Peer-reviewed version available at *Genes* **2020**, *11*, 987; doi:10.3390/genes11090987

*Review*

# A Systematic Review on Extreme Phenotype Strategies to Search for Rare Variants in Genetic Studies of Complex Disorders

**Sana Amanat[1], Teresa Requena[2] and Jose A. Lopez-Escamez[1,3,4] ***

[1] Otology & Neurotology Group CTS495, Department of Genomic Medicine, GENYO - Centre for Genomics and Oncological Research – Pfizer/University of Granada/ Junta de Andalucía, PTS, Granada, Spain
[2] Centre for Discovery Brain Sciences, Edinburgh Medical School: Biomedical Sciences, University of Edinburgh, Edinburgh, UK
[3] Department of Otolaryngology, Instituto de Investigación Biosanitaria ibs.GRANADA, Hospital Universitario Virgen de las Nieves, Universidad de Granada, Granada, Spain
[4] Department of Surgery, Division of Otolaryngology, Universidad de Granada, Granada, Spain
**Corresponding author**. Phone. +34 958 715 500-160 E-mail: antonio.lopezescamez@genyo.es

**Abstract.** Exome sequencing has been commonly used in rare diseases by selecting multiplex families or singletons with an extreme phenotype (EP) to search for rare variants in coding regions. The EP strategy covers both extreme ends of a disease spectrum and it has been also used to investigate the contribution of rare variants to heritability in complex clinical traits. We have conducted a systematic review to find evidence supporting the use of EP strategies to search for rare variants in genetic studies of complex diseases, to highlight the contribution of rare variation to the genetic structure of multiallelic conditions. After performing the quality assessment of the retrieved records, we selected 19 genetic studies considering EP to demonstrate genetic association. All the studies successfully identified several rare variants, *de novo* mutations and many novel candidate genes were also identified by selecting an EP. There is enough evidence to support that the EP approach in patients with an early onset of the disease can contribute to the identification of rare variants in candidate genes or pathways involved in complex diseases. EP patients may contribute to a better understanding of the underlying genetic architecture of common heterogeneous disorders such as tinnitus or age-related hearing loss.

**Keywords:** genetic association studies; extreme phenotype; genetic epidemiology; tinnitus

## 1. Introduction

A clinical phenotype is the set of observable signs, symptoms and behavioural features associated with a human disorder. The phenotype includes multiple features or traits and it may be categorical (male or female sex) or quantitative (glucose levels or hearing thresholds). These observable variations in the phenotype for a disorder is known in genetics as expressivity and it may range from mild to severe [1,2]. Phenotypic variation in quantitative traits can be represented by a bell shape graph where mild and severe phenotypes are located in both tails of the distribution. However, the majority of the subjects show an intermediate phenotype (Fig. 1).

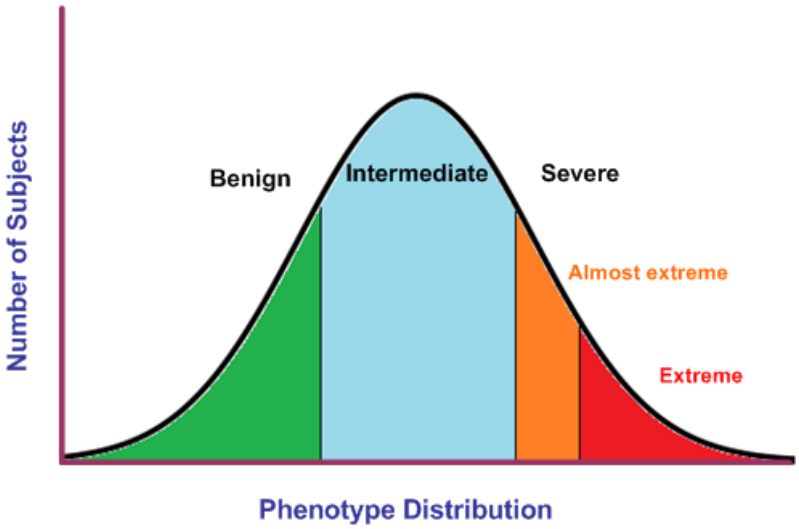


**Figure 1.** Phenotypic variation in quantitative traits. Individuals can be classified as benign, intermediate or severe according to general and disease-specific criteria. Extreme phenotype includes both ends of the normal distribution (green, orange and red areas).

The genetic architecture in human diseases plays an important to better understand the genetic variants that can influence the phenotype in complex diseases[3]. To uncover the missing heritability, Next Generation Sequencing (NGS) technology has been used to elucidate the genetic contribution to common and rare diseases with underlying heterogeneity. In particular, Whole Exome Sequencing (WES) provides an opportunity to capture rare and ultra-rare alleles, residing in protein-coding genes, influencing disease risk with a high effect size. So, in the last few years, several novel genes have been identified in various neurological diseases by utilizing WES, such as epileptic encephalopathies (*KCNQ2, STXBP1, and KCNB1)* and Parkinson's disease (*VPS13C, ARSB, PTPRH, GPATCH2L,and UHRF1BP1L*) [4,5,6.]

A significant increase in the prevalence of complex diseases has been reported the last decades such as bipolar disorder, coronary artery disease [7], type 2 diabetes, hypertension, obesity and cancer [8]. This increase could be related to environmental factors such as diet or lifestyle changes. However, the genetic contribution to complex conditions is still largely unknown, since the contribution of rare variation to heritability is still missing. There are several factors that limit the power of gene-discovery approaches such as phenotypic variance [9], the overlap of clinical features with similar conditions, the minor allelic frequency (MAF), heterogeneous nature of loci, and the low effect size of the potential risk alleles [10]. The underlying hypothesis is that extreme phenotype (EP) will occur in extreme cases with an excess of rare variants as an additive effect on common variants for the trait of interest. The EP strategy aims to identify rare genetic variants causing a large effect on disease risk [11,12.] The EP study design includes the selection of individuals that covers the extreme ends of a disease phenotype distribution. These extreme subjects may include early or late age of onset, benign or severe form of disease, family history, fast progression of symptoms, very high or very low scores on psychometric tests or extreme levels of a biomarker [13,14,15.] This strategy may identify rare genetic variants by sequencing a relatively small sample size, and it can target novel candidate genes, since rare variants that contribute to a particular trait are enriched in both extremes of a disease distribution [10]. The combination of EP with WES has successfully identified several rare mutations and candidate genes in diabetic retinopathy [16], bipolar disorder [17] and cystic fibrosis [18] across diverse ethnic groups. The aim of this systematic review is to critically analyse the contribution of extreme phenotype strategies to uncover novel mutations or candidate genes in genetic studies of complex disorders.

## 2. Materials and Methods

### *2.1. Study design*

This is a systematic review of genetic studies in complex diseases and it follows Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines (Table S1) [19] and recommendations from Human Genome Epidemiology Network (HuGENet) review handbook (https://www.cdc.gov/genomics/hugenet/).

### *2.2. Search strategies*

Literature search was performed on 12th December, 2019 for EP strategy using two bibliographic databases (PubMed and Embase). For EP strategy the following keywords "phenotypic extreme", "extreme phenotype", "rare variant" and "genetics" were used to formulate the search string. The selected keywords could appear in the title, abstract, text word, author keywords or MeSH Terms of the articles. The keyword string used for literature search in PubMed was: ((((("phenotypic extreme"[Title/Abstract] OR "extreme phenotype")[Title/Abstract] AND ("rare variant"[Title/Abstract] OR "genetics")[Title/Abstract])) OR (("phenotypic extreme"[Text Word] OR "extreme phenotype")[Text Word] AND ("rare variant"[Text Word] OR "genetics")[Text Word])) OR (("phenotypic extreme" OR "extreme phenotype") AND ("rare variant" OR "genetics") [MeSH Terms]) and for Embase it was: ('phenotypic extreme': ti, ab, kw OR 'extreme phenotype': ti, ab, kw) AND ('rare variant': ti, ab, kw OR 'genetics': ti, ab, kw) AND [2009-2019]/py AND [english]/lim. The records with a publication date < 10 years, literature in English language and only human studies were included during literature search by configuring filters if available e.g. on PubMed.

### *2.3. Research question and selection criteria*

The objective of this systematic review is to assess the evidence to support the design of genetic studies using extreme phenotype strategies to find novel mutations or genes in complex disorders. According to this hypothesis, we formulated the following research question: "are EP strategies useful to establish the genetic contribution in complex diseases?". To answer this question, we followed the PICO strategy:

- Population: Patients with a complex disease or condition.
- Intervention: Selection of individuals according to any extreme phenotype criteria (i.e., early onset, fast progression of disease, very high or very low scores on psychometric tests or extreme levels of a biomarker).
- Comparison: Genetic association studies (GWAS, WES, genotyping, Sanger sequencing or targeted sequencing).
- Outcome: genetic finding reported (rare variants, candidate genes or pathways associated with the condition of interest).
- Study design: case-control, case-report, case-cohort or trios.

### *2.4. Exclusion criteria*

- Studies in non-human models.
- Studies not published in English.

### *2.5. Quality assessment of selected studies*

The extracted records were screened to remove the duplicate entries. The title and abstract of all articles were reviewed to exclude reviews, meta-analysis, and irrelevant records (non-genetic studies, pharmacogenomics or clinical studies). The search was conducted primarily for rare variants, but any type of variants were retained and included in this systematic review. After screening, the obtained records were considered for full-text assessment in the next step. To assess the quality of these articles we formulated 8 questions for EP studies (Table 1). For each question, a positive answer was scored

as 1 and a negative answer as 0. Each author classified and rated each record independently of each other. Differences in the scores were discussed to get a final consensus score .If a record had achieved ≥ 60% of the total score, the response to Q8 was "yes" and the reported rare variants (MAF<0.05), then the record was selected for synthesis. So, only studies with significant results were included". Two of the authors carried out the synthesis (SA, JALE). The outcome for each selected study was assessed according to Q8 and the following criteria: if a given study had found any rare variant, common variant, *de novo* mutation, copy number variants, candidate genes or pathways for EP subjects then the major outcome was considered as positive.

**Table 1.** Criteria used to assess the quality of the genetic studies using an extreme phenotype approach.

| No. | Question | Answer |
|---|---|---|
| Q1 | Is there enough description of the study design? | Yes/No |
| Q2 | Has the study described the method of sequencing/genotyping? | Yes/No |
| Q3 | Has the study provided information about population ancestry? | Yes/No |
| Q4 | Is there any information on sex of selected individuals? | Yes/No |
| Q5 | Is there any information on age of onset? | Yes/No |
| Q6 | Has the study used extreme phenotype criteria for sample recruitment? | Yes/No |
| Q7 | Has the study performed sex specific analysis for genetic associations? | Yes/No |
| Q8 | Has the study reported significant genetic findings? | Yes/No |

*2.6. Data extraction and synthesis*

The following information was extracted from each article for the studies selected for synthesis: first author's last name, publication year, disease/disorder name, population ancestry, study design, sequencing method, EP/disease phenotype criteria, sample size for cases, age of onset, sex of individuals, MAF and the main genetic findings. Moreover, the phenotype criteria and the main genetic findings for EP were of great interest for synthesis.

*2.7. Risk of bias*

The Cochrane collaboration tool[20] was used to assess the risk of bias for each selected studies, Table S2.

## 3. Results

*3.1. Selection and characteristics of EP studies*

For the EP strategy, we retrieved 106 records in total, 66 records from PubMed and 40 from Embase by using the search strings discussed in the search strategy section. After duplicate removal, we retained 89/106 records aggregated from both databases. Next, after screening by title and abstract of the articles, we retrieved 30/89 records that were included for full text assessment. The discarded records were reviews, meta-analyses, non-genetic studies, pharmacogenomics, posters or abstracts presented at scientific meetings. All studies including variants with MAF>0.05, single case or < 5 patients with EP were excluded. We performed the quality assessment on 30 articles and 19/30 records surpassed the minimum quality assessment score and were considered for synthesis. (Fig. 2, Table S3).

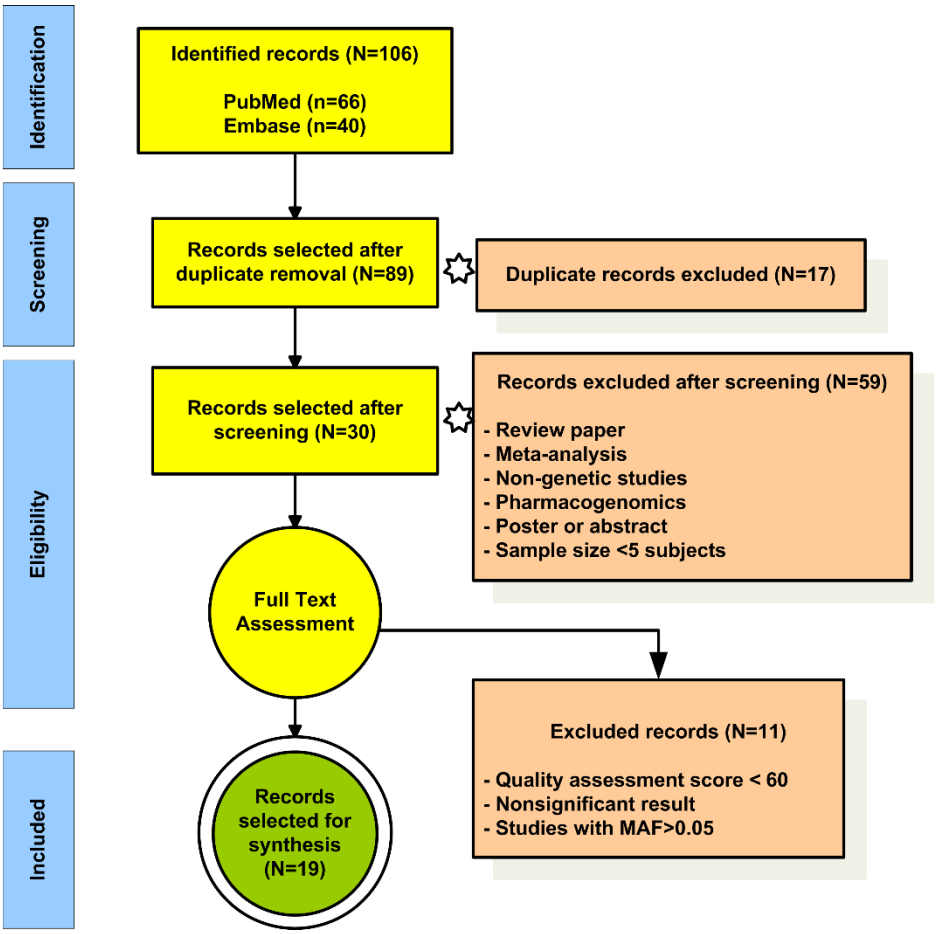


**Figure 2.** Flowchart to select extreme phenotype records for synthesis.

Among the 19 studies selected for synthesis, 16 records were related to physical conditions, 1 on bipolar disorder and 2 related to neurological disorders including epilepsy, and Alzheimer's disease. All of these studies reported rare variants, candidate genes or potential pathways associated with a particular trait using an EP approach. These 19 EP studies covered 18 complex diseases.

Information about population ancestry and sample size of cases was available for all 19 studies. Only 11/19 studies reported the age of onset and 18/19 records reported the sex of individuals. The most common criteria to define EP included early-onset, late-onset, family history, acute form, and/or fast progression of a disease. In addition, disease specific features were also considered to define an EP such as the worst score in biomarkers levels including Bone Mass Density (BMD) or spirometry-based severity using Global Initiative for Chronic Obstructive Lung Disease (GOLD) grade. The reported sample size had a range between 12-32,965 individuals. A summary of the characteristics of these 19 EP studies are shown in Table 2.

### *3.2. Synthesized findings of EP studies*

In 19 EP selected studies the combination of general and disease specific EP criteria were used to select individuals. Information on the study design, sequencing technique and ancestry population was available for all 19 studies. The reported sample size ranged according to the design and sequencing method; 1711±2513 (mean ± SD) for GWAS, 929 ± 2389 for genotyping, 1274 ±9380 for WES, 29 ± 9 for targeted sequencing and 949 ±8742 for Sanger sequencing. All 19 selected studies using EP to select individuals reported significant findings including several rare variants, copy number variants, potential candidate genes or pathways associated with the condition of interest. WES was able to find rare variants in 13/19 studies (MAF= 0.00-0.05) in identified variants. It also

helped in the identification of several novel candidate genes including *TACC2*[21], *PRKCD, C1QTNF4, DNMT3A* [22], *LOC728699*, and *FASTK* [16]. GWAS identified a rare variant in 1/19 study (MAF= 0.04). In addition, genotyping, targeted and Sanger sequencing also contributed in the identification of many candidate genes and micro-deletions.

**Table 2.** Summary of the 19 genetic studies using an extreme phenotype approach selected for synthesis.

| Reference | Disease | EP criteria | Study design | Sequencing Method | Ancestry | Number of patients | Onset | Not Sex | Genetic findings | | AF (Ancestry dependent) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Gene/pathway | SNP/mutation | |
| **Pullabhatla et al. (2017)[22]** | **Systemic lupus erythematosus** | **Proband with early onset and clinical features with poor outcome** | **Family trios, Replication cohort** | **WES** | **EU** | **30 trios, *10995*** | **<25 yr** | **reported** | *PRKCD* | *3: 53223122 G>A* | *De novo mutation and novel genes* |
| | | | | | | | | | *C1QTNF4* | *11: 47611769 G> C* | |
| | | | | | | | | | *DNMT3A* | *2: 25457236 G> A* | |
| Johar et al.(2016)[23] | Polyautoimmunity | Polyautoimmunity and familial autoimmunity | Case-control, Cross-Sectional | WES | Colombian | 47 | *Not reported* | M,F | *PLAUR* | *rs4760* | *0.1* |
| | | | | | | | | | *DHX34* | *rs151213663* | *0.004* |
| | | | | | | | | | *SRA1* | rs5871740, rs202193903 | *Not found* |
| | | | | | | | | | *ABCB8* | 7:150744528:G>T, 7:150744370: CGT/- | *Not found* |
| | | | | | | | | | *MLL4* | rs186268702 | 0.0007 |
| Kunkle et al. (2017)[24] | Alzheimer disease | Early onset Alzheimer disease, familial or sporadic | Case-control, Replication cohort | WES | NHW and Caribbean Hispanic | 93, *8570* | <65 yr | M,F | *RUFY1* | *5:179036506:T>G* | *0.001* |
| | | | | | | | | | *RIN3* | *14:93022240:G>T* | *0.0005* |
| | | | | | | | | | *TCIRG* | *11:67810477:C>T* | *0.0007* |
| | | | | | | | | | *PSD2* | *5:139216541:G>A, 5:139216759:G>A* | *0.0006, 0.00005* |
| Emond et al.(2012)[13] | Cystic fibrosis (CF) | CF with early onset of persistent P. aeruginosa infection | Case-control, Replication cohort | WES | EU America, African American, White Hispanic, NHW, Asian, Aleut | 43, *696* | ≤2.5 yr | M,F | *DCTN4* | *rs11954652, rs35772018* | 0.048, 0.017 |
| Shtir et al. (2016)[16] | Diabetes | Diabetes for at least 10 years without diabetic retinopathy | Case-control, Cross-Sectional | WES | Saudi | 43 | *Not reported* | M,F | *FASTK* | *7:150774771:C>T, 7:150777859:A>T* | *0, 0* |
| | | | | | | | | | *LOC728699* | *rs149540491, rs117616768, 12:20704520:C>A* | *0.05, 0.01, 0.02* |
| Liu et al. (2016)[25] | Lung cancer | Familial or sporadic lung cancer cases, ever smokers or severe chronic obstructive pulmonary disease | Case-control, Cross-Sectional | WES | NHW | 48 sporadic and 54 familial | mean: 56 yr(familial) and 61 | M,F | *DBH* | *rs76856960* | 0.0034 |
| | | | | | | | | | *CCDC147* | *rs41291850* | 0.0026 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | yr(sporadic) | | | | |
| Husson et al.(2018)[17] | Bipolar I disorder | Family history of mood disorder and early onset | Case-control, Cross-Sectional | WES | EU | 92 | mean:24 yr | M,F | *>13 genes* | *>100 SNPs* | *0.000015-0.009* |
| Johar et al.(2015)[26] | Multiple autoimmune syndrome | Multiple autoimmune syndrome with Sjögren's syndrome | Case-control, Cross-Sectional | WES | Colombian | 12 | 28-67 yr | F | *LRP1/STAT6* | *12:57522754:A>C* | *Novel mutation* |
| Hiekkala et al.(2018)[27] | Hemiplegic migraine | ≥2 migraine attacks, completely reversible motor weakness | Case report, Cross sectional | WES | Finnish | 293 | median:12 yr | M,F | *ATP1A2* | *rs765909830, 1:160100376:G>A* | 0, 0 |
| | | | | | | | | | *CACNA1A* | *rs121908212* | 0 |
| Qiao et al.(2018)[28] | Chronic obstructive pulmonary disease(COPD) | COPD cases with GOLD grade 3 or 4 | Case-control, Cross-Sectional | WES | EU, NHW, African American | ≈1769 | >45 yr , ≤65 yr | M/F | *jak-stat signaling pathway* | - | *Not reported* |
| | | | | | | | | | *TBC1D10A, RFPL1* | *Not reported* | |
| Bruse et al.(2016)[21] | COPD | COPD cases with GOLD grade 3 or 4 | Case-control, Cross-Sectional | WES | NHW | 62 | *Not reported* | M/F | *TACC2* | chr10:123842508, 10:123844900, 10:123903149, 10:123970638, 10:123987443, 10:123996970, 10:124009124 | *0.000008901, 0.000008796, 0.001851, 0.000008999, Not found 0.03476, 0.07* |
| Nuytemans et al.(2018)[29] | Thrombotic storm (TS) | Severe onset of ≥2 arterial, unusual clot location, refractory, reoccurrence | Case report, Cross sectional | WES, Targeted sequencing | White and Indian | 26(13 trios) | *Not reported* | M,F | *STAB2* | *rs779748342, rs758868186, rs201799617, rs17034336, rs149382223* | *Not found, Not found, 0.0002, 0.0441, 0.0008* |
| | | | | | | | | | *CHPF* | *2:220405189:C>T* | *Not found* |
| | | | | | | | | | *CHST3* | *rs145384892* | *Not found* |
| | | | | | | | | | *SLC26A2* | *rs104893919, rs78676079* | *Not found, 0.0076* |
| | | | | | | | | | *CHST12* | *rs17132399* | *Not found* |
| | | | | | | | | | *CHPF2* | *rs776052782, rs117332591, rs377232422* | *Not found, 0.0028, Not found* |
| | | | | | | | | | *CHST15* | *rs34639461* | *0.011* |
| | | | | | | | | | *PAPSS2* | *rs45467596* | *0.0219* |
| Aubart et al.(2018)[30] | Marfan syndrome | Severe aortic features(dissection or preventive thoracic aortic | Case-control, Cross sectional | WES | EU | 51 EP and 8 sib-pairs | ≈10-30 yr | M,F | *COL4A1* | *c.4615C>T, c.1630G>C, c.4453T>C,* | *0.02, 0.04, 0.003* |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | aneurysm rupture surgery at a young age) or sib- pairs | | | | | | | *FBN1* | *c.1585C>T* | *0* |
| | | | | | | | | | *SMAD3* | *c.6424T>C* | *0* |
| Gregson et al. (2018)[31] | Bone mass density | Extreme high or moderate high bone mass density | Case-control, Replication cohort | GWAS | EU | 1258, *32965* | *Not reported* | M,F | *WNT4/ZBTB40* | *rs113784679* | 0.04 |
| Lee et al. (2018)[32] | Ulcerative colitis | Ulcerative colitis patients with good or poor prognosis | Case-control, Replication cohort | Genotyping | Korean | 881, *274* | 35.6 ± 13.9 yr | M,F | *HLA-DRA and HLA-DRB* | *rs9268877* | *0.000* |
| Tomaiuolo et al. (2012)[33] | Acute myocardial infarction (AMI) | AMI patients with first episode before or after 45 years of age | Case-control, Replication cohort | Genotyping | EU | 1653, *909* | *Not reported* | M,F | *MTHFR C677T, FII G20210A, Factor V Leiden* | *-455G>A* | - |
| Goldberg-Stern et al. (2013)[34] | Epilepsy with febrile seizures plus | Generalized epilepsy with febrile seizures plus, a proband with Dravet syndrome | Case-control, Cross sectional | *Sanger sequencing* | Ashkenazi Jewish | 14 familial cases | infancy to 7 yr | M,F | *SCN1A* | *c.4114A>G: p.K1372E; exon 21* | - |
| Shen et al. (2017)[35] | Spermatogenic failure | Spermatogenic failure with azoospermia, mild oligozoospermia or severe oligozoospermia | Case-control, Cross sectional | *Sanger sequencing* | Chinese Han | 884 | *Not reported* | M | *MAGEA9* | *Deletion (chrX:149580739-149580850)* | - |
| Uzun et al. (2016)[36] | Preterm birth | Patients delivering < 34 weeks | Case report, Cross sectional | Targeted Sequencing of 329 genes | African-American; Asian; Hispanic; White; Native American | 32 | *Not reported* | F | *WASF3* | *rs17084492* | *0.01357(NFE), 0.07(African)* |
| | | | | | | | | | *AZU1* | *rs28626600* | *0.1(NFE), 0.01662(African)* |

Legend: Non-Hispanic White (NHW), European (EU), Whole Exome Sequencing (WES).

## 4. Discussion

### *4.1. Summary of main findings*

Our systematic review shows that individuals with an EP have enough potential to reveal rare variants that may influence genetic susceptibility in most complex disorders. Complex disorders are considered as a heterogeneous spectrum of symptoms with variable expressivity observed on each patient. By cluster analysis, it is possible to identify subgroups of patients, and by selecting patients with EP (high expressivity), we would expect to find an enrichment of rare variation associated with the EP [37]. However, we cannot recommend a particular EP strategy to select patients, although the selection of individuals with an early onset and/or severe phenotype (genetic anticipation) will probably help in the search of rare variation. In contrast, elderly patients can show mutations associated with the exposure to environmental factors along life (ultraviolet radiation, chemical agents, pollutants) [38]. In general, the criteria to define EP was a combination of common and disease specific features such as chronic state of a disease, very high or low biomarker levels such as Bone Mass Density (BMD), spirometry-based severity level using Global Initiative for Chronic Obstructive Lung Disease (GOLD), family history and early/late onset age.

Of note, a large sample size was not required in WES studies for the discovery cohort and 10/19 records had a number of cases < 100. So, a moderate sample size of individuals with EP was enough to identify candidate mutations or genes. These individuals with EP were carriers of rare variants with a high effect size to target new candidate genes. The EP approach was reproducible across different populations, since the selected studies recruited cases with different ethnic backgrounds including Asian, African and European ancestry, and with monogenic disease such as cystic fibrosis [13] with an extreme phenotype (persistent tracheobronquial infection with early onset) [39]. So, the information about the age of onset and sex of selected individuals is essential to define an EP [40].

### *4.2. Selection of EP in quantitative traits*

Individuals with EP are characterized by extreme clinically relevant attributes, toxic effect or extreme responses to a treatment[1]. From a theoretical perspective, a very EP is more informative than an almost EP, but in practice there are several limitations associated with the very EP such as vulnerability to phenotype heterogeneity and measurement errors. If a significant proportion from both sides of an extreme is discarded, the almost EP can still be more powerful than random sampling of the same size. The benefits of EP sampling were demonstrated by proposing power calculation methods with the help of maximum likelihood approach [11,41]. It was also indicated that EP sampling to detect rare variants is more cost efficient as compared to traditional study designs [42]. Replication in a second independent EP cohort to enhance the power of the study is highly recommended, but it is unlikely to obtain a large sample size of EP subjects from a single region[43]. However; the EP approach is considered as an added value to the detection of rare variants associated with the trait over a random sampling[11]

### *4.3. Familial disorders as EP strategy*

Some common disorders show rare familial phenotypes with Mendelian inheritance associated with rare variants with large effect size. There are many studies on complex disorders using EP strategy in familial cases such as Alzheimer disease (AD) [24], polyautoimmunity disorder [23] and congenital hypothyroidism [44]. For example, a recent study using linkage analysis has demonstrated that selecting individuals with familial autoimmunity and polyautoimmunity as EP, it was possible to identify the *SRA1* gene (LOD score= 5.48) [23]. Furthermore, a WES study on AD analysed non-Hispanic White patients and Caribbean Hispanic families to find genes associated with early onset AD. Heterozygous non-synonymous variants with global MAF<0.001 were selected for variant prioritization and showed autosomal dominant segregation in these families. Several genes such as *RUFY, TCIRG1, PSD2* and *RIN3* were identified that could be involved in endolysosomal transport for both early and late onset AD [24]. In some complex diseases such as Meniere disease (MD), a

syndrome characterized by hearing loss, episodic vertigo and tinnitus, there is also a strong evidence of genetic predisposition with most of the families showing an autosomal dominant inheritance with almost 60% penetrance. By using WES in familial MD, a burden of multiplex rare missense variant in the *OTOG* gene was reported in 30% of familial cases [45], which illustrates the success of familial cases as EP. Furthermore, a study on genetic epilepsy with hay febrile seizures plus (Dravet syndrome) has reported *SCN1A* missense mutation in a large Jewish family (14/17 cases) with epilepsy syndrome from both extremes (low and high) [15] and a study on thyroid dysgenesis with congenital hypothyroidism found a familial *PAX8* mutation associated with EP [44].

### *4.4. An EP strategy to investigate the genetic contribution to tinnitus*

Tinnitus is the perception of noise in the absence of an external acoustic stimulation, affecting more than 15% of population with a decrease in health-related quality of life [46]. Several specific instruments have been defined to characterize chronic or severe tinnitus, and these instruments have been proposed to measure tinnitus annoyance to define EP for genetic studies [47]. Epidemiological evidence to support genetic contribution to tinnitus is still weak because of the heterogeneous nature of this condition. So, tinnitus can occur together with multiple comorbidities including hearing loss, migraine, sleep disorders, anxiety, other psychological conditions and some rare monogenic disorders [48]. The careful selection of the phenotype for genetic studies is crucial. The inclusion criteria should consider young individuals with severe forms of bilateral tinnitus to investigate the genetic contribution of rare variation to tinnitus. These individuals may carry a greater susceptibility and lower environmental load; however, severe forms of tinnitus in young individuals are rare [49].

### *4.5. Limitations*

Some weaknesses were found in the design of EP and they deserve further research. The replication of the genetic studies across different populations with different ethnic backgrounds has enough potential to validate the genetic associations [36,13]; however, the frequency of allelic variants is different across different populations, and specific reference data for allelic frequencies are needed for each population. The rare variants reported in simplex families with EP should be validated in more patients with severe phenotype [23]. Most of the studies have used WES rather WGS that can lead towards the loss of useful genetic information and erroneous results in calculating effect size of rare variants at individual level across a particular phenotype[17].

## 5. Conclusions

Genetic studies have confirmed the effectiveness of the EP strategies to establish the genetic contribution of rare variation to complex diseases.

**Acknowledgments:** We would like to thank Patricia Perez-Carpena, MD, PhD for the critical review and the useful comments to improve the manuscript and Marisa Flook for the English language editing.

**Author contributions:** JALE conceived the study design and develop the scientific arguments. JALE and SA performed literature search, quality assessment of the studies, interpretation of data, drafting the manuscript and revised the final version. TR also helped in the interpretation of the data, developing the scientific arguments and revised the final draft.

**Funding:** This study has been funded by H2020 MSCA-ITN-2016–722046 Grant (JALE). The project leading to these results has received funding from "la Caixa" Foundation (ID 100010434), under agreement LCF/PR/DE18/52010002 (JALE). This project is a part of European School of Interdisciplinary Tinnitus (ESIT) research and Sana Amanat is a PhD student in Biomedicine Program at the University of Granada.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pérez-Gracia, J. L. *et al.* Selection of extreme phenotypes: The role of clinical observation in translational research. *Clin. Transl. Oncol.* **12**, 174–180 (2010).

2. Zimmermann, E., Gamborg, M., Sørensen, T. I. A. & Baker, J. L. Sex differences in the association between birth weight and adult type 2 diabetes. *Diabetes* **64**, 4220–4225 (2015).
3. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
4. Lohmann, K. & Klein, C. Next Generation Sequencing and the Future of Genetic Diagnosis. *Neurotherapeutics* **11**, 699–707 (2014).
5. Srivastava, S. *et al.* Clinical whole exome sequencing in child neurology practice. *Ann. Neurol.* **76**, 473–483 (2014).
6. Jansen, I. E. *et al.* Discovery and functional prioritization of Parkinson ' s disease candidate genes from large-scale whole exome sequencing. 1–26 (2017). doi:10.1186/s13059-017-1147-9
7. Johnson, M. B. *et al.* A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. *Diabetologia* **61**, 862–869 (2018).
8. Verma, S. S. & Ritchie, M. D. Another round of "clue" to uncover the mystery of complex traits. *Genes (Basel).* **9**, (2018).
9. Craddock, N. *et al.* Dissecting the phenotype in genome-wide association studies of psychiatric illness. *Br. J. Psychiatry* **195**, 97–99 (2009).
10. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).
11. Bjørnland, T., Bye, A., Ryeng, E., Wisløff, U. & Langaas, M. Improving power of genetic association studies by extreme phenotype sampling: a review and some new results. 1–26 (2017).
12. Barnett, I. J., Lee, S. & Lin, X. Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. *Genet. Epidemiol.* **37**, 142–151 (2013).
13. Emond, M. J. *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat. Genet.* **44**, 886–889 (2012).
14. Bjørnland, T., Langaas, M., Grill, V. & Mostad, I. L. Assessing gene-environment interaction effects of FTO, MC4R and lifestyle factors on obesity using an extreme phenotype sampling design: Results from the HUNT study. *PLoS One* **12**, 1–16 (2017).
15. Goldberg-Stern, H. *et al.* Broad phenotypic heterogeneity due to a novel SCN1A mutation in a family with genetic epilepsy with febrile seizures plus. *J. Child Neurol.* **29**, 221–226 (2014).
16. Shtir, C. *et al.* Exome - based case – control association study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. *Hum. Genet.* **135**, 193–200 (2016).
17. Husson, T. *et al.* Identification of potential genetic risk factors for bipolar disorder by whole-exome sequencing. *Transl. Psychiatry* **8**, (2018).
18. Blue, E. *et al.* Variation in cilia protein genes and progression of lung disease in cystic fibrosis. *Ann. Am. Thorac. Soc.* **15**, 440–448 (2018).
19. Moher, D. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement David. *Syst. Rev.* **207**, 1–9 (2015).
20. Higgins, J. P. T. *et al.* The Cochrane Collaboration ' s tool for assessing risk of bias in randomised trials. 1–9 (2011). doi:10.1136/bmj.d5928
21. Bruse, S. *et al.* Whole exome sequencing identifies novel candidate genes that modify chronic obstructive pulmonary disease susceptibility. *Hum. Genomics* **10**, 1–12 (2016).
22. Pullabhatla, V. *et al.* De novo mutations implicate novel genes in systemic lupus erythematosus. **27**, 421–429 (2018).
23. Johar, A. *et al.* Definition of mutations in polyautoimmunity. *J. Autoimmun.* **72**, 65–72 (2016).
24. Kunkle, B. W. *et al.* Early-onset Alzheimer disease and candidate risk genes involved in endolysosomal transport. *JAMA Neurol.* **74**, 1113–1122 (2017).
25. Liu, Y. *et al.* Focused analysis of exome sequencing data for rare germline mutations in familial and sporadic lung cancer. *J. Thorac. Oncol.* **11**, 52–61 (2016).
26. Johar, A. S. *et al.* Novel and rare functional genomic variants in multiple autoimmune syndrome and Sjögren's syndrome. *J. Transl. Med.* **13**, 1–11 (2015).
27. Hiekkala, M. E. *et al.* The contribution of CACNA1A, ATP1A2 and SCN1A mutations in hemiplegic migraine: A clinical and genetic study in Finnish migraine families. *Cephalalgia* **38**, 1849–1863 (2018).
28. Qiao, D. *et al.* Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *Hum. Mol. Genet.* **27**, 3801–3812 (2018).

29. Nuytemans, K. *et al.* Variants in chondroitin sulfate metabolism genes in thrombotic storm. *Thromb. Res.* **161**, 43–51 (2018).
30. Aubart, M. *et al.* Association of modifiers and other genetic factors explain Marfan syndrome clinical variability. *Eur. J. Hum. Genet.* **26**, 1759–1772 (2018).
31. Gregson, C. L. *et al.* Genome-wide association study of extreme high bone mass: Contribution of common genetic variation to extreme BMD phenotypes and potential novel BMD-associated genes. *Bone* **114**, 62–71 (2018).
32. Lee, H. S. *et al.* An intergenic variant rs9268877 between HLA-DRA and HLA-DRB contributes to the clinical course and long-term outcome of ulcerative colitis. *J. Crohn's Colitis* **12**, 1113–1121 (2018).
33. Tomaiuolo, R. *et al.* Prothrombotic gene variants as risk factors of acute myocardial infarction in young women. *J. Transl. Med.* **10**, 1–5 (2012).
34. Goldberg-Stern, H. *et al.* Broad phenotypic heterogeneity due to a novel SCN1A mutation in a family with genetic epilepsy with febrile seizures plus. *J. Child Neurol.* **29**, 221–226 (2014).
35. Shen, Y. *et al.* Evidence for the involvement of the proximal copy of the MAGEA9 gene in Xq28-linked CNV67 specific to spermatogenic failure. *Biol. Reprod.* **96**, 610–616 (2017).
36. Uzun, A. *et al.* Targeted sequencing and meta-analysis of preterm birth. *PLoS One* **11**, 1–17 (2016).
37. Allen, A. S. *et al.* Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. *Lancet Neurol.* **16**, 135–143 (2017).
38. Forsberg, L. A. *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* **90**, 217–228 (2012).
39. Castellani, C. & Assael, B. M. Cystic fibrosis : a clinical view. *Cell. Mol. Life Sci.* (2016). doi:10.1007/s00018-016-2393-9
40. Fuseini, H. & Newcomb, D. C. Mechanisms Driving Gender Differences in Asthma. *Curr. Allergy Asthma Rep.* **17**, (2017).
41. Kang, G., Lin, D., Hakonarson, H. & Chen, J. Two-stage extreme phenotype sequencing design for discovering and testing common and rare genetic variants: Efficiency and power. *Hum. Hered.* **73**, 139–147 (2012).
42. Li, D., Lewinger, J. P., Gauderman, W. J., Murcray, C. E. & Conti, D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet. Epidemiol.* **35**, 790–799 (2011).
43. Guey, L. T. *et al.* Power in the phenotypic extremes: A simulation study of power in discovery and replication of rare variants. *Genet. Epidemiol.* **35**, 236–246 (2011).
44. De Sanctis, L. *et al.* Familial PAX8 small deletion (c.989_992delACCC) associated with extreme phenotype variability. *J. Clin. Endocrinol. Metab.* **89**, 5669–5674 (2004).
45. Roman-naranjo, P. *et al.* Rare Variants in the OTOG Gene Are a Frequent Cause of Familial Meniere ' s Disease. (2019).
46. Gilles, A., Camp, G., Van de Heyning, P. & Fransen, E. A pilot genome-wide association study identifies potential metabolic pathways involved in tinnitus. *Front. Neurosci.* **11**, 1–10 (2017).
47. Lopez-Escamez, J. A. *et al.* Genetics of tinnitus: An emerging area for molecular diagnosis and drug development. *Front. Neurosci.* **10**, 1–13 (2016).
48. Vona, B., Nanda, I., Shehata-Dieler, W. & Haaf, T. Genetics of tinnitus: Still in its infancy. *Frontiers in Neuroscience* (2017). doi:10.3389/fnins.2017.00236
49. Maas, I. L. *et al.* Genetic susceptibility to bilateral tinnitus in a Swedish twin cohort. *Genet. Med.* **19**, 1007–1012 (2017).