzyme activity is common practice[280,327]. Yet, predictions are not evidence, they need to be confirmed. An allelic imbalance assay in individuals heterozygous for one genetic variant is thought to be a powerful tool to distinguish differences in gene expression – within each individual – that can be attributed to *cis*-

regulatory elements[179,180,333,335,336,351,352]. As a functional effect from less substantial allelic imbalances was previously reported[350], the considerable effect sizes found here suggest that polymorphisms could play an important role in differences in drug efficacy in-between patients. However, the allelic imbalance approach is only robust when there is no genetic imprinting affecting the gene expression[352]. This body of work did not study the *CES2* gene imprinting and, to our knowledge, no data is available on the subject. To rule it out, a luciferase reporter assay was performed (data not shown), to express in hepatocytes this reporter enzyme under *CES2* promoter or 3'-UTR, carrying the major or the minor variant for rs11075646 (for the promoter plasmid) or rs8192925 (for the 3'-UTR plasmid). After a few months of troubleshooting the first step of this cloning experiment (using site-directed mutagenesis), both DNA constructs – received independently from respiratory collaborators and a renowned company – were revealed to be flawed (data not shown). Given the budget and time constraints of this thesis, this cloning experiment was aborted, and another approach was chosen: designing the clinical study presented in this chapter. The very limited sample size of patients included in the study at the time of thesis submission (and, most importantly, the current lack of information about the rifampicin-treated patients responding state) did not allow for a proper analysis of the potential association between their genotype for rs11075646 and rs8192925 and their response to treatment.

This study focuses 3 "common" *CES2* SNPs, their commonality being important here to be able to observe, if there is any, an impact on gene expression or enzyme activity which has the potential both to be relevant to a large proportion of the world population, and to be found in the limited sample size of our clinical study. Yet, it was reported that less common genetic variants could have a considerably larger effect on gene transcription or stability[353]. The selected SNPs were all exonic, although reports support the functional impact of intronic ones[166]. Moreover, even if rs11075646 and/or rs8192925 were the only *CES2* genetic variants capable of impacting CES2 enzyme, there is no data about the potential effect of genotypes harbouring minor variants for both rs11075646 and rs8192925. Due to linkage disequilibrium, the allelic imbalance cohort did include individuals being double heterozygous for the variants of interest, rendering the findings less clear and less indisputable – the investigation could not be limited to an allelic imbalance observation. An unsuccessful cloning experiment and a

pilot clinical study were therefore performed. Several other weaknesses could be attributed to the clinical part of this body of work. As this is a pilot study, the observation of the following limitations was beneficial to improve the future of this ongoing clinical study. Initially, it was observed that participant rate data should be retrieved, in case specific genotype would impact on patients' certainty and compliance to join the study. Then, despite the very preliminary data available at the time of submitting this thesis, two limitations were highlighted and considered of primary importance. Firstly, given that the responder groups for MMF patients focused on the side-effects they experience, additional samples may be required to normalise their side-effects depending on their treatment length (the longer they were under MMF, the more likely adverse effects can develop and be reported). Secondly, and most importantly, this pilot study shows that most of the patients in both cohorts are heavily medicated (due to their tuberculosis or kidney transplant, but also to concomitant diseases), this may impact on their xenobiotic metabolism[108]. This is particularly important for the MMF cohort as their experienced side-effects could be mistakenly attributed to MMF instead of their other treatments (for example adoport and/or prednisolone which are prescribed to most MMF-treated patients). This could bias any potential correlation (or lack thereof), advocating for the use of a bigger sample size. As *CES2* has been particularly overlooked in studies[108,146,339], the influence of their concomitant treatments on this enzyme - and therefore on the metabolism of both rifampicin and MMF - was not yet established and would need to be clarified in order to analyse the results of this pilot study properly. This is particularly important because of contradictory studies on the possible impact of *CES2* polymorphisms on the enzyme activity or drug concentrations[138,144,146,149,233].

The CES2 enzyme has been particularly overlooked over the last decades, its importance in the xenobiotic metabolism is poorly characterized[139,146,147,321,339]. Polymorphism in the *CES2* gene was rarely studied, and mainly focused on rare genetic variants[138]. In a recent article[142], the studied SNPs were only selected given their presence in tuberculosis patients from South Asia, while the *CES2* SNPs have been shown to be differentially distributed around the world[147]. rs11075646 and rs28382828 were not mentioned, and the potential effect of rs8192925 was just observed through differences in drug concentration in the organisms of the patients – discarding the potential influence of other *CES2* genetic variants, *trans*-regulatory

elements and environmental factors. The preliminary data of the currently on-going clinical study being not analysable at the time of submitting this thesis, our study show only indirectly that rs11075646 and rs8192925 are associated to differences in the transcription or stability of *CES2* mRNA. Yet differences at the mRNA levels have been correlated to differences at the protein levels for this enzyme[325,326], therefore such differences could be associated potentially to differences in rifampicin or MMF metabolism. If the presence of the minor variant (homozygous-ly or not) for rs11075646 and/or rs8192925 has indeed an effect on the CES2 enzyme concentration in patient organism, this could impact the circulating concentration of their CES2-metabolised drugs and therefore their response to treatment. In the case of rifampicin and MMF, the current prescribed doses only consider body mass[82,83,152], discarding the influence of genetic polymorphisms and their differential distribution between ethnic groups. Thus, the presence of the minor allele for rs11075646 could induce a decrease in drug concentrations and therefore a poor treatment efficiency in some patients, while the presence of the minor allele for rs8192925 could imply an increase in drug concentrations, associated to severe side-effects (hepatotoxicity[129,151,152,324,343]). Focussing on changes regarding rifampicin concentration, both scenarii could affect the drug-drug interactions associated to rifampicin[45], could weaken a potential synergistic interaction to vancomycin towards MRSA infections - triggering *de-novo* antibiotic resistance[27].

As previously mentioned, allelic imbalance assay is a powerful tool to study the potential effect of one particular genetic variant on the gene expression. Assuming both i) that there is no gene imprinting for *CES2*, ii) that rs11075646 and rs8192925 were validated as affecting the levels of mRNA, and iii) that there is indeed a strong correlation between *CES2* mRNA and CES2 enzyme, rs1075646 and rs8192925 could have a considerable impact on the metabolism of drugs metabolised by CES2. What would happen to patients carrying minor variants of both SNP is still to be elucidated. In a first step, the clinical study is still on-going, and its dataset will be finalised and analysed in due time. Given the magnitude of the allelic imbalances, the frequency of these SNPs and the Hardy-Weinburg equilibrium[349] , it is expected that the sample size of 20 patients per treatment cohort could allow the discrimination – if any - of two groups of responders for both rifampicin and MMF, permitting the

observation of a potential association between the patient's genotype for rs11075646 and rs8192925 and their treatment response. Then, future work would be to analyse "*in vitro*" – as planned with the aborted cloning experiment – the impact of rs11075646 on *CES2* transactivation, and of rs8191925 on *CES2* silencing. Using high-performance liquid chromatography, it could also be possible to correlate directly the presence of the minor variant of each SNP (separately or not) to the levels of rifampicin and its metabolites in patients. Similarly, a systematic dosage of MMF absorbance in blood could be undertaken on the MMF-treated patients, this procedure being rarely performed at the RD&E hospital (personal communication). The observed clinical study limitations were considered, and changes have already been made – for example, patients with critically treatment inefficacy were included. Overall, this could help evidence if *CES2* genetic variants impact on drug metabolism. If this is the case, tailoring the dosage of administered drugs to the patient genetic background could help optimise treatment efficacy, and - in the case of antibiotherapy - it could help fight against antibiotic resistance, which is predicted to cause 10 million deaths per year by 2050[181].

In summary, this thesis chapter brings an insight on how the polymorphism in xenobiotic metabolism genes could alter the metabolism of clinically relevant drugs. We selected 3 *CES2* genetic variants which could have an impact on rifampicin or MMF treatment experience and efficacy, through differences in its metabolism. *In silico*, rs11075646, rs8192925 and rs28382828 were predicted to affect such metabolism, and both the minor alleles for rs11075646 and rs8192925 were significantly associated to differences in *CES2* mRNA levels. A pilot clinical study was designed to associate a phenotypical effect on patients, if there is any, to these 3 SNPs. Preliminary data is shared and discussed. In case our hypothesis is confirmed, systematic genotyping could be performed before prescription, to tailor patients' regimen to their genetic profile – as part of personalised medicine[90,106,120]. This would be beneficial to the patients: they would experience less adverse reactions; their treatment would be more efficient. In the case of antibiotherapies, this would help assessing our prescriptions, making sure we use the right antibiotic, at the right dose and for the right length of time, to reduce the probability of antibiotic resistance emergence[27].

5.5. Reflection on the research

This study acknowledged the potential metabolism-altering nature of 2 *CES2* SNPs: rs11075646 and rs8192925. Both their minor alleles were significantly associated to differences in *CES2* mRNA levels and a pilot clinical study was designed to associate a potential phenotypical effect on patients, if there is any, to these SNPs. Unfortunately, due to time constraints, only preliminary data from the clinical study were retrieved and discussed. Given more time, we could have perhaps reached a higher number of participants to our study, and obtained all important baseline data about them, allowing us to conclude on the possible association between their genotype and their response to treatment. It would have been also a good opportunity to find individuals carrying the minor allele for both SNPs, permitting their simultaneous study.

In the event more time was allocated to this project, another time-consuming approach would have been to study all 1,977 genetic variants accounted for on *CES2.* We did select *in silico* the most common ones, yet rare SNPs could also potentially impact greatly individuals carrying their minor alleles homozygous-ly.

CHAPTER 6 – Discussion and future directions

In this thesis, three novel studies were presented to investigate both the nature of the interaction of two clinically relevant antibiotics (vancomycin and rifampicin) and the potential impact xenobiotic metabolism polymorphisms could have on antibiotic bioavailability and interactions. Each empirical chapter (Chapters 3-5) was aimed to answer a distinct research question in this field and was therefore presented with its own context and conclusions. Throughout this final chapter, the key findings of this body of work are briefly summarized and considered in their framework, discussing their implications, strengths and limitations. Finally, the future work needed to improve our understanding in this field is highlighted, and perspectives on the importance of expanding our knowledge on this topic are shared.

6.1. Summary of the key findings

In light of the current literature on both the global spread of antibiotic resistance (particularly regarding Methicillin-Resistant *Staphylococcus aureus* – MRSA) and the clinical impact of genetic variations on xenobiotic genes, several unanswered questions were acknowledged and studied.

The combination of vancomycin and rifampicin to treat MRSA infections, while avoiding the emergence of resistance, has shown critical discrepancies about its effectiveness, *in vitro*, *in vivo* and in clinical studies[55,71-81]. Some hypotheses have been drawn to explain the heterogeneity in the outcomes, such as the influence of differences in strains, in the nature of the infection, in the experimental techniques or between patients baseline parameters (age, immunity, xenobiotic metabolism genetics, etc)[33,45]. To elucidate the influence of such variables on laboratory practices, as reported in Chapter 3, one operator studied *in vitro,* in the same laboratory and using the same stock of supplies, the interaction of vancomycin and rifampicin towards

MRSA in various experimental settings. This study demonstrated that the type of interaction of these two antibiotics was particularly unclear: it seemed to depend on the experimental settings, and no synergy was accounted for. Planktonic and biofilm experiments showed significant antagonism (Chapter 3, pages 78-89), while two structured environments using soft or solid agar failed to deliver a clear answer (Chapter 3, pages 90-100). The usage of Etests was also questioned as this clinically-used antibiotic susceptibility test concluded in an indifference between vancomycin and rifampicin, but also between a combination of two strips of the same antibiotic (Chapter 3, pages 100-107). Overall, no sign of synergy was observed between vancomycin and rifampicin towards MRSA. The unstable nature of vancomycin and rifampicin Minimum Inhibitory Concentrations (MICs) were acknowledged.

Differences in the outcomes of clinical studies about the combination of vancomycin and rifampicin towards MRSA infections could also be potentially attributed to inter-individual differences[91,92,95-98]. Indeed, literature reported that genetic variations (such as Single Nucleotide Polymorphisms - SNPs) on xenobiotic metabolism genes can possibly have an impact on enzyme activity[98,110], affecting therefore drug concentrations and treatment effectiveness. Rifampicin is particularly interesting here because it is both metabolised by the Carboxylesterase 2 (CES2) enzyme[136,138] and inducing its activity[130,132,137,139]. The *CES2* gene being subjected to common polymorphisms, *Chapter 5* focuses on the potential impact of 3 “common” SNPs: rs11075646, rs8192925 and rs28382828. With an allelic imbalance assay, it was demonstrated that the minor variant for rs11075646 was associated to a significant 3-fold increase in the CES2 messenger RNA (mRNA) levels (Chapter 5, pages 189-190), while the minor variant for rs8192925 was associated to a significant 1.13-fold decrease in the mRNA levels (Chapter 5, pages 191-192). No association was found between the minor variant for rs28382828 and mRNA levels (Chapter 5, pages 187-189). This gave rise to a clinical study focusing on rs11075646 and rs8192925 and their potential phenotypical impact. Two cohorts were recruited: tuberculosis patients treated with rifampicin, and kidney-transplant patients submitted to Mycophenolate MoFetil (MMF) - a CES2-metabolised pro-immuno-suppressor[149,150] (Chapter 5, pages 193-196). This clinical study aims to observe an association between, on one hand, the clinical phenotype of these patients (differences in treatment response and in the experience of side-effects) and their genotype for rs11075646 and rs8192925. At the

time of submitting this thesis, most of the important baseline data about these patients was lacking, but extremely preliminary results could indicate a lack of association between those genetic polymorphisms and side-effects experienced by MMF-treated patients (Chapter 5, pages 197-205).

More generally, it seems crucial to develop an extensive knowledge about both i) which enzymes are metabolising the clinically relevant drugs, and ii) which SNPs can potentially affect their gene expression or enzyme activity, especially because of the differential distribution of their variants in different ethnic groups. An *in silico* study was therefore performed in *Chapter 4* to find the most "common" SNPs on the gene coding for CYP3A4, a xenobiotic enzyme of critical clinical importance[106], and to investigate their potential effect on the *CYP3A4* gene expression or enzyme activity. rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 were acknowledged (Chapter 4, pages 125-136), while the open access literature only reported the potential impact of rs2242480. The 8 SNPs of interest were classified according to their potential impact score (Chapter 4, pages 136-139) and then their distribution in the world was considered (Chapter 4, pages 139-146). Given the importance of the xenobiotic metabolism by CYP3A4, the high frequencies of those SNPs and their differential distribution in the world, it would be vital to take these SNPs into account before prescribing any of the numerous CYP3A4-metabolised drugs.

6.2. General discussion of the empirical chapters and future perspectives

This thesis aimed to provide both clarity about the nature of the interaction of vancomycin and rifampicin towards MRSA, and insights on the potential impact xenobiotic metabolism genetic variations could have on clinically-relevant antibiotic bioavailability and their interactions.

6.2.1. No sign of synergy between vancomycin and rifampicin towards Methicillin-Resistant Staphylococcus aureus

With the same operator performing experiments in the same laboratory using the same materials, most of the variabilities between reported *in vitro* testing of the combination of vancomycin and rifampicin towards MRSA were eluded in *Chapter 3*. This is a significant improvement compared to the numerous articles from the literature which focused on one technique and compared directly their results to those provided by another one[71,79]. Here, one could observe that differences in experimental settings (planktonic, biofilm, cultures on and in agar, Etests) are associated to differences in the reported nature of the interaction of vancomycin and rifampicin. Depending on the type of experiments, the interaction of vancomycin and rifampicin towards MRSA is considered antagonist, indifferent or even undetermined. There was no sign of the synergy required to propose an efficient treatment, despite its past reports[71,73,75,77,79,81]. This conclusion is supported by studies demonstrating the influence of *in vitro* experimental conditions in their experimental outcomes[195,196]. With clinical settings already not following the same practices in term of antibiotic susceptibility testing and treatment selection (for example, there are no guidelines on the appropriate rifampicin dose, time of initiation and length of treatment[45]), and with vancomycin and rifampicin MICs proven particularly unstable, this could also shed light on why the conclusions from clinical studies were inconsistent.

The limitations of this study are mostly related to the techniques. Several experiments were proven challenging to be performed, particularly because of extreme variability in the MICs of monotherapies in planktonic culture, but also because of difficulties in handling the Etests in combination (the validity of this technique was even questioned in Chapter 3, pages 100-107). Interestingly, most of the handling issues reported here could also be impacting operators in clinical diagnostic laboratories – another possible explanation for the published discrepancies. Improvements could be proposed, regarding the addition of an agar plating step before overnight culture, or the addition of a final day of planktonic culture in antibiotic-free media. Finally, the clinically relevant MRSA252 strain was used in this work and, yet given that inconsistencies in-between experimental settings have been attributed to differences in strains in the past, another MRSA strain could have been tested. The role of persister or tolerant bacterial cells was not studied here[354], it could be elucidated using single-cell techniques for

example[242-244]. A first step in future work should therefore consider replicating these experiments with another MRSA strain, and perhaps to study sequential or cycling combinations instead of mixed combinations[155,156]. This could be done following a strict scheduling to avoid changes in circadian rhythm, as this was reported to possibly impact on laboratory experiment outcomes[366].

This finding corroborates reports stating that the combination of vancomycin and rifampicin may not be the best therapy to treat any type of MRSA infections[55]. Most of the recommendations involving vancomycin and rifampicin were made because of a lack of good alternatives (the other antibiotics not meeting the expected outcomes in some particular MRSA infections)[82,83], an alternative and effective treatment needs therefore to be developed. Recent reports deemed sequential therapy of particular interest to avoid the emergence of resistance in bacterial infections[155], yet this therapy is not recommended with non-synergistic antibiotics. Especially given antagonistic combinations may require higher drug dosages[49,52,54] and both high doses of vancomycin and rifampicin are deemed toxic for the patients[45,53,67]. New antibiotics discovery[6,182] or the development of MRSA vaccines[355,356] should have better chances of success to fight against MRSA infections.

6.2.2. In silico identification of eight common CYP3A4 variants with the potential to affect xenobiotic metabolism

In *Chapter 4*, 8 common and potentially important SNPs were identified on the *CYP3A4* gene. rs2242480 being already categorized as “important” by the scientific community[232,254,301-311], our findings extend the scope of *CYP3A4* polymorphisms that may be interesting, if not critical for some patients, to consider while prescribing one of the various CYP3A4-metabolised drugs. The potential impact of SNPs affecting the xenobiotic metabolism has been studied for decades[98,110], with researches focussing particularly on *CYP3A4* which code for a major enzyme of this metabolism[101]. It is therefore of significant interest that 7 of the uncovered “common” SNPs in our study (rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620 and rs2687116) have been overlooked in the past.

However, the potential impact of these SNPs has yet to be proven. Some *in silico* predictions have indeed being refuted in the past[373]. Can the presence of the minor variant of one only SNP actually alter the *CYP3A4* gene expression or enzyme activity to the point of altering its target concentration? As CYP3A4 and CYP3A5 are often sharing drug specificity, could an alteration of CYP3A4 concentrations be counter-balanced by the activity of CYP3A5? Despite these questions being left unanswered, these findings remain of significant interest to highlight the need to focus studies on the potential impact of xenobiotic metabolism gene polymorphisms on drug efficiency. Especially because of the differential distribution of those variants in different ethnic groups.

It will be crucial to prove *in vitro*, *in vivo* and with clinical studies if there is a correlation between the minor variant of all those 8 identified SNPs and changes in *CYP3A4* gene expression or enzyme activity. If there is a correlation, investigations should focus on the potential impact rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 can have, alone or together, on drug concentrations and clinical outcomes. 30% of the clinically relevant drug metabolism being attributed to CYP3A4/5, the potential impact of such polymorphisms could be critical for various types of diseases - especially given differential distribution of those variants in different ethnic groups.

6.2.3. CES2 variants rs11074656 and rs8192925 could impact on patients' xenobiotic metabolism, treatment efficacy and adverse effects

In the past, the metabolism of rifampicin has been wrongfully attributed to the cytochrome P450 enzymes[33,45], instead of the CES2 enzyme[138]. In the first part of *Chapter 5*, 3 SNPs on the *CES*2 gene (rs11075646, rs8192925 and rs28382828) were investigated and the findings provide first evidence of an association between the presence of the minor variants for either rs11075646 or rs8192925 and changes at the CES2 mRNA levels. It is unsurprising that the overlooked *CES2* gene presents potentially important polymorphisms, as it is the case for most Phases I and II xenobiotic metabolism genes[101,117,124,319,357]. However, the size of the observed effects on the mRNA levels was unprecedented: using an allelic imbalance assay, a

significant 3-fold increase was associated with the presence of the minor variant for rs11075646. As this polymorphism is particularly common, it could have a critical phenotypical impact. Indeed, if *CES2* mRNA levels are proportional to the CES2 activity (as it has been reported recently[325]), patients carrying the minor variant for rs11075646, homozygous-ly or not, could potentially be quick metabolisers, who eliminate many clinically relevant drugs readily. These patients would need a higher dosage to improve treatment efficacy, or a different medicine when the prescribed drugs are considered toxic at high doses. The significant 1.13-fold decrease of *CES2* mRNA levels associated to the presence of the minor variant for rs8192925 could also be noteworthy, as similar effect size have been reported to have a phenotypical impact[350]. This would make the carrier of the minor variant for this SNP slow metabolisers, who are prone to develop more side-effects due to the toxicity of the drugs. In the case of rifampicin, both scenarii could lead to the emergence of antibiotic resistance[27], impacting the patient recovery and the spread of resistances in the community. Given the differential distribution of those 2 variants in different ethnic groups, quick and slow metabolisers could be found at different frequencies in different populations.

If allelic imbalance experiments are now particularly recognized, the findings from *Chapter 5* need to be validated. An association between genotype and mRNA levels is not necessarily followed by an association between genotype and enzyme activity, let alone a causality between genotype and drug concentrations, or genotype and clinical outcome. Other genetic or transcriptomic processes can happen and impact on the effect of an observed alteration of the mRNA levels. Moreover, the fact that both rs11075646 and rs8192925 are showing divergent associations to *CES2* mRNA levels will complicate the interpretations for patients carrying the minor variants for both SNPs, homozygous-ly or not. Future work includes the cloning experiments which were attempted before, with no success (data not shown). Indeed, it seemed important to validate the effect of rs11075646 and rs8192925 on the transactivation of the *CES2* gene by their disruption of the binding sites of transcription factors (for rs11075646) and microRNAs (for rs8192925). Cloning experiments were performed with the aim to express, with a luciferase reporter gene, the minor variant for each SNP in hepatocytes. These experiments were not successful due to flawed genetic plasmids, from both our collaborators and a registered company. Apart from this (still required)

validation, it would be interesting to investigate the other SNPs carried by the *CES2* gene. Then it would be possible to analyse the direct relationship between on, one hand, the *CES2* genotype and, on the other hand, the drug metabolism, through drug concentrations. Even if only rs11075646 and rs8192925 have a possible impact on the *CES2* gene expression or enzyme activity, nothing is known yet about the potential xenobiotic metabolism of patients carrying the minor variants for both SNPs – this would need to be elucidated. Another idea for the future, currently in development, would be to look at the association between the host genotype and bacterial survival outcome, with a co-culture of MRSA with hepatocytes in a media supplemented with antibiotics. Amongst the numerous drug metabolized by CES2, the case of rifampicin is of particular importance and would need to be closely investigated as this anti-tuberculous agent is also inducing the CES2 metabolism[139].

In the second part of *Chapter 5*, the potential effect of rs11075646 and rs8192925 on the *CES2* gene expression or enzyme activity was investigated indirectly, by developing a pilot clinical study in two cohorts of patients being treated with CES2-metabolised drugs (rifampicin and MMF). Only the preliminary samples and data were available at the time of thesis submission, limiting the reported findings in this chapter. It was notably pointed out that the 12 MMF-treated patients seemed to experience side-effects independently of their genotype. This would need to be confirmed with a bigger sample size, in Hardy-Weinburg equilibrium[349] (20 samples minimum), and a correction for the side-effects associated to their other medications. Another complication was the selection of patients: when they are deemed unresponsive to their treatment, they are readily redirected to another one, and could therefore not be selected in this study. Consequently, the cohorts were expected here to be of mildly-impacting genotypes, if any. Importantly, by looking at the experience of side-effects in MMF patients, the factor “time” should be considered in the analysis, as the emergence of side-effects in patients can be missed in patients treated with MMF for only 6 weeks, compared to patients treated with it for 6 years. Another crucial limitation to our study, while observing liver function test results for rifampicin patients and side-effects in MMF patients, was that these patients are heavily medicated, their other drugs either being potentially metabolised by CES2 as well, or being able to induce the CES2 metabolism, or being able to affect liver function test results and/or the experience of side-effects in patients. Finally, as developed in the *Thesis Introduction*,

other inter-individual differences could impact on drug concentrations and efficacy, such as the age, the sex, the body mass index, the immune system or concomitant diseases[82,92,95-98]. All these personal baseline data was differing between the recruited patients.

Despite these limitations, it is important to note that clinicians were really interested to participate to this study, because they face patients being unresponsive with no proper cause (personal communication). Knowing if the xenobiotic metabolism of those patients is causing their lack of or inadequate response would allow clinicians to treat more efficiently their patients, whether the genotype for *CES2* rs11075646 and rs8192925 is implicated or not. Future work would be therefore to expand this study to a bigger cohort, including patients having failed to respond to their treatment, which have been redirected to another regimen. By increasing the sample size, it would be easier to observe associations, if any, between the emergences of side-effects and the treatment length, as well as to distinguish the potential effect on side-effects and liver function tests between those due to the tested drugs and those due to the concomitant treatments.

CES2 may be less well-known than the cytochromes p450, but this enzyme is involved in the metabolism of clinically-relevant drugs of all types: the potential impact of such polymorphisms could be life-saving.

6.3. Implications and conclusions on the key findings

In summary, this thesis has provided insights on why the combination of vancomycin and rifampicin may not be the best therapy to treat any type of MRSA infections. It has also significantly advanced our current knowledge on how the xenobiotic metabolism gene polymorphisms could have a potential impact on drug concentrations and clinical outcomes. It has highlighted the underestimated role of CES2 and identified in total 10 common metabolism gene SNPs having such a potentiality: rs11075646 and rs8192925 on the *CES2* gene, and rs28988603, rs28988604, rs28969391, rs28371763, rs28988606, rs12721620, rs2242480 and rs2687116 on the *CYP3A4*

gene. These SNPs are differentially distributed in different ethnic groups, their potential impact on drug metabolism could therefore be found at different frequencies in different descent populations.

If future *in vitro* or clinical studies demonstrates, once and for all, that the combination of vancomycin and rifampicin is not beneficial to fight against MRSA infections, or if they show their type of interaction is still undetermined or variable depending on experiment settings, a new regimen will have to be recommended in practices – in order to treat the patients more effectively and decrease the selective pressure on the bacteria. Currently, alternatives can be prescribed for most types of MRSA infections (guidelines mention notably monotherapies of vancomycin, linezolid, daptomycin or tigecycline[82,83]), clear recommendations would need to be published to help the clinicians in their choice of prescriptions to avoid allergic reactions and unsuitability for children in some cases[358]. Vaccination could even be considered, if the current research on MRSA vaccines succeed. This would help the infected patients, but also their clinicians (who are following contradictory guidelines), other hospitalised patients (as MRSA is a common nosocomial bacterium) and the community (because of the cost of MRSA infections, amongst other reasons).

On a bigger scale, the observed disparities in the interaction between two antibiotics depending on the experimental settings could be applied to other bacterial infections. This would be something to keep in mind while studying, for example, the combination of beta-lactam and/or carbapenem and/or aminoglycoside and/or ciprofloxacin against *Pseudomonas aeruginosa* infections[359]. The same type of selective pressure could be applied on this clinically-relevant bacteria.

Both the studied *CYP3A4* and *CES2* genes coding for enzymes responsible for the metabolism of numerous clinically relevant drugs, future work is needed to elucidate

the actual impact of their common SNPs. More generally, it is required to expand our knowledge on which enzymes are metabolising the clinically relevant drugs, and which genetic variants can affect these enzyme concentrations or their activities. By clarifying this, it would then be possible to systematically qualify the nature of each drug-drug interaction, whether these drugs are used in combination to treat the same disease, or their concomitant use is purely fortuitous. This would allow the development of prescriptions that are tailored to the patients, which will doubtfully benefit them – they will be treated more efficiently while avoiding side-effects - and also the community. The final aim would be to develop the systematic genotyping of the patients for key SNPs in order to tailor their dosages to their genetic background - another step towards personalized medicine. This would have clinical implications for the patients, who will be treated more efficiently and with less experiences of side-effects. On a community perspective, a personalised prescription of antibiotics could lower the selective pressure on bacteria and help fighting against antibiotic resistance. This would benefit the health all the potential hospital in-patients, and the out-patients who are currently avoiding visiting a hospital because of those nosocomial infections. A more economical point of view would be that these advances could allow a decline of the cost of antibiotic resistance (less drug waste, shorter hospitalisations, and better productivity of the professionally active patients)[360,361]. Obviously, a strict ethical regulation would need to be designed in order to genotype the patients for thousands of SNPs before they are prescribed a medicine. This could even have political implications if we ask pharmaceutical companies to provide data on how exactly their drugs are metabolized by patients from various genetic backgrounds, during their research and development phases.

The evidences presented in this thesis highlight the need to make sure the right drugs and the right dosages of said drugs are prescribed to patients. Given the global threat of antibiotic resistance, this is necessary to treat infectious diseases efficiently, but this can be also critical for other types of diseases (cancers, mental illnesses …) for which the clinical outcomes depend on the bioavailability of the administrated drugs too.

Altogether, these studies have provided a strong case on how important the study of drug interactions is, and how the xenobiotic metabolism gene polymorphism can act on it. Clearly, much more work is needed in this field, to both provide a suitable treatment for MRSA infections currently treated with vancomycin and rifampicin, and to make sure the scientific knowledge is comprehensive about which enzymes are metabolising the clinically relevant drugs, and which genetic variants can affect these enzyme concentrations or their activities. The findings in this thesis demonstrate the importance of addressing these issues in depth.

APPENDIX A - Chapter 3 - Interaction between vancomycin and rifampicin towards Methicillin-Resistant *Staphylococcus aureus*, in various experimental settings.

A.1. Materials and Methods:

A.1.1. Bacterial culture in a planktonic environment (broth dilutions)

Hill fit:

Assuming the data decreases monotonically with dose, a Hill function was fitted to the data – allowing the estimation of standard errors and standard deviations, with the visualisation of a Hill curve and a 95% confidence interval along it. As published earlier[24], a Hill function takes the following form, with K the IC50 of the data, and H_0 and n two other parameters depending on the data fit:

$$H(A) = H_0 \times \frac{K^n}{K^n + A^n}$$

Using MatLab, Professor Robert Beardmore's best hill fit function (modelfun for non linear regression) was used, and adaptred as necessary:

```
modelfun = @(p,A)(p(1)+p(2)./((p(3)./A).^p(4) + 1));
   beta0=[0.1 2 3 4];
   mdl = NonLinearModel.fit(Dosages, DoseResponse,modelfun,beta0)
   R2=mdl.Rsquared.Ordinary;
   beta=mdl.Coefficients.Estimate;
   [beta,resid,J,Sigma] = nlinfit(Dosages, DoseResponse,modelfun,beta);
   xFit=linspace(0,1.1*max(Dosages),100);
   [yFit, delta] = nlpredci(modelfun,xFit,beta,resid,'Covar',Sigma);
   area(xFit',yFit+delta,yFit(end)-1.1*delta(end),'FaceColor',[0.9 0.9 0.9],'EdgeColor',[1 1 1]);
   hold on
   area(xFit',yFit-delta,yFit(end)-1.1*delta(end),'FaceColor',[1 1 1],'EdgeColor',[1 1 1]);
   plot(xFit, yFit+delta, '-','Color',[0.5 0.5 0.5]);
   plot(xFit, yFit-delta, '-','Color',[0.5 0.5 0.5]);
   plot(xFit, yFit, 'k-','LineWidth',2);
text(max(xFit),yFit(end)+1.1*delta(end),['R^2=',num2str(R2)],'FontSize',16,'HorizontalAlign','R
ight','VerticalAlign','bottom');%formatting
   Y=myMeans;
   ICs=[.5 .8 .95 .99];%to display the ICs on the graph
   for k=1:length(ICs)
      OD50=(1-ICs(k))*Y(1);
      fromDose=find(yFit>=OD50);
      IC50=interp1(yFit, xFit, OD50,'linear');
      plot([IC50 IC50], [-max(yFit) max(yFit)], ':', 'LineWidth', 1,'Color',[0.3 0.3 0.3])
      if ICs(k)==.7
         s=1.05;
      else
         s=1;
      end
      text(s*IC50, max(yFit), ['IC_{',num2str(100*ICs(k)),'}'], 'FontSize', 14,'Color', [0.3 0.3
0.3],'VerticalAlignment','bottom','HorizontalAlignment','center')
   end
   errorbar(dosages,myMeans,myStds,'.-k','markersize',30,'linewidth',1);
   xlabel('dosage (ug/ml)')
   ylabel('Density at 24h');
   axis([-.02*max(xFit) max(xFit) -1.3*max(delta) 1.1*max(yFit)]);
   title('Dose response of E.coli(Wcl)');
end
function y = ste(v)
   y = std(v)/sqrt(length(v));
end
```

A.1.2. Bacterial culture inside a "square dish" structure

Visual example of the methodology followed to analyse the halos of growth inhibition using ImageJ.

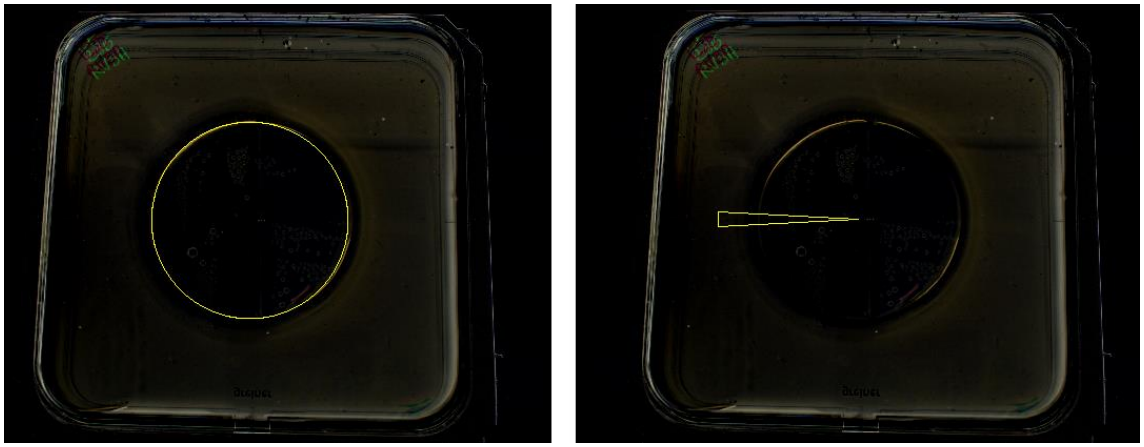


Figure A1

Visual example of the methodology followed on ImageJ to analyse the halos of growth inhibition for the "square dish" experiment, after removal of the background. Left – Localisation of the centre of the small dish, using a circle colocalising with the visible edges. Right – Creation of 3 angles of 3° (here, 1 is shown in vancomycin monotherapy) starting at the previously located centre of the small dish.

A.1.3. Bacterial culture on a "layer dish" structure

Visual example of the methodology followed to analyse the halos of growth inhibition using ImageJ.

Figure A2

Illustration of the image analysis on ImageJ to extract the profile of the bacterial growth from the pole A (up) to the pole B (down) of one streak on a Petri dish.

A.1.4. Plotting and statistics

Methodology for statistical test on biofilm Optical Density (OD) data (interval data), using Stata:

1- Is OD data normal? Shapiro–Wilk test was performed (swilk), p = 0.00060. OD data is normal.
2- Is OD data parametric? One-sample variance-comparison test was performed (sdtest), p = 0.0314 to 0.0025 depending on the conditions tested. OD data is non-parametric (inequal variances).
3- Is OD data ordinal data? Yes.
4- For 2-10 unpaired samples/groups, Mann-Whitney test was performed, per couple of samples (ranksum).

Methodology for statistical test on “square dish” data (interval data), using Stata:

Depending on the dataset:

1- Is data normal? Shapiro–Wilk test was performed (swilk), p = 0.00553. Data is normal.
2- Is data parametric? One-sample variance-comparison test was performed (sdtest), p = 0.0883 to 0.3598, depending on the conditions tested. Data is parametric (equal variances).
3- Is data ordinal data? Yes.
4- Two samples t-test was performed, for each couple of samples (ttest).

Or

1- Is data normal? Shapiro–Wilk test was performed (swilk), p = 0.009378. Data is not normal.
2- For 2-10 unpaired samples (groups of data), Mann-Whitney test was performed, for each couple of samples (ranksum).
3- Is data ordinal? Yes.
4- For the three groups of data studied together, with different distribution (shape of the data), Kruskal-Wallis test for mean rank was performed (kwallis).

A.2. Results

A.2.1. Bacterial culture in a planktonic environment (broth dilutions)

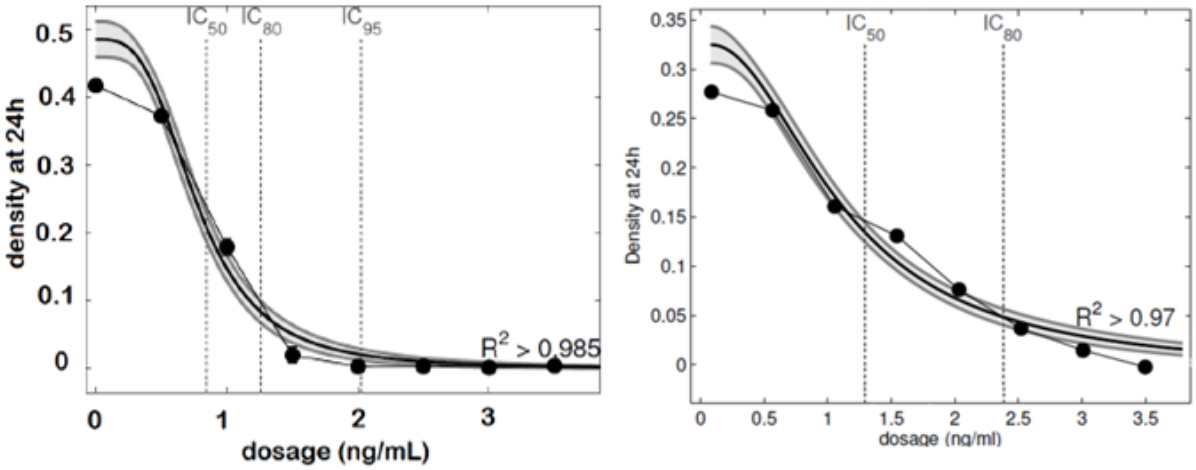


Figure A3

24-hour dose-response of rifampicin on Methicillin-Resistant *Staphylococcus aureus*, on two different days in dilution broth. The concentrations required to reach the Minimum Inhibitory Concentration (MIC) in monotherapy on Methicillin-Resistant *Staphylococcus aureus* (MRSA) were fluctuating from one day to another. Example of two rifampicin monotherapies, one is significantly associated to a MIC of 2ng/mL (left), and one to a MIC of 3ng/mL (right). *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data. A Hill equation is fitted to the data, with 95% confidence intervals around predicted mean, associated to the expected biochemical interaction of the bacteria and the antibiotic, in the grey area along the black curve (superimposed Hill curve). The data follows the Hill curve, with the correlation coefficient R^2 indicated on the graph. The Inhibitory Concentrations 50 (IC50), 80 (IC50) and 95 (IC95) are indicated.*

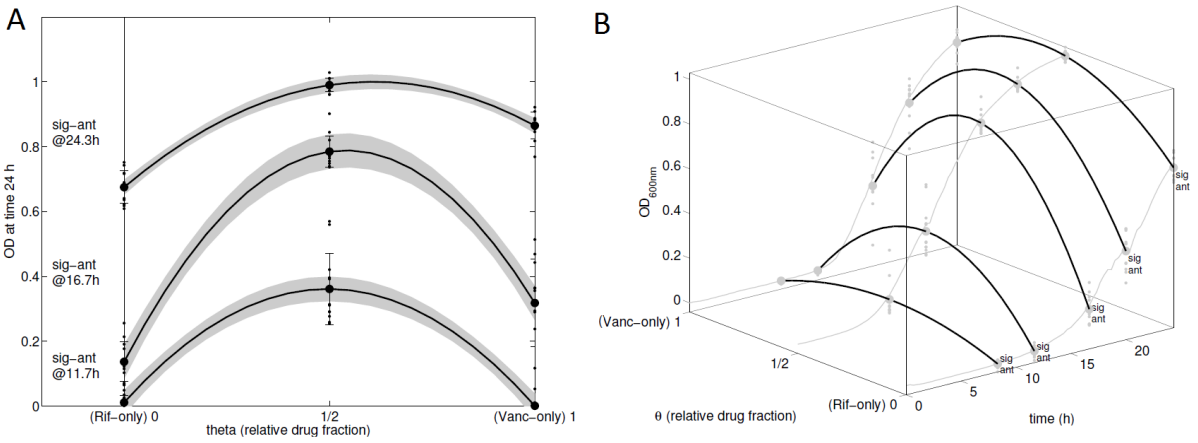


Figure A4

Stable antagonism for the combination of vancomycin and rifampicin on Methicillin-Resistant *Staphylococcus aureus*, in broth dilution. A frown is observed here in 2D (A) and 3D (B) visualisations of the combination therapy as a function of the drug concentration delta factor or relative drug fraction θ and time, on the first day of the experiment from 11.7h to 24.3h. The antagonism of vancomycin and rifampicin is stable and significant (sig-ant), over time on the first day of the experiment, against Methicillin-Resistant *Staphylococcus aureus* (MRSA). *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data. A Hill equation is fitted to the data, with 95% confidence intervals around predicted mean, associated to the expected biochemical interaction of the bacteria and the antibiotic, in the grey area along the black curve (superimposed Hill curve). The data follows the Hill curve, with the correlation coefficient R^2 indicated on the graph. Vanc stands for vancomycin and rif for rifampicin, and θ is expressed in vancomycin ratio in the combination (from 0 to 1).*

A.2.2. Bacterial culture in biofilms

Each antibiotic was tested as monotherapy on the same plate as the combination assay. The following plots (Figures A5-A6) represent their optical density reads at 600nm, before crystal staining. Well homogeneity was not insured before crystal staining; such dataset is therefore less quantitative than qualitative: “can the microplate reader detect a difference in absorbance from the negative control?”. If yes, it was attributed to bacterial growth. If no, there may be no growth, or there may be a

lack of homogeneity inside the well (bacteria growth could be inhibited locally to where the OD is measured). The data being less quantitative, no MIC or IC50 are provided.

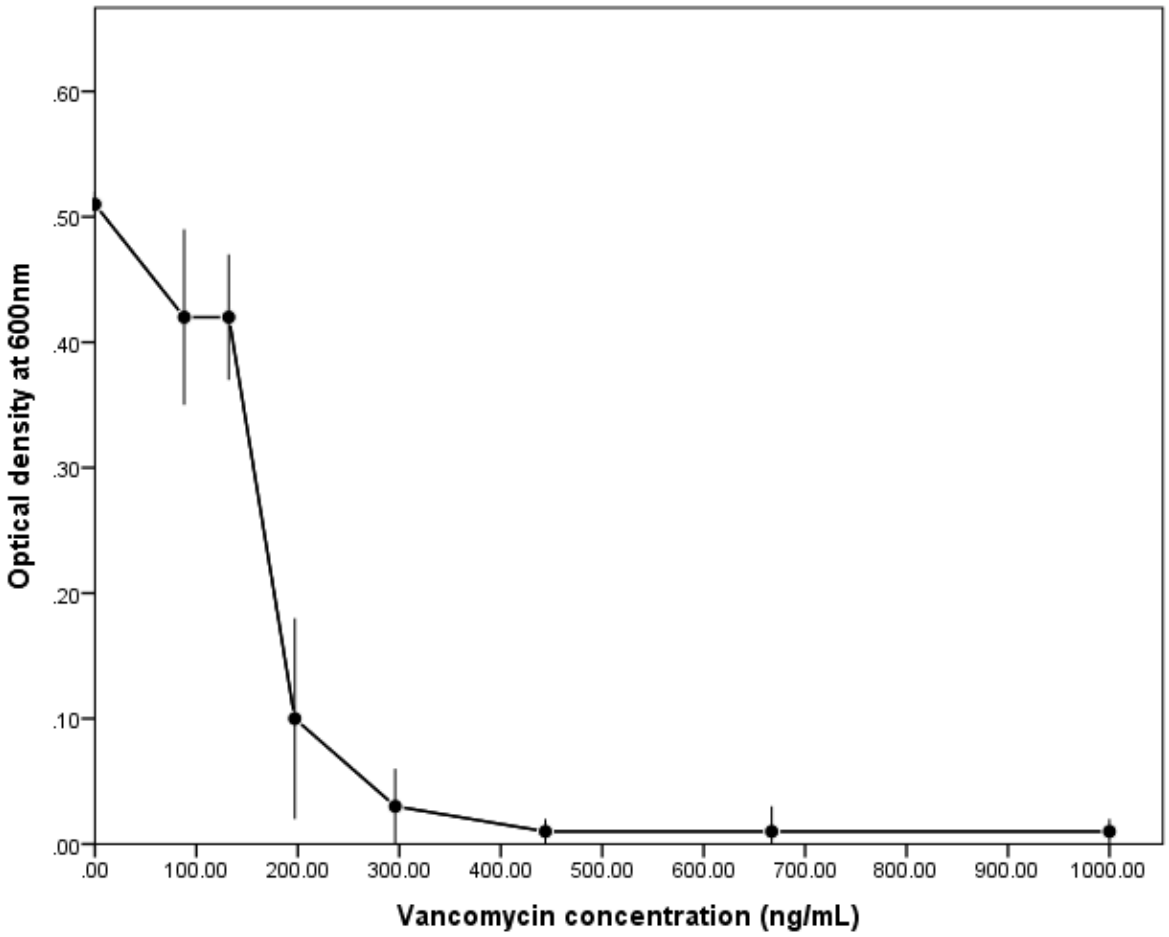


Figure A5

Dose-response of MRSA252 in different concentrations of vancomycin (ng/mL) in a biofilm. Data collected as “optical density” at 600nm before the crystal violet staining. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data (4 replicates).*

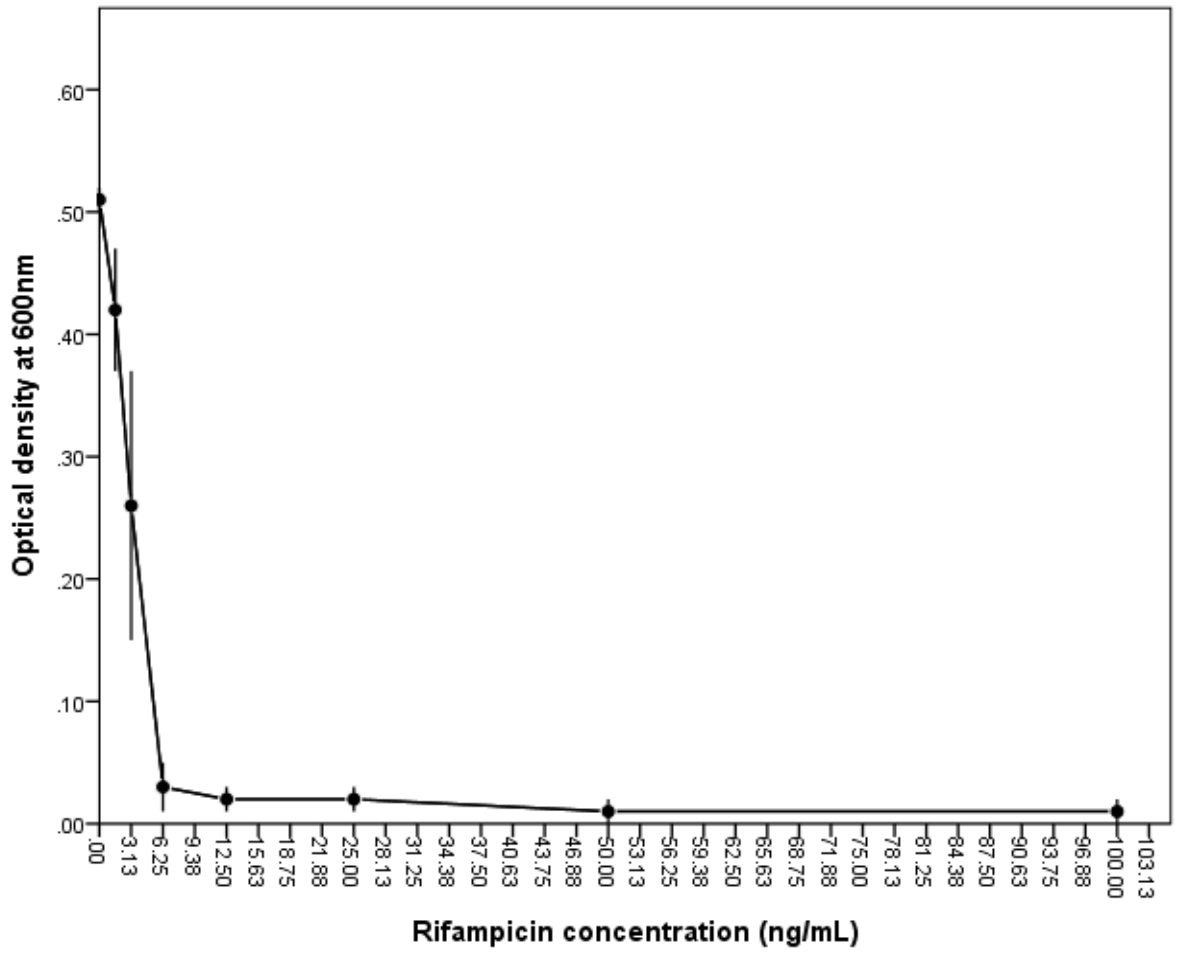


Figure A6

Dose-response of MRSA252 in different concentrations of rifampicin (ng/mL) in a biofilm. Data collected as "optical density" at 600nm before the crystal violet staining. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data (4 replicates).*

Recording the optical density after the crystal violet stain of the biofilm of MRSA252 allowed the quantification of the biofilm formation and removed the errors due to the difference in thickness of biofilm inside the wells of the microplate (by homogenising the disrupted biofilm inside the wells). The following plots (Figures A7-A8) represent the optical density reads at 600nm of both monotherapies, after crystal staining. This is the same experiment than in Figures A5-A6, therefore the concentrations used are the same.

Assuming the third data point in Figures A7-A8 were artefacts (based on Figures A5-A6, this seemed to be the case), MIC in vancomycin was not reached – despite an apparent plateau. MIC for rifampicin was estimated to 6.25ng/mL.

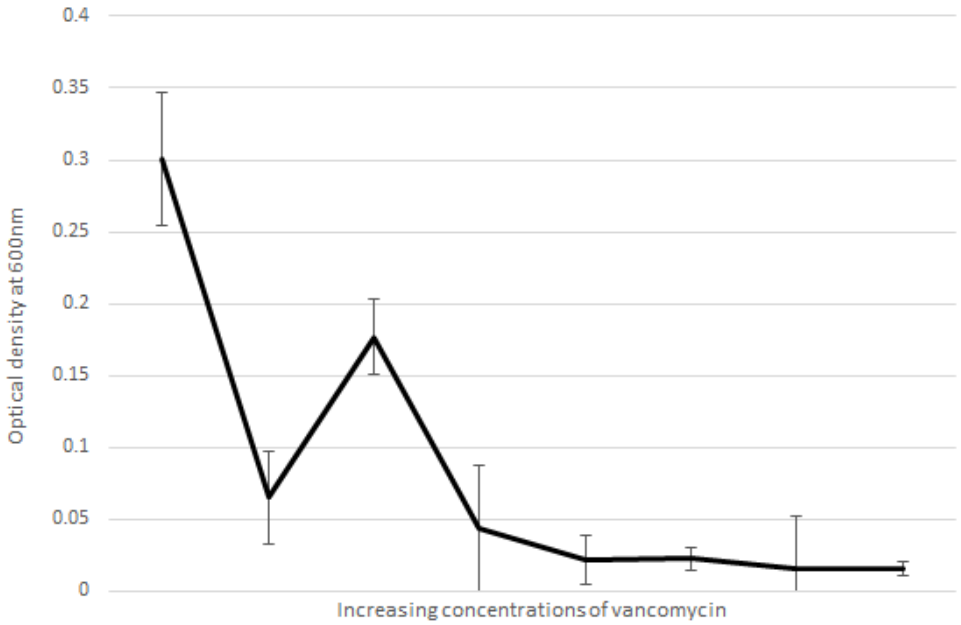


Figure A7

Biofilm formation of MRSA252 in increasing concentrations of vancomycin (ng/mL) in a biofilm. Data collected as "optical density" at 600nm after the crystal violet staining. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data (4 replicates).*

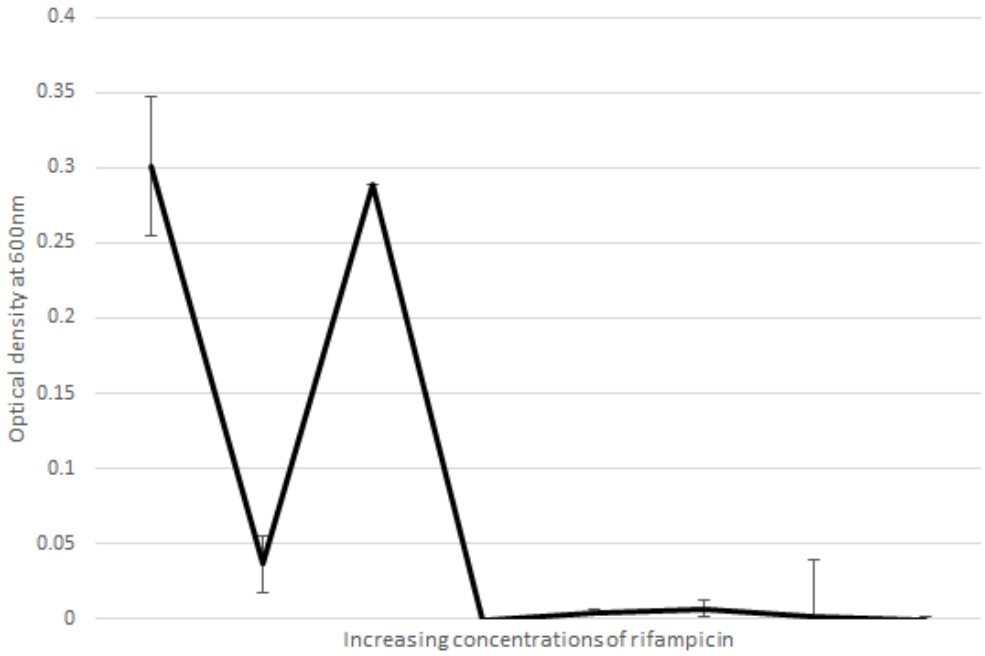


Figure A8

Biofilm formation of MRSA252 in increasing concentrations of rifampicin (ng/mL) in a biofilm. Data collected as “optical density” at 600nm after the crystal violet staining. *The full points represent the average of the observed data, and the black vertical error bars the standard deviation between the observed data (4 replicates).*

A.2.3. Bacterial culture on a "layer dish" structure

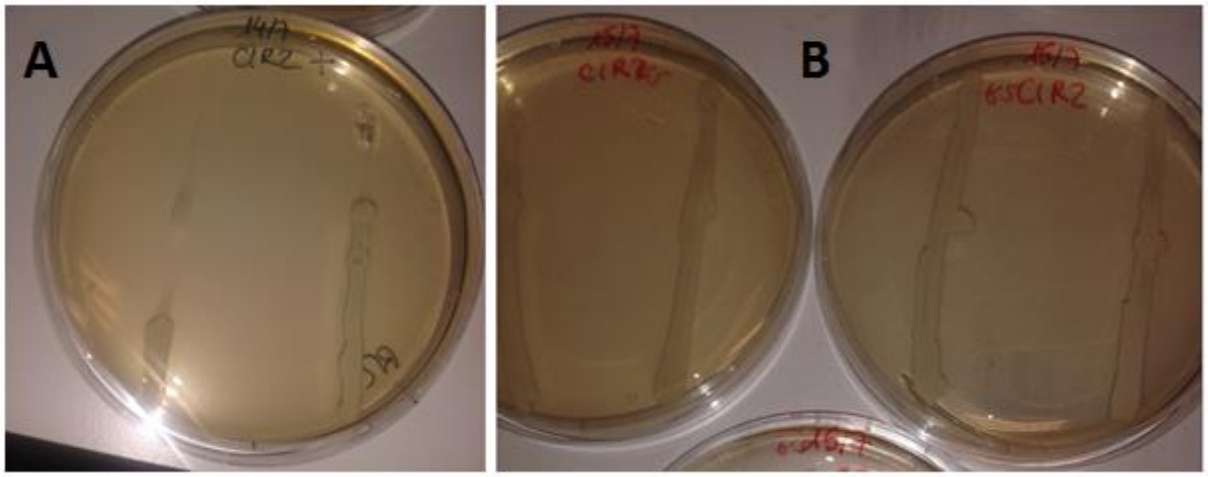


Figure A9

Examples of the discrepancies observed in the bacterial growth pattern on the antibiotic control dishes. On one day (A), the C1R2 control dish was inhibiting partially the growth of the bacteria, with 4.72MIC (Minimum Inhibitory Concentration) of rifampicin. The next day (B), there was no inhibition of the bacterial growth on the similar C1R2 control dish, despite increasing the concentration in rifampicin to 5.66MIC.

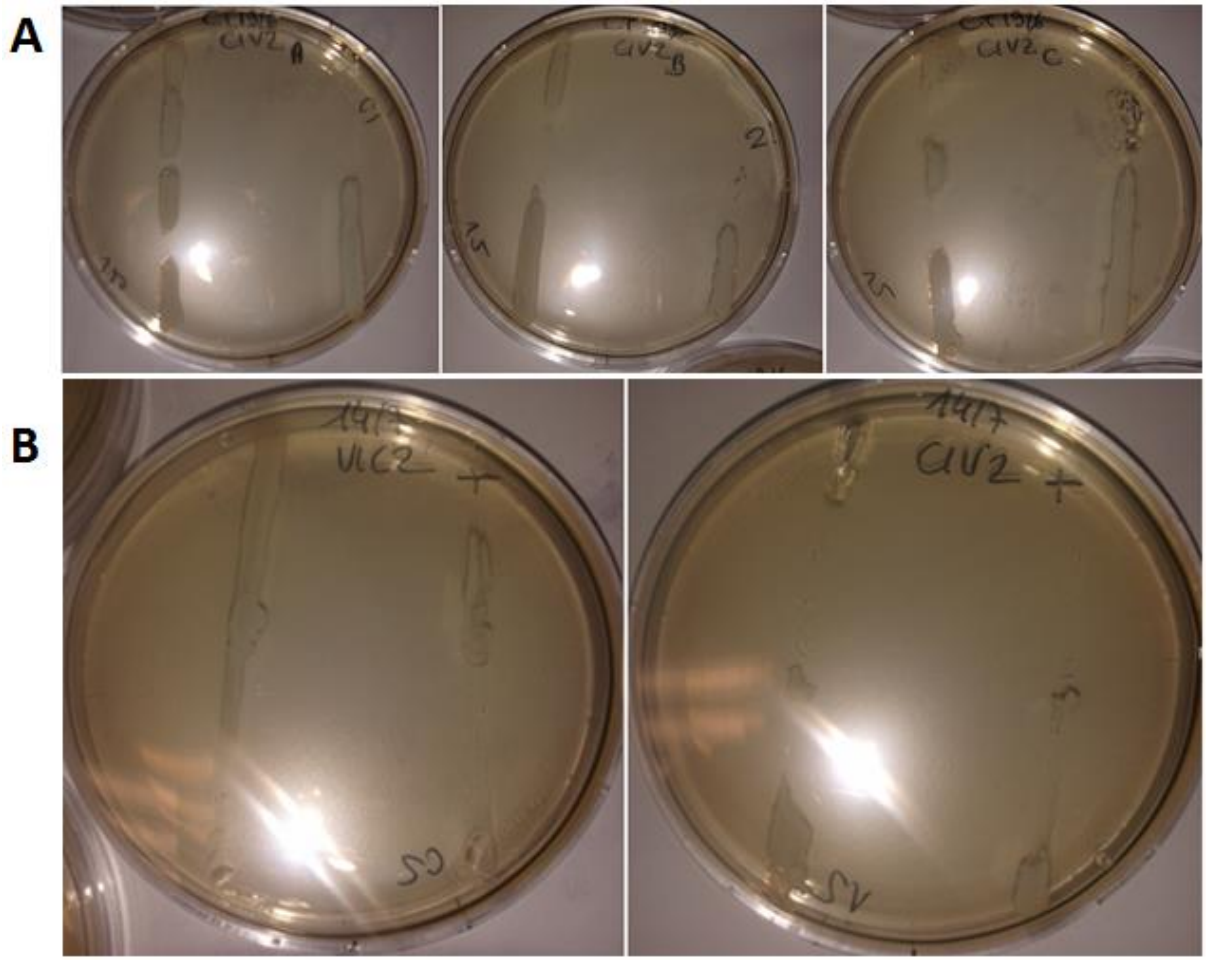


Figure A10

Examples of the discrepancies observed in the bacterial growth patterns on antibiotic control dishes produced on the same day, A - between replicated dishes, and B - between the two different control conditions of the same antibiotic (V1C2 and C1V2 being expected to provide a similar gradient of antibiotic inside the soft agar, allowing a similar bacterial growth pattern).

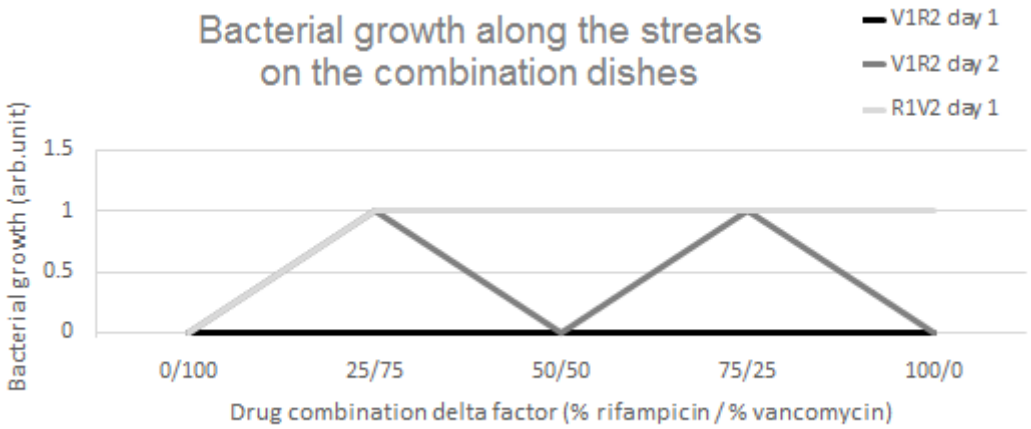


Figure A11

For low doses of antibiotics, the MRSA252 growth patterns on the combination "layer dishes" vary and do not evidence any type of interaction between vancomycin and rifampicin. Using 4.6-times the Minimum Inhibitory Concentration (MIC) for vancomycin and 2.76-times the MIC for rifampicin, the antibiotic concentrations along the streaks are expressed as drug concentration delta factor or relative drug fraction θ (percentage of rifampicin / percentage of vancomycin) and the bacterial growth on five points of the streaks (θ = 0/100, = 25/75, = 50/50, = 75/25 and = 100/0) is in arbitrary (arb.) unit (0 stands for inhibition and 1 for growth). *The dish "V1R2 day 1" (black) was produced on the same day as "R1V2 day 1" (light grey), and the dish "V1R2 day 2" was produced on the next day, using the same protocol – the same growth pattern was expected. Data represents the streaks from all replicates of the dishes (V1R2 or R1V2) produced on said-days.*

A.2.4. Bacterial culture using Etests

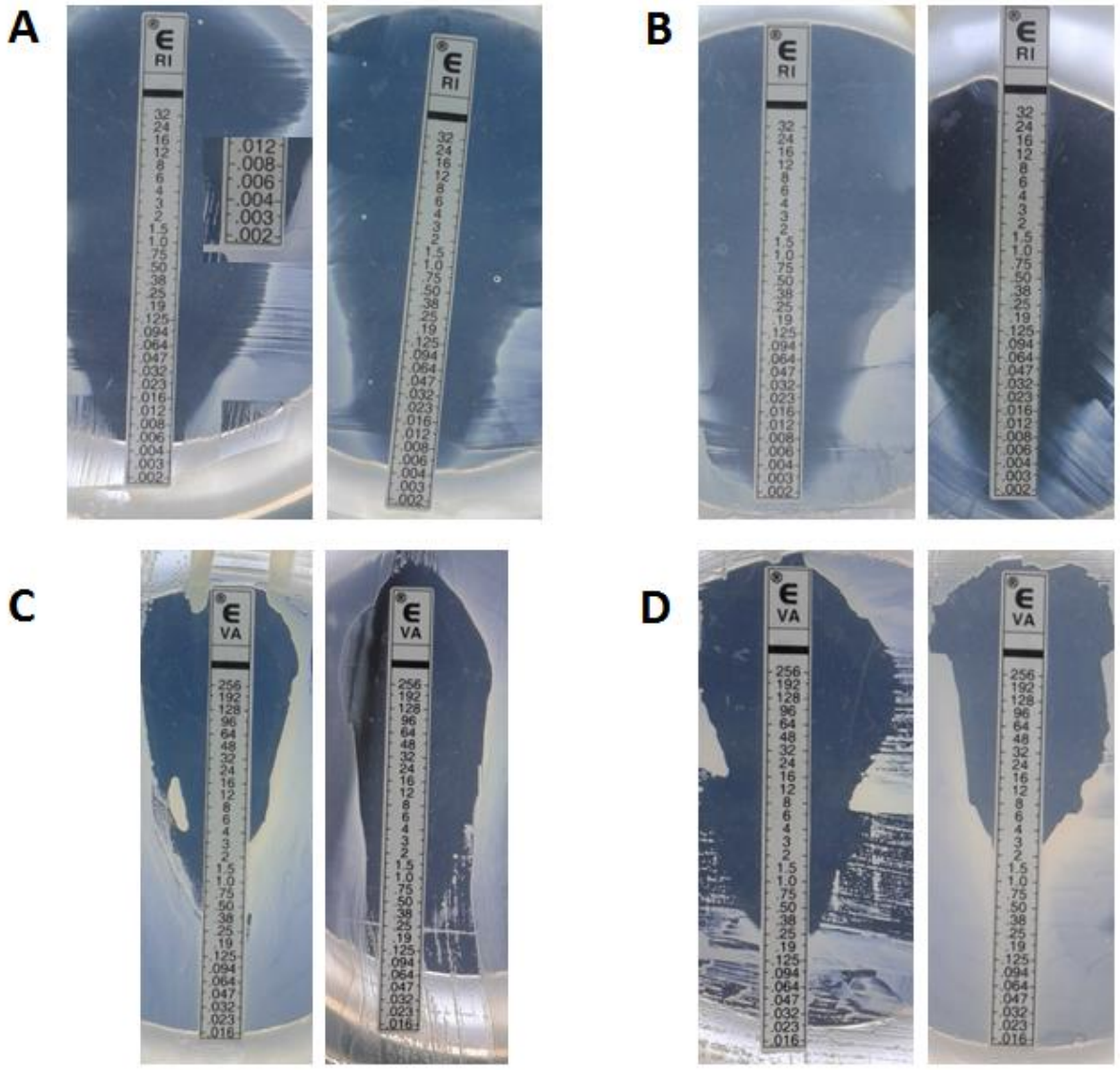


Figure A12

Examples of differences in the observed antimicrobial susceptibility testing determination of the Minimum Inhibitory Concentration (MIC) of antibiotics used in monotherapy, using Etests for rifampicin (A and B) and vancomycin (C and D) on different days, on both LB (A and C) and MH (B and D) agar plates previously streaked with MRSA252. For rifampicin, the determined MICs were read here as: (LB) 0.003, 0.012, (MH) 0.004 and 0.008 µg/mL. For vancomycin, the determined MICs were read as: (LB) 4, 0.038, (MH) 0.038 and 3 µg/mL. *LB stands for Luria Bertani and MH for Muller-Hinton.*

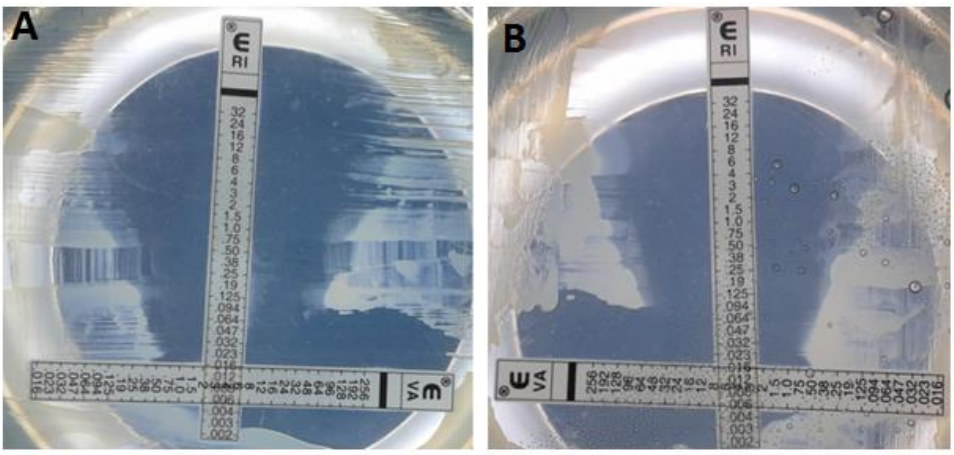


Figure A13

Example of two Mueller Hinton-agar plates streaked, on the same day, with MRSA252 before the application of Etests for the combination of vancomycin (at 2 µg/mL) and rifampicin (at 0.004 µg/mL), with vancomycin being applied first (A) or rifampicin being applied first (B). The determined Minimum Inhibitory Concentrations for both antibiotics used in combination were read here as: (A) 1 µg/mL for vancomycin and 0.004µg/mL for rifampicin, and (B) 1µg/mL for vancomycin and 0.003µg/mL for rifampicin.

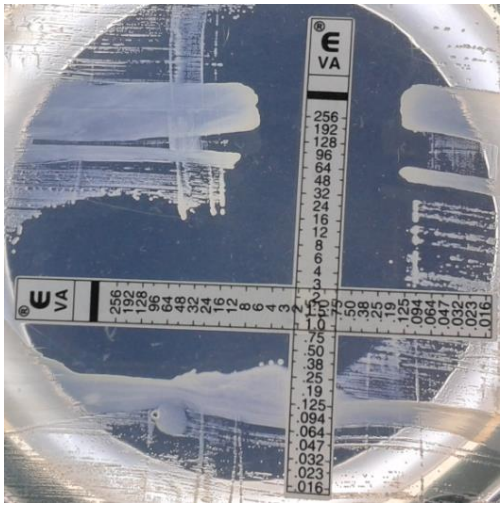


Figure A14

Example of a Luria Bertani-agar plate streaked with MRSA252 before the application of the combination of two vancomycin strips (at 0.75µg/mL). The determined Minimum Inhibitory Concentrations for this combination was read as 0.3µg/mL, and the Fractional Inhibitory Concentration Index (FICI) data was calculated as 1.01µg/mL, associating this combination to an indifference.

Computing FICI:

FICI data is provided in the main chapters as the sum of ratios of the Minimum Inhibitory Concentration (MIC) of the combination of Antibiotics A and B divided by the MIC of the antibiotic A (or B) monotherapy, using previous day monotherapy MIC data (as in Clinical Diagnostic laboratories). Contrarily to this “method 1”, we could consider another way to calculate FICI: using the same day monotherapy MIC data (method 2). It is particularly important given the daily variability of the MICs (Figure 3.23).

Table A1 shows the differences in FICI data between both methods, using the MIC data from Tables 3.5 and 3.7. The FICI were variating with MICs in monotherapy, to the point that 3 experiments (in bold) out of 13 had their antibiotic interpretation changed: a “real” combination with rifampicin on top (on MH-agar) and a double vancomycin combination (on LB-agar) were associated either to indifference (method 1) or synergy (method 2), depending on the selected monotherapy MIC data. And, likewise, a double vancomycin combination (on MH-agar) was associated to either indifference (method 1) or antagonism (method 2).

Experiments		FICI results, depending on the experiment (A-D) and the method (1-2)							
		A		B		C		D	
		1	2	1	2	1	2	1	2
On MH-agar plates	Vancomycin on rifampicin	0.83	2	NA					
	Rifampicin on vancomycin	1.16	1.75	**1.25**	**0.5**	0.58	1.56	1.26	0.68
On LB-agar plates	Double vancomycin combination	**1.17**	**0.44**	NA	NA	0.84	1.66	NA	
	Double rifampicin combination	NA		2	0.75	NA			
On MH-agar plates	Double vancomycin combination	NA				**1.16**	**4.6**	1.55	1.32
	Double rifampicin combination	0.58	1.67	NA				1.75	0.58

Table A1

Examples of several experiment Fractional Inhibitory Concentration Index (FICI) results from antimicrobial susceptibility testing using Etest combinations of two - different or not - antibiotics, on Luria Bertani (LB) and Mueller-Hinton (MH) agar plates streaked with MRSA252. Two methods to calculate FICI were used: “1” selects for the monotherapy MIC data from the previous day, “2” for those from the same day. These experiments were conducted during the same couple of weeks using the same stock material to prepare the media. FICI ≤ 0.5 is correlated to a synergy of both drugs against the microorganism, 0.5 > FICI > 4.0 to indifference and FICI ≥ 4.0 to antagonism. *NA indicates an absence of validated experimental condition; no MIC was retrieved and therefore no FICI could be calculated. Bold shows experiments in which the two methods provide different interpretations.*

APPENDIX B - Chapter 4 - *In silico* investigation of the potential impact of the most common variants on the *CYP3A4* gene expression or enzyme activity.

B.1. Materials and Methods

B.1.1. Sequences

Gene / Transcript variant	Sequence reference	Sequence length (bp)	Link
CYP3A4 gene	NG_008421.1 (GenBank) or Genomic Sequence (chr7:99756960-99784188) (UCSC)	34,205	FASTA or FASTA
CYP3A4 transcript variant 1 (mRNA)	NM_017460.5 (NCBI Reference Sequence)	2,792	FASTA
CYP3A4 transcript variant 2 (mRNA)	NM_001202855.2 (NCBI Reference Sequence)	2,789	FASTA

Table B1

Detail of the *CYP3A4* sequences used in this study. *mRNA stands for messenger RNA and bp for base pair.*

B.1.2. Identification of the CYP3A4 variants of interest

The UCSC Genome Browser was used to identify the most "common" *CYP3A4* variants. "Common SNPs (150)" under the "Variations" tab allowed to display only those associated to a world Minor Allele Frequency of 1%.

To select only the SNPs in i) exons, or ii) intronic and in vicinity (at maximum 100 nucleotides) of the closest exon boundary, the locus of all "common" genetic variants was accessed (by clicking on the SNP), recorded and compared to the exon boundary loci. *For example, the intronic rs2687107 (chr7:99,774,418-99,774,418) sits in-between exon 3 (closest boundary at: chr7:99,778,028) and exon 4 (closest boundary at: chr7:99,772,684), yet it was not considered in the vicinity of them.*

B.1.3. In silico searches to predict the potential effect of the CYP3A4 variants

Default parameters were used for the database searches, as advised by their manuals and as performed in the literature. Pre-set selections and parameters were kept, as follow:

RegRNA 2.0: Transfac TFBS in human (score ≥ 0), miRNA Target Sites in human (score ≥ 170 and free-energy ≤ -25).

Poly-miRTS: minimum conservation of 2.

Human Splice Finder 3.1: MaxEnt (with a threshold of 3), ESE Finder matrices (Table B2), ESE matrices (threshold of 59.245 for 9G8 and 75.964 for Tra2- β), silencer motifs from *Sironi et al.* (threshold of 60), hnRNP matrixe (threshold of 65.476 for hnRNP A1), etc.

	HSF scale	ESE Finder scale
ESE Finder - SF2/ASF	72.98	1.956
ESE Finder - SF2/ASF (IgM - BRCA1)	70.51	1.867
ESE Finder - SC35	75.05	2.383
ESE Finder - SRp40	78.08	2.67
ESE Finder - SRp55	73.86	2.676

Table B2
Thresholds associated to the enhancers in ESE Finder 3.0, directly or through Human Splicing Finder 3.1.

Fruit Fly: minimum score of 0.4 for both types of splice sites.

ESE Finder 3.0: SR proteins (Table B2), Branch Sites (threshold of 0), human 5'- splice sites (threshold of 6.67) and human 3'- splice site (threshold of 6.632).

SpliceAid 2: all available tissues were tested.

mfold Web Server: conformation at 37°C, in a concentration of sodium cation of 1.0M, in absence of magnesium cation.

B.2. Results

B.2.1. Identification of the most common and potentially altering CYP3A4 variants

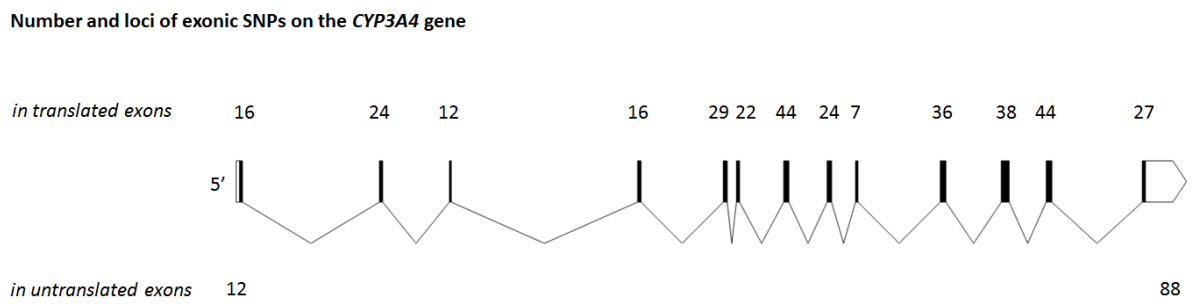


Figure B1

Distribution of all exonic Single Nucleotide Polymorphisms (SNPs) along the *CYP3A4* gene. All exonic *CYP3A4* SNPs are visualised from the 5'-Untranslated Region (UTR) to the 3'-UTR (both in white). The number of translated exonic SNPs are provided above the gene drawing, the position of the number indicating on which exon they are located. Similarly, the number of untranslated exonic SNPs are provided below the gene drawing, the position of the number indicating on which untranslated exon they are located. As 3,100 variants were counted on db150 and 439 exonic variants are presented here, the number of intronic *CYP3A4* SNPs and unreferenced variants is of 2,611*. Black vertical boxes represent the exons, black lines the introns.*

***In silico* prediction of the effect of the eight *CYP3A4* SNPs of interest**

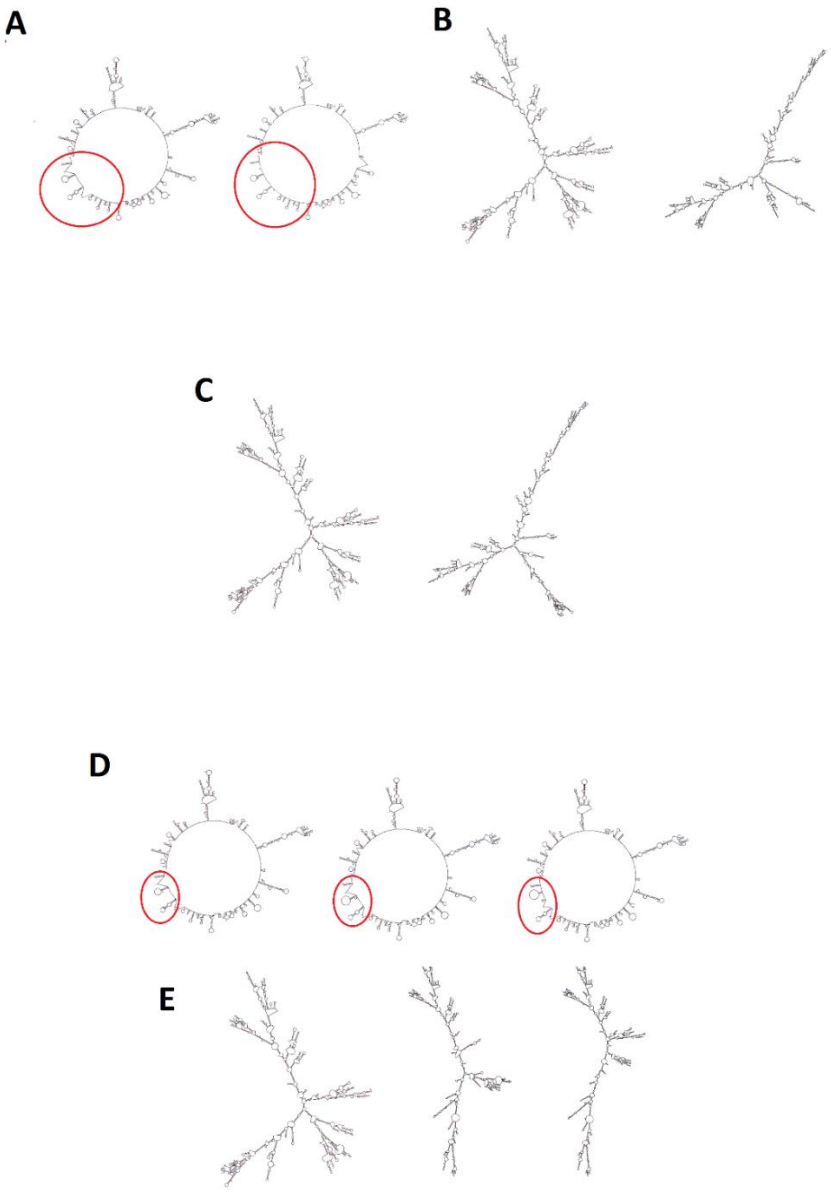


Figure B2

Conformation changes between the major and the minor variants of *CYP3A4* rs28988603, rs28988604 and rs28988606 at 37°C. AB- DNA (A) and RNA (B) folding conformation change between the major (left) and the minor variant (right) of rs28988603. C- RNA folding conformation change between the major (left) and the minor variant (right) of rs28988604. No DNA conformation change for rs28988604. DE- DNA (D) and RNA (E) folding conformation changes between the major (left), the minor variant T (middle) and the minor variant G (right) of rs28988606. *Conformations retrieved from mfold.*

B.2.2. Distribution of the genotypes for the eight variants of interest in the 1000Genome Project

rs28969391 A > -	rs28371763 T > A	rs28988604 G > A	rs28988606 G > C or T	rs28988603 A > C	rs12721620 A > G	rs2242480 T > C	rs2687116 C > A	#	%
-|-	T|T	A|A	C|C	C|C	A|A	T|T	A|C	1	0.04
-|-	T|T	A|A	C|C	C|C	A|A	T|T	C|C	1	0.04
-|-	T|T	A|A	G|G	A|A	G|G	T|T	C|C	2	0.08
-|-	T|T	A|G	C|G	A|C	A|A	T|T	A|A	1	0.04
-|-	T|T	A|G	C|G	A|C	A|A	T|T	A|C	10	0.40
-|-	T|T	A|G	C|G	A|C	A|A	T|T	C|C	2	0.08
-|-	T|T	A|G	C|G	A|C	A|G	T|T	A|A	4	0.16
-|-	T|T	A|G	C|G	A|C	A|G	T|T	A|C	9	0.36
-|-	T|T	A|G	C|G	A|C	A|G	T|T	C|C	5	0.20
-|-	T|T	A|G	G|G	A|A	A|G	T|T	C|C	2	0.08
-|-	T|T	A|G	G|G	A|A	G|G	C|T	A|A	1	0.04
-|-	T|T	A|G	G|G	A|A	G|G	C|T	A|C	1	0.04
-|-	T|T	A|G	G|G	A|A	G|G	T|T	A|A	1	0.04
-|-	T|T	A|G	G|G	A|A	G|G	T|T	A|C	3	0.12
-|-	T|T	A|G	G|G	A|A	G|G	T|T	C|C	6	0.24
-|-	T|T	G|G	G|G	A|A	A|A	T|T	A|A	6	0.24
-|-	T|T	G|G	G|G	A|A	A|A	T|T	A|C	20	0.80
-|-	T|T	G|G	G|G	A|A	A|A	T|T	C|C	26	1.04
-|-	T|T	G|G	G|G	A|A	A|G	T|T	A|A	9	0.36
-|-	T|T	G|G	G|G	A|A	A|G	T|T	A|C	38	1.52
-|-	T|T	G|G	G|G	A|A	A|G	T|T	C|C	67	2.68
-|-	T|T	G|G	G|G	A|A	G|G	C|C	A|A	2	0.08
-|-	T|T	G|G	G|G	A|A	G|G	C|T	A|A	22	0.88

-|-	T|T	G|G	G|G	A|A	G|G	T|T	A|A	93	3.71
-|-	T|T	G|G	G|G	A|A	G|G	T|T	A|C	29	1.16
-|-	T|T	G|G	G|G	A|A	G|G	T|T	C|C	38	1.52
-|A	A|T	A|G	C|G	A|C	A|G	C|T	A|A	1	0.04
-|A	A|T	G|G	G|G	A|A	A|G	C|T	A|C	3	0.12
-|A	A|T	G|G	G|G	A|A	G|G	C|T	A|A	14	0.56
-|A	T|T	A|G	C|G	A|C	A|G	C|T	A|A	3	0.12
-|A	T|T	A|G	C|G	A|C	A|G	C|T	A|C	4	0.16
-|A	T|T	A|G	C|G	A|C	A|G	C|T	C|C	1	0.04
-|A	T|T	A|G	C|G	A|C	A|G	T|T	A|A	1	0.04
-|A	T|T	A|G	C|G	A|C	A|G	T|T	A|C	7	0.28
-|A	T|T	A|G	C|G	A|C	A|G	T|T	C|C	7	0.28
-|A	T|T	A|G	C|G	A|C	G|G	T|T	A|C	1	0.04
-|A	T|T	A|G	G|G	A|A	A|G	T|T	A|A	1	0.04
-|A	T|T	A|G	G|G	A|A	G|G	C|C	A|A	19	0.76
-|A	T|T	A|G	G|G	A|A	G|G	C|C	A|C	1	0.04
-|A	T|T	A|G	G|G	A|A	G|G	C|T	A|A	9	0.36
-|A	T|T	A|G	G|G	A|A	G|G	C|T	A|C	8	0.32
-|A	T|T	A|G	G|G	A|A	G|G	C|T	C|C	1	0.04
-|A	T|T	A|G	G|G	A|A	G|G	T|T	A|C	2	0.08
-|A	T|T	A|G	G|G	A|A	G|G	T|T	C|C	7	0.28
-|A	T|T	G|G	G|G	A|A	A|G	C|T	A|A	14	0.56
-|A	T|T	G|G	G|G	A|A	A|G	C|T	A|C	50	2.00
-|A	T|T	G|G	G|G	A|A	A|G	C|T	C|C	20	0.80
-|A	T|T	G|G	G|G	A|A	A|G	T|T	A|A	5	0.20
-|A	T|T	G|G	G|G	A|A	A|G	T|T	A|C	25	1.00
-|A	T|T	G|G	G|G	A|A	A|G	T|T	C|C	61	2.44

-|A	T|T	G|G	G|G	A|A	G|G	C|C	A|A	51	2.04
-|A	T|T	G|G	G|G	A|A	G|G	C|T	A|A	436	17.41
-|A	T|T	G|G	G|G	A|A	G|G	C|T	A|C	70	2.80
-|A	T|T	G|G	G|G	A|A	G|G	C|T	C|C	22	0.88
-|A	T|T	G|G	G|G	A|A	G|G	T|T	A|A	23	0.92
-|A	T|T	G|G	G|G	A|A	G|G	T|T	A|C	36	1.44
-|A	T|T	G|G	G|G	A|A	G|G	T|T	C|C	56	2.24
A|A	A|A	G|G	G|G	A|A	G|G	C|C	A|A	1	0.04
A|A	A|T	G|G	G|G	A|A	G|G	C|C	A|A	36	1.44
A|A	A|T	G|G	G|G	A|A	G|G	C|T	A|A	1	0.04
A|A	T|T	G|G	A|G	A|A	G|G	C|C	A|A	2	0.08
A|A	T|T	G|G	G|G	A|A	G|G	C|C	A|A	919	36.70
A|A	T|T	G|G	G|G	A|A	G|G	C|C	A|C	13	0.52
A|A	T|T	G|G	G|G	A|A	G|G	C|C	C|C	3	0.12
A|A	T|T	G|G	G|G	A|A	G|G	C|T	A|A	65	2.60
A|A	T|T	G|G	G|G	A|A	G|G	C|T	A|C	41	1.64
A|A	T|T	G|G	G|G	A|A	G|G	C|T	C|C	15	0.60
A|A	T|T	G|G	G|G	A|A	G|G	T|T	A|A	21	0.84
A|A	T|T	G|G	G|G	A|A	G|G	T|T	A|C	11	0.44
A|A	T|T	G|G	G|G	A|A	G|G	T|T	C|C	18	0.72

Table B3

Genotypic distribution of the 2504 individuals from 1000Genome Project for rs28969391, rs28371763, rs28988604, rs28988606, rs28988603, rs12721620, rs2242480 and rs2687116. 70 different genotypes were retrieved. *"A > C" indicates that, for the studied variant, A is the major and ancestral variant, and C is the minor variant. The # column indicates the number of individuals from the world 1000Genome Project population associated to each genotype, % the percentage of these individuals in the project. A stands for Adenine, C for cytosine, G for guanine and T for thymine.*

APPENDIX C - Chapter 5 - Investigating the impact of *Carboxylesterase 2* genetic variants on host xenobiotic metabolism, treatment efficacy and adverse effects.

C.1. Material and Methods

C.1.1. Bioinformatics

C.1.1.1. Identification of gene sequences and genetic variation

The UCSC Genome Browser was used to identify the most "common" *CYP3A4* variants. "Common SNPs (150)" under the "Variations" tab allowed to display only those associated to a world Minor Allele Frequency of 1%.

To measure the distance of the SNPs from consensus splice sites, the locus of all "common" genetic variants was accessed (by clicking on the SNP), recorded and compared to the exon boundary loci. *For example, rs28382827 (chr16:66,943,992-66,943,992) is located in the translated exon 13, associated to the closest exon boundary at chr16:66,943,839-66,943,839 – it is not considered in the vicinity of this boundary.*

C.1.1.2. In silico assessment of potential effects on RNA transcription or processing

Default parameters were used for the database searches, as advised by their manuals and as performed in the literature. Pre-set selections and parameters were kept, as follow:

RAVEN: "Exclude low-scoring hits option", transcription factor (score threshold of 80%), conservation cut-off of 0.4, minimum SNP-caused score difference of 1.5 and minimum specificity of 10 bits. RAVEN was searched on all available mapping for genes (BX538086, BC032095, AK095522, AL713761, Y09616, D50579, U60553, AI601196).

RegRNA 2.0: Transfac TFBS in human (score ≥ 0), miRNA Target Sites in human (score ≥ 170 and free-energy ≤ -25).

Poly-miRTS: minimum conservation of 2.

mfold Web Server: conformation at 37°C, in a concentration of sodium cation of 1.0M, in absence of magnesium cation.

C.1.2. Measurement of relative allelic expression

C.1.2.1. Determination of genotypes

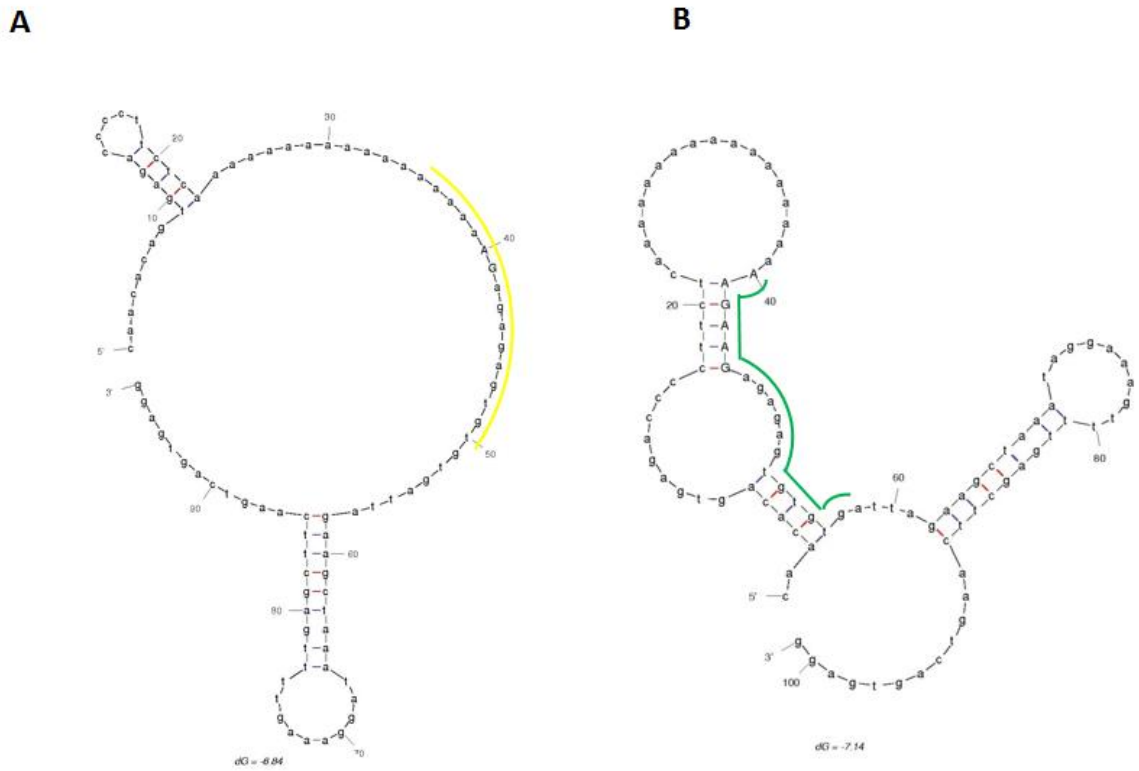


Figure C1

DNA conformation of the VIC-tagged allele (A) and the FAM-tagged allele (B) from the Custom Taqman Single Nucleotide Polymorphism Assay for rs28382828. *Data from mfold*.

Descriptions		Sequences (5'- to 3'-)
M13-cap for…	Forward Primer	TGTAAAACGACGGCCAGT
	Reverse Primer	CAGGAACACGCTATGACC
Primers as designed by Pyromark	Forward Primer	AGCCGAGGAGCCTGAAGA
	Reverse Primer	GGATTAGGGGCATGAGCCAC
Complete primers	Forward Primer	TGTAAAACGACGGCCAGTAGCCGAGGAGCCTGAAGA
	Reverse Primer	CAGGAACACGCTATGACCGGATTAGGGGCATGAGCCAC

Table C1

Primers to amplify both rs8192925 and rs28382828 by Polymerase Chain Reaction, selected with Pyromark Assay Design (Qiagen) and capped with a M13.

C.1.2.2. Allelic imbalance assay

Pyrosequencing allele quantification method

Pyrosequencing allows simultaneous DNA sequencing and complementary strand amplification. For every incorporated nucleotide, a pyrophosphate is released, leading to a cascade of reactions generating light (depending on the nucleotide incorporated). The Pyrosequencing platform quantifies the light emission and compares it to the value resulting from the incorporation of the expected nucleotide.

In a first step, a conventional Polymerase Chain Reaction (PCR) was performed on 20ng gDNA or cDNA samples, using HotStar Taq MasterMix (Qiagen) and 10µM of each of the Forward and Reverse primers (Forward being 5'-biotinylated and both being provided with a M13-cap). After running in the thermal cycler (Table C2), the PCR product was quantified on SYTO60 stained-agarose gel using the LI-COR scanner to validate PCR efficiency.

	Time	Temperature
Step 1	15min	95°C
Step 2	30sec	94°C (denaturation)
Step 3	1min	60°C (annealing)
Step 4	1min	72°C (synthesis)
Step 5	*Repeat steps 2-4 29 times*	
Step 6	10min	73°C

Table C2

Cycles and parameters for the conventional PCR for pyrosequencing. *Min and sec stand for minutes and seconds, respectively.*

8-38µL PCR products were immobilised with streptavidin beads in Binding Buffer, in the wells of 96-wells plates on a shaker for 10min. The Pyrosequencing Vacuum Prep Tool on the PyroMark station was used to pick the samples up. To separate DNA strands and release the samples into the PyroMark Q24 plate, the filter probes met successively 70% ethanol (EtOH), denaturing sodium hydroxide (NaOH) and 1x Wash Buffer. The samples were finally loaded in a PyroMark Q24 plate filled with Sequencing primer and Annealing Buffer. The plate was heated at 80°C for 90 sec and placed in the Pyrosequencer. A cartridge previously loaded with the enzyme, the substrate, and all four nucleotides, was placed inside the Pyrosequencer and it was set to dispense the nucleotides, to detect those included in the products and to record their accordance with the expected sequence (major variant). Divergences were reported by the accompanying software as percentages of sequences harbouring the major variant or not (from 0% - absence of the expected nucleotide, to 100% - only the expected nucleotide was introduced in the product at this position).

C.1.3. Clinical study

Sample donation and baseline data collection

Body Mass Index (BMI) was collected from recruited patients, or their height and weight as:

$$BMI = \frac{weight\ (kg)}{height\ (m) \times height\ (m)}.$$

According to the National Health Service (NHS, https://www.nhs.uk/live-well/healthy-weight/bmi-calculator/), BMI < 18.5 is associated to underweight patients, 18.5 ≤ BMI ≤ 24.9 to patients with an healthy weight, 25 ≤ BMI < 30 for overweight patients, and BMI > 30 for obese patients.

Baseline liver function test results were provided and included: ALanine Transaminase (ALT), Gamma-GlutamylTransferase (GGTP), BILirubin (BIL), ALkanine Phosphatase (ALP), ALBumin (ALB). Elevated levels of ALT, GGTP, BIL and ALP indicate a potential liver or bile duct damage. Higher-than-normal BIL are also indicative of a potential anaemia. Similarly, lower-than-normal levels of ALB are associated to potential liver damage. Normal ranges for these results are provided in Table C3[362].

Normal results range	Minimum normal value	Maximum normal value	Standard unit
ALT	7	55	units per liter
GGTP	8	61	units per liter
BIL	0.1	1.2	milligram per decilitre
	1.71	20.5	µmol per liter
ALP	40	129	units per liter
ALB	35	50	grams per liter

Table C3

Values of the normal ranges associated to the baseline liver function test results, with ALanine Transaminase (ALT), Gamma-GlutamylTransferase (GGTP), BILirubin (BIL), ALkanine Phosphatase (ALP), ALBumin (ALB).

C.2. Results

C.2.1. In silico predictions

C.2.1.1. Selecting bioinformatics-ly the CES2 genetic variants of interest

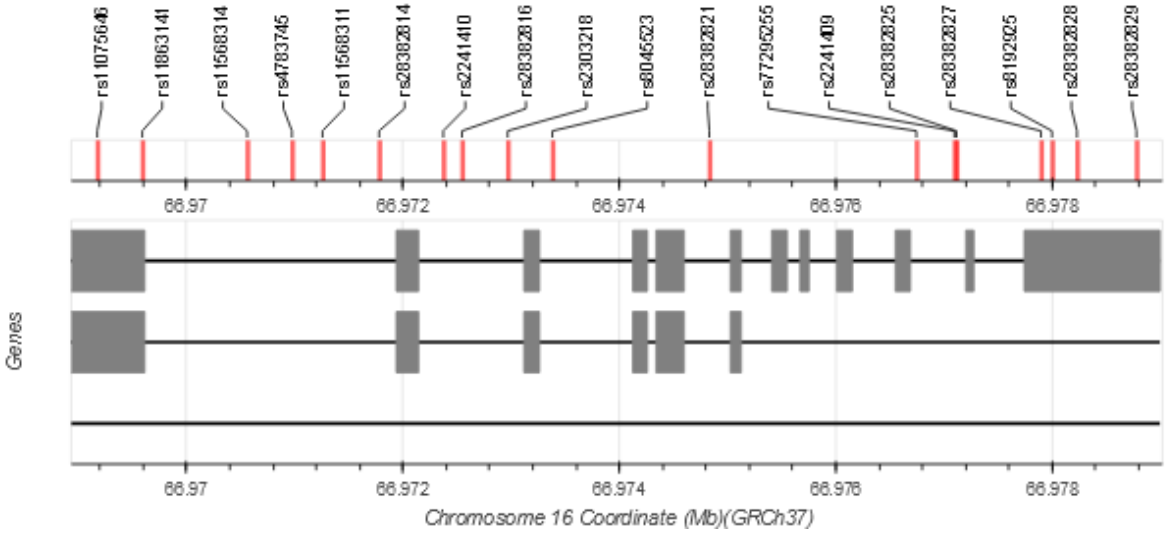


Figure C2

Distribution of the 18 "common" Single Nucleotide Polymorphisms on the *CES2* gene (visualised from 5'- to 3'-UnTranslated Region). *Drawing retrieved from LDLink.*

Variant	MAF (minor allele)	Allele count						Localisation on *CES2*
		World	Africa	South Asia	East Asia	America	Europe	
rs8045523	**0.37** (A)	G: 0.626 (3137) A: 0.374 (1871)	G: 0.222 (293) A: 0.778 (1029)	G: 0.781 (764) A: 0.219 (214)	G: 0.730 (736) A: 0.270 (272)	G: 0.733 (509) A: 0.267 (185)	G: 0.830 (835) A: 0.170 (171)	Intron 3
rs2241409	**0.33** (A)	G: 0.673 (3369) A: 0.327 (1639)	G: 0.396 (523) A: 0.604 (799)	G: 0.782 (765) A: 0.218 (213)	G: 0.730 (736) A: 0.270 (272)	G: 0.749 (520) A: 0.251 (174)	G: 0.820 (825) A: 0.180 (181)	Intron 10
rs4783745	**0.32** (G)	A: 0.676 (3384) G: 0.324 (1624)	A: 0.396 (523) G: 0.604 (799)	A: 0.781 (764) G: 0.219 (214)	A: 0.730 (736) G: 0.270 (272)	A: 0.754 (523) G: 0.246 (171)	A: 0.833 (838) G: 0.167 (168)	Intron 1
rs28382828	**0.25** (AGAA)	-: 0.749 (3752) AGAA: 0.251 (1256)	-: 0.464 (614) AGAA: 0.536 (708)	-: 0.945 (924) AGAA: 0.055 (54)	-: 0.777 (783) AGAA: 0.223 (225)	-: 0.784 (544) AGAA: 0.216 (150)	-: 0.882 (887) AGAA: 0.118 (119)	Exon 12 (in 3'-UTR)
rs11075646	**0.13** (G)	C: 0.867 (4341) G: 0.133 (667)	C: 0.685 (906) G: 0.315 (416)	C: 0.926 (906) G: 0.074 (72)	C: 0.960 (968) G: 0.040 (40)	C: 0.935 (649) G: 0.065 (45)	C: 0.907 (912) G: 0.093 (94)	Exon 1 (in 5'-UTR)
rs28382816	**0.10** (C)	T: 0.904 (4528)	T: 0.740 (978)	T: 0.895 (875)	T: 1.000 (1008)	T: 0.978 (679)	T: 0.982 (988)	Intron 2

		C: 0.096 (480)	C: 0.260 (344)	C: 0.105 (103)		C: 0.022 (15)	C: 0.018 (18)	
rs8192925	**0.09** (G)	A: 0.907 (4542) G: 0.093 (466)	A: 0.983 (1299) G: 0.017 (23)	A: 0.960 (939) G: 0.040 (39)	A: 0.763 (769) G: 0.237 (239)	A: 0.843 (585) G: 0.157 (109)	A: 0.944 (950) G: 0.056 (56)	Exon 12 (in 3'-UTR)
rs28382825	**0.08** (A)	G: 0.924 (4625) A: 0.076 (383)	G: 0.811 (1072) A: 0.189 (250)	G: 0.945 (924) A: 0.055 (54)	G: 0.960 (968) A: 0.040 (40)	G: 0.978 (679) A: 0.022 (15)	G: 0.976 (982) A: 0.024 (24)	Intron 10
rs2303218	**0.06** (C)	T: 0.940 (4706) C: 0.060 (302)	T: 0.999 (1321) C: 0.001 (1)	T: 0.993 (971) C: 0.007 (7)	T: 0.776 (782) C: 0.224 (226)	T: 0.912 (633) C: 0.088 (61)	T: 0.993 (999) C: 0.007 (7)	Intron 2
rs11568311	**0.05** (A)	G: 0.950 (4758) A: 0.050 (250)	G: 0.897 (1186) A: 0.103 (136)	G: 0.983 (961) A: 0.017 (17)	G: 1.000 (1008)	G: 0.961 (667) A: 0.039 (27)	G: 0.930 (936) A: 0.070 (70	Intron 1
rs11863141	**0.04** (A)	G: 0.956 (4790) A: 0.044 (218)	G: 0.843 (1115) A: 0.157 (207)	G: 1.000 (978)	G: 1.000 (1008)	G: 0.984 (683) A: 0.016 (11)	G: 1.000 (1006)	Exon 1 (transcribed part, synonymous amino acid: CT**G** > CT**A** coding for a leucine)

rs28382829	**0.04** (A)	G: 0.963 (4821) A: 0.037 (187)	G: 0.957 (1265) A: 0.043 (57)	G: 0.896 (876) A: 0.104 (102)	G: 1.000 (1008)	G: 0.987 (685) A: 0.013 (9)	G: 0.981 (987) A: 0.019 (19)	Exon 12 (in 3'-UTR)
rs2241410	**0.03** (A)	C: 0.972 (4869) A: 0.028 (139)	C: 1.000 (1322)	C: 0.997 (975) A: 0.003 (3)	C: 0.907 (914) A: 0.093 (94)	C: 0.947 (657) A: 0.053 (37)	C: 0.995 (1001) A: 0.005 (5)	Intron 2
rs28382827	**0.03** (T)	C: 0.966 (4837) T: 0.034 (171)	C: 0.967 (1278) T: 0.033 (44)	C: 0.896 (876) T: 0.104 (102)	C: 1.000 (1008)	C: 0.990 (687) T: 0.010 (7)	C: 0.982 (988) T: 0.018 (18)	Exon 12 (transcribed part, synonymous amino acid: CT**C** > CT**T** coding for a leucine)
rs11568314	**0.02** (T)	A: 0.984 (4927) T: 0.016 (81)	A: 0.996 (1317) T: 0.004 (5)	A: 0.989 (967) T: 0.011 (11)	A: 1.000 (1008)	A: 0.978 (679) T: 0.022 (15)	A: 0.950 (956) T: 0.050 (50)	Intron 1
rs28382821	**0.02** (T)	C: 0.984 (4926) T: 0.016 (82)	C: 0.939 (1241) T: 0.061 (81)	C: 1.000 (978)	C: 1.000 (1008)	C: 0.999 (693) T: 0.001 (1)	C: 1.000 (1006)	Intron 5
rs28382814	**0.01** (T)	C: 0.988 (4946) T: 0.012 (62)	C: 0.998 (1319) T: 0.002 (3)	C: 0.984 (962) T: 0.016 (16)	C: 1.000 (1008)	C: 0.991 (688) T: 0.009 (6)	C: 0.963 (969) T: 0.037 (37)	Intron 1

rs77295255	**0.01** (T)	**C**: 0.987 (4944) **T**: 0.013 (64)	**C**: 0.952 (1258) **T**: 0.048 (64)	**C**: 1.000 (978)	**C**: 1.000 (1008)	**C**: 1.000 (694)	**C**: 1.000 (1006)	Intron 10

Table C4

Minor Allele Frequencies (MAF), allele count in the world and in different ethnic groups (Africa, East Asia, South Asia, America and Europe) and localisation on the gene of the 18 most common *CES2* Single Nucloetide Polymorphisms (found in at least 1% of the world population). *Data retrieved from the 1000 Genome Project database and UCSC Genome Browser. UTR stands for UnTranslated Region, A for Adenine, C for cytosine, G for guanine and T for thymine.*

C.2.1.2. Investigating the potential impact of rs11075646, rs8192925 and rs28382828 with bioinformatics

Figure C3

Concordant sequences between the nucleotides neighbouring rs11075646 (highlighted in red) and the motif associated to the human CDP (Cux1) transcription factor (underlined), as reported in Jaspar.

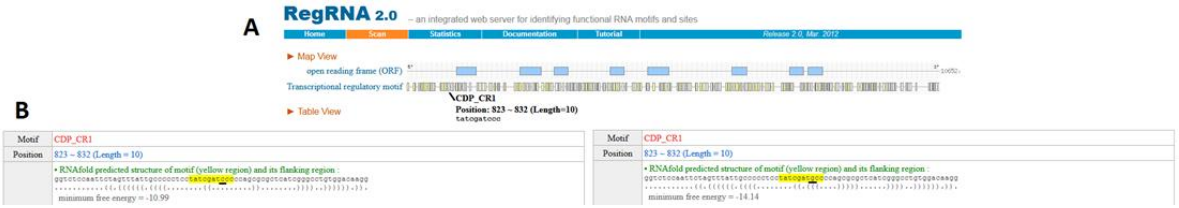


Figure C4

Bioinformatics evidences about the co-localisation of rs11075646 with the CDP_CR1 (and CDP_CR3 HD – data not shown) transcription factor. *Data retrieved from RegRNA 2.0. A showing where the CDP_CR1 can localise on the CES2 gene, B showing the RNAfold predicted structure (in yellow) and its flanking region.*

	rs11075646	rs8192925	rs28382828
Archives names	rs60326948	rs58407178	rs139332152
	rs16957087	16:g.66977997A>G	rs145299747
	16:g.66969176C>G	ENST00000566869.1:n.147+723A>G	rs66487958
	ENST00000317091.4:c.-171C>G	ENST00000317091.4:c.*69A>G	rs35735195
	ENST00000566182.1:c.-171C>G	ENST00000568470.1:c.*1837A>G	16:g.66978228_66978229insAGAA
	ENST00000568470.1:c.-171C>G	ENST00000417689.1:c.*69A>G	ENST00000566869.1:n.147+954_147+955insAGAA
	ENST00000561697.1:c.-85+360C>G		ENST00000317091.4:c.*300_*301insAGAA
	ENST00000417689.1:c.-171C>G		ENST00000417689.1:c.*300_*301insAGAA

Table C5

Names previously given to rs11075646, rs8192925 and rs28382828.

C.2.2. Measurement of relative allelic expression for rs11075646, rs8192925 and rs28382828

C.2.2.1. Finding heterozygous samples for the three variants of interest in the Exeter 10,000 cohort

Figures C5 and C6 (next pages) **– Matrixes of pairwise linkage disequilibrium statistics in the world population between the 18 "common" *CES2* Single Nucleotide Polymorphisms: R^2 followed by D'**. *Red shows the correlation in R^2 or D'. Matrix retrieved from the LDLink tool, based on the 1000Genome Project database.*

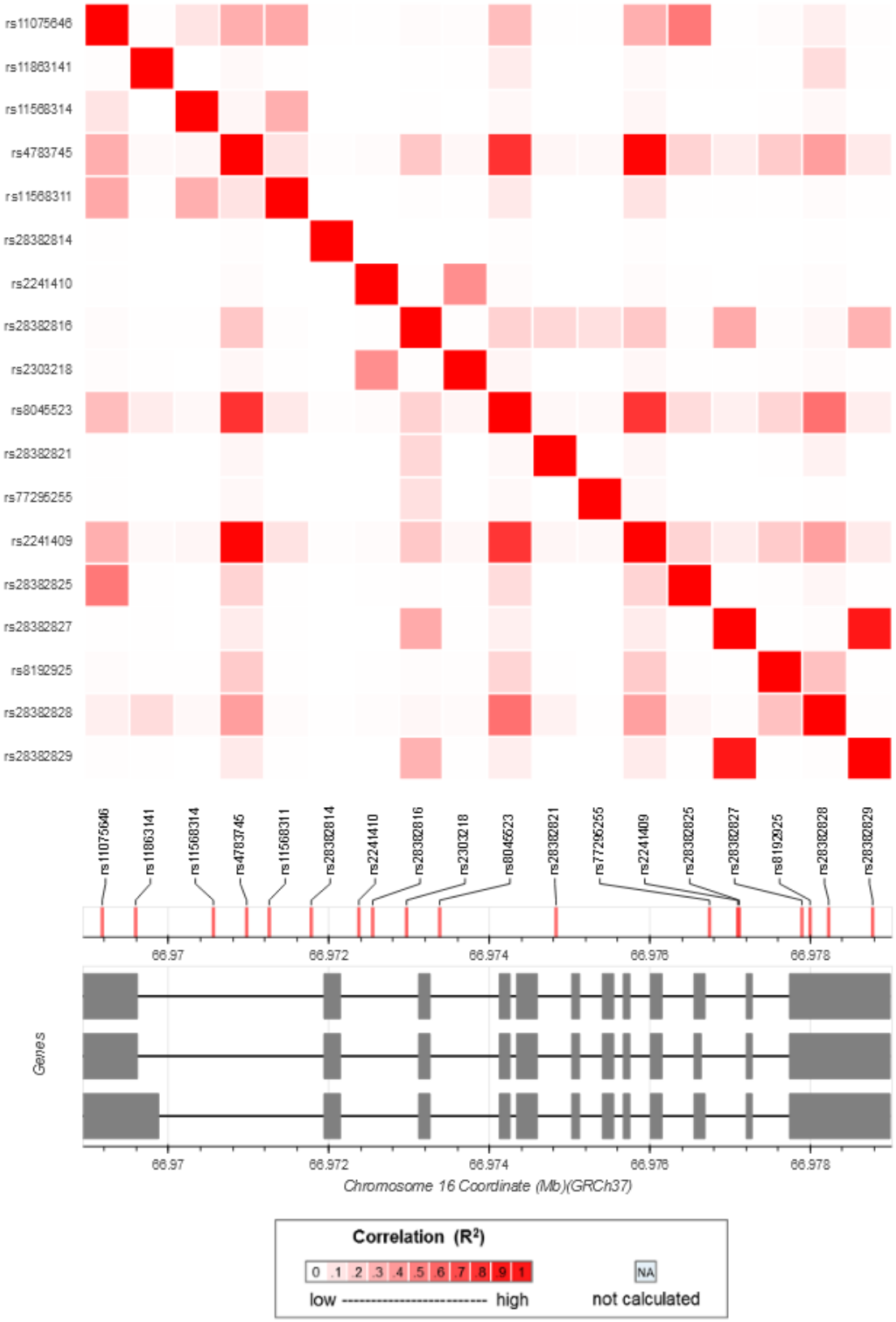

rs11075646
rs11863141
rs11568314
rs4783745
rs11568311
rs28382814
rs2241410
rs28382816
rs2303218
rs8045523
rs28382821
rs77295255
rs2241409
rs28382825
rs28382827
rs8192925
rs28382828
rs28382829
66.97
66.972
66.974
66.976
66.978
Genes
Chromosome 16 Coordinate (Mb)(GRCh37)
Correlation (R²)
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
low ---------- high
NA
not calculated

Figure C5

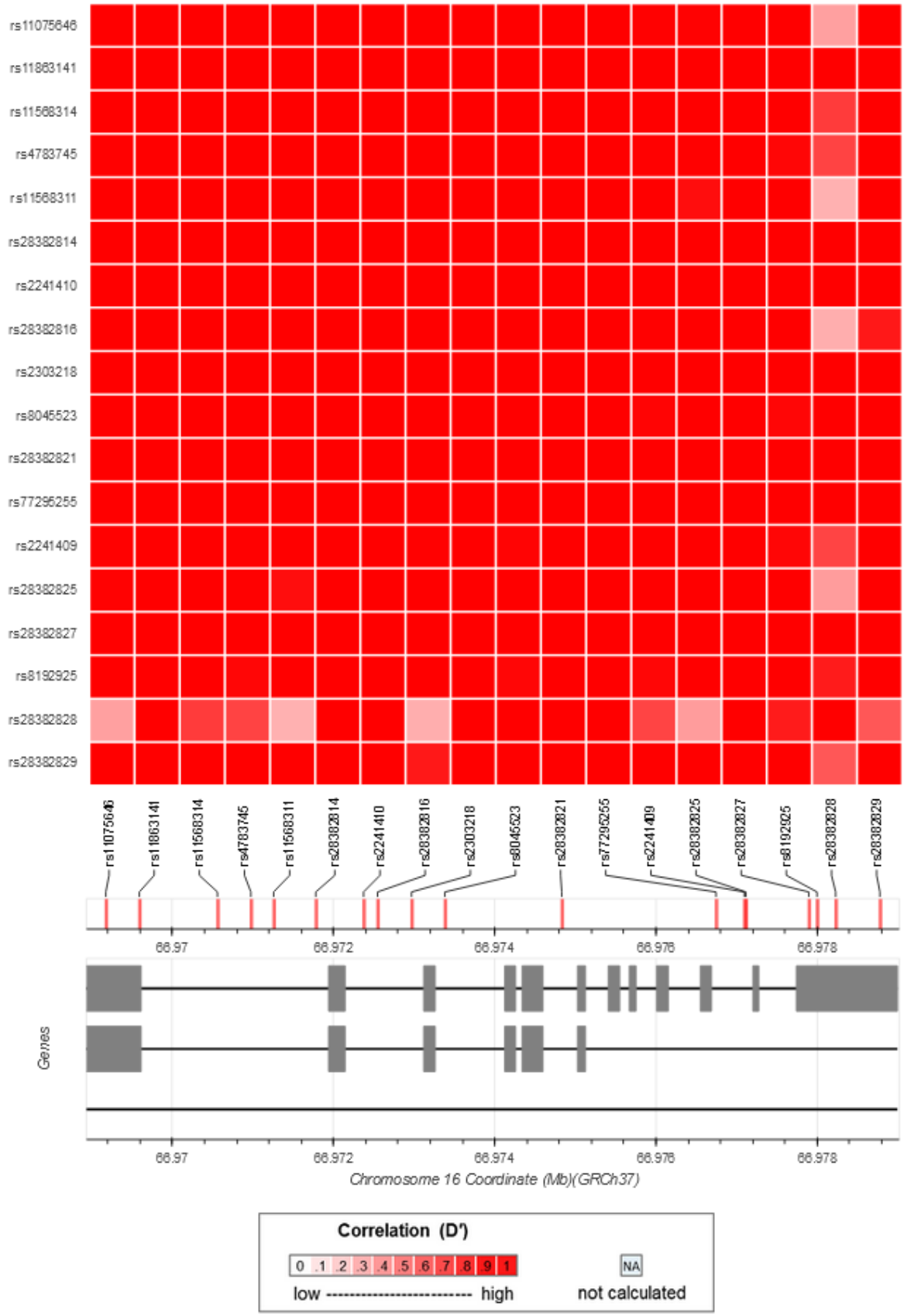

rs11075646
rs11863141
rs11568314
rs4783745
rs11568311
rs28382814
rs2241410
rs28382816
rs2303218
rs8045523
rs28382821
rs77295255
rs2241409
rs28382825
rs28382827
rs8192925
rs28382828
rs28382829
66.97
66.972
66.974
66.976
66.978
Genes
Chromosome 16 Coordinate (Mb)(GRCh37)
Correlation (D')
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
low ---------- high
NA
not calculated

Figure C6

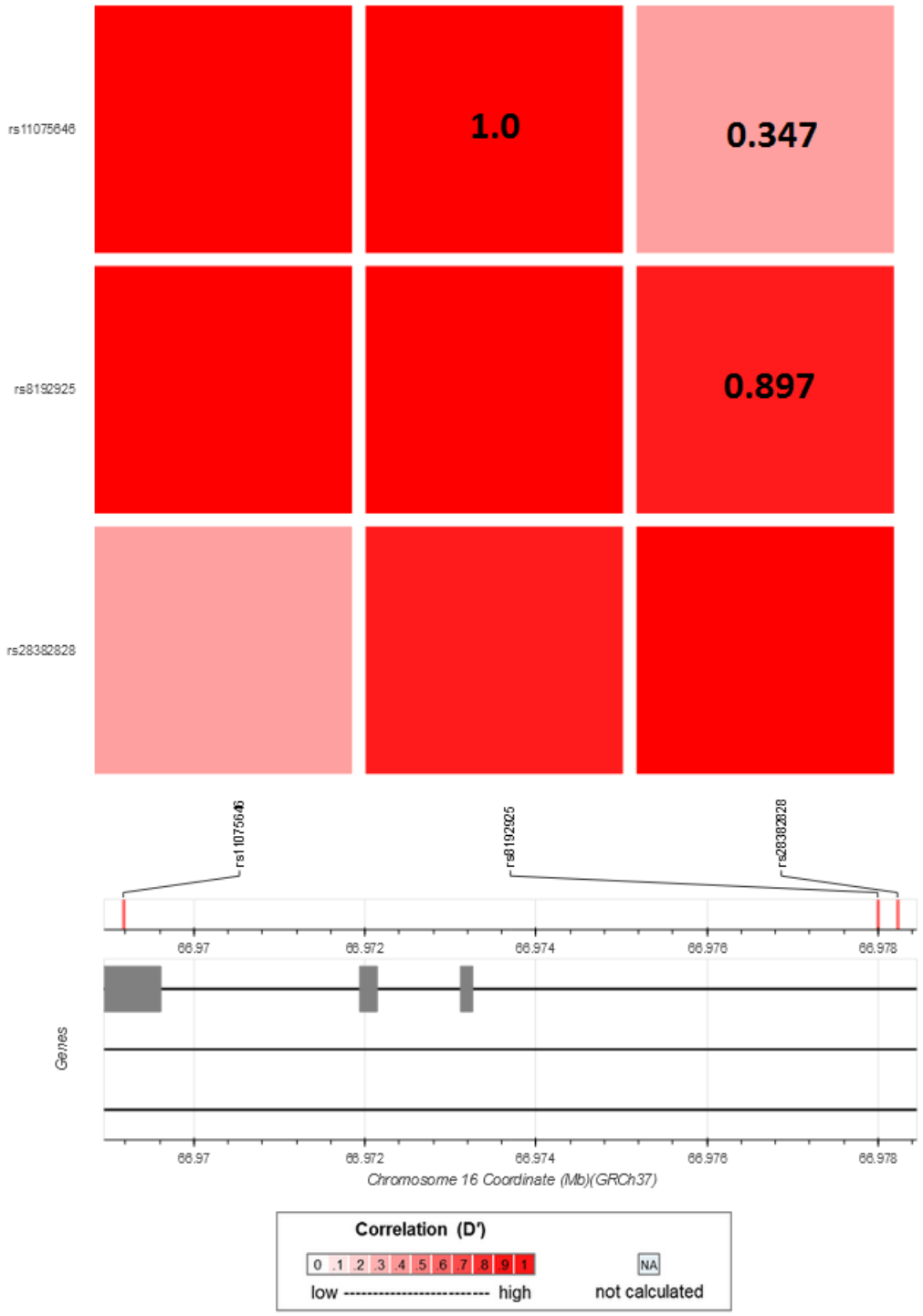


Figure C7

Matrix of pairwise linkage disequilibrium statistics in the world population between rs11075646, rs8192925 and rs28382828. *Red shows the correlation in D', the value is provided. Matrix retrieved from the LDLink tool, based on the 1000Genome Project database.*

C.2.2.2. Investigating the effect of rs11075646, rs8192925 and rs28382828 on CES2 expression

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
data	Equal variances assumed	1.573	.214	-1.633	68	.107	-1.52029	.93125	-3.37857	.33800
	Equal variances not assumed			-1.633	43.123	.110	-1.52029	.93125	-3.39818	.35761

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
corrected data	Equal variances assumed	5.044	.032	.528	32	.601	.27412	.51959	-.78426	1.33249
	Equal variances not assumed			.528	24.414	.603	.27412	.51959	-.79730	1.34554

Figure C8

rs28382828 allelic imbalance assay statistics. Top – Allelic balance for rs28382828 in samples heterozygous for both rs11075646 and rs28382828 (and homozygous for the major variant of rs8192925). Levene's test for equality of variances, sig. (significance) = 0.214, therefore variances are assumed unequal, therefore an independent samples t-test was performed, associated to a p-value = 0.101. **Bottom** - Allelic balance for rs28382828 in samples heterozygous for both rs8192925 and rs28382828 (and homozygous for the major variant of rs11075646). Levene's test for equality of variances, sig. = 0.032, therefore so variances are assumed equal, therefore an independent samples t-test was performed, associated to a p-value = 0.601.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Data	Equal variances assumed	36.714	.000	43.054	76	.000	2.099688972	.0487688368	2.002557416	2.196820529
	Equal variances not assumed			43.054	40.702	.000	2.099688972	.0487688368	2.001176423	2.198201521

Figure C9

rs11075646 allelic imbalance assay statistics. Allelic imbalance for rs11075646 in samples heterozygous for rs11075646. Levene's test for equality of variances, sig. (significance) ≤ 0.001, therefore variances are assumed equal, and an independent samples t-test was performed, associated to a p-value ≤ 0.001.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Data	Equal variances assumed	.087	.770	-4.916	34	.000	-.101897619	.0207288885	-.144023789	-.059771449
	Equal variances not assumed			-4.916	33.958	.000	-.101897619	.0207288885	-.144025688	-.059769549

Figure C10

rs8192925 allelic imbalance assay statistics. Allelic imbalance for rs8192925 in samples heterozygous for rs8192925. Levene's test for equality of variances, sig. (significance) = 0.770, therefore variances are assumed unequal, and an independent samples t-test was performed, associated to a p-value ≤ 0.001.

C.2.3. Clinical study of rs11075646 and rs8192925

Preliminary clinical data

Concomitant treatments administrated to patients from both cohorts				
adcal d3	betnorate	fortonir	nebivolol	rituximab infusion
Adoport	bisoprolol	fostair	nororapiol	romiticline
adoport erming	bisoprolol fumarate	fultium d3	nutritional supplements	salbutamol
alfacalciodol	calcidien d3	furosemide	oestrogen pessary	sodium bicarbonate
Amikacin	carbocisteine	gliclazide	**omeprazole**	spiriva respimat
amitriptyline	clofazimine	insulatorol	oxybutamin hydroclorin	tazocin
amplodopine	colomycin nebulizer	inhaled ciclesonide	paracetamol	tramadol
Apixaban	coracten	inhaled salmeterol	phospate-somdoz	ventolin
arames surrechic	cotrimoxazole	insuman basal	**prednisolone**	verapamil
Aspirin	denosumab	**isoniazid**	prograf	zopiclone
Atemolol	docusate sodium	isosorbide	pyrazinamide	
atorvastatin	doxazosin	lausoprasole	**pyridoxine**	
Autodipine	**ethambutol**	lencanidipre	ramitidine	
azithromycin	fluorexine	mebeverine	**rampiril**	
beclometasone	**folic acid**	metformin hydrochloriol	ranitidine	

Table C6

List of concomitant treatments administrated to tuberculosis or kidney transplant patients. The treatments in bold are the ones delivered to many of the patients from the same cohort.

Concomitant diseases and symptoms experienced by patients from both cohort	
Rifampicin cohort (tuberculosis)	**MMF cohort (kidney transplant)**
Acid fast bacilli positive	Alport syndrome
Asthma	Ankylosing spondylitis with NSAIDS
Basal cell carcinomas	Arthropathy
Bronchiectasis	BK virus nephropathy
Cerebral vasculitis	Coronary artery disease
Infarct	Deep vein thrombosis
Lymphadenopathy	Fenestration of lymphocele
Mycobacterium avian intracellular infection	Forearm vascularized composite allotransplantation
Mycobacterium avium infection	Gut
P-ANCA positive vasculitis	Haemodialysis
Previous hysterectomy with secondary bowel perforation and colostomy	Hemopoiesis
Pseudomonas infection	Hypertension
Superimposed lower respiratory tract infection	IgA neuropathy
Vasculitis	Infected collection around transplant pole
XXX chromosome syndrome	Liver unrelated transplant
	Neck lipoma
	Nephrectomy
	New onset diabetes mellitus after transplantation
	Non-ST segment elevation myocardial infarction
	Obstructive sleep apnea
	Pancreas transplant
	Papilledema

	Peritoneal dialysis
	Psoriasis
	Reiter's disease
	Supraventricular tachycardia
	Type I diabetis
	Type II diabetis
	Uretic stone

Table C7

List of past and concomitant diseases and symptoms experienced by tuberculosis (left) or kidney transplant (right) patients. *MMF stands for Mycophenolate Mofetil.*

Bibliography

1 Alberts, B. J., A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Molecular Biology of the Cell. 4th edition. *Garland Science* (2002).

2 Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* **109**, E2774-2783, doi:10.1073/pnas.1210309109 (2012).

3 Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* **99**, 803-808, doi:10.1073/pnas.022629899 (2002).

4 D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457-461, doi:10.1038/nature10388 (2011).

5 Jansen, G. B., C.; Schulenburg, H. Experimental evolution as an efficient tool to dissect adaptive paths to antibiotic resistance. *Drug Resistance Updates* **16**, 96-107 (2013).

6 Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277-283 (2015).

7 Peton, V. & Le Loir, Y. Staphylococcus aureus in veterinary medicine. *Infect Genet Evol* **21**, 602-615, doi:10.1016/j.meegid.2013.08.011 (2014).

8 Witte, W. Selective pressure by antibiotic use in livestock. *Int J of Antimicrobial Agents* **16**, 19-24 (2000).

9 Mehndiratta, P. L. & Bhalla, P. Use of antibiotics in animal agriculture & emergence of methicillin-resistant Staphylococcus aureus (MRSA) clones: need to assess the impact on public health. *Indian J Med Res* **140**, 339-344 (2014).

10 Mwangi, M. M. *et al.* Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. *Proc Natl Acad Sci U S A* **104**, 9451-9456, doi:10.1073/pnas.0609839104 (2007).

11 Blair, J. M. *et al.* AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. *Proc Natl Acad Sci U S A* **112**, 3511-3516, doi:10.1073/pnas.1419939112 (2015).

12 Komolafe, O. O. Antibiotic resistance in bacteria - an emerging public health problem. *Malawi Med J* **15**, 63-67 (2003).

13 Souli, M., Galani, I. & Giamarellou, H. Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Euro Surveill* **13** (2008).

14 Gandhi, N. R. *et al.* Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* **375**, 1830-1843, doi:10.1016/S0140-6736(10)60410-2 (2010).

15 World Health Organization. Antibacterial agents in clinical development – an analysis of the antibacterial clinical development pipeline, including Mycobacterium tuberculosis. *World Health Organization* (2017).

16 Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol* **5**, 175-186, doi:10.1038/nrmicro1614 (2007).

17 Ventola, C. L. The antibiotic resistance crisis: part 2: management strategies and new agents. *P T* **40**, 344-352 (2015).

18 Centers of Diseases Control and Prevention. Timeline of Antibiotic Resistance Compared to Antibiotic Development. (2013).

19 O'Neill, J. Tackling Drug-Resistant Infections Globally: final report and recommendations. (2016).

20 Viens, A. M. & Littmann, J. Is Antimicrobial Resistance a Slowly Emerging Disaster? *Public Health Ethics* **8**, 255-265, doi:10.1093/phe/phv015 (2015).

21 Andrei, S. V., L.; Chirvasuta, R.;Stefan, M.-G. New FDA approved antibacterial drugs: 2015-2017. *Discoveries* **6**, doi:10.15190/d.2018.1 (2018).

22 Spellberg, B., Powers, J. H., Brass, E. P., Miller, L. G. & Edwards, J. E., Jr. Trends in antimicrobial drug development: implications for the future. *Clin Infect Dis* **38**, 1279-1286, doi:10.1086/420937 (2004).

23 Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol* **8**, 423-435, doi:10.1038/nrmicro2333 (2010).

24 Hewlett, M. *The evolution of resistance to multidrug antibiotic therapies.* PhD thesis, University of Exeter, (2015).

25 Lombardino, J. G. & Lowe, J. A., 3rd. The role of the medicinal chemist in drug discovery--then and now. *Nat Rev Drug Discov* **3**, 853-862, doi:10.1038/nrd1523 (2004).

26 O'Neill, J. Antimicrobials in agriculture and the environment: reducing the unnecessary use and waste. (2015).

27 Dryden, M., Johnson, A. P., Ashiru-Oredope, D. & Sharland, M. Using antibiotics responsibly: right drug, right time, right dose, right duration. *J Antimicrob Chemother* **66**, 2441-2443, doi:10.1093/jac/dkr370 (2011).

28 Meropol, S. B. C., A.; Chen, Z.; Finkelstein, J.A.; Hennessy, S.; Lautenbach, E.; Platt, R.; Schech, S.D.; Shatin, D.; Metlay, J.P. Adverse Events Associated with Prolonged Antibiotic Use. *Pharmacoepidemiol Drug Saf.* **17**, 523-532 (2014).

29 Oz, T. *et al.* Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. *Mol Biol Evol* **31**, 2387-2401, doi:10.1093/molbev/msu191 (2014).

30 Andrade, R. J. & Tulkens, P. M. Hepatic safety of antibiotics used in primary care. *J Antimicrob Chemother* **66**, 1431-1446, doi:10.1093/jac/dkr159 (2011).

31 Dowling, H. F., Lepper, M. H. & Jackson, G. G. When should antibiotics be used in combination? *J Am Med Assoc* **151**, 813-815 (1953).

32 Lacey, R. W. Genetic basis, epidemiology, and future significance of antibiotic resistance in Staphylococcus aureus: a review. *J Clin Pathol* **26**, 899-913 (1973).

33 Coiffier, G., Albert, J. D., Arvieux, C. & Guggenbuhl, P. Optimizing combination rifampin therapy for staphylococcal osteoarticular infections. *Joint Bone Spine* **80**, 11-17, doi:10.1016/j.jbspin.2012.09.008 (2013).

34 Levin, S. & Harris, A. A. Principles of combination therapy. *Bull N Y Acad Med* **51**, 1020-1038 (1975).

35 Worthington, R. J. & Melander, C. Combination approaches to combat multidrug-resistant bacteria. *Trends Biotechnol* **31**, 177-184, doi:10.1016/j.tibtech.2012.12.006 (2013).

36 Clinical & Laboratory Standards Institute. Harmonized terminology database.

37 Mouton, J. W., Dudley, M. N., Cars, O., Derendorf, H. & Drusano, G. L. Standardization of pharmacokinetic/pharmacodynamic (PK/PD) terminology for anti-infective drugs: an update. *J Antimicrob Chemother* **55**, 601-607, doi:10.1093/jac/dki079 (2005).

38 Andrews, J. M. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother* **48 Suppl 1**, 5-16 (2001).

39 Vogelman, B. & Craig, W. A. Kinetics of antimicrobial activity. *J Pediatr* **108**, 835-840 (1986).

40 Drusano, G. L., Louie, A., Deziel, M. & Gumbo, T. The crisis of resistance: identifying drug exposures to suppress amplification of resistant mutant subpopulations. *Clin Infect Dis* **42**, 525-532, doi:10.1086/499046 (2006).

41 Pena-Miller, R. *et al.* When the most potent combination of antibiotics selects for the greatest bacterial load: the smile-frown transition. *PLoS Biol* **11**, e1001540, doi:10.1371/journal.pbio.1001540 (2013).

42 Gunnison, J. B., Kunishige, E., Coleman, V. R. & Jawetz, E. The mode of action of antibiotic synergism and antagonism: the effect in vitro on bacteria not actively multiplying. *J Gen Microbiol* **13**, 509-518, doi:10.1099/00221287-13-3-509 (1955).

43 Gunnison, J. B., Shevky, M. C., Bruff, J. A., Coleman, V. R. & Jawetz, E. Studies on antibiotic synergism and antagonism: the effect in vitro of combinations of antibiotics on bacteria of varying resistance to single antibiotics. *J Bacteriol* **66**, 150-158 (1953).

44 Renneberg, J. Definitions of antibacterial interactions in animal infection models. *J Antimicrob Chemother* **31 Suppl D**, 167-175 (1993).

45 Forrest, G. N. & Tamura, K. Rifampin combination therapy for nonmycobacterial infections. *Clin Microbiol Rev* **23**, 14-34, doi:10.1128/CMR.00034-09 (2010).

46 Chen, D., Liu, X., Yang, Y., Yang, H. & Lu, P. Systematic synergy modeling: understanding drug synergy from a systems biology perspective. *BMC Syst Biol* **9**, 56, doi:10.1186/s12918-015-0202-y (2015).

47 Bollenbach, T. Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Curr Opin Microbiol* **27**, 1-9, doi:10.1016/j.mib.2015.05.008 (2015).

48 Tam, V. H., Schilling, A. N., Lewis, R. E., Melnick, D. A. & Boucher, A. N. Novel approach to characterization of combined pharmacodynamic effects of antimicrobial agents. *Antimicrob Agents Chemother* **48**, 4315-4321, doi:10.1128/AAC.48.11.4315-4321.2004 (2004).

49 Hegreness, M., Shoresh, N., Damian, D., Hartl, D. & Kishony, R. Accelerated evolution of resistance in multidrug environments. *Proc Natl Acad Sci U S A* **105**, 13977-13981, doi:10.1073/pnas.0805965105 (2008).

50 Yeh, P. J., Hegreness, M. J., Aiden, A. P. & Kishony, R. Drug interactions and the evolution of antibiotic resistance. *Nat Rev Microbiol* **7**, 460-466, doi:10.1038/nrmicro2133 (2009).

51 Pena-Miller, R., Lahnemann, D., Schulenburg, H., Ackermann, M. & Beardmore, R. The optimal deployment of synergistic antibiotics: a control-theoretic approach. *J R Soc Interface* **9**, 2488-2502, doi:10.1098/rsif.2012.0279 (2012).

52 Chait, R., Craney, A. & Kishony, R. Antibiotic interactions that select against resistance. *Nature* **446**, 668-671, doi:10.1038/nature05685 (2007).

53 Wong-Beringer, A., Joo, J., Tse, E. & Beringer, P. Vancomycin-associated nephrotoxicity: a critical appraisal of risk with high-dose therapy. *Int J Antimicrob Agents* **37**, 95-101, doi:10.1016/j.ijantimicag.2010.10.013 (2011).

54 Baym, M., Stone, L. K. & Kishony, R. Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* **351**, aad3292, doi:10.1126/science.aad3292 (2016).

55 Tremblay, S., Lau, T. T. & Ensom, M. H. Addition of rifampin to vancomycin for methicillin-resistant Staphylococcus aureus infections: what is the evidence? *Ann Pharmacother* **47**, 1045-1054, doi:10.1345/aph.1R726 (2013).

56 Gould, I. M. *et al.* Management of serious meticillin-resistant Staphylococcus aureus infections: what are the limits? *Int J Antimicrob Agents* **37**, 202-209, doi:10.1016/j.ijantimicag.2010.10.030 (2011).

57 Hiramatsu, K. *et al.* Multi-drug-resistant Staphylococcus aureus and future chemotherapy. *J Infect Chemother* **20**, 593-601, doi:10.1016/j.jiac.2014.08.001 (2014).

58 Holten, K. B. & Onusko, E. M. Appropriate prescribing of oral beta-lactam antibiotics. *Am Fam Physician* **62**, 611-620 (2000).

59 Baron, S. Medical Microbiology. 4th Edition. *Univ. of Texas Medical Branch* **4th edition** (1996).

60 Boswihi, S. a. U., E. Methicillin-resistant Staphylococcus aureus: An update on the epidemiology, treatment options and infection control. *Current Medicine Research and Practice* **8**, 18-24 (2018).

61 Klevens, R. M. *et al.* Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA* **298**, 1763-1771, doi:10.1001/jama.298.15.1763 (2007).

62 World Health Organization. Antimicrobial Resistance - Global Report on Surveillance *World Health Organization* (2014).

63 Reimann, H. A. Infectious diseases. Annual review of significant publications. *Postgrad Med J* **45**, 428-445 (1969).

64 Hessling, B. *et al.* Global proteome analysis of vancomycin stress in Staphylococcus aureus. *Int J Med Microbiol* **303**, 624-634, doi:10.1016/j.ijmm.2013.08.014 (2013).

65 Sievert, D. M. *et al.* Vancomycin-resistant Staphylococcus aureus in the United States, 2002-2006. *Clin Infect Dis* **46**, 668-674, doi:10.1086/527392 (2008).

66 Howden, B. P., Davies, J. K., Johnson, P. D., Stinear, T. P. & Grayson, M. L. Reduced vancomycin susceptibility in Staphylococcus aureus, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. *Clin Microbiol Rev* **23**, 99-139, doi:10.1128/CMR.00042-09 (2010).

67 Falagas, M. E., Bliziotis, I. A. & Fragoulis, K. N. Oral rifampin for eradication of Staphylococcus aureus carriage from healthy and sick populations: a systematic review of the evidence from comparative trials. *Am J Infect Control* **35**, 106-114, doi:10.1016/j.ajic.2006.09.005 (2007).

68 Pazlarová, J. P., S.; BaBuliková, J.; and Demnerová, K. Effects of Ampicillin and Vancomycin on Staphylococcus aureus Biofilms. *Czech J. Food Sci.* **32**, 137-144 (2014).

69 Peterson, L. R., Gerding, D. N., Hall, W. H. & Schierl, E. A. Medium-dependent variation in bactericidal activity of antibiotics against susceptible Staphylococcus aureus. *Antimicrob Agents Chemother* **13**, 665-668 (1978).

70 Faville, R. J., Jr. *et al.* Staphylococcus aureus endocarditis. Combined therapy with vancomycin and rifampin. *JAMA* **240**, 1963-1965 (1978).

71 Bayer, A. S. & Morrison, J. O. Disparity between timed-kill and checkerboard methods for determination of in vitro bactericidal interactions of vancomycin plus rifampin versus methicillin-susceptible and -resistant Staphylococcus aureus. *Antimicrob Agents Chemother* **26**, 220-223 (1984).

72 Bayer, A. S. & Lam, K. Efficacy of vancomycin plus rifampin in experimental aortic-valve endocarditis due to methicillin-resistant Staphylococcus aureus: in vitro-in vivo correlations. *J Infect Dis* **151**, 157-165 (1985).

73 Van der Auwera, P. & Joly, P. Comparative in-vitro activities of teicoplanin, vancomycin, coumermycin and ciprofloxacin, alone and in combination with

rifampicin or LM 427, against Staphylococcus aureus. *J Antimicrob Chemother* **19**, 313-320 (1987).

74 Levine, D. P., Fromm, B. S. & Reddy, B. R. Slow response to vancomycin or vancomycin plus rifampin in methicillin-resistant Staphylococcus aureus endocarditis. *Ann Intern Med* **115**, 674-680 (1991).

75 Tan, T. Q., Mason, E. O., Jr., Ou, C. N. & Kaplan, S. L. Use of intravenous rifampin in neonates with persistent staphylococcal bacteremia. *Antimicrob Agents Chemother* **37**, 2401-2406 (1993).

76 Gang, R. K., Sanyal, S. C., Mokaddas, E. & Lari, A. R. Rifampicin as an adjunct to vancomycin therapy in MRSA septicaemia in burns. *Burns* **25**, 640-644 (1999).

77 Naval G. Daver, S. A. S., Robert L. Atmar, Thomas P. Giordano, Charles E. Stager, Charles A. Reitman, A. Clinton White Jr. Oral step-down therapy is comparable to intravenous therapy for Staphylococcus aureus osteomyelitis. *J Infect.* **54**, 539-544 (2007).

78 Riedel, D. J., Weekes, E. & Forrest, G. N. Addition of rifampin to standard therapy for treatment of native valve infective endocarditis caused by Staphylococcus aureus. *Antimicrob Agents Chemother* **52**, 2463-2467, doi:10.1128/AAC.00300-08 (2008).

79 Rose, W. E. & Poppens, P. T. Impact of biofilm on the in vitro activity of vancomycin alone and in combination with tigecycline and rifampicin against Staphylococcus aureus. *J Antimicrob Chemother* **63**, 485-488, doi:10.1093/jac/dkn513 (2009).

80 Jang, H. C. *et al.* Salvage treatment for persistent methicillin-resistant Staphylococcus aureus bacteremia: efficacy of linezolid with or without carbapenem. *Clin Infect Dis* **49**, 395-401, doi:10.1086/600295 (2009).

81 Jung, Y. J. *et al.* Effect of vancomycin plus rifampicin in the treatment of nosocomial methicillin-resistant Staphylococcus aureus pneumonia. *Crit Care Med* **38**, 175-180 (2010).

82 Liu, C. *et al.* Clinical practice guidelines by the infectious diseases society of america for the treatment of methicillin-resistant Staphylococcus aureus infections in adults and children. *Clin Infect Dis* **52**, e18-55, doi:10.1093/cid/ciq146 (2011).

83 Gould, F. K. *et al.* Guidelines (2008) for the prophylaxis and treatment of methicillin-resistant Staphylococcus aureus (MRSA) infections in the United Kingdom. *J Antimicrob Chemother* **63**, 849-861, doi:10.1093/jac/dkp065 (2009).

84 Schweizer, M. L. *et al.* Optimizing antimicrobial prescribing: Are clinicians following national trends in methicillin-resistant staphylococcus aureus (MRSA) infections rather than local data when treating MRSA wound infections. *Antimicrob Resist Infect Control* **2**, 28, doi:10.1186/2047-2994-2-28 (2013).

85 Watson, B. M. Pharmacogenetics. A review of current literature. *Anaesthesia* **24**, 230-248 (1969).

86 Karczewski, K. J., Daneshjou, R. & Altman, R. B. Chapter 7: Pharmacogenomics. *PLoS Comput Biol* **8**, e1002817, doi:10.1371/journal.pcbi.1002817 (2012).

87 Lam, S. W. Medication Administration/Critical Care Research - Pharmacokinetics, Pharmacodynamics, and Pharmacogenomic. *American College of Clinical Pharmacy* **Critical Care Self-Assessment Program (CCSAP)**, 7-25 (2016).

88 Ehrlich, P. Address in Pathology, ON CHEMIOTHERAPY: Delivered before the Seventeenth International Congress of Medicine. *Br Med J* **2**, 353-359 (1913).

89 Fleming, A. *Penicillin*, <https://www.nobelprize.org/uploads/2018/06/fleming-lecture.pdf> (1945).

90 Murugan, R. Movement towards personalised medicine in the ICU. *Lancet Respir Med* **3**, 10-12, doi:10.1016/S2213-2600(14)70310-8 (2015).

91 Wilkinson, G. R. Drug metabolism and variability among patients in drug response. *N Engl J Med* **352**, 2211-2221, doi:10.1056/NEJMra032424 (2005).

92 Shaw, A. C., Goldstein, D. R. & Montgomery, R. R. Age-dependent dysregulation of innate immunity. *Nat Rev Immunol* **13**, 875-887, doi:10.1038/nri3547 (2013).

93 Klotz, U. Pharmacokinetics and drug metabolism in the elderly. *Drug Metab Rev* **41**, 67-76, doi:10.1080/03602530902722679 (2009).

94 Pea, F. Pharmacokinetics and drug metabolism of antibiotics in the elderly. *Expert Opin Drug Metab Toxicol*, doi:10.1080/17425255.2018.1528226 (2018).

95 Pinsonneault, J. & Sadee, W. Pharmacogenomics of multigenic diseases: sex-specific differences in disease and treatment outcome. *AAPS PharmSci* **5**, E29, doi:10.1208/ps050429 (2003).

96 Acocella, G. & Angel, J. H. Short-course chemotherapy of pulmonary tuberculosis: a new approach to drug dosage in the initial intensive phase. *Am Rev Respir Dis* **134**, 1283-1286, doi:10.1164/arrd.1986.134.5.1283 (1986).

97 Felton, T. W., Hope, W. W. & Roberts, J. A. How severe is antibiotic pharmacokinetic variability in critically ill patients and what can be done about it? *Diagn Microbiol Infect Dis* **79**, 441-447, doi:10.1016/j.diagmicrobio.2014.04.007 (2014).

98 Belle, D. J. & Singh, H. Genetic factors in drug metabolism. *Am Fam Physician* **77**, 1553-1560 (2008).

99 Candon, S. *et al.* Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. *PLoS One* **10**, e0125448, doi:10.1371/journal.pone.0125448 (2015).

100 Berthelot, P. Mechanisms and prediction of drug-induced liver disease. *Gut* **14**, 332-339 (1973).

101 Zanger, U. M. & Schwab, M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* **138**, 103-141, doi:10.1016/j.pharmthera.2012.12.007 (2013).

102 van Ingen, J. A., R.E.; Donald, P.R.; Diacon, A.H.; Dawson, R.; Plemper van Balen, G.; Gillespie, S.H.; Boeree, M.J. Why Do We Use 600 mg of Rifampicin in Tuberculosis Treatment? *Clinical Infectious Diseases* **52**, e194–e199 (2011).

103 NICE antimicrobial stewardship: right drug, dose, and time? *Lancet* **386**, 717, doi:10.1016/S0140-6736(15)61522-7 (2015).

104 Weinshilboum, R. Inheritance and drug response. *N Engl J Med* **348**, 529-537, doi:10.1056/NEJMra020021 (2003).

105 Anzenbacher, P. Z., U. M. *Metabolism of drugs and other xenobiotics*. (Wiley-VCH ; John Wiley distributor, 2012).

106 Evans, W. E. & Relling, M. V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487-491 (1999).

107 Tanaka, E. Clinically important pharmacokinetic drug-drug interactions: role of cytochrome P450 enzymes. *J Clin Pharm Ther* **23**, 403-416 (1998).

108 Anzenbacher, P. & Zanger, U. M. *Metabolism of drugs and other xenobiotics*. (Wiley-VCH ; John Wiley distributor, 2012).

109 Yan-hui Yang, H. A., Daigo Inoyama, Yu Chen, Long-qin Hu. Enzyme-mediated hydrolytic activation of prodrugs. *Acta Pharmaceutica Sinica B* **1**, 143-159 (2011).

110 Fakunle, E. S. & Loring, J. F. Ethnically diverse pluripotent stem cells for drug development. *Trends Mol Med* **18**, 709-716, doi:10.1016/j.molmed.2012.10.007 (2012).

111 Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).

112 Spielman, R. S. *et al.* Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* **39**, 226-231, doi:10.1038/ng1955 (2007).

113 Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol Sci* **25**, 193-200, doi:10.1016/j.tips.2004.02.007 (2004).

114 Takahashi, H. & Echizen, H. Pharmacogenetics of CYP2C9 and interindividual variability in anticoagulant response to warfarin. *Pharmacogenomics J* **3**, 202-214, doi:10.1038/sj.tpj.6500182 (2003).

115 Nebert, D. W. & Vasiliou, V. Analysis of the glutathione S-transferase (GST) gene family. *Hum Genomics* **1**, 460-464 (2004).

116 Piacentini, S. *et al.* GSTT1 and GSTM1 gene polymorphisms in European and African populations. *Mol Biol Rep* **38**, 1225-1230, doi:10.1007/s11033-010-0221-0 (2011).

117 Leiro, V. *et al.* Influence of glutathione S-transferase M1 and T1 homozygous null mutations on the risk of antituberculosis drug-induced hepatotoxicity in a Caucasian population. *Liver Int* **28**, 835-839, doi:10.1111/j.1478-3231.2008.01700.x (2008).

118 Matimba, A., Del-Favero, J., Van Broeckhoven, C. & Masimirembwa, C. Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects. *Hum Genomics* **3**, 169-190 (2009).

119 Lynch, T. & Price, A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am Fam Physician* **76**, 391-396 (2007).

120 Preissner, S. C. *et al.* Polymorphic cytochrome P450 enzymes (CYPs) and their role in personalized therapy. *PLoS One* **8**, e82562, doi:10.1371/journal.pone.0082562 (2013).

121 Mizutani, T. PM frequencies of major CYPs in Asians and Caucasians. *Drug Metab Rev* **35**, 99-106, doi:10.1081/DMR-120023681 (2003).

122 Tsai, D., Jamal, J. A., Davis, J. S., Lipman, J. & Roberts, J. A. Interethnic differences in pharmacokinetics of antibacterials. *Clin Pharmacokinet* **54**, 243-260, doi:10.1007/s40262-014-0209-3 (2015).

123 Gaedigk, A. Interethnic differences of drug-metabolizing enzymes. *Int J Clin Pharmacol Ther* **38**, 61-68 (2000).

124 Sabbagh, A., Darlu, P., Crouau-Roy, B. & Poloni, E. S. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. *PLoS One* **6**, e18507, doi:10.1371/journal.pone.0018507 (2011).

125 Launay-Vacher, V., Izzedine, H., Mercadal, L. & Deray, G. Clinical review: use of vancomycin in haemodialysis patients. *Crit Care* **6**, 313-316 (2002).

126 Chaijamorn, W. & Wanakamanee, U. Pharmacokinetics of vancomycin in critically ill patients undergoing continuous venovenous haemodialysis. *Int J Antimicrob Agents* **44**, 367-368, doi:10.1016/j.ijantimicag.2014.06.002 (2014).

127 Pea, F., Viale, P. & Furlanut, M. Antimicrobial therapy in critically ill patients: a review of pathophysiological conditions responsible for altered disposition and pharmacokinetic variability. *Clin Pharmacokinet* **44**, 1009-1034, doi:10.2165/00003088-200544100-00002 (2005).

128 Acocella, G. Pharmacokinetics and metabolism of rifampin in humans. *Rev Infect Dis* **5 Suppl 3**, S428-432 (1983).

129 Inouye, B., Uchinomi, Y. & Wachi, T. In vitro effect of rifampicin and its derivatives on energy transfer reactions in mitochondria. *J Antibiot (Tokyo)* **27**, 192-198 (1974).

130 Acocella, G. *et al.* Kinetics of rifampicin and isoniazid administered alone and in combination to normal subjects and patients with liver disease. *Gut* **13**, 47-53 (1972).

131 Nitti, V., Ninni, A., Meola, G., Iuliano, A. & Curci, G. Comparative investigations of the enzyme-inducing activity of rifampicin and barbiturates in man. *Chemotherapy* **19**, 206-210, doi:10.1159/000221456 (1973).

132 Capelle, P., Dhumeaux, D., Mora, M., Feldmann, G. & Berthelot, P. Effect of rifampicin on liver function in man. *Gut* **13**, 366-371 (1972).

133 Acocella, G. Clinical pharmacokinetics of rifampicin. *Clin Pharmacokinet* **3**, 108-127 (1978).

134 Acocella, G. *et al.* Comparative bioavailability of isoniazid, rifampin, and pyrazinamide administered in free combination and in a fixed triple formulation designed for daily use in antituberculosis chemotherapy. II. Two-month, daily administration study. *Am Rev Respir Dis* **138**, 886-890, doi:10.1164/ajrccm/138.4.886 (1988).

135 Acocella, G., Conti, R., Luisetti, M., Pozzi, E. & Grassi, C. Pharmacokinetic studies on antituberculosis regimens in humans. I. Absorption and metabolism of the compounds used in the initial intensive phase of the short-course regimens: single administration study. *Am Rev Respir Dis* **132**, 510-515, doi:10.1164/arrd.1985.132.3.510 (1985).

136 Furesz, S. Recent investigations on the biological properties of rifampicin. *Acta Tuberc Pneumol Belg* **60**, 266-269 (1969).

137 Zilly, W., Breimer, D. D. & Richter, E. Pharmacokinetic interactions with rifampicin. *Clin Pharmacokinet* **2**, 61-70 (1977).

138 Song, S. H. *et al.* Relationship between CES2 genetic variations and rifampicin metabolism. *J Antimicrob Chemother* **68**, 1281-1284, doi:10.1093/jac/dkt036 (2013).

139 Wang, D. *et al.* Human carboxylesterases: a comprehensive review. *Acta Pharm Sin B* **8**, 699-712, doi:10.1016/j.apsb.2018.05.005 (2018).

140 Jamis-Dow, C. A., Katki, A. G., Collins, J. M. & Klecker, R. W. Rifampin and rifabutin and their metabolism by human liver esterases. *Xenobiotica* **27**, 1015-1024, doi:10.1080/004982597239994 (1997).

141 Tong, S. Y., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G., Jr. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* **28**, 603-661, doi:10.1128/CMR.00134-14 (2015).

142 Wu, M. H. *et al.* Determination and analysis of single nucleotide polymorphisms and haplotype structure of the human carboxylesterase 2 gene. *Pharmacogenetics* **14**, 595-605 (2004).

143 Dompreh, A. *et al.* Effect of Genetic Variation of NAT2 on Isoniazid and SLCO1B1 and CES2 on Rifampin Pharmacokinetics in Ghanaian Children with Tuberculosis. *Antimicrob Agents Chemother* **62**, doi:10.1128/AAC.02099-17 (2018).

144 Kubo, T. *et al.* Functional characterization of three naturally occurring single nucleotide polymorphisms in the CES2 gene encoding carboxylesterase 2 (HCE-2). *Drug Metab Dispos* **33**, 1482-1487, doi:10.1124/dmd.105.005587 (2005).

145 Kim, S. R. *et al.* Haplotypes and a novel defective allele of CES2 found in a Japanese population. *Drug Metab Dispos* **35**, 1865-1872, doi:10.1124/dmd.107.015339 (2007).

146 Merali, Z., Ross, S. & Pare, G. The pharmacogenetics of carboxylesterases: CES1 and CES2 genetic variants and their clinical effect. *Drug Metabol Drug Interact* **29**, 143-151, doi:10.1515/dmdi-2014-0009 (2014).

147 Marsh, S. *et al.* Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* **84**, 661-668, doi:10.1016/j.ygeno.2004.07.008 (2004).

148 Kim, S. R. *et al.* Twelve novel single nucleotide polymorphisms in the CES2 gene encoding human carboxylesterase 2 (hCE-2). *Drug Metab Pharmacokinet* **18**, 327-332 (2003).

149 Fujiyama, N. *et al.* Involvement of carboxylesterase 1 and 2 in the hydrolysis of mycophenolate mofetil. *Drug Metab Dispos* **38**, 2210-2217, doi:10.1124/dmd.110.034249 (2010).

150 Satoh, T. *et al.* Current progress on esterases: from molecular structure to function. *Drug Metab Dispos* **30**, 488-493 (2002).

151 Cooper, M. *et al.* Enteric-coated mycophenolate sodium immunosuppression in renal transplant patients: efficacy and dosing. *Transplant Rev (Orlando)* **26**, 233-240, doi:10.1016/j.trre.2012.02.001 (2012).

152 Khosroshahi, H. T., Shoja, M. M., Peyrovifar, A., Hashemi, S. R. & Amjadi, M. Mycophenolate mofetil dose reduction in renal transplant recipients: a 5-year follow-up study. *Transplant Proc* **41**, 2797-2799, doi:10.1016/j.transproceed.2009.07.036 (2009).

153 Sathyapalan, T., Atkin, S. L. & Kilpatrick, E. S. Disparate effects of atorvastatin compared with simvastatin on C-reactive protein concentrations in patients with type 2 diabetes. *Diabetes Care* **33**, 1948-1950, doi:10.2337/dc10-0201 (2010).

154 Andrews, L. M. *et al.* Pharmacokinetic considerations related to therapeutic drug monitoring of tacrolimus in kidney transplant patients. *Expert Opin Drug Metab Toxicol* **13**, 1225-1236, doi:10.1080/17425255.2017.1395413 (2017).

155 Fuentes-Hernandez, A. *et al.* Using a sequential regimen to eliminate bacteria at sublethal antibiotic dosages. *PLoS Biol* **13**, e1002104, doi:10.1371/journal.pbio.1002104 (2015).

156 Beardmore, R. E., Pena-Miller, R., Gori, F. & Iredell, J. Antibiotic Cycling and Antibiotic Mixing: Which One Best Mitigates Antibiotic Resistance? *Mol Biol Evol* **34**, 802-817, doi:10.1093/molbev/msw292 (2017).

157 Meyer, J. R. G., I.; Beardmore, R. Biophysical mechanisms that maintain biodiversity through trade-offs. *Nat. Commun.* **6** (2015).

158 Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298, doi:10.1038/nrg775 (2002).

159 Majewski, J. & Ott, J. Distribution and characterization of regulatory elements in the human genome. *Genome Res* **12**, 1827-1836, doi:10.1101/gr.606402 (2002).

160 Liu, C. *et al.* Effects of genetic variations on microRNA: target interactions. *Nucleic Acids Res* **42**, 9543-9552, doi:10.1093/nar/gku675 (2014).

161 Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell* **22**, 769-781, doi:10.1016/j.molcel.2006.05.008 (2006).

162 Fang, Z. & Rajewsky, N. The impact of miRNA target sites in coding sequences and in 3'UTRs. *PLoS One* **6**, e18067, doi:10.1371/journal.pone.0018067 (2011).

163 Wu, Y., Zhang, Y. & Zhang, J. Distribution of exonic splicing enhancer elements in human genes. *Genomics* **86**, 329-336, doi:10.1016/j.ygeno.2005.05.011 (2005).

164 Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1-30, doi:10.1016/j.gene.2012.07.083 (2013).

165 Chorev, M. & Carmel, L. The function of introns. *Front Genet* **3**, 55, doi:10.3389/fgene.2012.00055 (2012).

166 Woolfe, A., Mullikin, J. C. & Elnitski, L. Genomic features defining exonic variants that modulate splicing. *Genome Biol* **11**, R20, doi:10.1186/gb-2010-11-2-r20 (2010).

167 Havlioglu, N. *et al.* An intronic signal for alternative splicing in the human genome. *PLoS One* **2**, e1246, doi:10.1371/journal.pone.0001246 (2007).

168 Hube, F. & Francastel, C. Mammalian introns: when the junk generates molecular diversity. *Int J Mol Sci* **16**, 4429-4452, doi:10.3390/ijms16034429 (2015).

169 Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**, 291-323, doi:10.1146/annurev-biochem-060614-034316 (2015).

170 Yeo, G., Hoon, S., Venkatesh, B. & Burge, C. B. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc Natl Acad Sci U S A* **101**, 15700-15705, doi:10.1073/pnas.0404901101 (2004).

171 Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414 e1324, doi:10.1016/j.cell.2016.10.026 (2016).

172 Zaina, S., Perez-Luque, E. L. & Lund, G. Genetics talks to epigenetics? The interplay between sequence variants and chromatin structure. *Curr Genomics* **11**, 359-367, doi:10.2174/138920210791616662 (2010).

173 Zanger, U. M. *et al.* Genetics, epigenetics, and regulation of drug-metabolizing cytochrome p450 enzymes. *Clin Pharmacol Ther* **95**, 258-261, doi:10.1038/clpt.2013.220 (2014).

174 Sadee, W. & Dai, Z. Pharmacogenetics/genomics and personalized medicine. *Hum Mol Genet* **14 Spec No. 2**, R207-214, doi:10.1093/hmg/ddi261 (2005).

175 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).

176 Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-485, doi:10.1038/nrg2361 (2008).

177 Lakiotaki, K. *et al.* Exploring public genomics data for population pharmacogenomics. *PLoS One* **12**, e0182138, doi:10.1371/journal.pone.0182138 (2017).

178 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

179 Loeuillet, C. *et al.* Promoter polymorphisms and allelic imbalance in ABCB1 expression. *Pharmacogenet Genomics* **17**, 951-959, doi:10.1097/FPC.0b013e3282eff934 (2007).

180 Locke, J. M., Hysenaj, G., Wood, A. R., Weedon, M. N. & Harries, L. W. Targeted allelic expression profiling in human islets identifies cis-regulatory effects for multiple variants identified by type 2 diabetes genome-wide association studies. *Diabetes* **64**, 1484-1491, doi:10.2337/db14-0957 (2015).

181 O'Neill, J. Tackling Drug-Resistant Infections Globally: final report and recommendations. *London: The Review on Antimicrobial Resistance.* (2016).

182 Organization, W. H. Antibacterial agents in clinical development – an analysis of the antibacterial clinical development pipeline, including Mycobacterium tuberculosis. *World Health Organization* (2017).

183 Centers for Disease Control and Prevention. ANTIBIOTIC RESISTANCE THREATS in the United States, 2013. (2013).

184 World Health Organization. *Antimicrobial resistance - Key facts*, <http://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance> (2018).

185 Samar S. Boswihi, E. E. U. Methicillin-resistant Staphylococcus aureus: An update on the epidemiology, treatment options and infection control. *Current Medicine Research and Practice* **8**, 18-24 (2018).

186 White, R. L., Burgess, D. S., Manduru, M. & Bosso, J. A. Comparison of three different in vitro methods of detecting synergy: time-kill, checkerboard, and E test. *Antimicrob Agents Chemother* **40**, 1914-1918 (1996).

187 Orhan, G., Bayram, A., Zer, Y. & Balci, I. Synergy tests by E test and checkerboard methods of antimicrobial combinations against Brucella melitensis. *J Clin Microbiol* **43**, 140-143, doi:10.1128/JCM.43.1.140-143.2005 (2005).

188 Varela, N. P., Friendship, R., Dewey, C. & Valdivieso, A. Comparison of agar dilution and E-test for antimicrobial susceptibility testing of Campylobacter coil

isolates recovered from 80 Ontario swine farms. *Can J Vet Res* **72**, 168-174 (2008).

189 Cetin, E. S., Tekeli, A., Ozseven, A. G., Us, E. & Aridogan, B. C. Determination of in vitro activities of polymyxin B and rifampin in combination with ampicillin/sulbactam or cefoperazone/sulbactam against multidrug-resistant Acinetobacter baumannii by the E-test and checkerboard methods. *Jpn J Infect Dis* **66**, 463-468 (2013).

190 Lewis, R. E., Diekema, D. J., Messer, S. A., Pfaller, M. A. & Klepser, M. E. Comparison of Etest, chequerboard dilution and time-kill studies for the detection of synergy or antagonism between antifungal agents tested against Candida species. *J Antimicrob Chemother* **49**, 345-351 (2002).

191 Lepe, J. A., Dominguez-Herrera, J., Pachon, J. & Aznar, J. Determining accurate vancomycin MIC values for methicillin-resistant Staphylococcus aureus by the microdilution method. *J Antimicrob Chemother* **69**, 136-138, doi:10.1093/jac/dkt308 (2014).

192 Rybak, M. J. *et al.* Evaluation of vancomycin susceptibility testing for methicillin-resistant Staphylococcus aureus: comparison of Etest and three automated testing methods. *J Clin Microbiol* **51**, 2077-2081, doi:10.1128/JCM.00448-13 (2013).

193 Macias, E. A., Mason, E. O., Jr., Ocera, H. Y. & LaRocco, M. T. Comparison of E test with standard broth microdilution for determining antibiotic susceptibilities of penicillin-resistant strains of Streptococcus pneumoniae. *J Clin Microbiol* **32**, 430-432 (1994).

194 Rosser, S. J., Alfa, M. J., Hoban, S., Kennedy, J. & Harding, G. K. E test versus agar dilution for antimicrobial susceptibility testing of viridans group streptococci. *J Clin Microbiol* **37**, 26-30 (1999).

195 Riedel, T. E., Berelson, W. M., Nealson, K. H. & Finkel, S. E. Oxygen consumption rates of bacteria under nutrient-limited conditions. *Appl Environ Microbiol* **79**, 4921-4931, doi:10.1128/AEM.00756-13 (2013).

196 Kram, K. E. & Finkel, S. E. Culture volume and vessel affect long-term survival, mutation frequency, and oxidative stress of Escherichia coli. *Appl Environ Microbiol* **80**, 1732-1738, doi:10.1128/AEM.03150-13 (2014).

197 Roberts, A. E., Kragh, K. N., Bjarnsholt, T. & Diggle, S. P. The Limitations of In Vitro Experimentation in Understanding Biofilms and Chronic Infection. *J Mol Biol* **427**, 3646-3661, doi:10.1016/j.jmb.2015.09.002 (2015).

198 Zhou, A. *et al.* Synergistic interactions of vancomycin with different antibiotics against Escherichia coli: trimethoprim and nitrofurantoin display strong synergies with vancomycin against wild-type E. coli. *Antimicrob Agents Chemother* **59**, 276-281, doi:10.1128/AAC.03502-14 (2015).

199 Holden, M. T. *et al.* Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A* **101**, 9786-9791, doi:10.1073/pnas.0402521101 (2004).

200 O'Toole, G. A. Microtiter Dish Biofilm Formation Assay. *J Vis Exp* **47** (2011).

201 Steenackers, H. P., Parijs, I., Dubey, A., Foster, K. R. & Vanderleyden, J. Experimental evolution in biofilm populations. *FEMS Microbiol Rev* **40**, 373-397, doi:10.1093/femsre/fuw002 (2016).

202 Fuchs, S., Pane-Farre, J., Kohler, C., Hecker, M. & Engelmann, S. Anaerobic gene expression in Staphylococcus aureus. *J Bacteriol* **189**, 4275-4289, doi:10.1128/JB.00081-07 (2007).

203 Unden, G., Becker, S., Bongaerts, J., Schirawski, J. & Six, S. Oxygen regulated gene expression in facultatively anaerobic bacteria. *Antonie Van Leeuwenhoek* **66**, 3-22 (1994).

204 Metzenberg R. L., a. J. S. G. Disruption of essential genes in Neurospora by RIP. *Fungal Genet Newsletter* **39**, 37-49 (1992).

205 Szybalski, W. & Bryson, V. Genetic studies on microbial cross resistance to toxic agents. I. Cross resistance of Escherichia coli to fifteen antibiotics. *J Bacteriol* **64**, 489-499 (1952).

206 Liu, Y. *et al.* Accurate assessment of antibiotic susceptibility and screening resistant strains of a bacterial population by linear gradient plate. *Sci China Life Sci* **54**, 953-960, doi:10.1007/s11427-011-4230-6 (2011).

207 Mann, C. M. & Markham, J. L. A new method for determining the minimum inhibitory concentration of essential oils. *J Appl Microbiol* **84**, 538-544 (1998).

208 Moreno, S., Scheyer, T., Romano, C. S. & Vojnov, A. A. Antioxidant and antimicrobial activities of rosemary extracts linked to their polyphenol composition. *Free Radic Res* **40**, 223-231, doi:10.1080/10715760500473834 (2006).

209 Klancnik, A., Piskernik, S., Jersek, B. & Mozina, S. S. Evaluation of diffusion and dilution methods to determine the antibacterial activity of plant extracts. *J Microbiol Methods* **81**, 121-126, doi:10.1016/j.mimet.2010.02.004 (2010).

210 Barbee, L. A., Soge, O. O., Holmes, K. K. & Golden, M. R. In vitro synergy testing of novel antimicrobial combination therapies against Neisseria gonorrhoeae. *J Antimicrob Chemother* **69**, 1572-1578, doi:10.1093/jac/dkt540 (2014).

211 Sopirala, M. M. *et al.* Synergy testing by Etest, microdilution checkerboard, and time-kill methods for pan-drug-resistant Acinetobacter baumannii. *Antimicrob Agents Chemother* **54**, 4678-4683, doi:10.1128/AAC.00497-10 (2010).

212 Manno, G. *et al.* Use of the E test to assess synergy of antibiotic combinations against isolates of Burkholderia cepacia-complex from patients with cystic fibrosis. *Eur J Clin Microbiol Infect Dis* **22**, 28-34, doi:10.1007/s10096-002-0852-8 (2003).

213 Irith Wiegand, K. H. a. R. E. W. H. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols* **3**, 163-175 (2008).

214 Syal, K. *et al.* Current and emerging techniques for antibiotic susceptibility tests. *Theranostics* **7**, 1795-1805, doi:10.7150/thno.19217 (2017).

215 Jarmila Pazlarová, S. P., Jana BaBuliková and Kateřina Demnerová. Effects of Ampicillin and Vancomycin on Staphylococcus aureus Biofilms. *Czech J. Food Sci.* **32**, 137-144 (2014).

216 Sutter, V. L., Sugihara, P. T. & Finegold, S. M. Rifampin-blood-agar as a selective medium for the isolation of certain anaerobic bacteria. *Appl Microbiol* **22**, 777-780 (1971).

217 Yoo, J. H., Choi, N. Y., Bae, Y. M., Lee, J. S. & Lee, S. Y. Development of a selective agar plate for the detection of Campylobacter spp. in fresh produce. *Int J Food Microbiol* **189**, 67-74, doi:10.1016/j.ijfoodmicro.2014.07.032 (2014).

218 O'Neill, A. J., Cove, J. H. & Chopra, I. Mutation frequencies for resistance to fusidic acid and rifampicin in Staphylococcus aureus. *J Antimicrob Chemother* **47**, 647-650 (2001).

219 Van Horn, K. G., Gedris, C. A., Rodney, K. M. & Mitchell, J. B. Evaluation of commercial vancomycin agar screen plates for detection of vancomycin-resistant enterococci. *J Clin Microbiol* **34**, 2042-2044 (1996).

220 Hunt, D. E. & Sandham, H. J. Improved agar gradient-plate technique. *Appl Microbiol* **17**, 329-330 (1969).

221 Jorgensen, J. H. & Ferraro, M. J. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis* **49**, 1749-1755, doi:10.1086/647952 (2009).

222 Skulnick, M. *et al.* Evaluation of accuracy and reproducibility of E test for susceptibility testing of Streptococcus pneumoniae to penicillin, cefotaxime, and ceftriaxone. *J Clin Microbiol* **33**, 2334-2337 (1995).

223 Reiter, K. C. *et al.* Rifampicin fails to eradicate mature biofilm formed by methicillin-resistant Staphylococcus aureus. *Rev Soc Bras Med Trop* **45**, 471-474 (2012).

224 Bell, A. M., S.M. Assessing synergy or antagonism for drug combinations. *Nature Review Drug Discoveries* **3**, 982 (2004).

225 Doern, G. V. B., S.M. The Clinical Predictive Value (or Lack Thereof) of the Results of In Vitro Antimicrobial Susceptibility Tests. *Proceedings of Camp Clin Micro* **49**, S11-14 (2011).

226 Bonapace, C. R., Friedrich, L. V., Bosso, J. A. & White, R. L. Determination of antibiotic effect in an in vitro pharmacodynamic model: comparison with an established animal model of infection. *Antimicrob Agents Chemother* **46**, 3574-3579 (2002).

227 Baietto, L. *et al.* A 30-years review on pharmacokinetics of antibiotics: is the right time for pharmacogenetics? *Curr Drug Metab* **15**, 581-598 (2014).

228 Hu, Y. F. *et al.* Association of CYP3A4*18B polymorphisms with the pharmacokinetics of cyclosporine in healthy subjects. *Xenobiotica* **37**, 315-327, doi:10.1080/00498250601149206 (2007).

229 Meyer, M. R., Schutz, A. & Maurer, H. H. Contribution of human esterases to the metabolism of selected drugs of abuse. *Toxicol Lett* **232**, 159-166, doi:10.1016/j.toxlet.2014.10.026 (2015).

230 Lepper, E. R. *et al.* Effect of common CYP3A4 and CYP3A5 variants on the pharmacokinetics of the cytochrome P450 3A phenotyping probe midazolam in cancer patients. *Clin Cancer Res* **11**, 7398-7404, doi:10.1158/1078-0432.CCR-05-0520 (2005).

231 Liu, S. *et al.* Effect of CYP3A4(*)1G and CYP3A5(*)3 Polymorphisms on Pharmacokinetics and Pharmacodynamics of Ticagrelor in Healthy Chinese Subjects. *Front Pharmacol* **8**, 176, doi:10.3389/fphar.2017.00176 (2017).

232 Richards-Waugh, L. L., Primerano, D. A., Dementieva, Y., Kraner, J. C. & Rankin, G. O. Fatal methadone toxicity: potential role of CYP3A4 genetic polymorphism. *J Anal Toxicol* **38**, 541-547, doi:10.1093/jat/bku091 (2014).

233 Fujiyama, N. *et al.* Influence of carboxylesterase 2 genetic polymorphisms on mycophenolic acid pharmacokinetics in Japanese renal transplant recipients. *Xenobiotica* **39**, 407-414, doi:10.1080/00498250902807338 (2009).

234 Wang, D., Guo, Y., Wrighton, S. A., Cooke, G. E. & Sadee, W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J* **11**, 274-286, doi:10.1038/tpj.2010.28 (2011).

235 Pouplin, T., Phuong, P. N., Toi, P. V., Nguyen Pouplin, J. & Farrar, J. Isoniazid, pyrazinamide and rifampicin content variation in split fixed-dose combination tablets. *PLoS One* **9**, e102047, doi:10.1371/journal.pone.0102047 (2014).

236 Bhutani, H., Mariappan, T. T. & Singh, S. The physical and chemical stability of anti-tuberculosis fixed-dose combination products under accelerated climatic conditions. *Int J Tuberc Lung Dis* **8**, 1073-1080 (2004).

237 Jung, Y. *et al.* Area under the concentration-time curve to minimum inhibitory concentration ratio as a predictor of vancomycin treatment outcome in methicillin-resistant Staphylococcus aureus bacteraemia. *Int J Antimicrob Agents* **43**, 179-183, doi:10.1016/j.ijantimicag.2013.10.017 (2014).

238 Song, K. H. *et al.* Impact of area under the concentration-time curve to minimum inhibitory concentration ratio on vancomycin treatment outcomes in methicillin-resistant Staphylococcus aureus bacteraemia. *Int J Antimicrob Agents* **46**, 689-695, doi:10.1016/j.ijantimicag.2015.09.010 (2015).

239 Falcon, R. *et al.* High vancomycin MICs within the susceptible range in Staphylococcus aureus bacteraemia isolates are associated with increased cell wall thickness and reduced intracellular killing by human phagocytes. *Int J Antimicrob Agents* **47**, 343-350, doi:10.1016/j.ijantimicag.2016.01.014 (2016).

240 Dheda, K. *et al.* Drug-Penetration Gradients Associated with Acquired Drug Resistance in Patients with Tuberculosis. *Am J Respir Crit Care Med* **198**, 1208-1219, doi:10.1164/rccm.201711-2333OC (2018).

241 Stiefel, P. R., U.; Schneider, J.; Mauerhofer, S.; Maniura-Weber, K.; Ren, Q. Is biofilm removal properly assessed? Comparison of different quantification methods in a 96-well plate system. *Appl Microbiol Biotechnol* **100**, 4135-4145 (2016).

242 Choi, J. *et al.* A rapid antimicrobial susceptibility test based on single-cell morphological analysis. *Sci Transl Med* **6**, 267ra174, doi:10.1126/scitranslmed.3009650 (2014).

243 Wang, X. *et al.* Heteroresistance at the single-cell level: adapting to antibiotic stress through a population-based strategy and growth-controlled interphenotypic coordination. *MBio* **5**, e00942-00913, doi:10.1128/mBio.00942-13 (2014).

244 Baltekin, O., Boucharin, A., Tano, E., Andersson, D. I. & Elf, J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. *Proc Natl Acad Sci U S A* **114**, 9170-9175, doi:10.1073/pnas.1708558114 (2017).

245 Ahn, E. & Park, T. Analysis of population-specific pharmacogenomic variants using next-generation sequencing data. *Sci Rep* **7**, 8416, doi:10.1038/s41598-017-08468-y (2017).

246 Lamba, J., Hebert, J. M., Schuetz, E. G., Klein, T. E. & Altman, R. B. PharmGKB summary: very important pharmacogene information for CYP3A5. *Pharmacogenet Genomics* **22**, 555-558, doi:10.1097/FPC.0b013e328351d47f (2012).

247 Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B. & Klein, T. E. PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdiscip Rev Syst Biol Med* **10**, e1417, doi:10.1002/wsbm.1417 (2018).

248 Emoto, C. & Iwasaki, K. Enzymatic characteristics of CYP3A5 and CYP3A4: a comparison of in vitro kinetic and drug-drug interaction patterns. *Xenobiotica* **36**, 219-233, doi:10.1080/00498250500489968 (2006).

249 Li, X., Song, X., Kamenecka, T. M. & Cameron, M. D. Discovery of a highly selective CYP3A4 inhibitor suitable for reaction phenotyping studies and differentiation of CYP3A4 and CYP3A5. *Drug Metab Dispos* **40**, 1803-1809, doi:10.1124/dmd.112.046144 (2012).

250 Levison, M. E. & Levison, J. H. Pharmacokinetics and pharmacodynamics of antibacterial agents. *Infect Dis Clin North Am* **23**, 791-815, vii, doi:10.1016/j.idc.2009.06.008 (2009).

251 Lolodi, O., Wang, Y. M., Wright, W. C. & Chen, T. Differential Regulation of CYP3A4 and CYP3A5 and its Implication in Drug Discovery. *Curr Drug Metab* **18**, 1095-1105, doi:10.2174/1389200218666170531112038 (2017).

252 PharmGKB. *PharmGKB Pharmagenomics Prescribing Information on CYP3A4*, <https://www.pharmgkb.org/gene/PA130/guideline> (

253 NCBI Variation Viewer. *Variations on the CYP3A4 gene*, <https://www.ncbi.nlm.nih.gov/variation/view/?q=CYP3A4> (2018).

254 Liu, R. *et al.* Associations of CYP3A4, NR1I2, CYP2C19 and P2RY12 polymorphisms with clopidogrel resistance in Chinese patients with ischemic stroke. *Acta Pharmacol Sin* **37**, 882-888, doi:10.1038/aps.2016.41 (2016).

255 Lane, S. *et al.* The population pharmacokinetics of R- and S-warfarin: effect of genetic and clinical factors. *Br J Clin Pharmacol* **73**, 66-76, doi:10.1111/j.1365-2125.2011.04051.x (2012).

256 Qiu, X. Y. *et al.* Association of MDR1, CYP3A4*18B, and CYP3A5*3 polymorphisms with cyclosporine pharmacokinetics in Chinese renal transplant recipients. *Eur J Clin Pharmacol* **64**, 1069-1084, doi:10.1007/s00228-008-0520-8 (2008).

257 Zhang, W. *et al.* CYP3A4*1G genetic polymorphism influences CYP3A activity and response to fentanyl in Chinese gynecologic patients. *Eur J Clin Pharmacol* **66**, 61-66, doi:10.1007/s00228-009-0726-4 (2010).

258 Mupeng Li, Y. H., Huilan Li, Zhipeng Wen, Xiaolei Hu, Daoyu Zhang, Yanjiao Zhang,Jian Xiao, Jie Tang and Xiaoping Chen. No Effect of SLCO1B1 and CYP3A4/5 Polymorphisms on the Pharmacokinetics and Pharmacodynamics of Ticagrelor in Healthy Chinese Male Subjects. *Biol. Pharm. Bull.* (2017).

259 Bertrand, J. *et al.* Multiple genetic variants predict steady-state nevirapine clearance in HIV-infected Cambodians. *Pharmacogenet Genomics* **22**, 868-876, doi:10.1097/FPC.0b013e32835a5af2 (2012).

260 Levran, O. *et al.* CYP2B6 SNPs are associated with methadone dose required for effective treatment of opioid addiction. *Addict Biol* **18**, 709-716, doi:10.1111/j.1369-1600.2011.00349.x (2013).

261 Gao Y, Z. L., Fu Q. CYP3A4*1G polymorphism is associated with lipid-lowering efficacy of atorvastatin but not of simvastatin. *Eur J Clin Pharmacol.* **64**, 877–882 (2008).

262 Wacher, V. J., Wu, C. Y. & Benet, L. Z. Overlapping substrate specificities and tissue distribution of cytochrome P450 3A and P-glycoprotein: implications for drug delivery and activity in cancer chemotherapy. *Mol Carcinog* **13**, 129-134 (1995).

263 Zhang, Y., Guo, X., Lin, E. T. & Benet, L. Z. Overlapping substrate specificities of cytochrome P450 3A and P-glycoprotein for a novel cysteine protease inhibitor. *Drug Metab Dispos* **26**, 360-366 (1998).

264 Boros, E. *et al.* Overlapping Specificity of Duplicated Human Pancreatic Elastase 3 Isoforms and Archetypal Porcine Elastase 1 Provides Clues to Evolution of Digestive Enzymes. *J Biol Chem* **292**, 2690-2702, doi:10.1074/jbc.M116.770560 (2017).

265 Crofts, N., Sugimoto, K., Oitome, N. F., Nakamura, Y. & Fujita, N. Differences in specificity and compensatory functions among three major starch synthases determine the structure of amylopectin in rice endosperm. *Plant Mol Biol* **94**, 399-417, doi:10.1007/s11103-017-0614-8 (2017).

266 PharmVar. *CYP3A4 allele nomenclature* <https://www.pharmvar.org/gene/CYP3A4> (

267 Li, H., Chen, D. & Zhang, J. Analysis of intron sequence features associated with transcriptional regulation in human genes. *PLoS One* **7**, e46784, doi:10.1371/journal.pone.0046784 (2012).

268 Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J. & Fairbrother, W. G. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A* **108**, 11093-11098, doi:10.1073/pnas.1101135108 (2011).

269 McCarthy, E. M. & Phillips, J. A., 3rd. Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. *Hum Mol Genet* **7**, 1491-1496 (1998).

270 Swanson, M. S. & Dreyfuss, G. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J* **7**, 3519-3529 (1988).

271 Martinez-Contreras, R. *et al.* Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol* **4**, e21, doi:10.1371/journal.pbio.0040021 (2006).

272 Clancy, S. in *Nature Education* Vol. 1 (ed Nature) 187 (Nature, Nature Education, 2008).

273 Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802-813, doi:10.1261/rna.876308 (2008).

274 Estivill, X. Genetic variation and alternative splicing. *Nat Biotechnol* **33**, 357-359, doi:10.1038/nbt.3195 (2015).

275 Ryan, B. M., Robles, A. I. & Harris, C. C. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* **10**, 389-402, doi:10.1038/nrc2867 (2010).

276 Bartel, D. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).

277 Swart, M. & Dandara, C. Genetic variation in the 3'-UTR of CYP1A2, CYP2B6, CYP2D6, CYP3A4, NR1I2, and UGT2B7: potential effects on regulation by microRNA and pharmacogenomics relevance. *Front Genet* **5**, 167, doi:10.3389/fgene.2014.00167 (2014).

278 Niu, T. *et al.* Identification of a novel FGFRL1 MicroRNA target site polymorphism for bone mineral density in meta-analyses of genome-wide association studies. *Hum Mol Genet* **24**, 4710-4727, doi:10.1093/hmg/ddv144 (2015).

279 Chang, T. H. *et al.* An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* **14 Suppl 2**, S4, doi:10.1186/1471-2105-14-S2-S4 (2013).

280 Martinez, L. C. *et al.* In silico identification and experimental characterization of regulatory elements controlling the expression of the Salmonella csrB and csrC genes. *J Bacteriol* **196**, 325-336, doi:10.1128/JB.00806-13 (2014).

281 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).

282 Desmet, F. O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67, doi:10.1093/nar/gkp215 (2009).

283 Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**, 3568-3571 (2003).

284 Huang, H. Y., Chien, C. H., Jen, K. H. & Huang, H. D. RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* **34**, W429-434, doi:10.1093/nar/gkl333 (2006).

285 Radhakrishnan, P. *et al.* MicroRNA-200c modulates the expression of MUC4 and MUC16 by directly targeting their coding sequences in human pancreatic cancer. *PLoS One* **8**, e73356, doi:10.1371/journal.pone.0073356 (2013).

286 Bao, L. *et al.* PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res* **35**, D51-54, doi:10.1093/nar/gkl797 (2007).

287 Bhattacharya, A., Ziebarth, J. D. & Cui, Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res* **42**, D86-91, doi:10.1093/nar/gkt1028 (2014).

288 Ziebarth, J. D., Bhattacharya, A., Chen, A. & Cui, Y. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res* **40**, D216-221, doi:10.1093/nar/gkr1026 (2012).

289 de Almeida, R. C., Chagas, V. S., Castro, M. A. A. & Petzl-Erler, M. L. Integrative Analysis Identifies Genetic Variants Associated With Autoimmune Diseases Affecting Putative MicroRNA Binding Sites. *Front Genet* **9**, 139, doi:10.3389/fgene.2018.00139 (2018).

290 Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J Comput Biol* **4**, 311-323, doi:10.1089/cmb.1997.4.311 (1997).

291 Smith, P. J. *et al.* An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* **15**, 2490-2508, doi:10.1093/hmg/ddl171 (2006).

292 Piva, F., Giulietti, M., Burini, A. B. & Principato, G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* **33**, 81-85, doi:10.1002/humu.21609 (2012).

293 Svozil, D., Kalina, J., Omelka, M. & Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res* **36**, 3690-3706, doi:10.1093/nar/gkn260 (2008).

294 Bhattacharya, R., Rose, P. W., Burley, S. K. & Prlic, A. Impact of genetic variation on three dimensional structure and function of proteins. *PLoS One* **12**, e0171355, doi:10.1371/journal.pone.0171355 (2017).

295 Casci, T. Disease genetics: SNPs and the structural deficit. *Nat Rev Genet* **11**, 669, doi:10.1038/nrg2871 (2010).

297 Perera, M. A. *et al.* Prediction of CYP3A4 enzyme activity using haplotype tag SNPs in African Americans. *Pharmacogenomics J* **9**, 49-60, doi:10.1038/tpj.2008.13 (2009).

298 Kligun, E. & Mandel-Gutfreund, Y. The role of RNA conformation in RNA-protein recognition. *RNA Biol* **12**, 720-727, doi:10.1080/15476286.2015.1040977 (2015).

299 Lipp, J. J., Marvin, M. C., Shokat, K. M. & Guthrie, C. SR protein kinases promote splicing of nonconsensus introns. *Nat Struct Mol Biol* **22**, 611-617, doi:10.1038/nsmb.3057 (2015).

300 Drogemoller, B. *et al.* Characterization of the genetic variation present in CYP3A4 in three South African populations. *Front Genet* **4**, 17, doi:10.3389/fgene.2013.00017 (2013).

301 Jing Li, X. L., Chao Zhang, Lei Tian, Dongsheng Lu and Shuhua Xu. Increased genetic diversity of ADME genes in African Americans compared with their putative ancestral source populations and implications for Pharmacogenomics. *BMC Genetics* (2014).

302 Lopez-Garcia, M. A. *et al.* Influence of genetic variants of CYP2D6, CYP2C9, CYP2C19 and CYP3A4 on antiepileptic drug metabolism in pediatric patients with refractory epilepsy. *Pharmacol Rep* **69**, 504-511, doi:10.1016/j.pharep.2017.01.007 (2017).

303 Birdwell, K. A. *et al.* The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenet Genomics* **22**, 32-42, doi:10.1097/FPC.0b013e32834e1641 (2012).

304 Liu, M. Z. *et al.* IL-3 and CTLA4 gene polymorphisms may influence the tacrolimus dose requirement in Chinese kidney transplant recipients. *Acta Pharmacol Sin* **38**, 415-423, doi:10.1038/aps.2016.153 (2017).

305 de Velasco, G. *et al.* Pharmacogenomic Markers of Targeted Therapy Toxicity in Patients with Metastatic Renal Cell Carcinoma. *Eur Urol Focus* **2**, 633-639, doi:10.1016/j.euf.2016.03.017 (2016).

306 Liu, J. *et al.* Genetic Polymorphisms Contribute to the Individual Variations of Imatinib Mesylate Plasma Levels and Adverse Reactions in Chinese GIST Patients. *Int J Mol Sci* **18**, doi:10.3390/ijms18030603 (2017).

307 Li, J. *et al.* Correlation analysis of gene polymorphisms and beta-lactam allergy. *J Zhejiang Univ Sci B* **16**, 632-639, doi:10.1631/jzus.B1400309 (2015).

308 Robien, K. *et al.* Genetic and environmental predictors of serum 25-hydroxyvitamin D concentrations among middle-aged and elderly Chinese in Singapore. *Br J Nutr* **109**, 493-502, doi:10.1017/S0007114512001675 (2013).

309 Chau, C. H. *et al.* Finasteride concentrations and prostate cancer risk: results from the Prostate Cancer Prevention Trial. *PLoS One* **10**, e0126672, doi:10.1371/journal.pone.0126672 (2015).

310 Wong, M. L., Dong, C., Maestre-Mesa, J. & Licinio, J. Polymorphisms in inflammation-related genes are associated with susceptibility to major depression and antidepressant response. *Mol Psychiatry* **13**, 800-812, doi:10.1038/mp.2008.59 (2008).

311 Johnson, J. A. *et al.* Pharmacogenomics: application to the management of cardiovascular disease. *Clin Pharmacol Ther* **90**, 519-531, doi:10.1038/clpt.2011.179 (2011).

312 Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* **14**, 288-295, doi:10.1038/nrg3458 (2013).

313 Koudritsky, M. & Domany, E. Positional distribution of human transcription factor binding sites. *Nucleic Acids Res* **36**, 6795-6805, doi:10.1093/nar/gkn752 (2008).

314 Pedersen, R. S. *et al.* Linkage disequilibrium between the CYP2C19*17 allele and wildtype CYP2C8 and CYP2C9 alleles: identification of CYP2C haplotypes in healthy Nordic populations. *Eur J Clin Pharmacol* **66**, 1199-1205, doi:10.1007/s00228-010-0864-8 (2010).

315 Scott, S. A. *et al.* Identification of CYP2C19*4B: pharmacogenetic implications for drug metabolism including clopidogrel responsiveness. *Pharmacogenomics J* **12**, 297-305, doi:10.1038/tpj.2011.5 (2012).

316 Jin, T. *et al.* Polymorphisms and phenotypic analysis of cytochrome P450 3A4 in the Uygur population in northwest China. *Int J Clin Exp Pathol* **8**, 7083-7091 (2015).

317 Weinshilboum, R. & Wang, L. Pharmacogenomics: bench to bedside. *Nat Rev Drug Discov* **3**, 739-748, doi:10.1038/nrd1497 (2004).

318 Ingelman-Sundberg, M., Oscarson, M. & McLellan, R. A. Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. *Trends Pharmacol Sci* **20**, 342-349 (1999).

319 Li, C., Long, J., Hu, X. & Zhou, Y. GSTM1 and GSTT1 genetic polymorphisms and risk of anti-tuberculosis drug-induced hepatotoxicity: an updated meta-analysis. *Eur J Clin Microbiol Infect Dis* **32**, 859-868, doi:10.1007/s10096-013-1831-y (2013).

320 Laizure, S. C., Herring, V., Hu, Z., Witbrodt, K. & Parker, R. B. The role of human carboxylesterases in drug metabolism: have we overlooked their importance? *Pharmacotherapy* **33**, 210-222, doi:10.1002/phar.1194 (2013).

321 Sanghani, S. P., Sanghani, P. C., Schiel, M. A. & Bosron, W. F. Human carboxylesterases: an update on CES1, CES2 and CES3. *Protein Pept Lett* **16**, 1207-1214 (2009).

322 Ramappa, V. & Aithal, G. P. Hepatotoxicity Related to Anti-tuberculosis Drugs: Mechanisms and Management. *J Clin Exp Hepatol* **3**, 37-49, doi:10.1016/j.jceh.2012.12.001 (2013).

323 in *Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis WHO Guidelines Approved by the Guidelines Review Committee* (2014).

324 Allison, A. C. & Eugui, E. M. Mycophenolate mofetil and its mechanisms of action. *Immunopharmacology* **47**, 85-118 (2000).

325 Yang, J., Shi, D., Yang, D., Song, X. & Yan, B. Interleukin-6 alters the cellular responsiveness to clopidogrel, irinotecan, and oseltamivir by suppressing the expression of carboxylesterases HCE1 and HCE2. *Mol Pharmacol* **72**, 686-694, doi:10.1124/mol.107.036889 (2007).

326 Pope, C. N., Karanth, S., Liu, J. & Yan, B. Comparative carboxylesterase activities in infant and adult liver and their in vitro sensitivity to chlorpyrifos oxon. *Regul Toxicol Pharmacol* **42**, 64-69, doi:10.1016/j.yrtph.2005.01.004 (2005).

327 Andersen, M. C. *et al.* In silico detection of sequence variations modifying transcriptional regulation. *PLoS Comput Biol* **4**, e5, doi:10.1371/journal.pcbi.0040005 (2008).

328 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, doi:10.7554/eLife.05005 (2015).

329 Liu, C. *et al.* MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics* **13**, 661, doi:10.1186/1471-2164-13-661 (2012).

330 Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20, doi:10.1016/j.cell.2004.12.035 (2005).

331 Garcia, D. M. *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* **18**, 1139-1146, doi:10.1038/nsmb.2115 (2011).

332 Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68-73, doi:10.1093/nar/gkt1181 (2014).

333 Wang, H. E., S.C. in *Pyrosequencing protocols* Vol. 373 (ed J.M.; Marsh Walker, S.) (Humana Press, 2007).

334 Jones, A. V. *et al.* Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood* **106**, 2162-2168, doi:10.1182/blood-2005-03-1320 (2005).

335 Llavona, P. *et al.* Allelic Expression Imbalance in the Human Retinal Transcriptome and Potential Impact on Inherited Retinal Diseases. *Genes (Basel)* **8**, doi:10.3390/genes8100283 (2017).

336 Kwok, C. T., Ward, R. L., Hawkins, N. J. & Hitchins, M. P. Detection of allelic imbalance in MLH1 expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Fam Cancer* **9**, 345-356, doi:10.1007/s10689-009-9314-0 (2010).

337 Royo, J. L. H., M.; Ruiz, A. Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nature Protocols* **2**, 1734-1739 (2007).

338 Ruslami, R. *et al.* Pharmacokinetics and tolerability of a higher rifampin dose versus the standard dose in pulmonary tuberculosis patients. *Antimicrob Agents Chemother* **51**, 2546-2551, doi:10.1128/AAC.01550-06 (2007).

339 Di, L. The Impact of Carboxylesterases in Drug Metabolism and Pharmacokinetics. *Curr Drug Metab*, doi:10.2174/1389200219666180821094502 (2018).

340 Bowman, B. P. *et al.* Effects of sex and gonadectomy on cocaine metabolism in the rat. *J Pharmacol Exp Ther* **290**, 1316-1323 (1999).

341 Kinirons, M. T. & O'Mahony, M. S. Drug metabolism and ageing. *Br J Clin Pharmacol* **57**, 540-544, doi:10.1111/j.1365-2125.2004.02096.x (2004).

342 Corrigan, D. & Paton, J. Hepatic enzyme abnormalities in children on triple therapy for tuberculosis. *Pediatr Pulmonol* **27**, 37-42 (1999).

343 Subedi, A., Magder, L. S. & Petri, M. Effect of mycophenolate mofetil on the white blood cell count and the frequency of infection in systemic lupus erythematosus. *Rheumatol Int* **35**, 1687-1692, doi:10.1007/s00296-015-3265-6 (2015).

344 Boudreau, F. *et al.* A novel colonic repressor element regulates intestinal gene expression by interacting with Cux/CDP. *Mol Cell Biol* **22**, 5467-5478 (2002).

345 Mailly, F. *et al.* The human cut homeodomain protein can repress gene expression by two distinct mechanisms: active repression and competition for binding site occupancy. *Mol Cell Biol* **16**, 5346-5357 (1996).

346 Liu, X., Yu, X., Zack, D. J., Zhu, H. & Qian, J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* **9**, 271, doi:10.1186/1471-2105-9-271 (2008).

347 Devon County Council. *Race and Ethnicity*, <http://www.devonhealthandwellbeing.org.uk/jsna/overview/archive/equality-and-diversity/race-and-ethnicity/> (2011).

348 Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* **519**, 309-314, doi:10.1038/nature14230 (2015).

349 Stark, A. E. & Seneta, E. A reality check on Hardy-Weinberg. *Twin Res Hum Genet* **16**, 782-789, doi:10.1017/thg.2013.40 (2013).

350 Lee, B. P. *et al.* Functional characterisation of ADIPOQ variants using individuals recruited by genotype. *Mol Cell Endocrinol* **428**, 49-57, doi:10.1016/j.mce.2016.03.020 (2016).

351 Chen, X. *et al.* Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. *Hum Mol Genet* **17**, 1336-1348, doi:10.1093/hmg/ddn022 (2008).

352 Ghazanfar, S. *et al.* Gene expression allelic imbalance in ovine brown adipose tissue impacts energy homeostasis. *PLoS One* **12**, e0180378, doi:10.1371/journal.pone.0180378 (2017).

353 Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol* **18**, 77, doi:10.1186/s13059-017-1212-4 (2017).

354 Conlon, B. P. Staphylococcus aureus chronic and relapsing infections: Evidence of a role for persister cells: An investigation of persister cells, their formation and their role in S. aureus disease. *Bioessays* **36**, 991-996, doi:10.1002/bies.201400080 (2014).

355 Pier, G. B. Will there ever be a universal Staphylococcus aureus vaccine? *Hum Vaccin Immunother* **9**, 1865-1876, doi:10.4161/hv.25182 (2013).

356 Proctor, R. A. Challenges for a universal Staphylococcus aureus vaccine. *Clin Infect Dis* **54**, 1179-1186, doi:10.1093/cid/cis033 (2012).

357 Chan, S. L. *et al.* Association and clinical utility of NAT2 in the prediction of isoniazid-induced liver injury in Singaporean patients. *PLoS One* **12**, e0186200, doi:10.1371/journal.pone.0186200 (2017).

358 Stevens, D. L. *et al.* Practice guidelines for the diagnosis and management of skin and soft tissue infections: 2014 update by the infectious diseases society of America. *Clin Infect Dis* **59**, 147-159, doi:10.1093/cid/ciu296 (2014).

359 Safdar, N. H., J.; Maki, D.G. Does combination antimicrobial therapy reduce mortality in Gram-negative bacteraemia? A meta-analysis. *Infectious Diseases* **4**, 519-528 (2004).

360 Noskin, G. A. *et al.* The burden of Staphylococcus aureus infections on hospitals in the United States: an analysis of the 2000 and 2001 Nationwide Inpatient Sample Database. *Arch Intern Med* **165**, 1756-1761, doi:10.1001/archinte.165.15.1756 (2005).

361 Kock, R. *et al.* Methicillin-resistant Staphylococcus aureus (MRSA): burden of disease and control challenges in Europe. *Euro Surveill* **15**, 19688 (2010).

362 Mayo Clinic. *Liver tests functions*, <https://www.mayoclinic.org/tests-procedures/liver-function-tests/about/pac-20394595> (2018).

363 Alghamdi, W. A. *et al*. Protein Binding of First-Line Antituberculosis Drugs. *Antimicrob. Agents Chemother*. **62**, (2018).

364 Thabit, A. K. *et al*. Antibiotic penetration into bone and joints: An updated review. *Int. J. Infect. Dis*. **81**, 128–136 (2019).

365 Nau, R. *et al*. Penetration of drugs through the blood-cerebrospinal fluid/blood-brain barrier for treatment of central nervous system infections. *Clin. Microbiol. Rev*. **23**, 858–83 (2010).

366 Sartor, F. *et al*. Are There Circadian Clocks in Non-Photosynthetic Bacteria? *Biology* (Basel). **8**, 41 (2019).

367 Cijaudhary, K. C. J. R. S. *et al*. Effect of buffers and pH on Rifampicin stability. *Die Pharmazeutische Industrie*. **57**(5), 420-422 (2018).

368 Bethke, L. *et al*. Polymorphisms in the cytochrome P450 genes CYP1A2, CYP1B1, CYP3A4, CYP3A5, CYP11A1, CYP17A1, CYP19A1 and colorectal cancer risk. *BMC Cancer.* **7**, 123 (2007).

369 Berno, G. *et al*. Potential implications of CYP3A4, CYP3A5 and MDR-1 genetic variants on the efficacy of Lopinavir/Ritonavir (LPV/r) monotherapy in HIV-1 patients. *J. Int. AIDS Soc*. **17**, 19589 (2014).

370 Parada, H. *et al*. Genetic polymorphisms of phase I metabolizing enzyme genes, their interaction with lifetime grilled and smoked meat intake, and breast cancer incidence. *Ann. Epidemiol*. **27**, 208-214.e1 (2017).

371 Danielak, D. *et al*. Impact of CYP3A4*1G Allele on Clinical Pharmacokinetics and Pharmacodynamics of Clopidogrel. *Eur. J. Drug Metab. Pharmacokinet*. **42**, 99–107 (2017).

372 He, B. xia *et al*. A Functional Polymorphism in the CYP3A4 Gene is Associated with Increased Risk of Coronary Heart Disease in the Chinese Han Population. *Basic Clin. Pharmacol. Toxicol*. **108**, 208–213 (2011).

373 Becker, S. A. & Palsson, B. O. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst. Biol*. **2**, 14 (2008).