Applications of genomic research in pediatric endocrine diseases

Ja Hye Kim, MD, PhD, Jin-Ho Choi, MD, PhD

Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Running title: Genomic research in pediatric endocrine diseases

Correspondence author: Jin-Ho Choi, MD, PhD

Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-Gil, Songpa-Gu, Seoul 05505, Korea

Tel: +82-2-3010-3991, Fax: +82-2-473-3725

E-mail: jhc@amc.seoul.kr, http://orcid.org/0000-0003-1196-7826
Key message

- Recent advances in molecular genetics have advanced our understanding of the molecular mechanisms of pediatric endocrine disorders and now play a role in mainstream medical practice.

- Genome-wide association studies can offer better understanding of the biological mechanisms of disease and new therapeutic options.

- The identification of founder mutations has powerful advantages for efficiently localizing the genes that underlie Mendelian disorders.

- Next-generation sequencing technologies are beneficial both for clinical practice and for research on pediatric endocrinology.
Abstract

Recent advances in molecular genetics have advanced our understanding of the molecular mechanisms involved in pediatric endocrine disorders and now play a mainstream role in mainstream medical practice. The spectrum of endocrine genetic disorders has two extremes: Mendelian and polygenic disorders. Mendelian or monogenic diseases are caused by rare variants in a single gene, each exerting a strong effect on disease risk. Polygenic diseases or common traits are caused by the combined effects of multiple genetic variants in conjunction with environmental and lifestyle factors. Testing a single gene is preferable if the disease is phenotypically and/or genetically homogeneous. However, next-generation sequencing (NGS) can be applied to phenotypically and genetically heterogeneous conditions. Genome-wide association studies (GWASs) examine genetic variants across the entire genome in a large number of individuals who have been matched for population ancestry and assessed for a disease or trait of interest. Common endocrine diseases or traits, such as type 2 diabetes mellitus (DM), obesity, height, and pubertal timing, are the result of the combined effects of multiple variants in various genes, frequently found in the general population, and with each variant contributing a small individual effect. Isolated founder mutations can result either from a true founder effect or from an extreme reduction in population size. Studies on founder mutations offer powerful advantages for efficiently localizing the genes that underlie Mendelian disorders. The Korean population has been settled in the Korean peninsula for thousands of years, and several recurrent mutations have been identified as founder mutations. The application of molecular technology has allowed us to further our understanding of endocrine diseases, which has had an impact on the practice of pediatric endocrinology related to diagnosis and genetic counseling. This review focuses on the application of genomic research to pediatric endocrine diseases using GWASs and NGS technology for diagnosis and treatment.
**Key words:** Founder effect, Genome-wide association study, Next-generation sequencing
Introduction

Pediatric endocrine diseases have a substantial genetic component. Identifying the genetic etiologies of these diseases helps in identifying their molecular pathophysiology and providing customized treatment and disease prevention. Many pediatric endocrine disorders are caused by genetic defects in genes for hormone synthesis, binding proteins, transcriptional activity, channels, membrane receptors, and signal transductions [1]. In addition, genetic syndromes in endocrine diseases are based upon unique genetic pathophysiology.

There are two extremes of the spectrum of endocrine genetic disorders: Mendelian and polygenic disorders. Mendelian or monogenic diseases represent the extreme of possible genetic architecture; they are caused by rare sequence variants in a single gene, each with a large effect on disease risk (Fig. 1) [1]. Monogenic disorders, such as congenital adrenal hyperplasia, are defined as inherited conditions arising from mutations in a single gene that disrupt a physiologic pathway that has a large effect on the disease [2, 3]. The genetic variants causing Mendelian diseases are found in a small number of genes, can be highly penetrant, and are nearly unaffected by the environment. The wide spectrum of phenotypes that characterize many Mendelian endocrine diseases is often reflected in genetic heterogeneity. Some Mendelian disorders demonstrate the same phenotype due to multiple different mutations in the same gene or locus, which is called allelic heterogeneity. Others present as the same disease in different individuals resulting from mutations in multiple different genes, which is called locus heterogeneity and occurs in Kallmann and Noonan syndromes [2].

At the other extreme of the spectrum are common or polygenic diseases and traits, such as type 2 DM, obesity, or human height, which are caused by the combined effects of multiple common variants and are frequently observed in the population, with small to modest effects in conjunction with environmental and lifestyle factors (Fig. 1) [3].
The identification of disease genes in rare, monogenic, or syndromic disorders has been mostly driven by linkage analysis within multiple pedigrees [3]. This approach has been successful for the high-penetrant variants responsible for Mendelian disease [3]. In contrast, the variants for common polygenic disorders have been identified through association studies in the general population [4].

Diagnosing endocrine diseases requires comprehensive tools, including medical information, biochemical tests, dynamic hormone tests, and imaging studies. Currently, genetic testing is used as a confirmatory diagnostic tool in many pediatric endocrine diseases, particularly if the biochemical findings are vague. During the last two decades, the advancements of molecular genetic technologies have achieved profound development with respect to the diagnostics and research of pediatric endocrine diseases. In this regard, molecular techniques for genetic diseases have developed from traditional methods to modern genetic testing [2]. Traditional genetic testing includes routine karyotype analysis or tests for single genes or chromosomal regions, such as Sanger sequencing or fluorescent in situ hybridization. Modern genetic testing analyzes the entire genome, with technologies such as chromosomal microarray and next-generation sequencing (NGS) technologies [2].

The knowledge of the contribution of genetic and epigenetic alterations to endocrine disorders has expanded massively and furthered our understanding of their molecular pathophysiology, the provision of adequate genetic counseling and prenatal diagnosis, and the development of new therapeutic strategies based on the understanding of the molecular mechanisms of such diseases. This review focuses on the application of genomic technologies to pediatric endocrine diseases, such as genome-wide association studies (GWASs) and founder mutations. In addition, we highlight the benefits of moving from traditional genetic testing to NGS.
technologies to diagnose and establish treatment strategies for pediatric endocrine diseases.

Genome-wide association studies (GWASs) for pediatric endocrine diseases

GWASs aim to determine statistically associated genetic variants across the entire genome in a large number of individuals who have been matched by population ancestry and evaluate the disease or trait of interest [4]. GWASs can provide genetic information for use as a screening tool to identify individuals at risk for certain diseases or conditions. The most commonly studied genetic variants in GWASs are single-nucleotide polymorphisms (SNPs) [4]. The primary goal of these studies is to better understand the biology of disease and provide prevention or treatment strategies. Although a genetic variant at a certain locus and trait is not directly informative, GWASs have been successfully implemented to predict the relative roles of genes and the environment in disease risk [5].

GWASs are based on the common-variant, common-disease hypothesis, which holds that the genetic basis for common diseases and traits is driven by common genetic variants [4]. The statistical power to detect associations between genetic variants and a trait requires very large sample sizes because of the small effect sizes of the variants in GWASs. Using millions of association tests, a strict statistical threshold must be set to establish significance at a genome-wide level, which corresponds to a $p$-value of $< 5 \times 10^{-8}$ [5]. However, the vast majority of SNPs are novel and located in the non-coding regions, making it difficult to interpret the functional consequences of the variant on the phenotype [5]. In addition, there is a lack of well-validated cell and animal models in which to test functional impact. Therefore, a diverse set of approaches has been developed to infer the functional impacts of variants identified by GWASs. For the coding variants, multiple annotation tools, such as ANNOVAR or Variant Effect
Predictor [6], can be used to infer their potential impact on genes; however, only 2–3% of the loci are present in this region. An approach to identifying functional impact on the regulatory region is expression quantitative trait loci (eQTL) analysis, which identifies loci associated with RNA expression [7]. The various projects for mapping the effects of regulatory variations, such as ENCODE, Roadmap Epigenomics, and the Genotype-Tissue Expression (GTEx) project, provide an essential dissection of the characterization and interpretation of the non-coding variants [7, 8].

Common endocrine diseases and traits, such as type 2 DM, obesity, height, and puberty, are the result of the combined and simultaneous effects of multiple variants in various genes, commonly found in the general population, with each variant contributing a small effect. Polygenic risk scores (PRSs) can be calculated using the small effects of a number of genetic variants discovered in GWASs; these scores can predict the risk of complex diseases, such as coronary artery disease, atrial fibrillation, type 2 DM, inflammatory bowel disease, and breast cancer [9]. PRSs provide an overall estimate of an individual’s risk for different diseases [9]. Although individual variants typically have small to moderate effects on risk, when combined into a polygenic score, they offer increased power. GWASs are able to find genetic information that can be used as a screening tool to identify individuals at risk, predict the risks of certain conditions, and diagnose the diseases.

1. GWASs for type 2 diabetes mellitus

A number of common genetic variants are associated with energy balance, appetite, insulin resistance, and insulin secretion [10]. Type 2 DM is a multifactorial disorder for which manifestation depends on multiple interacting environmental and genetic factors. Type 2 DM
is a genetically heterogeneous disease, with several rare monogenic variants of large effect and a number of common variants with small to moderate effects, involving complex interaction of genetic and environmental factors.

More than 400 loci for type 2 DM can explain only 10% of the heritability, and all variants affect risk modestly, with relative risk usually around 1.1–1.2 [11]. GWASs have identified that the SNP (rs7903146) at TCF7L2 has the largest effect on risk, conferring a 1.4-fold increase per allele [12]. TCF7L2 is a Wnt signaling transcription factor and plays an important role in coordinating the expression of proinsulin and insulin [12]. A risk variant of MTNR1B, encoding melatonin receptor 1B, modifies the RNA expression of MTNR1B by increased enhancer binding for neurogenic differentiation factor 1 (NEUROD1) in human islets, which highlights the modulating role of melatonin in glucose homeostasis [13].

GWASs have also identified variants associated with obesity and high body mass index (BMI). An SNP in the second intron of FTO was identified in GWASs for both type 2 DM and obesity [14]. The association signal for type 2 DM entirely disappeared with correction for BMI, indicating that the SNP increased type 2 DM risk by increasing BMI. A splice acceptor site variant that generates premature stop codon and loss of function in ADCY3, which is highly expressed in visceral adipose tissues, has been identified as the cause of high risk of BMI and type 2 DM [15]. Functional studies in mice suggest that ADCY3 may be a new therapeutic target [15]. In addition, a combination of risk alleles, such as MNTR1B, G6PC2, and GCK manifested fasting hyperglycemia and decreased insulin secretion [13], whereas those with variants in SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGFBP2, and TCF7L2 manifested decreased insulin production and secretion [16].

GWASs conducted in various populations have revealed differences in the genetic risk factors
contributing to DM among different ancestral groups. A common SNP at a locus containing
SLC16A11/13 in Latino, Mexican, and Japanese ancestry confers a 1.25-fold increased risk of
DM; however, this has not been detected in European ancestry populations [17, 18]. Similarly,
a common variant in TBC1D4 in individuals from Greenland strongly increases the risk of type
2 DM with an allele frequency of 17%, but this variant is extremely rare in continental Europe
[19].

The Genetics of Type 2 Diabetes (GoT2D) consortium collected comprehensive genome-wide
sequence data from 2,657 cases with type 2 DM and controls in Europe. The Type 2 Diabetes
Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples (T2D-GENES)
consortium focused on exome sequence variants from 12,940 individuals from five ancestry
groups including GoT2D exomes [10]. This study employing NGS has identified variants
within the regions previously found in GWASs by genotyping arrays and imputation,
suggesting that low-frequency variants contribute much less to the heritability of type 2 DM
than do common variants [10]. UK Biobank has transformed population research into a new
phase with the announcement of exome sequences for 500,000 participants by 2019 [20]. The
first report found that rare loss-of-function variants in FAM234A significantly reduces the risk
of self-reported diabetes (odds ratio [OR] = 0.64, 95% confidence intervals [CI] 0.52–0.80, p
= 10^{-4}). In addition, protein-altering variants in MAP3K15 were associated with lower serum
glucose and protection from type 2 DM (OR = 0.85, 95% CI 0.79–0.91, P = 2.8 X 10^{-6}) [20].

The generation of PRS can identify individuals’ future risk of DM, which can benefit from
early interventions. A recent study using 18,197 cases and 423,697 controls from the UK
Biobank showed that maximal discrimination (area under the curve C statistic of 66%) was
obtained from a PRS of 136,795 variants [11]. These risk estimates are also conducted by the
direct-to-consumer company, 23andMe (https://www.23andme.com/), and they have shared results from 1244 SNP-related PRS, with a recommendation of lifestyle interventions for high-risk customers [9].

The genes implicated in monogenic DM also contribute to polygenic forms; thus, they can be utilized as important therapeutic targets and for prognosis prediction in type 2 DM [21]. Genes associated with monogenic DM, for instance, in patients with permanent neonatal diabetes caused by mutations in \textit{ABCC8} or \textit{KCNJ11}, can be treated with high-dose sulfonylureas instead of insulin [21]. Individuals with maturity-onset diabetes of the young (MODY) types 1 and 3, caused by mutations in \textit{HNF4A} and \textit{HNF1A}, have shown the superiority of sulfonylureas over metformin or insulin [21]. The \textit{GLP1R} gene encoding the receptor for glucagon-like peptide 1 (GLP-1) has a known association with fasting glucose and type 2 DM [22], and this receptor is the target for GLP-1 analogs [22].

Collectively, the effect sizes and clinical significance are only small to modest; however, identifying risk loci provides better understanding of the biological mechanisms of disease and offers new therapeutic options.

2. GWASs for human height

GWAS results have demonstrated that height is highly polygenic, which means many thousands of genetic variants contribute to an individual’s height. An early GWAS on adult height was performed in 4,921 individuals of European ancestry; and this study identified a strong association between \textit{HMGA2} and height [23]. Although the genetic variant in \textit{HMGA2} explains approximately 0.4 cm of final height gain, it was the first study to demonstrate the
genes relevant to growth through GWASs. Subsequently, the Genetic Investigation of Anthropometric Traits (GIANT) consortium conducted a GWAS in ~180,000 individuals for height. They reported that at least 180 loci influence adult height and that these loci are not random but rather an enriched biological pathway for skeletal growth defects, such as FGFR4, ECM2, and STAT2. In addition, they found that the 180 loci could explain, on average, 10.5% of the adult height variance [24]. An expanded study of the GIANT consortium found 697 variants clustered in 423 loci with ~250,000 samples of European ancestry, and they continued with more than two million samples with the release of the UK Biobank [25]. This study indicates that human height is highly polygenic and that these SNPs explained more than 20% of phenotypic variance [25].

Most GWASs conducted by the GIANT consortium have been focused on European ancestry. GWASs have also been performed in Asian populations and identified 98 regions of the genome associated with height; 17 of these were unique to Asians [26]. In about 8,000 samples from Korea, 15 loci were found to be associated with height; many of them had already been reported in the European population [27]. These findings suggest that most GWAS associations are shared across populations and that the genetic regulation of height is similar across the world.

A recent study using GWAS of 5.4 million individuals from diverse ancestries, 12,111 independent SNPs were significantly associated with height and accounted for 40% of phenotypic variance in European populations [28]. This study demonstrated a strong genetic overlap of GWAS signals across ancestries, however, moderate heterogeneity of SNPs between ancestries was observed. Prediction results using GWAS revealed apparent attenuation in non-European ancestry, highlighting the need to further increase the sample size of GWAS in non-European populations.
GWASs are designed to map the polygenic architecture of common diseases, but they have also benefited the detection of candidate genes of Mendelian growth disorders, typically caused by rare genetic variants in a single gene. Variants in the same genes can have variable effect sizes and minor allele frequencies. For example, a highly infrequent variant in the Indian hedgehog (IHH) gene causes the extremely rare syndrome of acrocapitofemoral dysplasia. Individuals with acrocapitofemoral dysplasia carrying highly deleterious mutations in IHH are 2.3 to 8.6 standard deviation scores (SDSs) below mean height. A rare missense variant in IHH is present in 0.2% of people and decreases height by 0.294 SDSs, while a common variant in IHH alters height by 0.05 SDSs per allele [29]. ACAN encodes the protein aggrecan, which is crucial for structure and function of growth palate and cartilage [30]. Homozygous mutations in ACAN cause a severe phenotype of spondyloepimetafphyseal dysplasia, whereas heterozygous mutations result in a milder phenotype of proportionate short stature with osteoarthritis. A recent GWAS results demonstrated that the most significant loci associated with height was observed on chromosome 15 near ACAN, and multiple types of common variants in ACAN, involving enhancers, missense variants, and tandem repeat polymorphisms, affect height [28]. In addition, recent studies showed pathogenic variants in ACAN established as one of the common monogenic conditions in idiopathic short stature [30]. Such overlap in regions highlighted by GWASs and Mendelian genetics illustrates an important aspect of the genetic architecture of height.

Using the UK Biobank-based GWAS of 253,299 European ancestry, PRS generated by 33,938 SNPs demonstrated the best predictive power [31]. In combination with age, sex, recruitment center, genotyping, and population stratification, PRS was able to capture 71.1% (95% CI 70.8%–71.4%) of the total variance in adult height. This value was similar to the 72.6% (95%
CI 69.6%–75.6%) variance estimated by mid-parental height. When predicting adult short stature with sex and PRS, PRS achieved an area under the receiver operating characteristic curve (AUROC) of 0.843 (95% CI 0.796–0.890) to identify children who would have adult short stature, which was also similar prediction by mid-parental height (AUROC 0.879, 95% CI 0.84–0.919). Further, combining mid-parental height and PRS provided better prediction accuracy than either metric alone. Therefore, combination with traditional adult height predictors with genetic predictors enhance screening power for children at risk for short stature in adulthood.

3. GWASs for timing of puberty

The timing of puberty is a complex process influenced by genetic, nutritional, and environmental factors. The high correlation of pubertal timing within racial/ethnic groups, families, and between monozygotic twins compared with dizygotic twins suggests that this timing is regulated by genetic factors [32]. Age at pubertal initiation and menarche was found to be younger in African-American girls than white girls [32, 33]. Mothers’ age at menarche was associated with that of daughters [34]. The constitutional delay of growth and puberty often has a familial component [35].

Studies of epidemiologic and intra-familial tools suggest that 50–95% of the variation in pubertal timing is determined by genetic control [32]. More than 100 genomic regions are associated with pubertal timing, and these genes are implicated in the hypothalamic-pituitary-ovarian axis involved in gonadotropin-releasing hormone (GnRH) secretion, pituitary development and function, hormone synthesis and bioactivity, energy homeostasis and growth, and peripheral feedback [32].
A GWAS for age at menarche performed in 2009 reported the first loci at LIN28B in the chromosome 6q21 region, which were reproducibly associated with population variations at the timing of puberty [36]. LIN28 is an evolutionarily conserved RNA-binding protein that regulates mRNA translation and miRNA let-7 maturation in onset of puberty; however, the detailed mechanism of the initiation of puberty still remained unknown [36]. A recent GWAS in about 180,000 women of European descent found 106 genomic loci associated with age at menarche. Most variants showed very small effect sizes between one week and five months for each allele on the timing of menarche and significant enrichment of age at menarche-associated variants in rare puberty disorders, such as LEPR, TACR3, and GNRH1, as well as imprinted regions with parent-of-origin-specific manner, such as DLK1, MKRN3, and KCNK9 [32]. In the largest study of 1000 Genomes Projects genomic data in 329,345 women of European ancestry, 389 independent genome-wide signals for age at menarche were identified [37]. These signals can explain ~7.4% of the variation in age at menarche, corresponding to ~25% of the estimated heritability [37]. This study also identified a significant role of imprinted genes in the regulation of pubertal timing with parent-of-origin-specific association, such as MKRN3 and DLK1 [32]. In males, genetic studies of puberty are much fewer and smaller in scale because of lack of data on male pubertal milestone, however, recent study using data of UK Biobank and 23andMe reported that genes implicated in pubertal timing include ALX4, SRD5A2, and INHBB [38]. They also found correlation between several adverse health outcomes and earlier male pubertal timing [38].

In case of central precocious puberty (CPP), monogenic mutations in MKRN3, DLK1, KISS1, KISSIR, and other candidate genes, such as genes biologically linked to gonadotropin signaling and genes encoding steroidogenesis enzymes, have been reported [32]. A recent study
characterized the genetic predisposition to CPP in girls of Han Chinese ancestry in Taiwan, and 105 loci were newly identified as genetic risk factors for early puberty [39]. They also found 33 SNPs from previous GWASs of pubertal timing [39]. Therefore, rare monogenic disease and common SNPs in puberty shared a genetic basis.

Early GWASs for age at menarche found significant signals in four loci previously associated with BMI, such as FTO, SEC16B, TRA2B, and TMEM18, and three in or near genes implicated in energy homeostasis, such as BSX, CRTC1, and MCHR2 [40]. Similarly, in GWASs that expanded the sample size and number of identified loci, several loci both pubertal timing and BMI overlapped [37, 39]. Both energy balance and reproduction are modulated by the peripheral signals, such as leptin and ghrelin [40].

**Founder effects in pediatric endocrine diseases**

1. **Founder effects and genetic drift**

Conducting GWASs in isolated populations is helpful to identify the founder effect owing to geographic or cultural barriers that have isolated genetic flows in neighboring populations. Founder mutation can result either from a true founder event or from a bottleneck effect. The former occurs when a subset of individuals separated from a larger population to establish new populations, and the latter is caused by a marked reduction in population size due to migration or isolation, which cause changes in allele frequency [41]. Founder populations can increase the frequency of certain autosomal recessively inherited Mendelian disorders by genetic drift. Genetic drift causes random fluctuations in the number of alleles in a population, resulting in a dramatic reduction in population size before recovery (bottleneck effect). Therefore,
populations with strong genetic drift tend to have one predominant mutant allele. The allele frequency of common and rare genetic variants in the population is influenced by natural selection through the evolution and demographic history [41].

The “Out of Africa” theory suggests that modern humans originated from a small population residing in Africa [42]. Accordingly, the most common genetic variants can be traced back to the ancient African population, from whence they were shared across the world, while rare variants are typically restricted to closely related populations [42]. A great demographic expansion began approximately 45,000–60,000 years ago in Africa, when highly rapid migration of populations occurred across the Eurasian continents [35]. Accordingly, there is a continuous loss of genetic diversity within populations living outside of Africa, resulting in “serial founder effects”.

2. Founder mutations in endocrine diseases

Several founder mutations have been identified in endocrine diseases (Table 1). Isolated GnRH deficiency (IGD) is a rare reproductive disorder with remarkable allelic and locus heterogeneity. Over the last three decades, genetic approaches have identified > 60 genes for IGD with various modes of inheritance [43]. Given the typical phenotype of IGD, it is the consequence of loss-of-function mutations, and most of which are private and non-recurrent [44]. However, haplotype analyses demonstrated eight mutations in five genes, including GNRHR, TACR3, PROKR2, FGFR1, and HS6ST1 founder mutations [44]. The estimated age of the four mutant alleles in GNRHR, TACR3, FGFR1, and HS6ST1 were approximately 5,000 years, corresponding to a time of rapid population expansion. In contrast, the PROKR2 p.L173R founder mutation is approximately 9,000 years old, which may confer a heterozygote
advantage. This finding suggests that the persistence of the loss-of-function alleles in diverse populations might reflect a potential heterozygote advantage to the carriers.

As the Korean population has been settled in the Korean peninsula as a single nation for thousands of years, several recurrent mutations have been identified as founder mutations. The p.Q258* of STAR in congenital lipoid adrenal hyperplasia (CLAH) is another example of a founder mutation [45]. This mutation is relatively common in China, Japan, and Korea. These countries are geographically adjacent and are influenced culturally by the movement of people. The age of the mutation is estimated to be about 5,000 years, which corresponds to the time when the Korean people settled in the Korean peninsula. Therefore, studying the founder effect can provide information about a population’s evolution and its migration pathways, thus enabling the screening of at-risk individuals.

Pathogenic variants in RET cause multiple endocrine neoplasia type 2 (MEN2), an autosomal dominantly inherited cancer syndrome. The codon-specific pathogenic variants in RET present strong correlations with the MEN2 phenotype [46]. The cysteine-rich extracellular domain in exons 10 and 11 frequently undergoes genetic alterations and is associated with pheochromocytoma, primary hyperparathyroidism, or both. In a recent comprehensive study of the geographical profile of RET variants, codon 634 mutations seem to be the most prevalent worldwide; however, founder effects of uncommon variants are observed in some countries [46-50]. The other cancer predisposing genes encoding succinate dehydrogenase (SDH) subunits cause paragangliomas (PGL). National datasets have identified high prevalence of particular pathogenic variants in SDHx according to geographical regions [51, 52]. The SDHD p.Y114C variant in the Mocheni valley close to Trentino, Italy, which is probably originated from Germany in 600–700 years ago [52]. The SDHC p.R133* mutation has been reported in
PGLs in French Canadians [53]. In the Netherlands, the SDHD founder mutation, p.D92Y, accounted for almost 70% of all SDHx carriers/cases [54]. Dutch founder effect for PGL was also found in South Africa, indicating potentially derived from historical Dutch emigration [51]. This molecular epidemiology can be helpful in screening for genetic risks and preventing the occurrence of cancer in individuals at risk.

Seven unrelated patients of the Perm Tatar ethnic group presented with hypoglycemia and excessive weight gain, low plasma adrenocorticotropin and cortisol levels [55]. These patients showed unusual clinical features of chronic obstructive pulmonary disease. Using whole exome sequencing (WES) and functional assay, a homozygous variant of the 5’-untranslated region in POMC (c.-71+1G>A) led to significant decrease in the POMC mRNA expression, and patients’ haplotype analysis suggested founder effect, which occurred at least 106.7 years ago. Another rare monogenic form of diabetes, Wolfram syndrome type 1, has also been reported to be associated with the founder effect of WFS1 in Southern India [56]. This study found 11 patients carrying WFS1 variants using NGS and identified recurrent WFS1 mutation of p.A370Sfs*173. Base on the haplotype analysis, microsatellite markers were shared in patients harboring the mutation. These rare disorders can lead to severe comorbidities, therefore, early diagnosis and genetic counseling are crucial.

Founder mutations have generated considerable interest in human genetics because their study may facilitate tracing ancestry, their migration, and the history of human populations. Indeed, screening for one or a few prevalent founder mutations is more efficient than testing for many rare mutations. Additional advantages of founder mutation screening are to find simple ways to identify at-risk groups, come up with new ideas for preventing and treating conditions associated with these mutations, and study their prevalence and penetrance in the population.
Next-generation sequencing (NGS) for pediatric endocrine diseases

1. Development of next-generation sequencing

Testing a single gene or small number of genes may be preferable if a disease is phenotypically and/or genetically highly homogeneous. However, NGS can be applied to phenotypically and genetically heterogeneous conditions. The availability of NGS allows for not only the identification of novel candidate genes but also an in-depth understanding of the architecture of several endocrine diseases. Several different NGS approaches are available that allow the sequencing of several regions of interest or the entire exome or genome.

NGS is another genomic approach for pediatric endocrine disorders. NGS platforms perform massively parallel sequencing of multiple genes or entire exomes or genomes. NGS is efficient and cost effective compared to sequential gene testing by Sanger sequencing [2]. The diagnostic efficiency of clinical WES for rare Mendelian disorders ranged from 20 to 30% [57]. However, in patients with distinctive phenotypes, such as neonatal diabetes or primary adrenal insufficiency, diagnostic yield increased up to 80% [58, 59]. Therefore, we have to select optimal sequencing platforms according to the patients’ phenotype and genetic heterogeneity.

Testing a single gene or small number of targeted genes may be preferable if the disease is phenotypically and/or genetically homogeneous. However, for phenotypically and genetically heterogeneous conditions, whole-exome or whole-genome sequencing can be applied.

2. Applications of NGS for pediatric endocrine diseases

The first discovery of gene mutations with respect to endocrine diseases by NGS was
published in 2011 [60]. This study identified two recurrent somatic mutations in KCNJ5 in eight of 22 patients with aldosterone-producing adenomas. The patients presented with hypertension and primary hyperaldosteronism. *In vitro* study demonstrated increased sodium conductance through the mutant channels, leading to the membrane depolarization of adrenal cortical cells and the stimulation of aldosterone release and cell proliferation [60].

In this context, NGS can be applied to many pediatric endocrine disorders with genetic heterogeneity, such as hypogonadotropic hypogonadism, disorders of sex development (DSD), skeletal dysplasia, Noonan syndrome and related disorders, congenital pituitary hormone deficiency, and monogenic diabetes (Table 2). Representatively, loss-of-function mutations in two causative genes, *MKRN3*, and *DLK1*, were identified by NGS in patients with CPP. *MKRN3* plays a role in the inhibition of factors that stimulate pubertal pulsatile GnRH secretion, and *DLK1* acts as an adipogenesis gatekeeper by preventing adipocyte differentiation. A recent meta-analysis suggested that mutations in *MKRN3* were more frequently found in familial cases (prevalence 19%, 95% CI, 0.05–0.36) than in sporadic cases (prevalence 2%, 95% CI, 0.01–0.04) [61]. Subgroup analysis showed that prevalence were higher in males, familial cases, and non-Asian countries [61]. In the case of *DLK1* and *KISS1*, although loss-of-function mutations have been shown to cause CPP, it has been rarely reported [32]. Interestingly, *MKRN3* and *DLK1* are both maternally imprinted and paternally expressed genes [32]. These findings suggest a role of genomic imprinting in regulating the timing of human puberty.

DSD is caused by a number of genetic etiologies with varying phenotypes, and genetic diagnosis using Sanger sequencing can be achieved in only 20% of patients. The diagnostic efficiency of targeted gene panel sequencing of 67 genes was about 30% in 44 patients with DSD [62]. NGS can be considered as a first-tier diagnostic tool for DSD with diverse genetic
heterogeneity and wide phenotypic spectrum. However, the diagnostic yield is not more than 50% due to key novel genes and genomic changes in enhancers or regulatory regions, copy-number variants, somatic changes during early embryonic life, or epigenetic or environmental factors [2], necessitating whole-genome sequencing or mapping technologies.

IGD is another prototype of endocrine disorders with genetic heterogeneity. In our group, the targeted gene panel sequencing of 69 genes identified pathogenic or likely pathogenic variants in 37% of patients [63, 64]. Among them, mutations in the FGFR1 gene were the most common [63]. However, genetic defects of IGD cannot be identified in over 50% of cases because of technical limitations of WES; mutations in non-coding areas; or cryptic, structural defects in the genome; or epigenetic factors.

Monogenic diabetes is characterized by a Mendelian inheritance pattern with a large effect size of causal variants and minimal environmental contributions [21]. Molecular diagnosis by NGS is also useful for monogenic diabetes, as described previously [21]. Genetically confirmed monogenic diabetes accounted for 5.1% of patients in a single-center study in Korea [65]. In targeted gene panel sequencing in 109 Korean patients with suspected monogenic DM, 21% of patients harbored pathogenic or likely pathogenic variants [66]. Characteristics of patients with pathogenic or likely pathogenic variants included lower BMI, higher MODY probability, and lower C-peptide levels [21]. Molecular genetic approaches for monogenic DM have clinical implications that include providing therapeutic strategies, such as insulin or sulfonylureas, and predicting outcomes, such as transient or permanent DM [21].

Conclusions
Recent advances in molecular genetics have advanced our understanding of the molecular mechanisms of pediatric endocrine disorders and now play a role in mainstream medical practice. The improved predictive power of GWAS variants makes the application of GWAS data a useful clinical tool for the near future. The identification of founder effects within a population offers powerful advantages for efficiently localizing the genes that underlie Mendelian disorders. NGS technology is beneficial both for clinical practice and for research on pediatric endocrinology. In addition, genome-wide sequencing can offer the benefit of reanalysis over time to incorporate advances in knowledge. In summary, the application of molecular technology has allowed us to further our understanding of endocrine diseases, and to have an impact on the practice of pediatric endocrinology related to diagnosis and genetic counseling.
Conflicts of interest

No potential conflict of interest relevant to this article was reported.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. NRF2021R1F1A104593011).
References


**Figure legends**

**Fig. 1.** Spectrum of genetic impacts/architectures in Mendelian and polygenic disorders. Mendelian diseases are caused by rare variants in a single gene, each with a strong effect on disease risk. Common diseases and traits are caused by the combined effects of multiple variants observed frequently in the population, each with a modest effect. Common variants are mostly ancient and typically have relatively modest clinical effects, whereas rare variants tend to have arisen more recently and can exert larger clinical effects.

**Fig. 2.** Effects of genetic drift and evolutionary force in isolated population. Marked reductions in population size due to migration or isolation cause changes of allele frequency, followed by reduction in diversity and founder effects.

**Fig. 3.** Comparison of sequencing platforms. Compared to Sanger sequencing, targeted gene panel or whole-exome sequencing utilize sequence reads concentrated over the coding portions of genes. In contrast, whole-genome sequencing analyzes almost the entire genomic sequence.
Table 1. Founder mutations in endocrine diseases

<table>
<thead>
<tr>
<th>Disease (references)</th>
<th>Gene</th>
<th>Mutation</th>
<th>Prevalence of mutation among affected</th>
<th>Region or ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p.R262Q</td>
<td>12/12</td>
<td>Caucasian, South Asian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p.R139H</td>
<td>15/15</td>
<td>Caucasian</td>
</tr>
<tr>
<td></td>
<td><strong>TACR3</strong></td>
<td>p.W275*</td>
<td>3/7</td>
<td>Caucasian</td>
</tr>
<tr>
<td></td>
<td><strong>PROKR2</strong></td>
<td>p.R85H</td>
<td>3/3</td>
<td>Caucasian</td>
</tr>
<tr>
<td></td>
<td><strong>FGFR1</strong></td>
<td>p.R250Q</td>
<td>4/5</td>
<td>Caucasian, South Asian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p.G687R</td>
<td>4/4</td>
<td>Caucasian, African American, South Asian</td>
</tr>
<tr>
<td></td>
<td><strong>HS6ST1</strong></td>
<td>p.R382W</td>
<td>5/5</td>
<td>Caucasian, South Asian</td>
</tr>
<tr>
<td>Congenital lipoid adrenal hyperplasia [45]</td>
<td><strong>STAR</strong></td>
<td>p.Q258*</td>
<td>74/84 alleles</td>
<td>Korean</td>
</tr>
<tr>
<td>Medullary thyroid cancer [47, 48]</td>
<td><strong>RET</strong></td>
<td>p.S891A</td>
<td>28/51</td>
<td>Northern Italy</td>
</tr>
<tr>
<td>Disorder</td>
<td>Gene</td>
<td>Mutation</td>
<td>Geographic Origin</td>
<td>Cases</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>------</td>
<td>----------</td>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Familial medullary thyroid cancer</td>
<td>RET</td>
<td>p.C618R</td>
<td>Cyprus</td>
<td>9/11</td>
</tr>
<tr>
<td>Multiple endocrine neoplasia type 2A</td>
<td>RET</td>
<td>p.C611Y</td>
<td>Denmark</td>
<td>12/21</td>
</tr>
<tr>
<td>Multiple endocrine neoplasia type 2B</td>
<td>RET</td>
<td>p.M918V</td>
<td>Brazil</td>
<td>42/80</td>
</tr>
<tr>
<td>Pheochromocytomas and paragangliomas</td>
<td>SDHD</td>
<td>p.D92Y</td>
<td>Netherland</td>
<td>24/32</td>
</tr>
<tr>
<td></td>
<td>SDHC</td>
<td>p.L139P</td>
<td></td>
<td>6/32</td>
</tr>
<tr>
<td></td>
<td>SDHD</td>
<td>p.R133*</td>
<td>French Canadian</td>
<td>9/13</td>
</tr>
<tr>
<td>Proopiomelanocortin deficiency</td>
<td>POMC</td>
<td>c.-71+1G&gt;A</td>
<td>Perm Tatar ancestry</td>
<td>Index cases</td>
</tr>
</tbody>
</table>
Table 2. Applications of next-generation sequencing for genetic diagnosis of pediatric endocrine diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Biologic mechanisms of involved genes</th>
<th>Involved genes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypogonadotropic hypogonadism</td>
<td>Development and/or migration of GnRH neurons</td>
<td>ANOS1, AXL, CHD7, DUSP6, FEZF1, FGF8, FGF17, FGFR1, FLRT3, HS6ST1, IL17RD, NSMF, PROK2, PROKR2, SPRY4, SEMA3A, SEMA3E, SEMA7A, TUBB3</td>
<td>[43, 63]</td>
</tr>
<tr>
<td>Hypothalamus/pituitary development</td>
<td>Hypothalamus/pituitary development</td>
<td>FSHB, GLCE, HESX1, LHB, LHX3, PROP1, SOX10, SOX2, SOX3</td>
<td>[43, 63]</td>
</tr>
<tr>
<td>Neuroendocrine regulation</td>
<td>Neuroendocrine regulation</td>
<td>GNRH1, GNRHR, IRF2BPL (EAP1), KISS1, KISS1R, LEP, LEPR, NDN, SRA1, TAC3, TACR3</td>
<td>[43, 63]</td>
</tr>
<tr>
<td>As a part of diverse syndromic disease</td>
<td>As a part of diverse syndromic disease</td>
<td>BBS1, BBS2, ARL6, BBS4, BBS5, MKKS, BBS7, TTC8, BBS9, BBS10, TRIM32, BBS12, CPE, DCAF17, DMXL2, RAB18, RAB3GAP1, RAB3GAP2, RBM28, RNF216</td>
<td>[43, 63]</td>
</tr>
<tr>
<td>Disorders of sex development</td>
<td>Development and function of gonad</td>
<td>AR, HSD17B3, LHCGR, MAP3K1, MYRF, NR0B1, SOX9, SRD5A2, SRY, WAGR, WT1, WNT4</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Noonan syndrome and related disorders</td>
<td>Regulation of the RAS/MAPK pathway</td>
<td>BRAF, HRAS, KRAS, LZTR1, MAP2K1, MAP2K2, MRAS, NRAS, PT, PN11, RAF1, RASA2, RIT1, RRAS2, SOS1, SOS2, SHOC2</td>
<td></td>
</tr>
<tr>
<td>Monogenic diabetes</td>
<td>Maturity onset diabetes of the young</td>
<td>GCK, HNF1A, HNF4A, HNF1B, INS, NEUROD1, PDX1, PAX4, AB, CC8, KCNJ11, KLF11, CEL, BLK</td>
<td></td>
</tr>
<tr>
<td>Neonatal diabetes</td>
<td>Rare forms of diabetes</td>
<td>ABCC8, KCNJ11, GCK, GATA6, INS, PTF1A, EIF2AK3, RFX6, GLIS3, GATA4, GATA6, IER3IP1, INSR, KCNJ11, PTF1A, RFX6, SLC19A2, SLC2A2, WFS1, mitochondrial genes</td>
<td></td>
</tr>
<tr>
<td>Central precocious puberty</td>
<td>Release of HPG axis inhibition</td>
<td>DLK1, MKRN3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stimulation of the HPG axis</td>
<td>KISS1, KISS1R</td>
<td></td>
</tr>
<tr>
<td>Severity</td>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mendelian diseases</td>
<td>Mendelian oligogenic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare variants in a single gene <em>(monogenic)</em></td>
<td>Common variants in many genes <em>(polygenic)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely rare (&lt;1:1000)</td>
<td>High frequency (&gt;1:20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High penetrance <em>Strong effects</em> (&gt;50-fold risk)</td>
<td>Low penetrance <em>Typically modest effects</em> (&lt;1.5-fold risk)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arisen more <em>recently</em> from spontaneous mutation</td>
<td>Mostly <em>ancient</em> Evolutionary selective pressures</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The **number** of genetic variants/genes

Their **frequency** in the population

Their respective contributions to **risk** (i.e., penetrance)

The periods the genetic effects arise
Fig. 2

**Founder Effect**
- Founder population
- Originating population

**Bottleneck Effect**
- Post-bottleneck population
- Time
Fig. 3

- **Sanger sequencing**
- **NGS gene panel (Only targeted genes)**
- **Whole-exome sequencing**
- **Whole-genome sequencing**