

doi: <http://dx.doi.org/10.11606/issn.1679-9836.v.94i4p201-210>

Molecular diagnosis in the clinical practice: an endocrinologist's perspective

Diagnóstico molecular na prática clínica: visão do endocrinologista

Bruno Ferraz-de-Souza

Ferraz-de-Souza B. Molecular diagnosis in the clinical practice: an endocrinologist's perspective / *Diagnóstico molecular na prática clínica: visão do endocrinologista*. Rev Med (São Paulo). 2015 out.-dez.;94(4):201-10.

ABSTRACT: Endocrinology is one of several medical specialities that have been gradually transformed by a deeper understanding of the molecular bases of disorders. Genetic testing with the purposes of defining a precise molecular diagnosis has increasingly gained space in the routine assessment of patients with endocrinopathies, and the advent of massive parallel sequencing (MPS) is boosting the incorporation of molecular information in the clinic. The main benefit of genetic testing is diagnostic precision, resulting in improved and individualized care for patients and family members, and better disease prevention. However, genetic tests are not infallible and may bear several potential risks, being thus indicated when clinical suspicion is strong and the benefit of determining a molecular diagnosis is unambiguous. In this review, these evolving concepts and current indications for molecular diagnosis in endocrinology will be explored. Molecular tools will be revised and contextualised, including those aimed at identifying changes in gene dosage (karyotype, FISH, MLPA, aCGH, SNParray) or in the DNA nucleotide sequence (allele-specific PCR, RFLP, Sanger sequencing, MPS or NGS). Finally, matters surrounding the complex attribution of biologically relevant functional impact to identified DNA variants will be explored, together with the challenges brought by high throughput molecular analysis. These are exciting times for molecular endocrinology, and hopefully soon a translation to multiple benefits for patients will be self-evident.

Keywords: Genetic testing; Endocrinology; Mutation; DNA mutation analysis; Transcription, genetic; Genetic markers; Sequence analysis, DNA/methods; Molecular sequence data.

RESUMO: A endocrinologia é uma de muitas especialidades médicas que vem sendo transformada pelo maior conhecimento das bases moleculares das doenças. O teste genético com o propósito de definir o diagnóstico molecular vem ganhando espaço na avaliação rotineira de pacientes com endocrinopatias, e o advento do sequenciamento paralelo em larga escala (SPLE) está ampliando o potencial de incorporação da informação molecular na prática clínica. O principal benefício do teste genético é a precisão diagnóstica, resultando em melhora e individualização do tratamento de pacientes e seus familiares, e da prevenção de doenças. Entretanto, testes genéticos não são infalíveis e podem trazer alguns riscos potenciais, estando portanto indicados quando a suspeita clínica é forte e o benefício do diagnóstico molecular é claro. Nesta revisão serão explorados conceitos genéticos em evolução e as atuais indicações de diagnóstico molecular em endocrinologia. Ferramentas moleculares serão revisadas, incluindo aquelas visando a identificação de alterações no número de cópias gênicas (cariótipo, FISH, MLPA, aCGH, SNParray) ou na sequência nucleotídica do DNA (PCR alelo-específica, RFLP, sequenciamento Sanger, SPLE ou NGS). Ainda, serão discutidos os percalços da difícil atribuição de impacto funcional e relevância biológica a variantes de DNA identificadas, e os desafios atuais das tecnologias de análise molecular de larga escala. A endocrinologia molecular vive um momento transformador e empolgante, com o potencial de, em breve, se traduzir em múltiplos benefícios no cuidado de nossos pacientes.

Descritores: Testes genéticos; Endocrinologia; Mutação; Análise mutacional de DNA; Transcrição genética; Marcadores genéticos; Análise de sequência de DNA/métodos; Dados de sequência molecular.

PhD in Endocrinology, University College London, UK. Attending Physician and Principal Investigator, Endocrinology/LIM-18, Hospital das Clínicas, University of Sao Paulo School of Medicine.

Correspondence to: Av. Dr. Arnaldo, 455. Sala 3324 (LIM-18). Cerqueira Cesar. São Paulo, SP. CEP: 01246-903.

E-mail: bferrazd@gmail.com or bruno.ferraz@hc.fm.usp.br

INTRODUCTION

The genetic bases of endocrinopathies have been progressively unravelled over the past decades, resulting in an increasing availability and clinical relevance of molecular diagnostics. Initially, a defined molecular pathogenesis was established for rarer endocrine conditions such as developmental defects and hereditary disorders, but more recently the genetic contribution to common endocrine diseases, such as obesity and osteoporosis, have begun to be understood¹⁻⁴. The advent of automated Sanger sequencing in the 1990s and the recent emergence of massive parallel sequencing have driven these advances^{5,6}. Indeed, while it may be still early days to consider including molecular tests in the routine assessment of every patient presenting with an endocrine complaint, the future endocrinologist will certainly need to be able to aptly construe the significance of molecular diagnostic reports.

In certain scenarios, which will be detailed throughout this review, a precise molecular diagnosis has a clear benefit and allows for better care, impacting on therapeutic decisions, on the prevention of complications and on accurate genetic counselling of affected individuals and family members. Nevertheless, genetic testing is not infallible and has technical limitations, and therefore should only be pursued when the clinical suspicion of a molecular defect is high and the benefit of the test is clear. More importantly, it is becoming apparent that most allelic variants found on individual genomes are devoid of an immediately clear functional effect, and thus the finding of sequence variants on candidate genes for endocrine diseases should always be followed by an exhaustive appraisal of its potential biological relevance^{7,8}.

In this review, the current applicability of genetic testing in endocrinology and the most common molecular tools used for this intent will be discussed. Of note, 'genetic testing' will be used throughout this text as an umbrella definition for any procedure aiming at establishing a molecular diagnosis.

CLINICAL SUSPICION OF A MOLECULAR DEFECT

Several clinical, laboratory and image findings might point towards a genetic cause for an endocrinopathy. For example, hyperparathyroidism is a fairly common disorder in postmenopausal women, but if manifesting in a young individual (< 30 years-old) it should arise the suspicion for type 1 multiple endocrine neoplasia (MEN1) due to *MEN1* mutations, prompting investigation for pituitary and enteropancreatic tumours and additional familial cases^{9,10}. Conversely, if imaging reveals jaw tumours in such patient, a differential diagnosis with

hyperparathyroidism jaw tumour syndrome caused by *CDC73* (HRPT2, parafibromin) mutations is important, due to the increased risk of parathyroid carcinoma¹¹. Thus, the combination of history, symptoms and signals, and traditional diagnostic workup will point to the suspicion of a genetic cause in most cases.

Besides such disease-specific indicators, actively inquiring and obtaining a fully detailed family history is essential, even if similar familial cases are not spontaneously brought up by the patient. Indeed, many genetic disorders may present with variable expressivity within families, justifying the investigation of associated features even if the main phenotypic characteristic is absent in family members¹². Returning to the MEN1 example given above, a patient presenting with hyperparathyroidism may fail to recollect precisely similar cases in the family, but might mention relatives with a history of recurrent peptic ulcer or milk discharge, which could be manifestations of MEN1-associated tumours.

While obtaining the family history, it is also important to explore the possibility of consanguinity, which could favour the occurrence of autosomal recessive disorders. Finally, when obtaining a sample for genetic testing from an individual, it is wise to try to obtain samples from 1st degree relatives as well, especially if obtaining such samples in the future would be logistically challenging – the availability of related samples is paramount for analysing phenotypic segregation of identified genetic variants. As it will be discussed later, evidence of co-segregation helps to ascertain the biological impact of genetic variants⁸.

INDICATIONS FOR GENETIC TESTING

Any diagnostic procedure based on an individual's DNA, RNA, chromosomal or protein samples aiming to detect genotypes associated to genetic or hereditary disorders can be considered a genetic test. As detailed in major guidelines, genetic testing has several benefits but may also have risks^{13,14}.

The main benefit of a genetic test is allowing for a specific and precise molecular diagnosis. Important additional benefits stem from this diagnostic precision, such as improved and individualized follow-up, the possibility of offering a specific treatment, with better prediction of response, and prevention of complications. Further benefits of a molecular diagnosis include the anticipation of hereditary transmission and the appropriate vigilance and care of identifiable family members. A few examples of endocrine scenarios where a genetic test is beneficial include:

- Identifying a *RET* mutation in a patient with medullary thyroid carcinoma allows early recognition of potentially affected family members and better disease control¹⁵;

- Identifying a *MEN1* mutation in a patient with hyperparathyroidism enables vigilance for pituitary and enteropancreatic tumours, and screening of relatives⁹;

- Identifying *HNF1A* or *HNF4A* defects in a patient with diabetes could shift their treatment to sulfonylureas, with better response and compliance in comparison to daily insulin administration¹⁶;

- Identifying a *PROPI* mutation in a child with short stature will prompt vigilance of adrenal insufficiency, a late but potentially lethal manifestation¹⁷.

On the other hand, genetic tests may bear several potential risks. Amongst these, the most serious ones are related to the potential breach of ethical, moral and legal principles. In order to avoid such risks, it is essential to offer appropriate pre- and post-test genetic counselling, making sure that the patient understands the nature of genetic testing, its advantages and limitations, and exploring the potential consequences of a molecular diagnosis in the personal and familial realms, taking into account the individual's social and cultural contexts. It is also important to obtain informed consent, signed by the patient or their legal representative and by the attending physician, recording all aspects discussed during counselling. These precautions are particularly important when a molecular

diagnosis carries a heavy burden in terms of prognosis: for example, identifying a germline *TP53* mutation in an adolescent means that tumour vigilance will be necessary throughout life, and, for some, such perspective might result in overwhelming anxiety¹⁸. Finally, as with all aspects of doctor-patient relationship, protecting the confidentiality of the individual's genetic identity is paramount.

Other potential risks of genetic testing derive from the nature of the laboratory test employed: technical limitations, errors in its interpretation (e.g., a negative result in a mutational analysis directed at hot spots in a candidate gene does not exclude the possibility of damaging variants in other parts of the gene), and lack of technical quality.

Hence, taking into account benefits and risks, genetic tests are currently indicated when clinical suspicion is strong and the benefit of determining a molecular diagnosis is unambiguous. While compiling a comprehensive list of all indications for genetic testing in endocrinology is unfeasible due to the fast-paced evolution of knowledge in the field, a selection of endocrinopathies for which a molecular diagnosis is currently considered to be beneficial is shown on Table 1. Several more indications may exist, being increasingly recognised in light of their potential benefits in personalising care, and therefore indicating a genetic test must be an individualised decision.

Table 1. Selected endocrinopathies for which genetic testing is indicated due to clear benefits of a molecular diagnosis

Disease (OMIM identifier)	Gene	When to suspect and test
Familial medullary thyroid carcinoma (#155240)	<i>RET</i>	All patients with medullary thyroid carcinoma
Multiple endocrine neoplasia type 2 (MEN2, #171400 and #162300)	<i>RET</i>	Association of MEN2-related tumours (medullary thyroid carcinoma, pheochromocytoma and others), see Wells et al. ¹⁵
Multiple endocrine neoplasia type 1 (MEN1, #131100)	<i>MEN1</i>	Association of two or more MEN1-related tumours (parathyroid, enteropancreatic, pituitary), see Thakker et al. ⁹
Pheochromocytoma and paragangliomas (#171300)	<i>VHL, RET, TMEM127, MAX, SDHB, SDHD</i> , etc	Early-onset, familial, bilateral or malignant pheochromocytomas; early-onset paragangliomas
Li-Fraumeni syndrome (#151623)	<i>TP53</i>	Early onset of multiple tumours, such as soft tissue sarcomas and osteosarcomas, breast cancer and adrenocortical carcinoma, with autosomal dominant inheritance
Glucocorticoid-remediable aldosteronism (#103900)	<i>CYP11B1 /CYP11B2</i> chimeric gene	Early-onset hyperaldosteronism, autosomal dominant inheritance
Congenital adrenal hyperplasia due to 21-hydroxylase deficiency (#201910)	<i>CYP21A2</i>	When parents of an index case (classic form) plan a new gestation and would consider prenatal dexamethasone treatment

OMIM, online Mendelian inheritance in men database (<http://www.ncbi.nlm.nih.gov/omim>). Adapted from Ferraz-de-Souza et al.⁴⁷

TOOLS AND METHODS COMMONLY USED FOR ESTABLISHING A MOLECULAR DIAGNOSIS

Obtaining samples and choosing the appropriate tool for molecular diagnosis

Most frequently, a molecular diagnosis is made upon samples of genomic DNA, i.e. nuclear chromosomal DNA present in all cells and passed from one generation to the next. In certain situations, however, samples of tissue- or cell-specific DNA and/or RNA are analysed to identify somatic variants and patterns of gene expression (transcriptome). It is also possible to perform a molecular investigation by obtaining samples of live cells (e.g. skin fibroblasts), allowing the characterisation of protein expression and function, for example. While transcriptomics and proteomics are surging in certain areas of endocrinology such as in cancer care as a means to understand and predict disease behaviour, RNA- and protein-based diagnosis are beyond the scope of this review, which will focus hereafter on DNA-based techniques.

Pathogenic variants in genomic DNA are termed germline mutations, and therefore can be transmitted to offspring; these correspond to the main types of mutations discussed throughout this text. In contrast, somatic DNA variants are those restricted to specific cells and their progeny in their course of division and differentiation, other than germ cells. Thus, somatic variants cannot be passed on to offspring. Endocrine examples of non-inheritable genetic disorders due to somatic mutations are fibrous dysplasia of bone and McCune Albright syndrome caused by *GNAS* defects¹⁹.

Samples of genomic DNA can be easily obtained from peripheral blood leukocytes following a routine blood draw, generally yielding good amounts of high quality DNA after extraction. It is also possible to obtain good quality DNA from saliva samples, which can be particularly interesting for field work (analysis of large families or population, for example) and for obtaining DNA samples from children.

Once the sample is secured, it is important to choose molecular diagnostic procedures that are most appropriate for the suspected defect. This important step cannot be overlooked, because if the wrong molecular tool is chosen, a “normal” result will not exclude that a molecular defect, which was not appropriately investigated, is in fact causing the disease. Broadly, DNA defects can be simplified into two categories (at the expense of ill defining particular situations, of course): changes in the number of gene copies (gene dosage) or changes in the DNA nucleotide sequence. Changes in gene dosage include abnormalities in chromosomal numbers, chromosomal translocations, deletions or duplications, large insertions and deletions (indels) and gene copy number variants. In these situations, the nucleotide sequence is generally unchanged (except

where the breaking points for structural changes are located), but gene copies are in excess (duplication) or lacking (deletion) in the genome, resulting in altered gene dosage²⁰. Changes in the DNA nucleotide sequence, on the other hand, range from single nucleotide variants (SNVs) to small insertions and deletions, and may affect the coding frame resulting in defective proteins. As will be shown below, tools for investigating these two categories of DNA variation are largely different, and it is up to the clinician to individually choose which direction to follow, based on the knowledge of the genetic defects most commonly associated with each endocrinopathy.

Searching for chromosomal abnormalities, large indels and copy number variants

A first step to analysing large structural abnormalities in chromosomes is the karyotype, a genetic tool that has been around for decades and still is very useful, particularly in the approach to children with disorders of sex development, short stature and pubertal abnormalities. A karyotype is usually obtained from culturing peripheral blood leukocytes, and analysing the number and appearance of chromosomes under light microscopy. Examples of situations in which a karyotype can guide the diagnostic approach or establish the diagnosis itself include the investigation of ambiguous genitalia in the neonate, of girls with short stature due to Turner syndrome (45,X and variants) and of pubertal abnormalities in boys due to Klinefelter syndrome (47,XXY)²¹.

Refining cytogenetics in the 1980s led to the development of **Fluorescent *In Situ* Hybridization**, or FISH, based on the use of fluorescent probes that bind to specific chromosomal regions, allowing the detection of finer chromosomal deletions, insertions or translocations. Unlike a normal karyotype, FISH requires a pre-test knowledge of which chromosomal regions to analyze. In the context of endocrine disorders, FISH can be used in the investigation of 46,XX disorders of sex development to ascertain the presence of a translocated *SRY*, and to diagnose Prader-Willi syndrome (15q11-13 critical region).

In recent decades, major interest has been placed onto unravelling submicroscopic structural abnormalities. Copy number variants (CNVs) can result from deletions or duplications involving exons, whole genes or a group of genes, and can be investigated in an individual gene basis or in genomic scale.

Individual gene CNVs are best analysed using **Multiplex Ligation-dependent Probe Amplification** (MLPA), given its practicality and accuracy. MLPA employs fluorescent-labelled oligonucleotides that bind specific DNA regions and require physical proximity in order to be amplified, allowing the comparison of signal intensity for regions of interest versus neighbouring regions. Applications for MLPA in endocrinology include

the search for gene deletions in monogenic diabetes (e.g. *GCK* deletions in *MODY2*) and in short stature (e.g. *SHOX* deletion)²². More recently, the advent of microarray technologies has allowed the investigation of CNVs in genomic scale using techniques such as array Comparative Genomic Hybridization (aCGH) or Single Nucleotide Polymorphism (SNP) arrays. In aCGH, a whole proband genome is individually compared to a reference genome, allowing detection of gains or losses, and through the detailed genotyping generated by SNP arrays, duplicated or deleted regions also become apparent. These exciting techniques are somewhat limited to the research realm, but have been increasingly used for the diagnosis of complex syndromic phenotypes²³.

Bearing in mind that most genes located in autosomes are expressed in double dosage, from copies on both maternal and paternal alleles (with the notable exception of those under genomic imprinting), complete or partial deletion of genes can result in biologically relevant changes in the dosage of codified proteins. When a deletion affects only one of the copies, for example, resulting haploinsufficiency might have a functional manifestation. Therefore, when looking for sequence variants in a candidate gene for an autosomal dominant disorder, if sequencing results return absolutely normal, one might consider the possibility of a heterozygous deletion, which would go undetected on DNA sequencing. In these cases, using MLPA would allow the identification of haploinsufficiency as a potential cause for the observed phenotype. Situations like these alert the clinician to the realisation that “normal” results in specific genetic tests not always mean that a molecular defect is absent.

Searching for single nucleotide variants and small indels

Changes in the DNA nucleotide sequence, such as single nucleotide (point) substitutions or small insertions or deletions, have long been sought as the cause of genetic disorders. Indeed, an error in the DNA coding frame may reasonably lead to a dysfunctional protein with physiological consequences. Even though DNA sequencing using chain terminators had been devised by Frederick Sanger in 1977, it was the development of polymerase chain reaction (PCR) as a means to easily amplify DNA by Mullis and others, in 1988, that really propelled molecular genetics into a new era²⁴. In general, for traditional mutational analysis, several rounds of DNA amplification are necessary in order to obtain sufficient amounts of the target region. Hence, PCRs became a staple of every molecular biology lab from the 1990s onwards.

Polymerase chain reactions are easily performed on a thermocycler, using a thermostable DNA polymerase (most frequently Taq, from the bacteria *Thermus aquaticus*) and oligonucleotides (commonly referred to as primers) that recognise and hybridise specific DNA sequences at

the borders of the target region. Therefore, an important characteristic (and perhaps limitation) of PCR is that the target area, or at least its boundaries, has to be previously known. Oligonucleotides are specifically designed for each piece of DNA to be amplified, manually or using freely available online software platforms (e.g. Primer3Plus), based on a reference sequence for the human genome.

Of note, several online resources have become available that provide guidance navigating the increasing wealth of genetic and molecular knowledge, and some are worth mentioning. Genome browsers provide up-to-date validated information on a candidate gene’s location, structure and sequence (for example, Ensembl, UCSC genome browser, NCBI’s Entrez Gene) and therefore can provide a reference guiding oligonucleotide design. Official gene symbols are curated by the Human Genome Nomenclature Committee (HGNC) and can be checked at www.genenames.org to make sure the correct candidate gene is being selected, since gene symbols frequently overlap. Finally, the Human Genome Variation Society (HGVS) issues guidelines on how to accurately and uniformly describe sequence variants (found at www.hgvs.org/mutnomen/), informing the report of identified variants.

PCR-based techniques

Straightforward PCR-based techniques such as allele-specific PCR and restriction fragment length polymorphism (RFLP) analysis can provide quick and accurate molecular diagnosis when searching for a specific pre-known nucleotide variant. Thus, these techniques are particularly useful for confirming variants identified by sequencing, for genotyping control populations in order to infer allelic frequencies, and for screening affected family members following the identification of an index case.

Allele-specific PCR adds a twist to standard PCR: normally, oligonucleotides target an invariant part of the genome to secure the amplification of a in-between region where the variant might be; in allele-specific PCR, however, one of the oligonucleotides is directed at the variant itself. This way, detection of an amplification product on agarose gel electrophoresis indicates that the variant of interest is indeed present. Since the whole experiment can be done cheaply in a few hours, allele-specific PCR is particularly useful for rapid genotype screening of large populations.

RFLP employs differential DNA digestion by restriction endonucleases to discriminate between wild type and altered sequences. Restriction enzymes were originally discovered in bacteria, where they serve as a defence mechanism against viral infections, and have the ability to cut the DNA at very specific nucleotide sequences, known as recognition sites. A profusion of such enzymes are now available, both natural and recombinant, and therefore several nucleotide sequences can be targeted. Thus, in order for this technique to be applicable, it is necessary

that the DNA variant of interest introduces or abolishes an endonuclease recognition site – several online tools analyse input nucleotide sequences and display existing restriction sites (for example, NEBcutter) in order to aid experimental design. RFLP has been extensively used to confirm variants identified by sequencing and to screen family members.

Sanger sequencing

Detecting terminally labelled amplification products through capillary electrophoresis constitutes the basis of Sanger sequencing, the most widely used technique to sequence DNA for the purposes of molecular diagnosis (so far). Automated sequencing protocols involve a second PCR, termed the “sequencing reaction”, using labelled chain-terminating dideoxynucleotides that allow the identification of A, T, C or G after capillary electrophoresis. The result is graphically displayed as an electropherogram, which can be visually analysed in order to compare the sequence of interest to the reference (wild type). Given the nature of chain-terminating nucleotides used for sequencing, only one primer is used in the reaction, determining the reading direction, sense or antisense. Although sequencing can be routinely done only in one direction for the purposes of screening, if there are quality issues (weak peak intensity, too much background noise) or if a variant is found and needs confirmation, the process must be repeated using the opposite primer.

Examples of the application of Sanger sequencing for mutational analysis and genotyping in endocrinology are bountiful in the literature. Some practical tips may help beginners eyeballing several electropherograms: a) updated reference DNA sequences (to be used for comparison) can be obtained from the genome browsers mentioned above; b) considering that genome assemblies and gene annotations may vary from time to time, it is critical to take note and report the identifier of the reference sequence used; and c) be sure to obtain images of critical findings: when reporting Sanger sequencing results, electropherograms displaying the identified variants are customarily shown²⁵.

Massive parallel sequencing

For all the usefulness of Sanger sequencing, and the vast experience amassed with this methodology, it has two notable limitations hampering its application to the discovery of new genes in endocrinopathies: analysed regions have to be pre-selected, meaning that candidate genes have to be previously known or hypothesized, and it can become quite laborious if several sequences need to be analysed (it bears saying that, on average, good Sanger reading is obtained for stretches up to 500 base pairs). From 2005 onwards, high throughput sequencing technologies became available, revolutionising mutational analysis to a new era of genomic scale, big data and fast results; these

technologies are commonly referred to as “next generation sequencing” but are perhaps better described as massive parallel sequencing (MPS)²⁶.

A few MPS platforms exist, and methodology and throughput are variable. All share the ability to sequence several different stretches of DNA at the same time, the possibility of barcoding, allowing simultaneous analysis of more than one individual, and generation of large data²⁷. Therefore, while the hands-on experimental part of MPS might be relatively simple, the technology (platforms, consumables, etc) is still expensive and the bioinformatics analysis is still somewhat burdensome – both are fast improving, rendering MPS more accessible. Possible applications include the analysis of the whole genome (whole genome sequencing, or WGS), of all coding exons of known genes (whole exome sequencing, WES) and of selected gene panels.

This paradigm shift in molecular analysis has been fully embraced in the endocrine field, both in research and at the clinic²⁸. If the use of WGS in the clinical front has been so far limited to the diagnosis of severe complex phenotypes²⁹, WES is now easily available as a diagnostic test and can be very useful for obtaining a molecular diagnosis in endocrinopathies for which several candidate genes exist, such as disorders of sex development^{6,30}. Indeed, implementing MPS in endocrine research has allowed the discovery of new molecular mechanisms in rare and common disorders, such as precocious puberty and obesity, respectively^{2,31}. Due to cost-effectiveness, candidate gene panels are promising to take Sanger sequencing’s place as first choice test in the molecular diagnosis of genetically heterogeneous endocrine disorders, such as monogenic diabetes and osteogenesis imperfecta³².

CONFIRMING AND ASCERTAINING THE BIOLOGICAL IMPACT OF DNA VARIANTS

Identifying a sequence variant within a candidate gene is always an exciting moment in a molecular biology lab. Nevertheless, several rounds of subsequent analysis are necessary to confirm the finding and to ascertain its biological impact, i.e. to try to establish causality for the observed phenotype. This can be a particularly daunting task in view of the hundreds of thousands of DNA variants we all carry in our genomes, most of which have no identifiable repercussion^{33,34}.

Firstly, the sequence variant identified by Sanger or MPS must be confirmed. For that, it is good practice to repeat the experiment from the beginning, from the initial DNA amplification (some go as far as obtaining a new DNA sample to discard contamination). If the observed results are still the same, they need be confirmed by a different methodology - allele-specific PCR or RFLP can be used for this purpose.

Once confirmed, the potential biological impact of

the identified variant must be established. It is wise to start by checking if the variant is novel or has been previously described in association with the observed phenotype, or as a polymorphic variant. Several databases might aid this search, from PubMed, GoogleScholar and OMIM to Ensembl, dbSNA, HapMap, 1000Genomes, ExAC and Exome Variant Server. Frequently, pre-analysed MPS reports will already come with this information. For variants identified by Sanger sequencing, it is crucial to correctly describe variants according to HGVS guidelines in order to guarantee that this search is effective.

Whenever a DNA variant is postulated to be associated with a phenotypic trait or disorder, it becomes essential to investigate co-segregation within the family and its frequency in a control population of similar ethnic background. As discussed above, allele-specific PCR or RFLP can be a practical means for such familial and population screenings.

In general, attributing a functional impact for single nucleotide variants (SNVs, or point variants) is challenging, but their type and location within the structure of the gene may give important clues to potential functional effects. Traditionally, SNVs located inside the coding region of a gene are termed a) nonsense, if they introduce a premature stop codon (TAG, TAA or TGA); b) missense or nonsynonymous if they result in amino acid change; or c) synonymous, if the coded amino acid is unchanged. These concepts have been evolving with the advent of MPS, and the term ‘loss-of-function’ (LoF) has been increasingly used in the literature to describe variants more frequently expected to result in protein impairment, including nonsense or splice site-disrupting SNVs, indels disrupting the coding frame, or larger deletions³⁵.

Nonsense SNVs are expected to code for truncated proteins with reduced (or abolished) function; therefore, attributing a biological impact is usually more straightforward. For example, we have followed for several decades a patient with severe hereditary vitamin D-resistant rickets due to a nonsense mutation in the vitamin D receptor (VDR)²⁵. While wild type VDR has 427 amino acids, this patient’s homozygous mutation introduces a premature stop codon at residue 30, meaning that the resulting protein function is severely dysfunctional.

On the other hand, nonsynonymous SNVs may have different degrees of functional effect. The investigation of the potential functional effect of a nonsynonymous SNV involves gathering several lines of evidence supporting a causative role, such as: a) comparing the physicochemical properties of wild type and changed amino acids; b) analysing the degree of amino acid conservation amongst species (UCSC genome browser and ClustalW2 are useful to this purpose); c) locating the change amino acid within the protein functional domains (this information can be

found at Uniprot); d) *in silico* prediction of functional effect (several freely available web-based software tools can be used, such as PolyPhen, SIFT, MutationTaster and Mutalyzer); d) *in silico* protein modelling, aiming to identify changes in tertiary protein structure; and e) *in vitro* functional studies, which require previous knowledge of protein function and are often laborious, requiring considerable laboratory set up and expertise with the devised experiment, but are generally regarded as strong evidence of effect. For example, in the study of several different mutations in the pivotal nuclear receptor steroidogenic factor-1 (SF-1, NR5A1), many of these lines of evidence were gathered to establish a functional impact on adrenal, gonadal and reproductive phenotypes³⁶.

Finally, despite the lack of a direct effect on the protein amino acid sequence, synonymous SNVs or those located in untranslated, intronic or promoter regions may still have important biological repercussions, particularly if affecting gene expression or mRNA splicing. For example, somatic mutations in the promoter region of *TERT* may lead to thyroid cancer, with potential diagnostic implications³⁷, and homozygous splice site mutations in the GH-releasing hormone receptor (*GHRHR*) can lead to short stature due to growth hormone deficiency³⁸.

HIGH THROUGHPUT TECHNOLOGIES: CHALLENGES OF A NEW ERA

Right now, in endocrinology and throughout health sciences, we are fast paced towards Genomic or Personalised Medicine³⁹. Ever-improving MPS technologies are expected to bring down costs, and better navigation and understanding of accumulated genomic data will translate into biologically relevant information, so that in the not-so-far future individuals may choose to fully sequence their full genomes with the hopes of more accurately predicting and preventing diseases^{40,41}. It will certainly be challenging for clinicians to incorporate this information into their assessment and management of individual patients, but the potential for precision is seemingly undeniable.

As a result of widespread availability of MPS, several challenges are currently being met in research and clinical settings. For example, ascertaining the impact of the hundreds of thousands of DNA variants identified in individual genomes is not easy; the term ‘variants of uncertain (or unknown) significance’ (VUS) has arisen to collectively describe variants without an immediately obvious causal role for a trait or disorder⁴². Additionally, there is an ongoing debate on the ethical implications of disclosing incidental genomic findings to patients, i.e. DNA variants known to predispose to diseases other than the one

being investigated at that particular moment⁴³⁻⁴⁵. Equally important is the concern of how to best protect the privacy of an individual's genomic data – as with any genetic testing, confidentiality and data protection are paramount.

Thousands of variants identified in individual genomes may indeed not have a detectable biological effect. In order to differentiate truly silent variants from those with a functional impact, huge efforts are being put into sequencing entire genomes of large populations, trying to ascertain both genetic variability and the relationship of variants, isolated or in combination, with phenotypic traits and diseases⁴⁶. To accompany the identification of all these DNA variants, new technologies will certainly need to be developed and fine-tuned to allow equally massive means of interpretation of functional effect, which will hopefully permit translation of this knowledge to the clinical setting and a true benefit to our patients.

REFERENCES

1. Arboleda VA, Lee H, Parnaik R, Fleming A, Banerjee A, Ferraz-De-Souza B, Et Al. Mutations in the PCNA-binding domain Of CDKN1C Cause image syndrome. *Nat Genet.* 2012;44(7):788-92. doi: 10.1038/ng.2275.
2. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med.* 2013;368(26):2467-75. doi: 10.1056/NEJMoa1302160.
3. Ramachandrapa S, Farooqi IS. Genetic approaches to understanding human obesity. *J Clin Invest.* 2011;121(6):2080-6. doi: 10.1172/JCI46044.
4. Richards JB, Zheng HF, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Genet.* 2012;13(8):576-88. doi: 10.1038/nrg3228.
5. Katsanis SH, Katsanis N. Molecular Genetic testing and the future of clinical genomics. *Nat Rev Genet.* 2013;14(6):415-26. doi: 10.1038/nrg3493.
6. De Bruin C, Dauber A. Insights from exome sequencing for endocrine disorders. *Nat Rev Endocrinol.* 2015;11(8):455-64. doi: 10.1038/nrendo.2015.72.
7. Collins FS, Hamburg MA. First FDA authorization for next-generation sequencer. *N Engl J Med.* 2013;369(25):2369-71. doi: 10.1056/NEJMp1314561.
8. Macarthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature.* 2014;508(7497):469-76. doi: 10.1038/nature13127.
9. Thakker RV, Newey PJ, Walls GV, Bilezikian J, Dralle H, Ebeling PR, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab.* 2012;97(9):2990-3011. doi: 10.1210/jc.2012-1230.

CONCLUSION

Molecular diagnosis in endocrinology is fast moving from rare disorders confined to academic institutions to common endocrinopathies present in the everyday clinic. At the moment, genetic testing is warranted when the benefits of a molecular definition are clear and outweighs potential risks, which always should be assessed and discussed with the patients. Examples of such situations where benefits are clear include the management of multiple endocrine neoplasia syndromes and some isolated endocrine cancers. Several types of genetic tests can be done, and their choice depends on the disorder under study and technical availability. Soon, increased accessibility to massive parallel sequencing technologies and improved interpretation of the functional impact of DNA variants identified this way will bring molecular information to the clinical setting with full force, helping us offer a more personalised care to our patients.

10. Eastell R, Brandi ML, Costa AG, D'Amour P, Shoback DM, Thakker RV. Diagnosis of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. *J Clin Endocrinol Metab.* 2014;99(10):3570-9. doi: 10.1210/jc.2014-1414.
11. Carpten JD, Robbins CM, Villablanca A, Forsberg L, Presciuttini S, Bailey-Wilson J, et al. HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet.* 2002;32(4):676-80. doi: 10.1038/ng1048.
12. Congdon T, Nguyen LQ, Nogueira CR, Habiby RL, Medeiros-Neto G, Kopp P. A novel mutation (Q40P) In PAX8 associated with congenital hypothyroidism and thyroid hypoplasia: evidence for phenotypic variability in mother and child. *J Clin Endocrinol Metab.* 2001;86(8):3962-7. Available from: <http://press.endocrine.org/doi/pdf/10.1210/jcem.86.8.7765>.
13. Sociedade Brasileira de Genética Clínica. Genética médica: teste laboratorial para diagnóstico de doenças sintomáticas. Projeto diretrizes da Associação Médica Brasileira e Conselho Federal De Medicina. São Paulo: AMB/CFM; 2004. Disponível em: <http://www.sbgm.org.br/cfm/testelab.pdf>.
14. Ross LF, Saal HM, David KL, Anderson RR. Technical report: ethical and policy issues in genetic testing and screening of children. *Genet Med.* 2013;15(3):234-45. doi: 10.1038/gim.2012.176.
15. Wells SA, Jr., Asa SL, Dralle H, Elisei R, Evans DB, Gagel RF, et al. Revised American Thyroid Association Guidelines for the management of medullary thyroid carcinoma. *Thyroid.* 2015;25(6):567-610. doi: 10.1089/thy.2014.0335.
16. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell

- diabetes. *Nat Clin Pract Endocrinol Metab.* 2008;4(4):200-13. doi: 10.1038/ncpendmet0778.
17. Mendonca BB, Osorio MG, Latronico AC, Estefan V, Lo LS, Arnhold IJ. Longitudinal hormonal and pituitary imaging changes in two females with combined pituitary hormone deficiency due to deletion of A301,G302 in the PROP1 gene. *J Clin Endocrinol Metab.* 1999;84(3):942-5. <http://dx.doi.org/10.1210/jcem.84.3.5537>.
 18. McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MH, Eccles RA, et al. Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nat Rev Clin Oncol.* 2014;11(5):260-71. doi: 10.1038/nrclinonc.2014.41.
 19. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating Mutations of the stimulatory g protein in the McCune-Albright syndrome. *N Engl J Med.* 1991;325(24):1688-95. doi: 10.1056/NEJM199112123252403.
 20. Carson AR, Feuk L, Mohammed M, Scherer SW. Strategies for the detection of copy number and other structural variants in the human genome. *Hum Genomics.* 2006 Jun;2(6):403-14. doi: 10.1186/1479-7364-2-6-403.
 21. Lambert SM, Vilain EJ, Kolon TF. A practical approach to ambiguous genitalia in the newborn period. *Urol Clin North Am.* 2010;37(2):195-205. doi: 10.1016/j.ucl.2010.03.014.
 22. Funari MF, Jorge AA, Pinto EM, Arnhold IJ, Mendonca BB, Nishi MY. Cryptic intragenic deletion of the SHOX Gene in a family with Leri-Weill dyschondrosteosis detected by Multiplex Ligation-Dependent Probe Amplification (MLPA). *Arq Bras Endocrinol Metabol.* 2008;52(8):1382-7. <http://dx.doi.org/10.1590/S0004-27302008000800029>
 23. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet.* 2010;86(5):749-64. doi: 10.1016/j.ajhg.2010.04.006.
 24. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51(Pt 1):263-73. doi: 10.1101/SQB.1986.051.01.032.
 25. Damiani FM, Martin RM, Latronico AC, Ferraz-De-Souza B. Normal bone mass and normocalcemia in adulthood despite homozygous vitamin D receptor mutations. *Osteoporos Int.* 2015;26(6):1819-23. doi: 10.1007/s00198-015-3076-3.
 26. Metzker ML. Sequencing technologies - The next generation. *Nat Rev Genet.* 2009;11(1):31-46. doi: 10.1038/nrg2626.
 27. Feero WG, Guttmacher AE, Collins FS. Genomic medicine - an updated primer. *N Engl J Med.* 2010;362(21):2001-11. doi: 10.1056/NEJMra0907175.
 28. Jacob HJ, Abrams K, Bick DP, Brodie K, Dimmock DP, Farrell M, et al. Genomics in clinical practice: lessons from the front lines. *Sci Transl Med.* 2013;5(194):194cm5. doi: 10.1126/scitranslmed.3006468.
 29. Jacob HJ. Genetic diagnosis through whole-exome sequencing. *N Engl J Med.* 2014;370(11):1069.
 30. Achermann JC, Domenice S, Bachega TA, Nishi MY, Mendonca BB. Disorders of sex development: effect of molecular diagnostics. *Nat Rev Endocrinol.* 2015;11(8):478-88. doi: 10.1038/nrendo.2015.69.
 31. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature.* 2014;507(7492):371-5. doi: 10.1038/nature13138.
 32. Sule G, Campeau PM, Zhang VW, Nagamani SC, Dawson BC, Grover M, et al. next-generation sequencing for disorders of low and high bone mineral density. *Osteoporos Int.* 2013;24(8):2253-9. doi: 10.1007/s00198-013-2290-0.
 33. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-73. doi: 10.1038/nature09534.
 34. Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, et al. clinical interpretation and implications of whole-genome sequencing. *JAMA.* 2014;311(10):1035-45. doi: 10.1001/jama.2014.1717.
 35. Macarthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic survey of loss-of-function variants in human protein-coding genes. *Science.* 2012;335(6070):823-8. doi: 10.1126/science.1215040.
 36. Ferraz-De-Souza B, Lin L, Achermann JC. Steroidogenic factor-1 (SF-1, NR5A1) and human disease. *Mol Cell Endocrinol.* 2011;336(1-2):198-205. doi: 10.1016/j.mce.2010.11.006.
 37. Melo M, Da Rocha AG, Vinagre J, Sobrinho-Simoes M, Soares P. Coexistence of TERT promoter and BRAF mutations in papillary thyroid carcinoma: added value in patient prognosis? *J Clin Oncol.* 2015;33(6):667-8. doi: 10.1200/JCO.2014.59.4614.
 38. Marui S, Trarbach EB, Boguszewski MC, Franca MM, Jorge AA, Inoue H, Et Al. GH-releasing hormone receptor gene: a novel splice-disrupting mutation and study of founder effects. *Horm Res Paediatr.* 2012;78(3):165-72. doi: 10.1159/000342760.
 39. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med.* 2010;363(4):301-4. doi: 10.1056/NEJMp1006304.
 40. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Et Al. Clinical assessment incorporating a personal genome. *Lancet.* 2010;375(9725):1525-35. doi: 10.1016/S0140-6736(10)60452-7.
 41. Green RC, Farahany NA. Regulation: the FDA is overcautious on consumer genomics. *Nature.* 2014;505(7483):286-7. doi: 10.1038/505286a
 42. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet.* 2013;93(4):631-40. doi: 10.1016/j.ajhg.2013.08.006.

43. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 2013;15(7):565-74. doi: 10.1038/gim.2013.73.
44. Middleton A, Morley KI, Bragin E, Firth HV, Hurles ME, Wright CF, et al. No expectation to share incidental findings in genomic research. *Lancet.* 2015 Apr 4;385(9975):1289-90. doi: 10.1016/S0140-6736(14)62119-X.
45. Knoppers BM, Zawati MH, Senecal K. Return of genetic testing results in the era of whole-genome sequencing. *Nat Rev Genet.* 2015;16(9):553-9. doi: 10.1038/nrg3960.
46. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. doi: 10.1038/nature15393.
47. Ferraz-de-Souza B, Dias-da-Silva MR, Almeida MQ, Jorge AA. Aplicação clínica dos testes genético-moleculares em endocrinologia. In: Saad MJA, Maciel RMB, Mendonca BB, editores. *Endocrinologia.* 2nd ed. São Paulo: Atheneu; 2015.