Evaluation of phenotypic abnormalities among patients with intellectual disability

Claudia J.M. Hellebrekers

Supervisors: Sandra Jansen[1], Jayne Y. Hehir-Kwa and Bert B.A. de Vries

Affiliations

[1]Department of Human Genetics, Donders Centre for Neuroscience, Radboud University Medical Center, P.O. Box 9101, 6500 HB, Nijmegen, the Netherlands.

Corresponding author: Claudia J.M. Hellebrekers, claudia.hellebrekers@gmail.com

Abstract: The development of more advanced genetic techniques like the SNP-array and genome wide sequencing techniques contributed largely to the identifying of genetic causes of intellectual disability. Interpretation of the genetic variations found with these techniques, however, remains difficult. Multiple patients with the same phenotype and genotype are needed to decide about a variant's pathogenicity. It is although difficult to determine how many patients are needed for a statistic significant conclusion. More information about the specificity of phenotypic features will greatly improve these considerations, as genetic variants in combination with rare features can be declared pathogenic when fewer patients are found than variants occurring in combination with common features. However, at the moment no comprehensive overview of the phenotype of patients with ID is available.

This article provides an overview of the specificity of phenotypic features by an extensive evaluation of the phenotype of a large cohort of 7407 well-defined patients with ID. From 1320 of these patients very detailed information was available, these patients formed a best-defined group used for the analyses, the other 6087 a well-defined replication cohort

An overview of the frequency of all Human Phenotype Ontology (HPO) features in these patients is presented. In addition we found 1534 combinations of positively associated features. These associations consisted of expected associations like Cleft lip (HP:0000204) and Cleft palate (HP:0000175) (P<0,001) but also unexpected associations like Short stature (HP:0004322) and Abnormality of the forebrain HP:0100547) (P<0,001). Most of these 1534 combinations were also associated in the replication-cohort. We also proofed that half of the major anomalies registered by Eurocat occur more in patients with intellectual disability than in the normal population and that patients with more congenital anomalies have a higher chance of having facial dysmorphism.

This extensive description of phenotypes of intellectual disability patients provides an overview for clinicians that can contribute to the deliberation about an intellectual disability patient's diagnosis, which is valuable for the patients and their parents.

Keywords: intellectual disability, genetics, phenotype

Introduction

Intellectual disability (ID) is a neurodevelopmental disorder that can be defined by an IQ score less than 70 together with significant deficits in adaptive behaviour and an onset before 18 year (Leonard & Wen, 2002; van Bokhoven, 2011). The world health organization approximates that about 1-3% of the world population has ID (Schalock et al., 2012).

Patients with ID can be separated in two groups: syndromic and nonsyndromic ID (van Bokhoven, 2011). In the case of syndromic ID there are additional clinical features present leading to either patients with well defined syndromes or patients that do not have a recognisable syndrome but still have phenotypic anomalies associated with the ID.

The origin of ID is heterogeneous. It can be induced by various genetic, prenatal, perinatal and postnatal causes although genetics is considered as the major cause in Western society. Severe ID (IQ < 50) has often a genetic cause (Willemsen & Kleefstra, 2014) such as chromosomal aberrations as well as de novo gene mutations (de Ligt et al., 2012; Vissers et al., 2010).

Over the years we made huge progress in identifying the genetic causes of ID by developing more advanced techniques like microarray to detect smaller genetic anomalies (Vulto-van Silfhout et al., 2013). Unfortunately remains in 55-60 % of the patients with severe ID the exact genetic cause still unknown (Topper, Ober, & Das, 2011). At the moment exome and genome sequencing are introduced in the diagnostics of ID (de Ligt et al., 2012; Rauch et al., 2012; Yang et al., 2013), allowing an unbiased look at the whole exome/genome and thereby detecting single nucleotide variants. Research in unsolved cases of ID (no cause found with previous methods like microarray) gave us promising results with up to 50% diagnostic rate (de Ligt et al., 2012; Gilissen et al., 2014).

In principle, it would be possible to find the cause for almost all patients with ID with a (pure) genetic origin when using high quality whole genome sequencing and extended analysis techniques. However, so far we are not even close to achieve this goal. Not every part of the DNA is covered well by the present techniques, we have to learn more about the significance of variants outside the exome and techniques for detecting pathogenic combinations of aberrations are still in development.

Another complicating factor, to the not (yet) achieved perfect quality, is the problem of ascertaining the causality of the mutations found. The novo mutations are likely causal in single cases with ID, however, at average we find 1,8 the novo Single Nucleotide Variations in each patient (Gilissen et al., 2014) and therefore we do not know which of these variations are actually causative.

The novo Single Nucleotide Variants in known ID genes (Kleefstra, Schenck, Kramer, & van Bokhoven, 2014) are very likely to be causative. When mutations are found with unknown clinical significance, for example, in genes not previously linked to ID, the phenotype is used to decide about the pathogenicity. We can imagine that in case we find more patients with similar mutations in the same gene or class of genes together with a comparable phenotype these mutations are probably the cause of the ID in these patients.

The latter approach is, for example, found in sibling studies (Schuurs-Hoeijmakers et al., 2013) and a comparable strategy proved to be fruitful in the detection of various novel syndromes such as the Kleefstra syndrome and the Koolen-De Vries syndrome (Koolen & de Vries, 2010; Willemsen et al., 2012).

Debate and research are on-going about the number of patients with the same aberration and clinical features needed for deciding about an aberration's pathogenicity. At the moment the number of patients we need for a statistical significant conclusion is estimated at six.

More common features have, however, less influence on the chance that an aberration is pathogenic than rare features. When in a number of patients with the same genetic aberration the same rare feature is present it is more likely that the aberration is pathogenic then when in all these patients a feature occurs that has a prevalence of, for instance, 10% in the population. Therefore, it would be a great advance if we could not only estimate the number of patients needed to proof the causality of gene variants but also calculate how many patients are needed to proof causality using the specificity of the involved phenotypic features. This calculation would help to establish earlier on, when only a couple of patients are found, the link between gene variants and very specific phenotypes and would prevent premature conclusions of linking more generic phenotypes to a genotype.

To perform these calculations we need to know the prevalence and thereby the specificity of the individual phenotypic features. However, in literature only prevalences are published for major congenital anomalies in the whole population as registered by Eurocat (Boyd et al., 2011) . Eurocat is a European wide network of population-based registries for major congenital anomalies. They register since 1979 the incidence of major anomalies in newborn children for epidemiologic reasons, for example, to get more information about risk factors in pregnancy. Doctors and midwifes are asked to inform Eurocat about all newborn children where they detect congenital anomalies. Eurocat registers the information of a large cohort with samples from European wide countries. They register the anomalies of 29% of the newborn children in Europe, this are more than 1,7 million births a year.

We are not aware of any extensive investigation of the incidences of congenital anomalies in the ID population except for the phenotypic evaluation of individual syndromes. For the incidence of facial features no comprehensive overview is available, not for the whole population as well as for the ID population.

It is not possible to get clinical information from an equally large ID cohort as the Eurocat cohort. Ten years of registration by clinicians provides us, however, with a database consisting of 7407 patients from the Netherlands and Belgium. Part of these patients are extensively described and provide us therefore with extra detailed information. The clinical information of all our ID patients is more detailed compared to Eurocat data, which only register major congenital anomalies which in general have clinical consequences.

The incidences of phenotypic features can also be used to examine if ID patients have more anomalies than the normal population and which of these anomalies are seen more in ID patients. It is suspected that children with ID have more congenital anomalies than normal

children because they are more likely to have a syndrome consisting of multiple anomalies also underlying their ID. However, until now this was never examined, because there is no overview of incidences available.

Distinct facial features are also often syndromic, therefore it is expected that these facial features are also more common in the ID population. Facial features have in most cases no direct effect on the health of patients, however, they are suspect of interest by clinical geneticists. This interest is due to the relevance of facial features for patient overlap and syndrome recognition. Patients with the same syndrome often look more like each other than they resemble their family. Facial features can thereby significantly contribute to the specificity of a phenotype and consequently to proving the causality of a genotype. Based on the importance of facial features in detection of disorders we hypothesize that patients with more facial features will also have more congenital abnormalities.

Some of the phenotypical features will have a common genetic background. This happens, for example, with some facial features. Clinical features with the same genetic background need to be clustered to perform up to date statistics for calculating the specificity of a phenotype. We investigated therefore if there are features, which occur more together than expected at random.

Taken together we will define the prevalence of all HPO features in patients with ID and we will compare the frequencies of phenotypic features in ID patients with the normal population. Additionally, we will investigate if there are significant associations between combinations of features.

This extensive evaluation of phenotypic abnormalities in ID patients will help to get more information about the specificity of phenotypic features, which is needed to proof the causality of gene variants. Using this information clinicians will be able to create a better diagnosis, which provides advantages for the patients and their parents.

Methods

Study population

Our study cohort consists of 7047 patients with ID. These patients were seen in a diagnostic setting at the clinical genetics department in the UMCN Nijmegen, VU Amsterdam or UVA Antwerpen over the last 10 years. We excluded patients with well-defined syndromes like Down syndrome. In all these patients a SNP-array was performed. An increasing part of the patients where no diagnosis was found with Micro-Array got Next Generation Sequencing to potentially find a diagnosis. As of writing about 800 patients of the cohort had their exome sequenced.

We divided the cohort in two subpopulations, one “well-defined” and a “best-defined group” alluding to how well they are phenotyped (**figure 1**). The well-defined group consisted of 6087 patients seen by a clinician who requested a NGS or Micro Array test. The form used for requesting these genetic tests comprises a question based phenotypical description. These summarized clinical descriptions are combined in a Microsoft Access database. The best-defined group are 1320 patients were also seen by a clinical geneticist and had a microarray and partially NGS. For these patients not only the compact clinical description from the request form was added to the phenotype-database but also additional features found in their clinical records.

For the general population we use the prevalence data available from Eurocat (Boyd et al., 2011). We used the European-wide prevelances registered between 1980-2012. We decided to use the data of living births without known genetic conditions. We opted for excluding the genetic conditions as Eurocat excludes only the genetic abnormalities known right after birth, for example, chromosomal abnormalities like Down syndrome, which are also excluded in our ID patients.

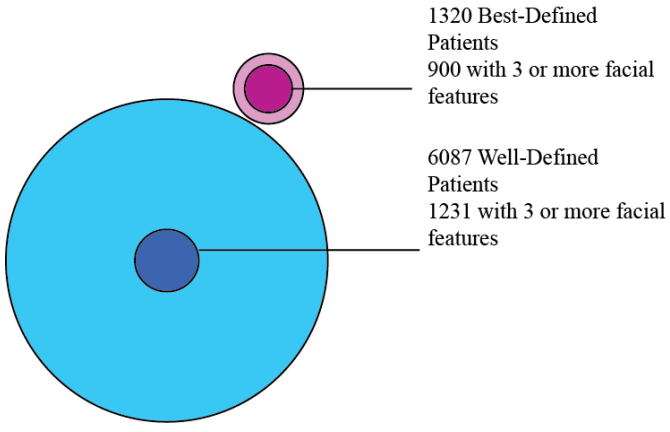

Figure 1, Description of patient cohort

Phenotype registration: HPO and Eurocat

The phenotype data of the patients were registered in the database using Human Phenotype Ontology terms (Kohler et al., 2014). All the features on the request forms and in the clinical

reports were standardised to the Human Phenotype Ontology Terms. The Human Phenotype Ontology consists of an ontology tree with 11000 standardized terms for phenotypic abnormalities. General features are divided in branches that becoming increasingly specific. The HPO is represented as Directed Acyclic Graph (DAG) with nodes and edges. HPO terms are related to their parental terms by “is a” relationships. It is possible for terms to have multiple parents, therefore a feature can be situated at different levels of the tree. This structure makes it possible that terms can be linked to multiple ancestors with different phenotypical aspects like, for example, Ataxia. (**figure 2**)

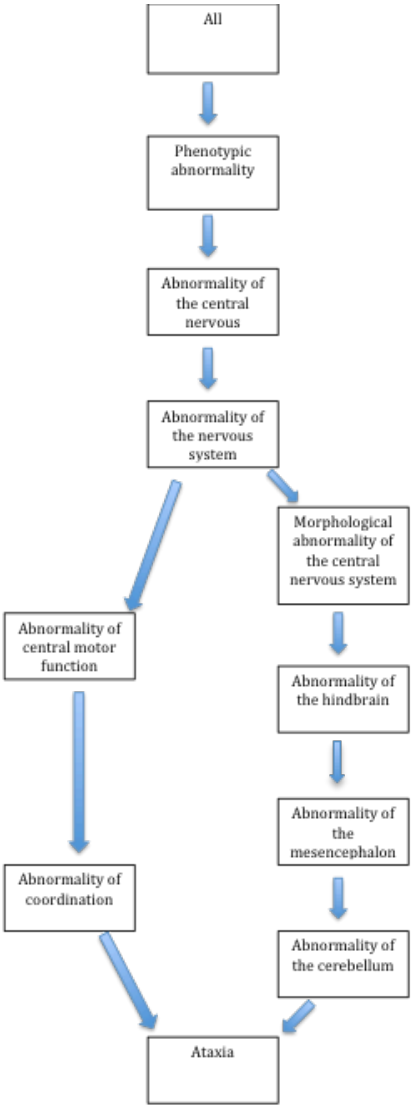

Figure 2 Structure of HPO

Our patient data was inserted into a SQL database and accessed through Microsoft Access for increased accessibility. The main tables used are a patient table with the personal information of the patients such as their research number and birthdate and on the second place a clinical table enclosing the clinical information of the patients. The research number was used to link these tables.

The features described in the prevalence tables of Eurocat are, however, not standardized to HPO terms. We translated the more specific features to HPO-terms (**supplementary table 1**). We took into account the list of minor features that are not registered by Eurocat. For some of the Eurocat terms we used multiple HPO codes to get the best translation. The more general terms as, for example, limb abnormalities could not be standardized to the HPO due to the high risk of mismatch between the included and excluded features between HPO and Eurocat. The genetic anomalies and birth problems were not translated, as these features should be absent in our database.

(Statistical) Analysis

The analysis of the phenotypic data was performed using Access (SQL), SPSS (Corp., 2011) and R (Team, 2014). Preliminary to answering the research questions we designed a method to connect the structure of the HPO ontology tree (Kohler et al., 2014) to the clinical data of the patients. The clinical data in the database only compromised specific features because the features were annotated to the most specific HPO term applicable. If a patient, for example, has a beaked nose (HP:0000444), then only this feature was registered in the database. However, this patient has also the features that are ancestors of the term “beaked nose” in the HPO ontology. In this case these are, for example, abnormality of the nose (HP:0000366) and phenotypic abnormality(HP:0000118).

To automatically add the ancestors of all the features for all of the patients we created a link file of the HPO ontology using the Linux version of OBO-Edit(Day-Richter, Harris, Haendel, Lewis, & Group, 2007). This link-file encloses each HPO-term and their parent terms in the ontology. We restructured this link file in a way that it could be added as table to our database using R. By joining this new HPO link table to the clinical and patient tables using SQL code an extensive clinical table of the patients was created which encloses not only the specific but also the more general, not by the clinician annotated, features.

First we evaluated the intern quality of our data. We investigated if our best-defined patient group was actually better phenotyped than our well-defined group. We calculated the influence of being in one of these two groups on the total number of clinical features a patient has. We calculated the number of features per patient by using the Access database and compared both groups by an independent Student’s T-test in SPSS. The number of features per patient was defined as the number of noted features in the database. So if a patient has a “beaked nose” [HP:0000444] we only counted this feature and not his ancestors in the ontology like nose abnormalities.

Additionally, a look was taken at differences in the kind of features in both patients group. This was done by comparing both groups by an independent Student’s T-test for as well the number of facial features per patient as the number of congenital anomalies defined as the number of Eurocat features per patient. We defined facial features as the features that are situated in the HPO ontology below the branches face and ear. For congenital anomalies we took the Eurocat features and defined the congenital anomalies as the HPO terms we mapped to the Eurocat terms and the features laying in the branches below these terms in the ontology (**figure 3**). In these groups we also only counted the features that were noted literally in the database, not the general features in higher branches of the ontology.

Total	**The total number of noted HPO terms per patient**
Facial	The number of noted HPO terms that are descendants of the terms Abnormality of the face: HP:0000271 Abnormality of the ear: HP:0000598 in the HPO ontology
Eurocat features	The number of noted HPO terms that are mapped to the Eurocat terms or children of these HPO terms
Non-facial features	The number of noted HPO terms that are not descendants of Abnormality of the face: HP:0000271 Abnormality of the ear: HP:0000598 in the HPO ontology

Figure 3 Description of feature selection

A independent two-sided t-test was performed and a *P*-value <0.05 was considered as statically significant for a confidence interval of 95%. The number of features or Eurocat features or facial features formed the continuous outcome of the independent variable group with the two different categories well-defined and best-defined patients. Expecting that a significant difference in the number of features in the preference of the best-defined group will be found we used the best-defined group as standard cohort for the rest of our research. The well-defined group was used as replication cohort. The larger numbers in this cohort will compensate partially for the lower quality.

Our main goal was to calculate the incidence for each HPO feature in the total group of patients. For calculating these incidences we used the extensive clinical data including the listed features and the features that are situated above the listed features in the HPO ontology. This approach is needed to get correct incidences of the more general features like abnormality of the face (HP:0000271) that are not often literally noted by clinicians who describe features as specific as possible.

We calculated the prevalence of all the present HPO features for the best-defined and the well-defined group. Additionally, we calculated the incidence of the congenital anomalies as registered by Eurocat for our population by using our HPO Eurocat. We compared the found incidences with the incidences of the normal population as found in Eurocat. A Fisher exact test was performed to calculate the difference in occurrence for each of the Eurocat feature between the normal population and our best-defined group of patients. We replicated this calculation by comparing the well-defined group and the Eurocat prevalences. We did not expect large differences between the prevalence Eurocat features in the well-defined and best-defined group as we expected that the congenital anomalies registered by Eurocat are so serious that they will be registered in almost all cases.

We calculated the Fisher exact test for each feature using R. We corrected for multiple testing using Bonferroni correction and considered a P-value less than 0,05 significant. In case of a P-value below 0,05 there is a significant difference between the incidence of a feature in the total population and the ID population. The Fisher exact test was used

because of low incidence of the Eurocat features. Chi-square would not be reliable due to the presence of low expectation values.

Additionally, we examined to which extent the face predicts congenital anomalies. Binary logistic regression in SPSS was performed to investigate if ID patients with more congenital anomalies were more likely to have facial dysmorphism. We looked at the relationship between the dependent categorical (not) having facial dysmorphism and the independent variable consisting of the continuous number of congenital anomalies.

We defined that patients with 3 or more facial features as noted by the clinician have facial dysmorphism. We defined facial features in the same way as mentioned in our first goal (**figure 3**). We defined congenital anomalies in two ways. We looked at the number of major Eurocat features a patient has and at the number of non facial features defined as the number of by the clinician noted features a patients has with the number of facial features subtracted.

Finally we took a look at associations between phenotypic features. We investigated if there were features that occurred more together than expected.

First we investigated combinations of facial features (**figure 3**). Because we were looking for associations between facial features we needed patients with multiple facial features therefore we selected the patients with 3 or more facial features from our best-defined population, this were 900 patients (**figure 1)**. We also investigated the association between the major congenital anomalies of Eurocat and the facial features and as a last step we looked at associations between all HPO features reciprocally. For all these steps, we used the patients from our best-defined group and replicated the analysis in our well-defined group.

We calculated first with use of SQL (Access) how often each combination of two facial features occurred. We used this information together with our previous defined prevalences for individual features to calculate the Fisher's exact for each existing feature combination. If the *P*-value was below 0,05 we defined the outcome as a significant connection between the features not expected to be caused at random.

We calculated the Fisher exact with 10000 Monte Carlo simulations to get more reliability with the large amount of features with in some cases small expectation values. We calculated the chi-square values with 10000 permutations using R. To give some more insight in the relationship between the features we calculated the odd's ratios for the feature combinations that were significantly associated.

We investigated a large number of combinations in each of these three steps. In total there were about 3000 HPO features in use in our dataset, so potentially 3000 *3000=9000000 combinations when we compared all HPO terms. The actual number of combinations, however, was lower because some of these features were very rare and did not occur in many combinations.

It was not useful to investigate all the existing combinations. We did not need to look at combinations of equal features like nose abnormalities and nose abnormalities (**figure 4**). Even so we only need one out of two reciprocal combinations. We need, for example, only the combination of mouth and nose abnormality and not also nose and mouth abnormality. We filtered out these combinations using a R script.

Furthermore we created a filter based on the HPO ontology (**figure 4**). The HPO ontology consists of branches. The ontology starts with general terms which divide in branches with more specific terms. Ancestor branches will therefore always be linked in some extent to the descendant branches in the ontology. An association between nose abnormality and a pointed nose is, for instance, expected. Therefore we created a filter in R that selects only feature combinations where both terms are not an ancestor of each other.

These filtering steps were not needed in the comparison between Eurocat and facial features as in this case two different group of features were compared so there are no reciprocal combinations or combinations of the same features or combinations where one feature is the ancestor of the other feature in the HPO tree.

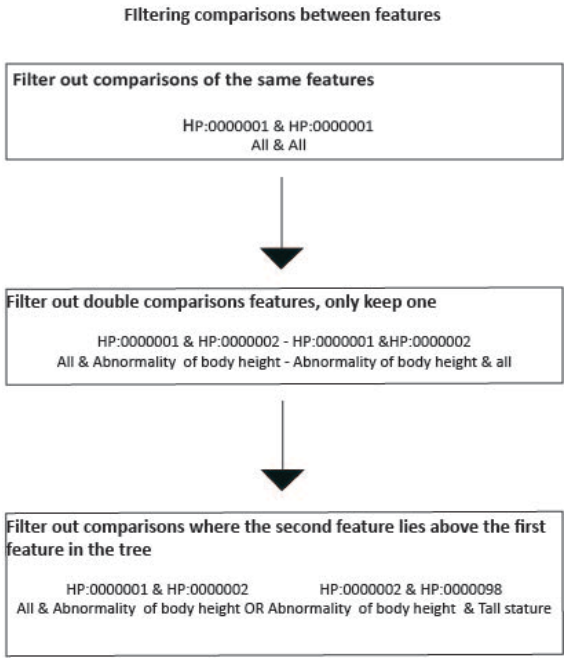

Figure 4 Filtering of feature combinations

After the filtering still a large amount of combinations was left. Therefore we had to correct our results for multiple testing. We decided to correct the data using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). We corrected the *P*-values after the filtering. This means that we used the number of combinations that were left after filtering as these are the combinations that are not automatically associated and therefore useful to compare.

Additionally, a clinical geneticist and a clinical geneticist in training looked independently from each other at the informative value of the combinations of associated features that were found. They investigated if there were associations consisting of two features that had roughly the same meaning. These were, for example, combinations where one of the features seemed to be the ancestor of the other feature although there is no ancestor-descendant relationship in the HPO tree or features that were approximately secondary terms of each other. An association like Deviated fingers (HP:0004097) and Radial deviation of the hand or finger of the hand (HP:0009485) was for instance seen as “non relevant”. We did not filter these associations out of the results. This selection was only performed to achieve a better understanding of the data.

Results

Internal quality of the data

The best-defined group of 1320 patients has 18,3 HPO features at average (SE=0,256), 1296 of these patients have 3 or more features. The replication cohort of 6047 well-defined patients had at average 6,7 HPO features (SE=0,071), 4446 of these patients had 3 or more features.

A two-sided independent t-test was performed and indicated a significant difference in the total number of HPO features between the best and well-defined groups, *t* (1528)=43,73, *P*<0,0001, with a large effect size (*d*=61,75). Levene's test for equality of variances was violated for this analysis, *F* = 659,33, *P* = 2,28E-139. Because the violation of this assumption we computed a *t*-statistic not assuming homogeneity of variance.

An independent two-sided independent t-test was also performed to compare the number of facial features and number of Eurocat anomalies between the well and best-defined group of patients.

There was a significant difference in the number of facial features between the well (M=1,36, SE=0,03) and best-defined group (M=5,15, SE=0,12), *t* (1490,1)=31,53, *P* < 0,001, with a large effect size (d=43,33). Levene's test for equality of variances was violated for this analysis, *F* = 1088,57, *P*< 0,001. Because the violation of this assumption we computed a t-statistic not assuming homogeneity of variance.

There was a significant difference in the number of Eurocat features between the well (M=0,38, SE=0,01) and best-defined group (M=0,77, SE=0,03), *t* (1615)=11,18, *p* < 0,001, with a large effect size (*d*=15,86). Levene's test for equality of variances was violated for this analysis, *F* = 248,80, *P* < 0,001. Because the violation of this assumption we computed a t-statistic not assuming homogeneity of variance.

Taken together there is a significant difference in the number of features as well facial and Eurocat features between the well and best-defined group in preference of the best-defined group.

Prevalence of features

In total there were 3078 HPO features present in our patients. In the best-defined group 2456 different HPO features were present of which 726 features were present in only one patient. In the well-defined group there were 2476 different HPO features of which 781 occur in only one patient. The prevalence of all HPO features in both groups is presented in **supplementary table 2**.

Comparison of the prevalence of the Eurocat features in the best-defined group and the normal population of Eurocat showed that 27 of the 64 (42,2%) congenital anomalies occur significantly more in the ID population than in the normal population (**supplementary table 3,** *P* < 0,05, Fisher exact test, corrected for multiple testing with Benjamini (Benjamini & Hochberg, 1995)). Replication of this comparison showed that 26 of these 27 congenital anomalies occur also significantly more in the well-defined group of patients. In the well-defined group of patients 9 additional features showed to be significant more prevalent than

in the normal population. In total were 35 features (54,7 %) more prevalent in the well-defined group than in the normal population of Eurocat. Taken together occurred about half of the Eurocat features significantly more in the group of ID patients than in the normal population.

Connection between facial features and congenital anomalies

Binary logistic regression was performed to predict the probability of having facial dysmorphism with the number of congenital anomalies as predictor. We performed this analysis with two different definitions of congenital anomalies, the number of Eurocat features and the number of non-facial anomalies (**figure 3**).

The logistic regression model was statistically significant (X^2(1)=45,316, P < 0,001) using the number of Eurocat features as predictor. The number of Eurocat features explained 7,4% (Nagelkerke R^2) of the variance in having facial features and the model correctly classified 68,2% of the patients. An increasing number of Eurocat features was associated with an increased likelihood of having facial features (*Wald*=36,379, P< 0,001). The Exp(B) value indicated that when the number of Eurocat features increases with one the Odds Ratio is 1,479 times as large. This means that a patient with one additional Eurocat feature is 1,479 times as likely to have facial features. We can conclude from this model that having more Eurocat features improves the chance of having facial dysmorphism (**figure 5**).

The Binary logistic regression was replicated in the well-defined group. The logistic regression model was statistically significant (X^2(1)=373,564, P < 0,001). The number of Eurocat features explained 9,4% (Nagelkerke R^2) of the variance in having facial features and the model correctly classified 80,2% of the patients. An increasing number of Eurocat features was associated with an increased likelihood of having facial features (*Wald*=332,544, P < 0,001). The Exp(B) value indicated that when the number of Eurocat features increases with one the Odds Ratio is 1,923 times as large. This means that a patient with one additional Eurocat feature is 1,923 times as likely to have facial features. These results confirm our conclusion that having more Eurocat features improves the chance of having facial dysmorphism (**figure 5**).

Using the number of non-facial features as predictor the logistic regression model was also statistically significant (X^2(1)=248,845, P < 0,001). The number of non-facial features explained 24,9% (Nagelkerke R^2) of the variance in having facial features and the model correctly classified 68,2% of the patients. An increased number of non-facial features was associated with an increased likelihood of having facial features (*Wald*=166,336, P < 0,001). The Exp(B) value indicated that when the number of non-facial features increases with one the Odds Ratio is 1,206 times as large. This means that a patient with one additional non-facial feature is 1,206 times as likely to have facial features. We can conclude from this model that having more non-facial features improves the chance of having facial dysmorphisms (**figure 5**).

The Binary logistic regression was replicated in the well-defined group. The logistic regression model was statistically significant (X^2(1)=928,124, P < 0,001). The number of non-facial features explained 22,3% (Nagelkerke R^2) of the variance in having facial features and the model correctly classified 79,8% of the patients. An increasing number of non-facial features was associated with an increased likelihood of having facial features

(*Wald*=759,268, $p < 0{,}001$). The Exp(B) value indicated that when the number of non-facial features increases with one the Odds Ratio is 1,264 times as large. This means that a patient with one additional non-facial feature is 1,264 times as likely to have facial features. These results confirm our conclusion that having more non-facial features improves the chance of having facial dysmorphisms (**figure 5**).

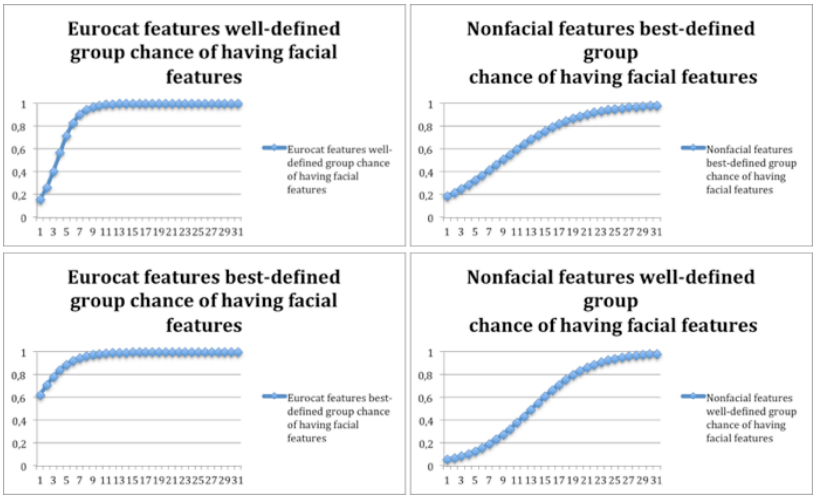

Figure 5 Plots presenting Binary Logistic Regression results

Feature associations

Fisher exact tests with Monte Carlo simulation were performed to investigate association between facial features. Significant association was found in 54 of the 23616 combinations that were left after the filtering steps. Before the filtering steps 53452 combinations were present. From the 54 associations 43 were found again in the replication cohort. In the replication cohort 11 additional associations were found that were not existent or not significant in our best-defined patients. None of the significant associations were prevalent in 5 or less patients. An overview of all the significant associations can be found in **supplementary table 4.**

Fisher exact tests with Monte Carlo simulation were also performed to investigate association between facial features and Eurocat anomalies. Significant association was found in 18 of the 5526 combinations. These 18 associations were also found in the replication cohort. In the replication cohort 11 additional associations were found that were non-existent or not significant in our best-defined patients. None of the significant associations were prevalent in 5 or less patients. An overview of all the significant associations can be found in **supplementary table 5.**

Finally Fisher exact tests with Monte Carlo simulation were performed to investigate association between all HPO-anomalies. Significant associations were found in 1559 of the 431968 combinations that were left after the filtering steps. Before the filtering steps 900994 combinations were present. From the 1559 significant associations there were 25 with a negative Odds Ratio indicating a negative association. There were 14 combinations of associated features that occurred in 5 or less patients. Both clinicians together called 937 associations relevant. Of the 1559 associations 1401 were found in the replication cohort. In the replication cohort 6399 additional associations were found that were not existent or not

significant in our best-defined patients. An overview of all the significant associations can be found in **supplementary table 6**

Discussion

In this article we presented the first overview of prevalences of all HPO features in a large cohort of extensively phenotyped ID patients. We proofed that half of the anomalies registered by Eurocat occurred more often in the ID population than in the normal population and that having more congenital anomalies increases the chance of having facial dysmorphism in patients with ID. We found 1534 combinations of features that occurred significantly more together than expected. Taken together we got results that met our expectations and could be replicated, for the largest part, in our replication cohort.

Internal quality of the data

We proofed that our best-defined group was better phenotyped than our well-defined group. However, it was remarkable that also the number of Eurocat features was significantly higher in the best-defined group. We did not expect a significant difference in the number of these major anomalies. It is possible that the patients from the well phenotyped group were somewhat less affected although group selection is not defined by severity of the phenotype. It is, however, also possible that the significant difference in the number of Eurocat features is caused by a difference in the number of less severe Eurocat Features. We investigated this by performing a Fisher exact test for the difference in prevalence of individual Eurocat anomalies in both groups. This Fisher exact test showed a significant difference between the well and best-defined group in preference for the best-defined group for the four features: limb reduction defects, hip dislocation and or dysplasia, microcephaly and congenital heart defects. These are features that can be less severe and it is therefore possible that they are not mentioned on an exome-sequencing request. This depends on the clinician.

Prevalence of features

The prevalances showed that almost one third of the present HPO features occurred in only one patient. About 8000 of the HPO features did not occur at all in our ID patients. This indicates that multiple features of the HPO are very rare in the group of ID patients. Comparing the found prevalences with the Eurocat data showed that around half of the Eurocat features occurred significantly more in the group with ID patients than in the normal population. This finding agrees with our expectations. The fact that we did not find more Eurocat features that occurred significantly often in the ID patients could partially be explained by the fact that many of the Eurocat features are very rare, 21 of the Eurocat features did not occur in our best-defined group. Our group of ID patients is large, however, multiple Eurocat features occur in less than 0,01% of the normal population. We can not exclude excluded that more Eurocat features occur significantly often in the ID population if a larger group of patients will be examined.

Related to these findings it is noteworthy that there were more Eurocat features in our replication cohort with a significantly higher prevalence than in our standard cohort (35 instead of 27). In the replication cohort only 8 Eurocat features were absent against 21 in our best-defined patients. The larger number of patients probably outreached the better description in the best-defined group for the major Eurocat features although the patients of the well-defined group have less Eurocat features per patient.

An additional explanation for the finding that not nearly all Eurocat features occur significantly more in the ID population is the fact that some of the Eurocat features are so serious that they decrease the chance of a long life expectancy. In order to obviate this we used only the data of live births from Eurocat. However, also part of these children will not have a long life. It is possible that children with severe anomalies lived for only a couple of hours or days. Children with such a short life will not be included in our data because we only included children where you can speak about having ID.

As Eurocat only consists of major features there will be no large difference between reporting by a general clinician to Eurocat and investigation by a clinical geneticist. However, it is possible that there is some additional reporting in our data of less severe Eurocat features like a mild microcephaly or a mild heart anomaly not detected shortly after birth.

If the Eurocat data would have provided information about which features occurred together in the same patient we could have performed a MANOVA instead of evaluating the individual features. MANOVA would have increased the statistical power and contributed to giving a more general answer to the question if major congenital anomalies occur more in the ID patients.

Unfortunately there was no opportunity to compare the incidences of “minor” congenital anomalies & facial features with the normal population as this data were not available. It would also be very difficult to collect this data because investigation of a large amount of “normal” people by a clinical geneticist is needed. It is not possible to collect this data in the manner used by Eurocat. In the future it will, although, be possible to use facial recognition to investigate automatically the face of the healthy parents of the patients. These data will create a normal reference population for the prevalence of facial features.

Connection between facial features and congenital anomalies

The results of the performed binary logistic regression confirmed our hypothesis that the number of nonfacial features and Eurocat features influences the chance that a patient has facial dysmorphism. The Odds showed that the models including the number of Eurocat features (Odds=1,479 or Odds=1,923) showed a steeper slope than the model including facial features (Odds=1,206 or Odds=1,264). This means that a patient needed less Eurocat features than non-facial features to increase the chance of having facial dysmorphism to almost 100%. The fact that the average patient only has 0,77 or 0,38 Eurocat features in the total of 6,7 or 18,3 features respectively will have influenced this finding. The small average number of Eurocat features will also have influenced the fact that only 7,4 and 9,4 % of the variance in having facial features is accounted to the number of Eurocat features. The non-facial features explained respectively 24,9 or 22,3% of the variance in the number of facial features.

Feature associations

We expected multiple associations between facial features as combinations of facial features are seen in many defined syndromes and it is suspected that multiple facial features have a related embryological development. We found 54 significant associations. Investigation of the found associations showed that a couple of these associations consisted of features where one of the features appears to be a secondary term or ancestor of the other feature in the HPO tree, for example, abnormality of the upper lip and thin lips (P<0,001) and

bulbous nose and abnormality of the external nose (*P*<0,001). These combinations have, however, no ancestor-descendant relation in HPO tree and were therefore not filter out by the R filter we created (**figure 4**).

We found multiple associations that were expected because these features, for example, have a known embryological link or were linked previously together like a cleft lip and cleft palate (*P*<0,001) and a thin lower- and upper lip (*P*<0,001). These logical associations proofed that our method worked and substantiated the association of these features in a large group of ID patients. Unfortunately we found only a small number of unexpected associations such as between different parts of the face like thin lips and a flat philtrum (*P*=0,002) or abnormality of the oral cavity and abnormality of the chin (*P*=0,007)

It is noteworthy that selection of the facial associations from the associations between all HPO features provided us with 135 instead of 55 significant facial associations. The difference in the number of associations can be explained by the fact that we investigated only the patients with 3 or more facial features for the facial associations. In our investigation of combinations between all HPO features we used all patients. A possible explanation for the fact that there were more significant facial associations when you investigate all patients is that the additional associations were often between general features like lip abnormality and abnormal nasal mophology (*P*<0,001). In the group with 3 or more facial features these features were very common, there were relative few patients that did not have both of these features, therefore these combinations did not became significantly associated.

The associations between facial features and Eurocat features did not show any unexpected results. We only found combinations of previously associated features like cleft lip and cleft palate. A possible reason for the absence of other associations is the rarity of most Eurocat features, in a larger cohort it is possible that other significant associations will be found.

The 1534 positive associations that were found during the investigation of all HPO terms were very interesting. We found expected associations, for example between Deviation of the 5th finger (HP:0009179) and Radial deviation of fingers (HP:0009466), however, also unexpected associations like Short stature (HP:0004322) and Abnormality of the forebrain (HP:0100547).

To get some more information about the 1534 significant associations we investigated if it was possible to find combinations of features that occur together in known syndromes although in our data only patients were included that had no diagnosis before a micro-array was performed. As example we took a look at the features of the Koolen-De Vries syndrome (OMIM 610433). We selected the 88 HPO terms (Kohler et al., 2014) linked to Koolen-De Vries syndrome (Koolen & de Vries, 2010) in the HPO-browser from our 1534 associations. We found only 4 combinations out of the 88*88=7744 possible combinations in our significant associations (**figure 5**). We found, for example, the combination of a Narrow palate (HP:0000189) and high palate (HP:0000218). This are two features that occur often together and are related to multiple syndromes like de Koolen-De Vries syndrome as combined HPO term High,Narrow palate (HP:0000189). Finding this small number of associations has different possible reasons, for example, that combinations of these features are caused only by Koolen-De Vries syndrome, that our sample size was too small for finding the associations or that only combinations of more than two of these features are

associated. Evaluation of more syndromes or collecting more patients can contribute to get more insight in these findings.

Feature1	HpoCode Feature1	Feature2	HpoCode Feature2	Uncorrected *P* -value best-defined patients	Corrected *P*-value best-defined patients	Uncorrected *P* -value well-defined patients	Corrected *P* -value well-defined patients
Cleft palate	HP:0000175	Cleft lip	HP:0000204	3,60E-25	1,59E-16	1,77E-68	2,68E-60
Narrow palate	HP:0000189	High palate	HP:0000218	1,35E-16	3,33E-08	1,44E-66	2,12E-58
Microcephaly	HP:0000252	Low birth weight	HP:0001518	8,08E-13	0,000138823	5,69E-47	5,13E-39
Microcephaly	HP:0000252	Feedingproblems in infancy	HP:0008872	1,78E-11	0,002640088	0,023442041	1

Figure 6 Feature associations occuring in Koolen-De Vries syndrome

A clinical geneticist and clinical geneticist in training investigated subjectively whether associations consisted of 2 approximately the same terms or one feature seems to be an ancestor of the other feature. According their rating 937 associations were relevant. We tried to come up with additional objective filters in order to select the most relevant associations out of the 1534 associations that were found. However, this proofed to be difficult.

We considered performing Multivariate analysis instead of Fisher exact tests. Multivariate analysis would have decreased the number of comparisons and thereby contributed to the visualisation of the results. This type of analysis would also have increased the statistical power. Performing this type of analysis, however, turned out to be complicated because there are multiple HPO terms that have more than one ancestor and are therefore, often, located at different levels of the HPO tree. In this situation, a HPO term has different distances to an ancestor (**figure 2**). We also considered filtering out the more general HPO terms, as associations including, for example, the general term Abnormality of the face (HP:0000271) are not very informative. We did not perform this filtering because it was impossible to designate a clear cut-off point because HPO features become specific at different levels of the HPO tree.

Finally, we thought about a filter based on the exclusion of features with a large overlap in ancestors. Features with a large overlap in ancestors like Deviation of the 5th finger (HP:0009179) and Radial deviation of fingers (HP:0009466) are often less informative as these features are frequently linked to each other in advance. However, terms with overlapping ancestors are not always associated in advance. There are contrary terms like microcephaly and macrocephaly that have a great overlap in parental terms, however, can not appear together in one patient. Another possibility are terms with overlapping ancestors that form interesting combination like between Wide mouth (HP:0000154) and Prominent lips(HP:0000184). It is possible to calculate a similarity score that indicates the extent to which the two HPO terms overlap each other (Deng, Gao, Wang, & Guo, 2015). As we will lose too many interesting associations if we filter using this similarity score we will only use this score to get more insight in the found associations. Calculation of these similarity scores will be performed in future research.

In the future, embryology can also help to get more insight in the etiology of feature associations. It is possible that there is an embryologic relationship between features due to,

for example, a mutation that plays a role in the development of both tissues or derivation of both features from the same embryonic germ layer.

Strengths and limitations

There are some general potential limitations that should be considered when interpreting the results of this research. First the fact that evaluation by a clinical geneticist is always less or more subjective. Clinical geneticists only write down remarkable features. They never judge for the complete list of about 11000 HPO features if they are present in a patient. The quality of the data also depends on how specific a clinician describes the features. In general is the examination of major congenital anomalies fairly objective, however, the assessment of facial features is always subject to some degree of subjectivity. New techniques like facial recognition can help to increase objectivity and completeness of this data in the future. The incidences of facial features assessed by clinicians provided in this article contribute on their side to the development of facial recognition techniques.

Furthermore is it possible that our translation of the Eurocat features was not optimal in all cases. In that case we could have included too many or too less patients, which could have influenced our results.

The final remark is that we used an older version of the HPO. In this version of the HPO a couple of errors occur like the situation of thick upper lip vermilion (**figure 7)** as parent of thin upper lip vermilion (**figure 7)**. We did not update the HPO version of our database because this would have involved too many manual corrections.

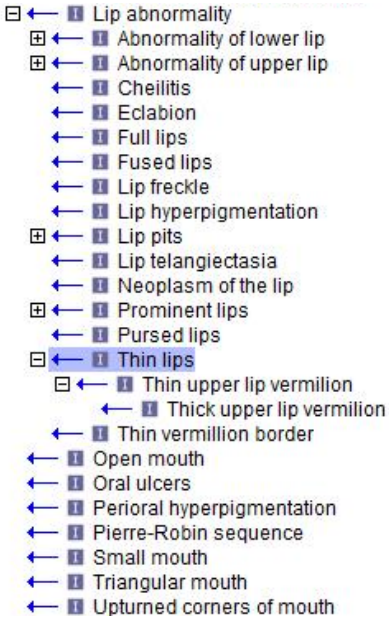

Figure 7 Part of HPO tree with aberration

This extensive phenotypic evaluation is despite the potential limitations very valuable for clinicians working with patients with ID and clinical geneticists in particular. There is no other source available comprising an overview of the prevalence of all HPO features and additional phenotypic evaluation of a large group of ID patients. The large international cohort of

patients that was investigated and the substantiating of the research with replications of the analyses in our replication cohort contributed to reliable results.

Conclusion

All together we need information about the prevalence of phenotypic features to decide about the specificity of phenotypic features. The specificity of the phenotype contributes to considerations about the number of patients with the same aberration needed to declare an aberration causative. Additional evaluation of the phenotype of ID patients like proving that ID patients have more major anomalies and searching for associations between features provides additional knowledge about the phenotype of ID patients, which also contributes to the considerations about a phenotype's specificity. This extensive evaluation of the phenotype of ID patients contributes to the deliberation about a genetic diagnosis, which is valuable for patients and their parents.

As a continuation of this study there are plans to evaluate the 1534 combinations of associated features more thoroughly by using similarity scores based on the HPO tree and embryological investigation. There are also plans to use facial recognition techniques to provide more objective prevalences of facial features and creating a control cohort with the prevalence of facial features in healthy parents of the ID patients. The overview of incidences of facial features we provided in this article will contribute to the development of these facial recognition techniques. Last of all there are plans to link the phenotypic information of these patients to their Whole Exome Sequencing data. This will provide multiple additional analysis opportunities like evaluating the influence of the number of the novo SNV's on the number of phenotypic features.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Boyd, P. A., Haeusler, M., Barisic, I., Loane, M., Garne, E., & Dolk, H. (2011). Paper 1: The EUROCAT network--organization and processes. *Birth Defects Res A Clin Mol Teratol, 91 Suppl 1*, S2-15. doi: 10.1002/bdra.20780

Corp., I. (2011). IBM SPSS Statistics for Windows. Armonk, NY:: IBM Corp.

Day-Richter, J., Harris, M. A., Haendel, M., Lewis, S., & Group, G. O. O.-E. W. (2007). OBO-Edit—an ontology editor for biologists. *Bioinformatics, 23*(16), 2198-2200.

de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med, 367*(20), 1921-1929. doi: 10.1056/NEJMoa1206524

Deng, Y., Gao, L., Wang, B., & Guo, X. (2015). HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. *PLoS One, 10*(2), e0115692. doi: 10.1371/journal.pone.0115692

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature, 511*(7509), 344-347. doi: 10.1038/nature13394

Kleefstra, T., Schenck, A., Kramer, J. M., & van Bokhoven, H. (2014). The genetics of cognitive epigenetics. *Neuropharmacology, 80*, 83-94. doi: 10.1016/j.neuropharm.2013.12.025

Kohler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., . . . Robinson, P. N. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res, 42*(Database issue), D966-974. doi: 10.1093/nar/gkt1026

Koolen, D. A., & de Vries, B. B. A. (2010). KANSL1-Related Intellectual Disability Syndrome. In R. A. Pagon, M. P. Adam, H. H. Ardinger, S. E. Wallace, A. Amemiya, L. J. H. Bean, T. D. Bird, C. T. Fong, H. C. Mefford, R. J. H. Smith, & K. Stephens (Eds.), *GeneReviews(R)*. Seattle (WA).

Leonard, H., & Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev, 8*(3), 117-134. doi: 10.1002/mrdd.10031

Minorafwijkingen. from http://www.rug.nl/research/genetics/eurocat/pdf/minorafwijkingen2010.pdf

Online Mendelian Inheritance in Man, OMIM® Retrieved 18/12/2015, from Johns Hopkins University World Wide Web 2011 http://omim.org/

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., . . . Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet, 380*(9854), 1674-1682. doi: 10.1016/S0140-6736(12)61480-9

Schalock, R., Borthwick-Duffy, S., Bradley, V., Buntinx, W., Coulter, D., Craig, E., . . . Snell, M. (2012). AAIDD's 11th Edition of Intellectual Disability: Definition, Classification, and Systems of Support.

Schuurs-Hoeijmakers, J. H., Vulto-van Silfhout, A. T., Vissers, L. E., van de, V., II, van Bon, B. W., de Ligt, J., . . . de Brouwer, A. P. (2013). Identification of pathogenic gene variants in small families with intellectually disabled siblings by exome sequencing. *J Med Genet, 50*(12), 802-811. doi: 10.1136/jmedgenet-2013-101644

Team, R. C. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013: ISBN 3-900051-07-0.

Topper, S., Ober, C., & Das, S. (2011). Exome sequencing and the genetics of intellectual disability. *Clin Genet, 80*(2), 117-126. doi: 10.1111/j.1399-0004.2011.01720.x

van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. *Annu Rev Genet, 45*, 81-104. doi: 10.1146/annurev-genet-110410-132512

Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., . . . Veltman, J. A. (2010). A de novo paradigm for mental retardation. *Nat Genet, 42*(12), 1109-1112. doi: 10.1038/ng.712

Vulto-van Silfhout, A. T., Hehir-Kwa, J. Y., van Bon, B. W., Schuurs-Hoeijmakers, J. H., Meader, S., Hellebrekers, C. J., . . . de Vries, B. B. (2013). Clinical significance of de novo and inherited copy-number variation. *Hum Mutat, 34*(12), 1679-1687. doi: 10.1002/humu.22442

Willemsen, M. H., & Kleefstra, T. (2014). [Genetic diagnostics in intellectual disability: what is the benefit?]. *Ned Tijdschr Geneeskd, 158*, A8098.

Willemsen, M. H., Vulto-van Silfhout, A. T., Nillesen, W. M., Wissink-Lindhout, W. M., van Bokhoven, H., Philip, N., . . . Kleefstra, T. (2012). Update on Kleefstra Syndrome. *Mol Syndromol, 2*(3-5), 202-212. doi: 000335648

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., . . . Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med, 369*(16), 1502-1511. doi: 10.1056/NEJMoa1306555