22 February 2018
EMA/CHMP/718998/2016
Committee for Medicinal Products for Human Use (CHMP)

Guideline on good pharmacogenomic practice

Agreed by Pharmacogenomics Working Party	21 March 2016
Adopted by CHMP for release for consultation	28 April 2016
Start of public consultation	1 May 2016
End of consultation (deadline for comments)	16 September 2016
Agreed by Pharmacogenomics Working Party	23 October 2017
Adopted by CHMP	22 February 2018
Date of coming into effect	1 September 2018

Keywords	***Pharmacogenomics, good practices, pharmacogenomic analyses, biomarkers, study design, pharmacokinetics, DNA sequencing***

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Guideline on good pharmacogenomic practice

Table of contents

Executive summary

Genomic data have become important in the evaluation of efficacy and safety of medicinal products for regulatory approval. Genomic information in the product information (PI) influences patient treatment decisions. Therefore, use of genomic biomarkers in drug development, should, in order to be of value, follow certain principles which are outlined in this guideline.

This guideline provides recommendations for the conduct of genomic studies in relation to medical therapy in order to provide high quality information on the impact of genomic variability on drug response. Primary focus is on the analysis of genomic germline DNA. The analysis of somatic DNA and genomic biomarkers for cancer treatment is not being discussed and might be developed as an Annex or in separate guidance.

Some methodological issues with respect to genomic clinical study design highlighted in this guidance is complementary to the what is described in the *Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection* (see section 3).

1. Introduction (background)

There has been an increase in our understanding of how inter-individual differences in DNA sequences, or 'genetic variants', link to drug response. As a consequence, there is now the opportunity to incorporate the genomic information as a basis for an estimated individual response to drug treatment, leading to a transition from population-based prescribing with a one-size-fits all dosing to a more individualized treatment.

The identification of genomic variability for explaining or predicting the response of a patient to a specific medicinal product has focused mainly on variation in genes encoding: (i) drug-metabolizing enzymes (ii) drug transporters, and (iii) drug targets. For prediction of adverse drug reactions, the analyses of specific (host) HLA haplotypes also play an important role for specific medicinal products.

Often, previous studies in this area have not been optimally designed e.g. retrospective in nature, often driven by association alone and not causality or functionality, low statistical power and use of inappropriate technology (see Annex I). Thus, there is a need for a guideline addressing factors of importance to consider in pharmacogenomics investigations.

2. Scope

The document provides guidance on methods of evaluation of genetic variations related to pharmacokinetics and response, where this could affect efficacy and/or safety

The guidance primarily focusses on the analysis of genomic germline DNA and does not discuss somatic DNA and genomic biomarkers for cancer treatment, RNA variations, proteomics or metabolomics, which are outside the scope of this document, however some principles might apply.

3. Legal basis and relevant guidelines

This guideline applies to Marketing Authorization Applications for medicines for human use and should be read in conjunction with other relevant EU and ICH guidelines as well as reflection papers. These include, but are not limited to:

- Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products - EMA/CHMP/37646/2009 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/02/WC500121954.pdf
- Guideline on key aspects for the use of pharmacogenomics in the pharmacovigilance of medicinal products - EMA/CHMP/281371/2013 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/11/WC500196800.pdf
- Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection - EMA/446337/2011 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf
- Reflection paper on co-development of pharmacogenomic biomarkers and Assays in the context of drug development- EMA/CHMP/641298/2008 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/07/WC500094445.pdf
- Guideline on the evaluation of anticancer medicinal products in man - EMA/CHMP/205/95/Rev.4 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137128.pdf
- Reflection paper on pharmacogenomics in oncology - EMEA/CHMP/PGxWP/128435/2006 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003866.pdf
- Reflection paper on pharmacogenomic samples, testing and data handling - EMEA/CHMP/PGxWP/201914/2006 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003864.pdf
- Note for guidance on definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories - EMEA/CHMP/ICH/437986/2006 (ICH Topic E15) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002880.pdf
- Position paper on terminology in pharmacogenetics - EMEA/CPMP/3070/01 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003889.pdf
- The Rules Governing Medicinal Products in the European Union Volume 2C Notice to Applicants; A guideline on summary of product characteristics (SmPC) September 2009 http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf
- Biomarkers Related to Drug or Biotechnology Product Development: Context, Structure, and Format of Qualification Submissions (ICHE16) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500097060.pdf
- Guideline on genomic sampling and management of genomic data (ICH E18) http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/md_pharmacogenomics/general_content_001300.jsp&mid=WC0b01ac058002958e For guidance relating to extrapolation of data

across populations, see ICH Topic E5: Ethnic Factors in the Acceptability of Foreign Clinical Data (CPMP/ICH/289/95).

4. Genetic variation and drug response

4.1. Common and rare genetic variants

For the purposes of this guideline, a genetic variant is defined as an allele carrying variation(s) as compared to the gene reference sequence. The term variant is preferably being used instead of the term mutation, which is reserved for rare variants (<1% frequency). Furthermore, the term polymorphism defined as variations having >1 % frequency in the population, is considered inappropriate due to the large inter-population differences in frequency.

Classical approaches in pharmacogenomics involved genotyping of a predefined panel of a selected number of functionally characterized variants. However, the predefined panel approach was mainly applicable to common variants and may not identify or classify rare variants; broader approaches such as next generation sequencing (NGS) are needed to identify rare variants of functional importance. Analyses of exomes from >60,000 individuals, reveal that every 8th nucleotide in the exome exhibits variation and 50 % of all rare genetic variants were indeed only seen in a single individual of the 60,000 analysed [1/]. Furthermore, phenotyping in monozygotic and dizygotic twins has shown that in case of two specific medicinal products, only about 40% of the inherited differences in metabolism could be explained by known gene variants exemplifying the large contribution of unknown rare mutations. Similar rare variability is seen in receptors e.g. mutations in the Cystic fibrosis transmembrane conductance regulator (CFTR) gene. The current routine analyses of the common allelic variants may thus not predict the full inter-individual variability in PK or PD.

In cases where the inter-individual differences in PK or PD are high and not explained by known genetic variants, it is recommended that genomic DNA sequencing is carried out using broader approaches such as whole exome sequencing (WES) or whole genome sequencing (WGS) with subsequent bioinformatic analyses of sequences in appropriate genomic regions. It is also crucial to define the functionality of the novel genetic variants identified (see below discussion on functionality). If variant and functionality such analysis is not immediately feasible, samples from PK or PD outliers should be stored for further exploration in line with recommendation of the ICH E-18 guideline on genomic sampling and management of genomic data (see section 3) The informed consent (for research and clinical studies) should therefore include the potential for later analysis. Handling of incidental findings or discoveries should also be detailed.

4.2. Interethnic differences in genetic variants

Genomic studies should consider inter-ethnic differences in allele frequency and distribution. This is true for the majority of rare variants, as well as common genetic variants in ADME genes 2/, HLA alleles 3-4/, and genes encoding drug targets. Rare variants in some populations might be very common in other specific geographical regions. When clinical trials are conducted in specific geographical regions, the regional specific distribution of genetic variants anticipated to influence PK or PD should be accounted in the analysis and conclusions. For guidance relating to extrapolation of data across populations, see ICH Topic E5 (see section 3).

4.3. Phenotype and genotype correlations

Genetic variants may influence the function of a gene (i) directly, by introduction of loss of function (LOF) mutations, or missense mutation causing amino acid exchange (ii) indirectly, by modifying the abundance of the gene product (mRNA expression), thereby affecting protein expression levels. The PK-phenotype can also be determined by multiple variations in different alleles, e.g. for medicinal products metabolized by multiple pathways. The detection of specific genetic variants in clinical routine helps to improve efficacy and safety of specific medicinal products in use, and for that, the genotype in question ultimately needs to be translated into a predicted phenotype. Whereas the gene analysis based prediction for the phenotype is usually straightforward for homozygous variants that result in a loss of function of the respective gene product, it can be more complex for heterozygous LOF genetic variations or missense mutations. Thus, phenotype assignment from genotypes where genetic variation(s) only modestly alter (increase or decrease) the respective enzyme activity is more difficult.

Additional factors that affect prediction of phenotype from genotype are, environmental as well as endogenous in nature i.e. (i) patient compliance, (ii) diet, (iii) bioavailability, (iv) hepatic blood flow and function, (v) renal function, (vi) co-medication and (vii) dependency on the particular metabolic pathway.

Phenotyping for a specific metabolizing enzyme by e.g. therapeutic drug monitoring at different time intervals after administration is a valid method for measuring the impact of genetic variance on metabolism. Compared to genotyping alone, this approach has the advantage of directly assessing functionality including contributions of several different gene variations or rare mutations of importance for the PK. The in vivo assessment of a phenotype relies on the use of specific probe drugs. For simultaneous determination of activities of multiple CYP enzymes, it is recommended to use a phenotyping drug (probe) cocktail, in which different metabolites formed by different CYP enzymes are individually quantified in blood or urine. Appropriate probe substances should be used, i.e. a compound must fulfill certain criteria such as i) selectivity for the enzyme in question, ii) having a specific, targeted, quantifiable metabolite, iii) should be relatively non-toxic and iv) not interacting with the parent substance or the metabolite from the enzymatic conversions.

The metabolic phenotype can vary substantially over time in the same individual due to physiological, pathological or dietary factors which must be taken into account when doing therapeutic drug monitoring. Genotyping gives a stable, lifelong insight into the functional status of the enzymes in question. However due to the important contribution of rare genetic variants in each individual (see Chapter 5.1) therapeutic drug monitoring takes into account all genetic variations of importance for the PK assessed, whereas Targeted genotyping can miss individual specific genetic variations, since these were not part of the testing SNP panel analyzed. For more discussion on this topic see the Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products EMA/CHMP/37646/2009guidance).

4.4. Specific issues for HLA alleles

Idiosyncratic adverse drug reactions (ADRs) include examples of strong associations between genotypes and drug responses. These reactions i) are potentially life-threatening, ii) they cannot be simply predicted from the known pharmacology of the medicinal product or its plasma concentration, iii) they are rare, therefore not usually detected in clinical trials during development and iv) thus only become apparent when many individuals are exposed to the culprit agent after marketing. It is therefore recommended to collect DNA for later HLA typing in a drug developmental program to

identify possible useful predictive HLA biomarkers. A comprehensive database listing ADRs associated with the HLA alleles and haplotypes is available [4].

HLA typing should be undertaken by accredited laboratories using DNA based methods. High resolution typing of the following six genes shall be performed: HLA-A, -B, -C, -DR, -DQ and –DP. The nomenclature for HLA alleles is given in Figure 1.

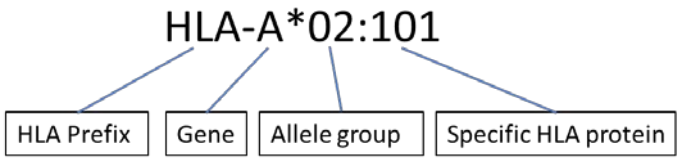

Figure 1. *Nomenclature of HLA alleles. (http://hla.alleles.org/nomenclature/naming.html)*

More than 15,000 different HLA alleles have been reported. Interpretation of HLA typing results is complex and it is not always possible to define the type to a single allele. Ideally a technology should be used that allows sequencing of the entire HLA gene, which removes ambiguity of HLA types produced. It also allows the analysis of the HLA haplotypes.

The advent of NGS and the development of computational tools capable of typing the HLA alleles using WGS or WES sequencing data, is likely to become the gold standard approach for HLA typing and current recommendation is to use this when possible.

Pre-emptive genotyping for some HLA alleles is mandatory for use of certain medicinal products either in clinical studies or in practice. These include testing for HLA-B*57:01 before starting abacavir, HLA-B*15:02 in Asians before initiating carbamazepine treatment and HLA-B*58:01 before starting allopurinol treatment. These tests have been shown to reduce the incidence of serious hypersensitivity reactions if individuals who test positive are given alternative medications. The development of HLA allele panels with integrated clinical decision support system, pre-emptive genotyping across the HLA region and reducing cost of next generation sequencing promise to deliver cost-effective and efficient HLA testing available at the time of treatment.

5. Quality aspects of pharmacogenomic analyses

5.1. Sample acquisition and handling (preanalytical)

Identification of genetic variants demands high quality genomic DNA (gDNA). Pre-analytical variations in (i) sample collection, (ii) stability, (iii) sample labelling, (iv) transport to the site of analysis, (v) tissue/sample processing and (vi) storage, should be minimized throughout the workflow) to guarantee the highest possible sample quality(see guideline ICH E18 cited in section 3). Procedures to ensure sample adequacy and quality must be in place in any genomic study particularly where multiple centres are involved. There is published advice regarding pre-analytical workflows encompassing e.g. isolation of DNA from snap frozen tissue as well as for isolation of gDNA, cell-free DNA (cfDNA) and circulating tumour DNA (ctDNA) from whole venous blood[5].

Barcode or radio frequency identification (RFID) labelling of samples has several advantages and the same label should follow the sample throughout all analyses. Coding and anonymization of stored samples should follow established protocols and quality management systems allowing for (i) the destruction of the samples if the patient withdraws the consent or (ii) further follow-up analyses if the patient consent is still valid (see ISO15189).

Sample quality is usually retained during long term storage of DNA samples in water at + 4 C and – 20°C with attention to avoid sample dry-out or repeated freeze- thaw cycles. The reader is here also referred to the guideline on genomic sampling and management of genomic data (see section 3).

5.1.1. Sample repository for retrospective studies

Retrospective PGx related studies using DNA analyses, including NGS, are often performed on stored samples. It is important that these biosamples are not limited by their quality and/or quantity. Increasingly, sophisticated genomic techniques for PGx analysis require the establishment of dedicated PGx sample repositories that employ scrupulous standards governing sample quality and usage.

Several national and European initiatives have led to the establishment of DNA repositories across Europe, with a very broad scope and access.

Detailed information for retrospective analyses of pharmacokinetic samples for PGx is provided in the *Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products (as well as ICH E18* (see section 3).

5.2. DNA Analytics

5.2.1. DNA extraction

A variety of protocols are available for the extraction of gDNA from different sources. It is important to employ a validated method leading to an isolation of gDNA of highest quality, suitable for the nucleotide variation analysis (single or array-based qPCR or end-point PCR), and for sequencing (Sanger, NGS). It is advised to test all DNA samples isolated for quality before any further work on DNA analysis. The routine analysis concerns the ratio of light absorption of the DNA solution at 260 nm compared to 280 nm where the ratio A 260/280 should be > 1.9. See also ICH E18.

5.2.2. Methods used for determination of nucleotide variations

Nucleotide variations can be determined with a variety of methods focusing (i) on targeted sequence areas or (ii) broader sequencing approaches. The latter encompass WES, whole gene sequencing by Sanger or NGS including promoter, introns and exons, on single or multiple genes or WGS covering the entire genome with exception of specific complex loci with high sequence homology.

It is important to test for known functionally relevant nucleotide variations, regardless of whether they are located in the coding or non-coding region of the gene. In the genes important for pharmacokinetics and dynamics, such intron mutations are functionally important in some cases (for CYP-genes about 2 % of all important SNPs) and this reduces the usefulness of plain WES.

Often it is important to validate the sequencing results, using either an independent analytically valid method or by re-sequencing a second amplicon from the same region. Furthermore, it is important, to validate the data obtained against samples known to lack the variation in question and samples have the genetic variation of interest.

Current practice for analyses of genetic variants include SNP analyses (single or array-based qPCR, hybridization etc.), PCR (endpoint or quantitative with or without restriction-enzyme digestion) and sequencing (Sanger, NGS). A variety of procedures with different technical and/or chemical approaches are currently used for genomic biomarker analytics involving subjects. The main difference in the testing approaches used lies in the number of variants tested per gene. It is expected that the analysis

or the test should provide unambiguous results and that “rare” variants affecting safety and efficacy of the medicinal product should not be excluded (see section 5.1).

Methods using primer-based technologies are prone for allele-drop-out artefacts due to failure of primer hybridization in case of specific mutations, leading to erroneous genotyping results and therefore to an inaccurate phenotype assignment. This should be avoided by identifying the respective allele-drop-outs or use tests able to avoid known allele drop outs.

Caution should be applied when proxy-SNPs are used for predicting the presence of functionally relevant SNPs, since there is not an absolute linkage between the proxy-SNP and the functionally important SNP. Preference should be given to the direct analysis of the respective functional relevant SNPs, either by sequencing, or an array-based or other approach directly analyzing the functional SNPs. Where proxy or tag-SNPs have to be used, a risk estimate for miscalls should be included in the analytical report.

Prior to use in clinical trials or in a diagnostic setting, the testing procedures need proper validation. The implementation of such test must comply with existing regional guidelines and being validated preferentially with respect to genetic variability, by two different methods for sequencing, e.g. NGS and Sanger techniques. Certain exemptions may apply for in-house tests.6.2.3 NGS specific issues.

NGS based genetic test workflows include DNA extraction, DNA processing, preparation of libraries, generation of sequence reads and base calling, sequence mapping, variant annotation and filtering, variant classification and interpretation. It is necessary that all these steps are carefully carried out using validated methods and continuously subjected to rigorous quality control. For NGS based sequencing the DNA quality must be very high. Before starting a new project, it is recommended to analyse a small number of representative samples using NGS. The chosen DNA isolation methods should be shown to yield satisfactory results before initiation of the full study. Quality Control (QC) steps necessary for the development of an in house diagnostic need to be followed.

A specific issue affecting reliability of NGS is the coverage the method provides for a specific DNA sequence. It is recommended that the technical predictive value should be at least 99.9%. For germline genetics, a minimum coverage of >30x seems to be a reasonable goal. If however, the allele frequency of the mutations analyzed is very low, a higher coverage is needed in order to ensure that also the rarer variants are detected by the sequencing.

NGS analysis of complex loci with high GC-content (guanine-cytosine content) or highly homologues genes and pseudogenes can contribute to miscalled variants due to sequencing artefacts. It is therefore recommended in such cases, to include methods that use substantially longer read lengths, i.e. fragments >1000 base pairs. This can be achieved with initial DNA amplification using long PCR techniques or synthetic long read methods, which use partitioning and barcoding of longer DNA molecules before standard library preparation, that allows the assembly of short reads into longer fragments. It is acknowledged that these techniques are not always technically possible, e.g. when using Formalin-Fixed Paraffin-Embedded (FFPE) samples.

5.3. Allele specificity

For purpose of the analysis of genotypes it is sometimes important to determine the specific allelic location of the variants and a thorough characterization of the entire haplotype (all mutations in the gene present on one allele). When two different genetic variations within the same gene with known functional implications are identified in heterozygosity in one individual, it is often important to know if the two variations are on the same allele (in cis) or segregated between the two alleles (in trans).

Such analysis can be performed following long allele-specific PCR amplification of the region of interest, followed by NGS or Sanger sequence analyses.

5.4. Copy number variations (CNV) and gene hybrids

CNVs of genes, when varying from the canonical two copies (i.e. fewer or more than 2) contribute to the phenotype prediction, especially of genes encoding metabolizing enzymes. It is required that for CNVs, only functional copies of the respective gene are taken into account when predicting the metabolizing phenotype. In the case a gene forms hybrids (chimeras), which share identity in the nucleotide sequence with another gene, the test used must be able to detect such gene hybrids. In the case of multiple copies (not hybrids), it is important that the assay employed is able to detect change of function mutations in the individual copies as well as to quantify the number of gene copies, therefore delivering unambiguous and correct i.e. allele-specific results.

In addition to full gene CNVs, the human genome has variable levels of partial gene deletions causing an inactive gene and partial gene duplications which are non-functional. The overall frequency of such recently discovered CNVs are 0-1 % of all genes affecting PK and are important to consider together with other types of genetic variations.

5.5. Quality assurance of DNA analytics

DNA analysis requires meticulous methodology and bioinformatics leading to an unambiguous genetic call in order to influence decision making about the safety and efficacy of the medicinal product. Therefore, appropriate quality assurance is recommended. This may require accreditation in the different member states or be subjected to validation using nationally accepted procedures for predictive biomarker analytics, including intra-laboratory proficiency testing. ISO15189, or corresponding certification, e.g. American standards set by the College of American Pathologists, CLIA, would be important to harmonize standards of good laboratory practice as well as use of the OECD standards. [6]

5.6. Reporting

It is recommended to primarily study genomic variations of functional importance for the phenotype to be predicted. For this it is important to carefully use published and well curated sequence databases and employ validated bioinformatics methods and algorithms.

Reporting should favor nucleotide variations with proven functional implication over nucleotide variations for whom the functional implication is only predicted but not proven. Predicted functional implications of Missense (amino acid substitution) mutations might be difficult to judge. Currently > 14 different functionality prediction algorithms have been published which have very different sensitivities and specificities. The best algorithms can predict the functional consequence of 75-85 % of the missense mutations on the gene product in question. The methods differ in attributes such as physiochemical properties, secondary structure, protein domain models or integrated functional residues, and how the results are interpreted. These algorithms may be suitable for screening of potential relevance of incidental findings but cannot accurately predict the functional consequences of missense mutations in genes encoding drug metabolism and transport. Further advances to functionally assess detected variants are necessary in order to generate clinically actionable recommendations.

In order to decrease the risk of finding incidental mutations, targeted sequencing of genomic regions of particular interest is recommended. Furthermore, during genome sequence interpretation it is

recommended to exclude known risk mutations causing specific diseases as continuously being updated. [7]

A good laboratory report from genetic analyses should indicate the name of the gene that was investigated, identify genetic variations including their Rs numbers (if known), the interpretation of the genetic variants to alleles (haplotype definition) and a prediction of the phenotypic consequences based on the variants detected.

6. Clinical study design

The impact of genomic variation on drug response may require investigation in clinical studies.

For genomic variation that may affect the PK of a medicinal product, clinical studies should be designed in accordance with the *Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products* (see section 3).

For genomic variation that may predict the PD of a medicinal product (affecting efficacy or safety), some considerations are given below.

6.1. General considerations

Genetic variants may be evaluated in clinical studies based on biological plausibility or non-clinical (*in vitro* or *in vivo*) findings. Alternatively, a predictive genetic variant may be identified for the first time during clinical studies.

A predictive genetic variant may be a single marker or be comprised of a multi-marker signature or an algorithm. Predictive genetic variants may be binary, categorical or continuous. For continuous genetic variants (e.g. gene expression levels), the test result is usually dichotomised using a pre-defined cut-off; in which case the predictive value is a function of the cut-off.

Selection bias may affect the apparent predictive value of the genetic variant at any stage of clinical development. Bias may arise due to non-availability of biological samples from all patients, the use of convenience samples (e.g. from certain centres or larger tumours), assay failure or the practice of pre-screening. Sources of selection bias should be minimized.

The chosen biomaterial should be appropriate for the study objective. All sources of variability (technical or biological) should be identified and controlled for (see Section 5.1). The analytical validity of the assay should be demonstrated as early as possible during clinical development. This is particularly important when the genetic variant is used to include/exclude patients or allocate treatment. Analytical validity at an early stage also optimizes the relevance of early clinical findings to later clinical development.

It is recommended to seek scientific advice on the proposed development strategy.

6.2. Exploratory stage

Exploration of candidate predictive genetic variants can be part of both early and late phases of clinical development. Early phase clinical studies are often small single arm cohort studies. Such study designs may provide evidence of an association between genetic variability and drug response. However, a control arm may be required to confirm that a genetic variant is predictive of drug response rather than prognostic. Selection bias may also be a particular problem for early phase studies (see section 6.1).

The design of clinical studies to explore candidate predictive genetic variants may depend on the type of genetic variant being investigated: single binary, continuous (requiring exploration of different cut-offs) or candidates for multi-marker signatures. In addition, exploration may be conducted in the context of an umbrella or basket design. Furthermore, flexibility may be incorporated by use of an adaptive design which allows for pre-planned interim analysis and design modification (see *Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design* - section 3).

Genetic variants predictive of drug response may also be identified by retrospective analysis of large clinical datasets including randomised controlled trials (RCTs), for example using a genome wide association study (GWAS) approach. In the case of GWAS, due to multiplicity issues, only very strong associations are likely to be reproducible during subsequent development. Furthermore, only relatively common genetic variants >1% can be identified using GWAS. This approach has been used to identify genetic variants predictive of toxicity.

6.3. Confirmatory stage

Confirmation of the clinical validity of a predictive genetic variant involves replication in an independent cohort of the observed genotype-dependent outcome. The frequency of the genetic variant in the target population and the outcomes of interest will drive the design of the study, including analyses to confirm the clinical relevance and utility of the genetic variant. In some cases, the pivotal trial for clinical efficacy and safety can provide sufficient evidence of clinical validity.

In some cases, confirmation of the clinical validity of the genetic variant could be based on retrospective data rather than prospective studies, provided the confirmatory dataset is distinct from that used in the exploratory stage.

To investigate the clinical validity of a genetic variant, a stratified RCT should be considered. Patients with or without the genetic variant are included, and randomization is stratified by variant status. This design allows measurement of the capacity of the genetic variant to separate patients into those more likely or less likely to demonstrate a drug response, within the same study. The analysis of treatment effect can be conducted in the overall pooled population as well as in the populations defined by the genetic variant; the primary analysis population should be pre-specified in the statistical analysis plan. For continuous genetic variants, the cut-off point for determination of variant status should be pre-specified and adequately justified.

In a targeted (enrichment) design, study inclusion is dependent on genetic variant status, and therefore it is not possible to confirm the clinical validity of the genetic variant. However, it may be possible to demonstrate a drug response in the presence of the genetic variant. This design is only appropriate if there is strong biological plausibility, and compelling non-clinical and/or clinical evidence, that drug response is likely to be variant-dependent. An enriched design may also refer to a variation on the stratified design in which a small number of variant negative patients are included; in this case, it may be possible to confirm the clinical validity of the genetic variant.

For a genetic variant that might predict a rare severe adverse drug reaction, a case control design may be more applicable than a prospective RCT.

For further details on appropriate study designs to confirm clinical validity, reference should be made to the *Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection* (see section 3).

7. Considerations on product information

The Summary of Product Characteristics (SmPC), sets out key elements of benefits and risks relevant to the clinical use of the product defined during the regulatory assessment process.

These include mandatory pre-emptive genotyping before prescription, or an indication that genotyping is recommended, or just that there is relevant pharmacogenomic consideration related to the use of that particular product. For details on product information and the SmPC, please also see *the Guideline on key aspects for the use of pharmacogenomics in the pharmacovigilance of medicinal products* (see section 3).

It is essential that the SmPCs include relevant pharmacogenomic data gathered through the life cycle of the product; both pre-authorization development and in the post-approval phases.. As science develops, increasing amount of information is continuously obtained regarding pharmacogenomic biomarkers of value for the specific drug treatment. Thus, the product description should continuously be updated with additional information as appropriate. Particular attention is often needed in describing the influence of gene variants on efficacy or safety of medical product and should include the specific variants to inform clinical decision making. Marketing Authorisation Holders should also aim to improve the SmPC based on pharmacovigilance data, new study results or information from other sources. Such updates of the SmPC may also be initiated by regulatory authorities. This will facilitate the appropriate use of pharmacogenomic information by prescribers and patients.

8. Conclusions

The main recommendations and conclusions of this guideline can be summarised as follows:

- Broader sequencing approaches are needed to identify rare variants of functional importance for drug pharmacokinetics and drug response.
- High quality DNA is needed for sequencing and validation of critical genetic elements should be done using another sequencing method. For specific mutations, gDNA samples from positive and negative carriers should be used as controls.
- The minimum level of Next Generation sequence (NGS) coverage should be adequate to ensure a high technical predictive value and higher coverage is needed for variants at low frequency.
- Incidental disease predicting mutations should be excluded from the analyses.
- For certain variants the allele specificity is important and here long allele-specific PCR amplification of the region of interest, followed by NGS or Sanger sequence analyses long PCR should be used.
- Appropriate methods should be used for determination of copy number variations and only functional copies of the respective gene should be considered for prediction of phenotype. Partial gene deletions should also be taken into account for predicting pharmacokinetics.
- Genomic studies in clinical trials should consider inter-ethnic differences in the distribution of genetic variants.
- Phenotyping is recommended in order to evaluate the functional significance of genetic variants in relation to pharmacokinetics where contribution of several different genetic variants are at hand or when it is expected that rare variants will contribute to the variability.
- When evaluating HLA alleles, the entire HLA gene should be sequenced when possible.

- Sample acquisition and handling procedures must be designed to guarantee sample adequacy and quality.
- For the genomic variation that may predict pharmacodynamics, investigation using stratified or enrichment trial designs may be appropriate. The analytical validity of the assay should be demonstrated as early as possible during clinical development. Selection bias should be avoided.
- It is essential that the product information includes relevant pharmacogenomic data gathered through the life cycle of the product.

9. References

[1/] http://exac.broadinstitute.org/about

[2/] Human Cytochrome P450 (CYP) Allele Nomenclature website now at www.PharmVar.org

[3/] www.allelefrequencies.net

[4/] http://www.allelefrequencies.net/hla-adr/default.asp;

[5/] http://www.bbmri-eric.eu/wp-content/uploads/2017/02/2017_Newsletter6_7_WEB.pdf

[6/] http://www.oecd.org/env/ehs/testing/goodlaboratorypracticeglp.htm

[7/] http://www.hgmd.org

Definitions and abbreviations

Accuracy	The degree of conformity of measured results in one test system with the true (actual) value.
ADRs	Adverse drug reactions
Allele	One of a number of alternative forms of the same gene or the same genetic locus
Allele specific SNP analytics	Determining on which allele (maternal or paternal) a certain genetic variation(s) is located
Analytical specificity	The ability to unequivocally assess the target nucleic acid in the presence of other nucleic acids or other components, which may be expected to be present.
Analytical sensitivity	The detection limit, which is the lowest amount of nucleic acid, which can be specifically detected by a PGx assay.
Biomarker	A characteristic that is measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacological responses to a therapeutic intervention.
CGI	Cancer Genome Interpreter
CFTR	Cystic fibrosis transmembrane conductance regulator
Clinical sensitivity	The proportion of individuals with a specified clinical disorder or clinical effect whose test values indicate that the disorder or clinical

	effect is present (e.g. the mutation associated with the disorder is identified).
Clinical specificity	The proportion of individuals who do not have a specified clinical disorder or effect and whose test results indicate that the disorder or clinical effect is not present.
CNV	Copy number variation
CTC	Circulating tumor cells
CYP	Cytochrome P450
Cross-validation of	Confirming a pharmacogenomic finding by repeating the experiment pharmacogenomic studies using an independent pharmacogenomics assay or with the same technique in another laboratory
cfDNA	Cell free DNA
ctDNA	Circulating tumour DNA
DDI	Drug-drug interactions
DNA	Deoxyribonucleic acid
Epigenetics	Changes to the genome that do not involve a change in the nucleotide sequence, e.g., DNA methylation or histone modification
Epigenomic biomarkers	DNA sequences where the extent of specific cytosine methylation or hydroxymethylation can provide useful information
FFPE	Formalin-Fixed Paraffin-Embedded Blocks
Functional polymorphism	A polymorphism that has been shown to alter enzyme or protein activity and/or the clinical disposition of drugs
Gene	A locatable region of genomic sequence, corresponding to a unit of inheritance
Genomic biomarker	A measurable DNA and/or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other interventions. (ICH15)
Genotype	The combination of alleles (maternal and paternal) that determine the expression of a particular phenotype
Genetic biomarker	A gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), or a long one, like mini satellites.
Genetic subpopulation	Subdivision of the whole population, with common, distinguishing genetic characteristics. These characteristics may include both the phenotype, e.g. poor metaboliser, as well as the genotype, e.g., CYP2D6*4

Germline DNA	The DNA in germ cells (egg and sperm cells that join to form an embryo). Germline DNA is the source of DNA for all other cells in the body. Also called constitutional DNA
gDNA	Genomic DNA
GST	Glutathione transferase
GWAS	Genome wide association study
Haplotype	A combination of alleles at different loci on the chromosome that are transmitted together
HLA	Human leukocyte antigen
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
Intermediate precision	The within-laboratory variation of repeat test results with one test system (different days, different analysts, different equipment).
IPV	Intra patient verification
Locus	The specific location of a gene or DNA sequence on a chromosome
NAT	N-acetyl transferase
NGS	Next generation sequencing, All post-Sanger sequencing methods, most commonly referring to massively parallel sequencing technology.
PD	Pharmacodynamics
PGx	Pharmacogenomics
Pharmacogenetics	(a subset of pharmacogenomics) The study of variations in DNA sequence as related to drug response (ICH15). CIOMs VII (2005): the study of interindividual variations in DNA sequence related to drug disposition (pharmacokinetics) or drug action (pharmacodynamics) that can influence clinical response.
Pharmacogenomics	The study of variations of DNA and RNA characteristics as related to drug response (ICH15). CIOMs VII (2005): the application of genomic technologies to elucidate disease susceptibility, drug discovery, pharmacological function, drug disposition and therapeutic response.
Phenotype	Observable characteristics influenced by genotype and by other additional factors, e.g., the environment.
Polymorphism	Occurrence of more than one form (or morph) of a (functional) phenotype in a frequency that is stable in different populations, and a frequency above 1%
PK	Pharmacokinetics
PM	Poor metaboliser

Precision	Expresses the closeness of agreement between a series of measurements obtained from multiple sampling of the same homogenous material under the prescribed conditions. Precision may be considered at three different levels: repeatability, intermediate precision and reproducibility.
Rare variant	Allelic variation with a frequency of <1 %
RCT	Randomized controlled trial
ROC	Receiver operating characteristics
Repeatability	the precision under the same operating conditions over a short interval of time (intra-assay precision).
Representative sample	A small group whose (pharmacogenomic) characteristics accurately reflect those of the larger population from which it is drawn.
Reproducibility	The inter-laboratory precision, which may be determined as the variation of repeat test results for one test system in different laboratories (intra platform variation) or for different test systems (inter-platform precision).
RNA	Ribonucleic acid
ROC	Receiver operating characteristic curves
RT-PCR	Real time polymerase chain reaction
SBT	Direct sequencing-based typing
SMRT	Pacific Biosciences Single Molecule Real Time
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SSOP	Sequence specific oligonucleotide probe
SSP	Sequence specific primer
ST	Sulphotransferase
Somatic DNA	DNA from cells that make up all the internal organs, like bones, blood and connective tissue
TDM	Therapeutic drug monitoring
TPMT	Thiopurine methyl transferase
UM	Ultrarapid metaboliser, individual exhibiting increased drug metabolising enzyme activity compared to extensive metabolisers
WES	Whole exome sequencing. Sequence by NGS all protein-coding exomes after capture utilizing hybridization to a whole-exome bait set designed to enrich DNAs in all protein coding portion of the genome.

WGS	Whole genome sequencing. Sequencing of the entire genome, usually via a random fragment (shotgun) and to sufficient coverage to ensure adequate representation of all alleles
UGT	UDP-glucuronosyltransferase

ANNEX I

Examples of pitfalls in genomic studies

- Differences in the ethnicity and frequencies of the relevant mutations among patients recruited in the different trials, also causing different outcomes.
- Different kinds of patient populations in different studies, leading to discrepant results.
- Pooling data derived from exposure to drugs with different pharmacology to achieve higher statistical power.
- Study on the influence by the polymorphism in the *CYP2C9* and *VKORC1* genes on two drugs, acenocoumarol and phenprocoumon, combined together in one group, despite the fact the drugs display very different half-lives, affecting the relative influence of the CYP2C9 polymorphism.
- Choice of inappropriate analyses of somatic DNA instead of germline DNA.
- Studies evaluating influence of *CYP2D6* polymorphisms on the response to tamoxifen in breast cancer leading to contradictory conclusions using somatic DNA.
- Studies with different doses of tamoxifen have been combined into the analyses, and lack of planned analysis to distinguish appropriate populations (premenopausal and post-menopausal women).