GENETICS IN ENDOCRINOLOGY

Genetic diagnosis of endocrine diseases by NGS: novel scenarios and unpredictable results and risks

Luca Persani1,2, Tiziana de Filippis2, Carla Colombo2 and Davide Gentilini3,4

1Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy, Labs of 2Endocrine and Metabolic Research, 3Molecular Biology Research, IRCCS Istituto Auxologico Italiano, Milan, Italy, and 4University of Pavia, Pavia, Italy

Abstract

The technological advancements in genetics produced a profound impact on the research and diagnostics of non-communicable diseases. The availability of next-generation sequencing (NGS) allowed the identification of novel candidate genes but also an in-depth modification of the understanding of the architecture of several endocrine diseases. Several different NGS approaches are available allowing the sequencing of several regions of interest or the whole exome or genome (WGS, WES or targeted NGS), with highly variable costs, potentials and limitations that should be clearly known before designing the experiment. Here, we illustrate the NGS scenario, describe the advantages and limitations of the different protocols and review some of the NGS results obtained in different endocrine conditions. We finally give insights on the terminology and requirements for the implementation of NGS in research and diagnostic labs.

Introduction

The technological advancements in genetics have had a profound impact on the research and diagnostics of non-communicable diseases. In several of these cases, the availability of next-generation sequencing (NGS) allowed the identification of novel candidate genes but also an in-depth modification of the understanding of the architecture of several diseases. Thanks to the power of these approaches and the progressive diminution of costs,
these changes have been occurring in a short period of time and a large portion of clinicians are not prepared for such revolution, and misinterpretation of the NGS information represents a real danger. This manuscript illustrates the new-generation approaches for DNA sequencing and the novel scenarios that we are currently facing in order to make the clinical endocrinologists aware of the potential advantages and risks of this revolution in genetics.

Variable approaches: NGS panel/targeted analyses vs WES or WGS

NGS is being adopted by genome diagnostics laboratories in different countries worldwide. However, implementing NGS-based tests according to diagnostic standards is a challenge for individual laboratories (1). One of the most important issues that need to be addressed regards the selection of the library to be adopted. The main question is: ‘Should I make a custom panel for this gene set, or should I do whole exome sequencing (WES), or is it better to perform whole genome sequencing (WGS)?’ The question is crucial since several issues of diagnostic or research workflow will depend on this choice (Fig. 1). While WGS approach can capture all possible mutations, WES or targeted gene panel sequencing (targeted NGS) are cost-effective approaches for capturing phenotype-altering mutations (2, 3). With unlimited resources and time, WGS is a clear winner as it allows you to interrogate SNVs, insertions or deletions (indels), structural variants (SVs) and copy number variants (CNVs) in both the 1% part of the genome that encodes protein sequences and the 99% of remaining non-coding regions. WES is focused on the detection of SNVs and indels in protein-coding genes and on other functional elements such as microRNA sequences; consequently, it omits regulatory regions such as promoters and enhancers. Although costs vary depending on the sequence capture solution, WES can be an order of magnitude less expensive than WGS to achieve an approximately equivalent breadth of coverage of protein-coding exons. These reduced costs offer the potential to greatly increase sample numbers, which is a key factor for many studies and clinical applications. On the other hand, targeted NGS represents the cheaper solution adopted in a great number of laboratories allowing the optimization of different features of molecular diagnostics workflow (reduced costs for library preparation steps and sequencing run, adaptable to different kind of samples, increased number of samples processed in each run and reduction of time needed for clinical counselling). In Table 1, we report the potential advantages and limitations of those approaches (1, 2, 3, 4, 5, 6).

Study design for different sequencing strategy

The choice of the sequencing strategy should be determined by the aim, the underlying biological hypothesis and size of the study. The targeted NGS, WES or WGS support the different testing strategies, going from the sequencing of candidate genes selected for a particular disorder or phenotype up to an unsupervised sequencing of all genes in the genome.

Studies of population genomics are mainly oriented on WGS strategy, they benefit from a trade-off between sample numbers and sequencing depth, in which many genomes are sequenced at low depth and variants are simultaneously called across all samples. Variant calls on individual low-depth genomes have a high false-positive rate, but this problem is overcome by combining results across a high number of samples (7, 8, 9).

When the aim of the study is to identify new genes involved in a particular disease, WES usually appears as the most convenient choice since it is considerably cheaper than WGS and allows the sequencing of a consistent number of subjects with a good detection quality of SNVs, SVs and indels, but not for CNVs (10, 11).

The analysis of gene panels isolated by targeted capture represents a valid alternative for diagnostic applications, when the lab should test only causative genes and give a response in a short time. The restricted targeting represents a relevant advantage as it reduces the possibility of incidental findings thus facilitating interpretation of genetic results and allowing higher coverage at lower cost than genome/exome-wide approaches (Table 1).

Notwithstanding differences among protocols and study design, the analytical approach is similar among strategies: sequencing step is followed by bioinformatics analysis and filtering of genetic variability aimed to select genetic variants that are considered clinically relevant. In such context, WES and WGS offer the advantage of expanding the search space to consider additional genes or genomic regions that have not been previously identified but that may potentially explain an individual’s complex or peculiar clinical presentation. However, a single exome can produce approximately 30K variants, while more than 3M variants can be found in a single genome. Both in diagnostic and in
research areas, the bioinformatics task is fundamental but is therefore often long and challenging. Apart from the sequencing strategy, the most important challenge today is represented by the analysis and classification of genomic variants that have not been previously reported in the medical literature or in public databases (see below). The assessment of the variant pathogenic impact should follow accepted guidelines for variant classification (1). The classification of genetic variability and the determination of its role on human phenotype represent one of the most important challenges for clinicians, geneticists and researchers.

**NGS application for rare or common complex diseases**

NGS can be applied to study both rare diseases and complex diseases. Different examples demonstrate the power of NGS in identifying causal variants for rare monogenic diseases even with very small sample size (12, 13, 14).

However, not all rare conditions reflect the scenario of typical monogenic diseases with Mendelian inheritance. Indeed, some rare conditions include similar manifestations, and the same manifestation can

---

**Figure 1**

Schematic illustration of the algorithm for the choice of the most suitable next-generation sequencing (NGS) approach. Targeted NGS is generally more suitable for clinical applications. The whole exome or whole genome sequencing (WES or WGS) approaches have mainly research applications, but positive results should expand the targeted NGS approaches. WES can also have a clinical application for the diagnosis of particular heritable conditions that are not clinically defined. VUS, variants of uncertain significance.
be induced by different mechanisms. In such context, the identification of causal variants generally requires larger sample sizes than classic monogenic diseases or affected families with multiple generations (15). In complex diseases, there is an extreme heterogeneity in both the clinical spectrum and underlying mechanisms: individuals with similar phenotypes may involve different causal variants from the same gene or multiple variants in different genes acting within the disease pathway(s). Conversely, patients with the same causal genetic factor may manifest a variable phenotype due to incomplete penetrance, as a consequence of the interaction with other genetic, epigenetic or environmental modifying factors. In this field, the characterization of genetic aetiology can become difficult.

A large survey of human genetic variation (16) shows that rare variants represent the great part (~70–80%) of genetic variability, and the UK10K project identified more than 42M SNVs over 3781 subjects (SNVs, 34.2M rare and 2.2M low frequency). Statistical genetics considerations of rare variant association analysis have been the focus of intensive method development over the last few years in the field of complex traits studies. Even if there is a substantial contribution from rare variants, it remains challenging to detect rare variant effects due to low statistical power. Deep WGS of large numbers of individuals would represent the most informative strategy for association studies of complex traits and diseases. However, large-scale WGS is generally unfeasible in the field of classical epidemiological designs, such as case-control and cohort studies, because of the high

### Table 1 Principal applications, advantages and limitations of the three different next-generation sequencing (NGS) strategies.

<table>
<thead>
<tr>
<th>NGS strategy</th>
<th>Principal application</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole genome sequencing (WGS)</td>
<td>Studies of population genomics</td>
<td>- Identification of SNVs, indels, SV and CNVs in coding and non-coding regions&lt;br&gt;- PCR amplification not required, reducing the potential of GC bias&lt;br&gt;- More consistent and uniform coverage compared to WES or Targeted NGS&lt;br&gt;- A lower average read depth is required to achieve the same coverage as WES&lt;br&gt;- No impact by sequencing read length&lt;br&gt;- Performance unaffected by capturing or amplification procedures&lt;br&gt;- Specific protocols of enrichment are not required&lt;br&gt;- Complete information on the genetic variability of each sample</td>
<td>- High cost&lt;br&gt;- Low number of samples&lt;br&gt;- Huge amount of genetic variants identified&lt;br&gt;- Greater effort in the data interpretation&lt;br&gt;- Limited applications in routine diagnostics</td>
</tr>
<tr>
<td>Whole exome sequencing (WES)</td>
<td>Identification of new causative genes</td>
<td>- Identification of SNVs, indels, SV and CNVs in coding regions&lt;br&gt;- Reduction of the cost in comparison with WGS&lt;br&gt;- Reduced number of identified genetic variants with a reduction of resources needed for their storage and interpretation&lt;br&gt;- Increased number of samples analysed&lt;br&gt;- More suitable than WGS for clinical applications</td>
<td>- Assessment of genetic variability only in exons&lt;br&gt;- Heterogeneous coverage influenced by library preparation procedures&lt;br&gt;- High risk of genetic incidental findings</td>
</tr>
<tr>
<td>Targeted NGS</td>
<td>Analysis of known causative genes</td>
<td>- Most suitable for clinical applications&lt;br&gt;- Higher coverage and sequencing depth than WES&lt;br&gt;- Customizable for different samples types, e.g. formalin-fixed paraffin-embedded tissues, cell free or circulating tumoural DNA, degraded samples&lt;br&gt;- Highest number of samples&lt;br&gt;- Reduced computational and storage resources&lt;br&gt;- Lower testing costs than WGS and WES&lt;br&gt;- Genetic variability is determined only for selected genes thus reducing the risk of incidental findings</td>
<td>- Genetic analysis restricted to selected regions&lt;br&gt;- Variable and heterogeneous coverage&lt;br&gt;- Problematic design of probes&lt;br&gt;- Biases of library preparation: false-positive variants due to PCR duplicates or false negative results due to allelic drop out&lt;br&gt;- Difficult identification of CNVs</td>
</tr>
</tbody>
</table>
cost. Several less costly sequencing strategies have been proposed and used in common traits studies, such as low-depth WGS, WES and targeted NGS. Many statistical methods have been then proposed to increase the signal or reduce the noise in testing variant-disease association using sequencing data. These statistical methods can be classified into three categories: (i) the BURDEN test (17, 18), (ii) the sequence kernel association test (SKAT) (19) and (iii) the $P$ values combination methods (20). Once a gene/region emerged as significantly associated, the next important step is to identify rare causal variants within these regions/genes. Identifying a small number of rare causal variants that contribute to complex diseases has become a major focus of investigation. Each of the above gene-based methods reports a $P$ value for the association of multiple rare variants and a specific phenotypic trait. However, the following identification of a small proportion of truly causal variants is an even more difficult challenge and often need other approaches such as in vitro or in vivo functional studies.

**Sequencing errors and confirmation of results**

More than 200,000 genomes and an even higher number of exomes have been sequenced to date. It is still widely established that variants found using NGS should be validated with the current ‘gold standard’ for DNA sequencing, Sanger sequencing (21), though several reports suggest that NGS results are at least as accurate or in some cases more accurate than Sanger sequencing (22, 23). Massively parallel sequencing technologies have revolutionized medical genetics, however, also NGS is prone to both negative and positive results. Problems may be generated during library preparation procedure, PCR artefacts for example, can introduce false-positive results in capturing-based libraries, while amplicon-based libraries are prone to allelic dropout problems due to presence of variation in the sequence that produce the selective amplification of a single allele. Errors can also be introduced during the bioinformatics analysis, extended insertions or deletions for example can be missed. To date, there are no validated procedures to detect sequencing errors.

**NGS for discovery of novel pathogenic mechanisms**

Conventionally, the diagnostic approach to endocrine diseases was based on physical findings, biochemical testing and imaging analysis, followed, secondly by molecular analysis. Until recently, DNA samples were analysed following a phenotype-driven strategy and using Sanger sequencing method of the coding regions of one candidate gene at a time, an operation that requires several weeks for a response.

The scenario has dramatically changed with the introduction of NGS. The outcomes, related consequences and risks are summarized in Table 2. With the NGS introduction, the time needed to systematically sequence a set of candidate genes has decreased from several weeks to few days. This finally has progressively brought down the costs for a massive parallel high-throughput DNA sequencing and a comprehensive genetic diagnosis. However, this advancement becomes possible with the introduction of new key-personnel for the design of NGS experiment and bioinformatics data analysis.

An increasing number of scientific reports (>600) using NGS for endocrine disease investigations were published in peer-reviewed journals since 2009 (Fig. 2). Here, we selected some examples showing how NGS appeared successful in identifying new causative genes or pathogenic mechanisms.

One of these studies is on familial central precocious puberty (CPP) (24). The Brazilian authors performed a WES in 40 members of 15 families affected by CPP of unknown aetiology. NGS analysis identified $MKRN3$ gene as a potential causative candidate gene for the disease. The important role of $MKRN3$ in human puberty initiation was reinforced by large genome-wide studies involving women of European descent from 57 studies (25). Subsequently, several studies confirmed that mutations in the paternally expressed imprinted $MKRN3$ gene constitute a major genetic cause of heritable or apparently sporadic CPP (26, 27). Very recently, a genomic defect in $DLK1$, a gene encoding a ligand of the Notch receptors, was discovered by NGS as being also associated with isolated familial CPP (28). Interestingly, $MKRN3$ and $DLK1$ are both paternally expressed imprinted genes. These findings suggest a role of genomic imprinting in regulating the timing of human puberty.

Importantly, $MKRN3$ falls in the locus on chromosome 15 that is known to be associated with Prader–Willi syndrome (characterized by hypothalamic and neurological dysfunctions associated with severe obesity due to a complete resistance to satiety signals), whereas $DLK1$ had been previously implicated in Temple syndrome (including precocious puberty with intrauterine growth retardation, postnatal short stature, hypotonia, small hands and mild facial dysmorphisms) (28). Therefore,
one additional consequence of the NGS approach is the unexpected discovery of the association between a single gene and two apparently distinct disorders. Another example of this kind concerns the JAG1 gene that encodes another ligand of the Notch receptors and was originally linked to Alagille syndrome (a disorder characterized by cholestasis, heart malformations, together with eye, facial and skeletal abnormalities, and by an extremely variable expressivity and penetrance of the heterozygous JAG1 mutations) (29). By an NGS approach, we observed that monoallelic JAG1 variants are frequently associated with rare variants affecting thyroid specific genes in several newborns affected with congenital hypothyroidism (CH), another disorder characterized by variable expressivity and penetrance of the candidate gene defects (30). A role for JAG1 in thyroid development was supported by in vivo studies in the zebrafish model, and it is therefore possible that JAG1 minor alleles could contribute to CH pathogenesis by acting as genetic modifiers and amplifying the partial loss of function associated with monoallelic defects in thyroid-specific genes (31).

Our NGS data obtained in a large Italian cohort of CH patients and in an ethnicity-matched population are consistent with an oligogenic model of CH pathogenesis (30). We found several variants of 11 genes with a relevant role in thyroid morphogenesis or function that have a rare or low frequency in the general population, but a significant enrichment in the CH population. When expressed alone in the first-degree relatives of CH patients, these rare variants were associated with minor

Table 2  Novel scenarios opened by the NGS approach offering massive parallel high-throughput DNA sequencing for research or diagnosis of endocrine conditions.

<table>
<thead>
<tr>
<th>NGS aftereffects (ref.)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced time and costs and high accuracy for genetic diagnosis (1, 2, 3, 39, 78, 79)</td>
<td>• Novel personnel requirements → bioinformatics for experiment planning and interpretation of results</td>
</tr>
<tr>
<td>Identification of novel causative genes (24, 28, 40, 46, 71, 76)</td>
<td>• Potential explanation of patients previously classified as phenocopies</td>
</tr>
<tr>
<td>Genetic classification of endocrine conditions (44, 50, 51, 52, 53, 54, 57, 61, 63, 64, 67, 68, 72, 75)</td>
<td>• Expanded understanding of endocrine conditions</td>
</tr>
<tr>
<td>Identification of multigenic involvement (30, 31, 32, 33, 34, 38, 41, 42)</td>
<td>• Novel pathogenic mechanisms</td>
</tr>
<tr>
<td>Genetic heterogeneity of one condition (34, 36, 38, 42, 43, 44, 57, 58, 59, 75)</td>
<td>• Precision medicine → targeted therapies and management, and accurate prognosis</td>
</tr>
<tr>
<td>One gene associated to multiple clinical conditions (26, 28, 29, 31, 37, 38, 40)</td>
<td>• Novel pathogenic mechanisms</td>
</tr>
<tr>
<td>Defined frequency of gene variants (46, 57, 61, 62, 63, 64, 67, 68, 77)</td>
<td>• Possible explanation for the variable expressivity and penetrance of certain candidate gene variations</td>
</tr>
<tr>
<td>Unexpected variants in genes unrelated to the investigated condition (incidental findings) (81)</td>
<td>• Expanded understanding of endocrine conditions</td>
</tr>
<tr>
<td>Identification of variants of unknown significance (78, 79, 80)</td>
<td>• Novel pathogenic mechanisms</td>
</tr>
<tr>
<td></td>
<td>• Allelic quantification → more defined prognosis for endocrine tumours</td>
</tr>
<tr>
<td></td>
<td>• Role of gene modifiers</td>
</tr>
<tr>
<td></td>
<td>• Expanded understanding of gene function</td>
</tr>
<tr>
<td></td>
<td>• Novel pathogenic mechanisms</td>
</tr>
<tr>
<td></td>
<td>• Challenging genetic counselling</td>
</tr>
<tr>
<td></td>
<td>• Classification of gene variants</td>
</tr>
<tr>
<td></td>
<td>• Improved genetic counselling</td>
</tr>
<tr>
<td></td>
<td>• Ethical issues</td>
</tr>
<tr>
<td></td>
<td>• Challenging genetic counselling</td>
</tr>
<tr>
<td></td>
<td>• Adequate information for patient’s consent</td>
</tr>
<tr>
<td></td>
<td>• Adequate bioinformatic and clinical classification</td>
</tr>
<tr>
<td></td>
<td>• Challenging genetic counselling</td>
</tr>
</tbody>
</table>

Figure 2
Timeline of articles in PubMed with the key words: Next-generation sequencing AND endocrine disease (updated March 2018).
thoracic defects, whereas the variable combinations of \( \geq 2 \) minor alleles were unexpectedly found in about 25% of CH patients. Other groups obtained similar NGS findings in smaller CH cohorts (32, 33). Such an oligogenic origin of CH provides a suitable explanation for the frequent sporadic appearance of CH in the population. In all these NGS studies targeted to a panel of known CH candidate genes, the frequency of variations found by their systematic and unsupervised analyses was unexpectedly higher than the rate of positive findings in previous studies performed by Sanger in phenotypically selected CH cohorts (5–20% depending upon the investigated phenotype). This might be explained both by an increased sensitivity of NGS protocols in the genetic variation detection as well as by the involvement of morphogenetic or functional genes independently of the observed CH phenotype. Noteworthy, the number of cases remaining unexplained after these targeted NGS analyses falls down below 40% thus representing a significant improvement for the genetic counselling in clinical practice (34).

In other cases, NGS allowed to link one gene to two endocrine conditions that were previously considered completely distinct. Izumi et al. identified by NGS a SOX3 polyalanine deletion in a patient with normosomic isolated hypogonadotropic hypogonadism (IHH) (35). This finding expanded the phenotypic spectrum of SOX3 polyalanine deletion to include IHH without other pituitary hormone deficiencies or mental retardation (36). Moreover, they detected a WDR11 splice-site mutation in a patient with multiple pituitary hormone deficiency (MPHD) (35). These data indicate that WDR11, a gene known to be involved in normosomic IHH, can also cause MPHDs. Similar results had been previously reported for other IHH and Kallmann syndrome (KS: IHH with anosmia) genes, such as PROKR2 or FGFR and FGFR1 that have been described in MPHDs also (37). The clinical presentation of heterozygous defects in candidate genes for IHH/ KS is known to be highly variable among and within the same families which, in combination with the >30 known candidate genes (38), makes the phenotype-driven genetic analyses very complicated and expensive, as well as disturbing and time consuming for the patients (39). The panel of candidate genes was further expanded by an NGS study conducted on 261 genes (known to be involved in hypothalamic, pituitary, and/or olfactory pathways, or suggested by chromosome rearrangements) that identified 18 new potential candidate genes for IHH/KS (40). In such heterogeneous-related conditions (IHH, KS, MPHDs) associated with a high number of candidate genes, the application of NGS protocols is now unavoidable. Indeed, the systematic NGS analyses will help to define the exact frequency of involvement of single genes and the clinical impact of oligogenic involvement (41).

This picture is very similar to that now seen in different clinical settings, such as the primary ovarian insufficiency (POI) or the disorders of sex development (DSD) (42, 43). The genetic diagnosis in POI patients may become useful for the preservation of fertility in the affected families or in the identification of patients with higher chances to obtain a fertilizable egg or at higher risk of extra-ovarian defects (44). The Vilain Lab conducted two studies on DSD patients (44, 45). In the first, they proposed a novel procedure, reversing the order of the diagnostic endocrine steps and starting from the genetic analysis. By this approach, they claim the possibility to eliminate non-indicated clinical tests, sparing the patient unnecessary stress and saving healthcare system’s resources. They designed a targeted panel sequencing of 35 DSD genes that revealed genetic defects in two out of seven patients not previously diagnosed and confirmed the diagnosis in another seven patients with known genetic causes (45). In the second study, they performed a WES expanding the list of genes to analyse 64 candidates. They reached a diagnostic yield of 35% in 40 patients with 46,XY DSD who had not previously received a genetic diagnosis (46).

The application of NGS recently expanded the view on the mechanisms underlying tumour formation. Until recently, activating mutations in the TSH receptor (TSHR) and in GNAS represented the principal cause of autonomous thyroid adenomas (ATAs) and were considered oncogenes sufficient to induce autonomous function and growth. In a study conducted on several ATAs by WES, we found that a relevant fraction of ATAs carry, beyond the well-known TSHR and GNAS variants, also a recurrent hot-spot mutation in EZH1, a key gene involved in the epigenetic regulation of cell differentiation and proliferation. Interestingly, the EZH1 variant was found to be associated with the TSHR or GNAS variants suggesting a 2-hit model for the pathogenesis of these benign tumours, whereby constitutive activation of the cAMP pathway and EZH1 mutations may cooperate to induce the hyperproliferation of thyroid cells (47).

### NGS in endocrine tumours: role in differential diagnosis and prediction of outcome

In the last 10 years, NGS technology led to a better knowledge in all human cancers, including endocrine,
sporadic and familial tumours. More importantly than in other clinical conditions, these new genetic knowledge will become more and more useful for clinicians in the personalized management of patients.

Cancer is a heterogeneous disease harbouring different subclonal cell populations that can be discriminated by their DNA mutations. The genome study of a cancer can help to better identify its heterogeneity and to support the clinicians in the choice of a more effective treatment for each patient. For this reason, NGS applications are becoming an integral part of the clinical routine diagnostics in different endocrine cancers.

In thyroid tumours, NGS represents the gold standard technique for the pre-surgical molecular diagnosis of thyroid nodules and for the molecular profiling of thyroid carcinoma. Fine-needle aspiration cytology is the standard pre-operative tool for thyroid nodule diagnosis, however, up to 25–30% of the samples are classified as indeterminate (48). In these cases, the guidelines of the American Thyroid Association recommend molecular testing in order to better identify malignant samples and to plan patient’s management (49). In order to improve diagnosis and optimize the management of thyroid nodules with indeterminate cytological diagnosis, Nikiforov et al. developed a targeted NGS test, ThyroSeq v2, which includes analyses of point mutations, gene fusions and abnormal gene expression in 56 thyroid-related genes (50). Subsequently, various groups validated the test (51, 52) that was very recently expanded to include 112 genes, the ThyroSeq v3 (53). This panel now includes also several genes whose role in thyroid carcinogenesis remains elusive (e.g. TSHR), thus making indefinite the interpretation of several potential variations. However, the integration of the genetic and cytology results is crucial to allow the stratification of patients according to their risk of malignancy, thus enabling an improved management (54). The clinical utility of molecular testing was demonstrated by the reduction of avoidable lobectomies in individuals with indeterminate cytology and a benign histological result (50, 55, 56).

Molecular evaluation through NGS techniques was also widely performed on histological samples. Indeed, new insights were obtained on papillary thyroid carcinoma (PTC), the most common type of thyroid carcinoma, by the Cancer Genome Atlas network’s study. The authors reported a comprehensive genomic landscape of 402 PTCs with different NGS technologies, reducing the fraction of PTC cases with unknown oncogenic driver from 25% to 3.5%. Based on their results, they propose a reclassification of thyroid cancers (as BRAF- or RAS-like) aimed to improve the management and the therapies of the patients (57). An accurate and defined diagnosis is needed also for rare and aggressive histotypes, the poorly differentiated thyroid carcinoma (PDTC) and anaplastic thyroid cancer (ATC): the use of advanced NGS could improve the molecular diagnosis of these cancers to better clarify their genetic origin and develop new drugs. Among the studies performed on the PDTCs and ATCs, Landa et al. examined the largest number of cases (58). They applied an ultra-deep sequencing strategy able to detect mutations in samples with low tumour purity, such as ATCs that are often infiltrated by macrophages, and proposed a model of tumorigenesis where PDTCs and ATCs derive from well-differentiated tumours through the accumulation of additional genetic abnormalities, many of which with prognostic and therapeutic relevance. Recently, another study focused attention on PDTCs (59): Gerber et al. analysed 25 PDTC cases through the Amplicon Cancer Panel (48 specific genes with 212 amplicons, Illumina). This study identified new potential genetic targets in PDTC cases, including for the first time four HER4 variants.

Important results were also obtained in medullary thyroid carcinoma (MTC), which is mainly caused by the RET and, rarely, RAS gene mutations. Many centres are implementing the use of NGS technology in MTC diagnostic routine, significantly improving the screening of familial MTC and the knowledge of somatic alterations in order to better plan patients’ management and treatments (52, 60, 61, 62). These studies confirmed that MTCs had mutually exclusive mutations in RET and RAS oncogenes and no other commonly recurrent driver mutations were found.

NGS technologies were applied with success also for the pheochromocytomas and paraganglioma tumours (PPGL). Genetic diagnosis is recommended for all PPGL cases, as driver mutations are identified in approximately 80% of the cases. The list of genes involved in the pathogenesis of PPGL is in constant expansion and targeted NGS has proven to be fast and cost-effective also for the genetic analysis of PPGL as demonstrated by several studies (63, 64, 65). Intriguingly, with the application of targeted NGS, it was possible to identify pathogenic germline NF1 mutations in three patients with pheochromocytomas without a prior diagnosis of neurofibromatosis type 1 (NF1) (66).

A relevant advancement has been provided by NGS in the study of neuroendocrine tumours (NETs), a very heterogeneous neoplastic category. Pancreas NETs (pNET), considered biologically more aggressive compared with
NET from other sites, are usually sporadic but can also occur as part of hereditary syndromes (67, 68). A recent study, performed on 90 patients with pNETs by means of a custom-targeted NGS panel, characterized the genetic signature of each tumour and pointed out the importance of a specific gastro-entero-pancreatic targeted NGS gene panel. Moreover, these data suggest that the MEN1 alterations represent a key event in the malignant progression (67). Indeed, germline MEN1 mutations have also been reported in five cases of apparently sporadic pNETs (with no family history or other MEN1 tumours) (68).

Key advancements were made by NGS also in several other rare endocrine tumours. Some studies analysed the complete genomic and transcriptomic landscape of primary and recurrent parathyroid tumour specimens using high-throughput sequencing technologies; the data confirmed the important role of oncogenes CCND1 and RET, and of tumour suppressors MEN1 and HRPT2, but revealed also mutations in well-characterized cancer genes such as mTOR, MLL2, CDKN2C, PIK3CA, POT1 and PRUNE2 (69, 70, 71). Involvement of multiple genes were also found in adrenocortical carcinoma (72, 73, 74), which may lead to a novel classification (75). The exome sequencing of corticotroph adenomas identified somatic mutations in the USP8 deubiquitinase gene (76, 77).

All these data suggest the potential application of NGS in defining the pathogenic mechanisms and in discovering targets to develop new therapeutics.

Risks associated with research or clinical applications of NGS technology

By adopting NGS, clinical laboratories are now performing an ever-increasing catalogue of genetic testing for genetic disorders. This shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context, the American College of Medical Genetics and Genomics (ACMG) convened a workshop in 2013 comprised representatives from the ACMG, the Association for Molecular Pathology and the College of American Pathologists to revisit and revise the standards and guidelines for the interpretation of sequence variants (1, 78, 79). A mutation was previously defined as a rare change in the nucleotide sequence, while a polymorphism was defined as a variant with a frequency >1%. However, the terms ‘mutation’ and ‘polymorphism’ often lead to confusion due to incorrect postulations of pathogenic and benign effects, respectively. Thus, the ACMG recommends that both terms be replaced by the term ‘variant’ with the following modifiers: (a) pathogenic, (b) likely pathogenic, (c) uncertain significance (VUS), (d) likely benign or (e) benign. While these modifiers may not address all human phenotypes, they comprise a 5-degree system of classification for variants relevant to Mendelian disease. It is recommended that all assertions of pathogenicity (including ‘likely pathogenic’) be reported with respect to a condition and inheritance pattern. Cases (a) and (b) include nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single exon or multi-exon deletion. The variants, including the missense ones, can also be classified as pathogenic (or likely) based on functional studies or by co-segregation with phenotype in familial settings, as well as if they appear de novo in the index case. The missense variants that do not have this support to the potential pathogenic role should be classified as VUS. One of the most accurate methods to discern the impact of a VUS is the co-segregation studies in the affected family. The synonymous or deep intronic variations should be classified as likely benign if their impact is not supported by functional or inheritance studies. Data on the frequency of variants may also be relevant for the variant classification, as common variants (>1% in the general population) are being generally classified as benign.

In silico analyses may be of support in the interpretation but the use of multiple databases is highly recommended (30, 79). When a variant has been identified, a bioinformatics approach should be routinely applied (79). Appropriate and dedicated personnel in the Lab should certainly check the existing databases for its frequency. Population databases (e.g., Exome Aggregation Consortium (ExAC)) are useful in obtaining the frequencies of variants in large populations. However, these databases cannot be assumed to include only healthy individuals, as they do not contain extensive information regarding any possible associated phenotypes and they can contain pathogenic variants. When using population databases, one should check whether healthy or disease cohorts were used, and if possible, whether more than one individual in a family was included, and the age range of the subjects. Furthermore, disease- and gene-specific database can be of help (e.g., Online Mendelian Inheritance in Man (OMIM) or the NCBI), but these databases often contain variants that are incorrectly classified (78). The classification of an identified variant should also include searching the scientific and medical literature.

As an internal method of assessment, clinical laboratories should also track all sequence variants
identified in each gene together with the related clinical questions. This is important for tracking genotype-phenotype correlations and the frequency of variants in affected and unaffected cohorts and to uncover variants with a particular frequency in the local population.

Pitfalls of exome sequencing are always around the corner, as in the case of the likely overestimated association of HABP2 variant with familial non-medullary thyroid cancer. This result may occur despite the development of robust computational algorithms, the accrued experience of analysing exome data sets and published guidelines. The bioinformatics approach thus remains a fundamental process with relevant decisions and interpretations that require an accurate supervision (80).

One of the most relevant outcomes of NGS analyses is the occasional detection of unexpected variants in genes that are totally unrelated to the investigated condition, but are associated with clinically occult conditions. The incidental finding is obviously possible in the NGS protocols not exclusively including the disease-specific genes. This gives rise to several ethical concerns. It is however not uncommon to find patients that although willing to know the cause of an already identified condition and refuse to receive predictions on the future health risks for their own or other family members. Therefore, the consent to NGS analyses should contain the possibility to deny the information on these off-target results. Notwithstanding, certain incidental findings (e.g., the identification of a 100% penetrant and strongly pathogenic mutation in RET oncogene predisposing to aggressive malignancy) may cause serious ethical concerns because disclosing the identification of a hereditary cancer predisposition would be highly relevant to the clinical care of these patients and have important implications for their relatives’ medical management. For clinical guidance, we refer the reader to the ACMG recommendations (81).

Clinical reports are the final product of laboratory testing and are often integrated into the patient’s electronic health record. Therefore, effective reports should be concise, but easy to understand. Reports should be written in clear language avoiding medical genetics jargon or defining such terms when used. The report should contain all of the essential elements of the test performed, including structured results with an interpretation, the supporting scientific references, the methodology and the appropriate qualifications of the personnel.

Conclusions

In the past decade, sequencing technology has evolved rapidly with the advent of high-throughput NGS. This technology generated a titanic advancement in the genomics with the identification of new disease mechanisms and the possibility to expand the role of predictive and personalized medicine. In this scenario, an adequate use of NGS protocols can now give a great support to the clinicians in endocrine diseases. For this reason, awareness on the advantages and risks of such a powerful technology should be further promoted by the national health systems.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Author contribution statement

T d F, C C and D G made the PubMed search and contributed the first draft. L P completed the draft and revised and finalised the entire manuscript.

Acknowledgements

Apologies to those whose publications in the field were not cited due to space limitations.

References


18 Madsen BE & Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. Plaus Genetics 2009 5 e1000384. (https://doi.org/10.1371/journal.pgen.1000384)


20 Lin WY, Lou XY, Gao G & Liu N. Rare variant association testing by adaptive combination of P-values. Plas ONE 2014 9 e85728. (https://doi.org/10.1371/journal.pone.0085728)


34 Persani L, Rurale G, de Filippis T, Galazzi E, Muzza M & Fugazzola L. Genetics and management of congenital hypothyroidism. Best Practice & Research Clinical Endocrinology & Metabolism 2018 32 75–81. (https://doi.org/10.1016/j.beem.2018.05.004)


36 Persani L & Bonomi M. The multiple genetic causes of central hypothyroidism. Best Practice and Research Clinical Endocrinology


80 Gerhard GS, Bann DV, Broach J & Goldenberg D. Pitfalls of exome sequencing: a case study of the attribution of HAP2 rs7080536 in familial non-medullary thyroid cancer. *NPJ Genomic Medicine* 2017 2 8. (https://doi.org/10.1038/s41525-017-0011-x)


Received 2 May 2018
Revised version received 31 May 2018
Accepted 6 June 2018