



Article A Gene-Based Algorithm for Identifying Factors That May Affect a Speaker's Voice

Rita Singh

Center for Voice Intelligence and Security, Carnegie Mellon University, Pittsburgh, PA 15213, USA; rsingh@cs.cmu.edu

Abstract: Over the past decades, many machine-learning- and artificial-intelligence-based technologies have been created to deduce biometric or bio-relevant parameters of speakers from their voice. These voice profiling technologies have targeted a wide range of parameters, from diseases to environmental factors, based largely on the fact that they are known to influence voice. Recently, some have also explored the prediction of parameters whose influence on voice is not easily observable through data-opportunistic biomarker discovery techniques. However, given the enormous range of factors that can possibly influence voice, more informed methods for selecting those that may be potentially deducible from voice are needed. To this end, this paper proposes a simple path-finding algorithm that attempts to find links between vocal characteristics and perturbing factors using cytogenetic and genomic data. The links represent reasonable selection criteria for use by computational by profiling technologies only, and are not intended to establish any unknown biological facts. The proposed algorithm is validated using a simple example from medical literature—that of the clinically observed effects of specific chromosomal microdeletion syndromes on the vocal characteristics of affected people. In this example, the algorithm attempts to link the genes involved in these syndromes to a single example gene (FOXP2) that is known to play a broad role in voice production. We show that in cases where strong links are exposed, vocal characteristics of the patients are indeed reported to be correspondingly affected. Validation experiments and subsequent analyses confirm that the methodology could be potentially useful in predicting the existence of vocal signatures in naïve cases where their existence has not been otherwise observed.

Keywords: FOXP2; genetic microdeletion syndromes; voice biomarkers; voice chains; voice profiling

1. Introduction

Aside from diseases that affect the biological structures and processes involved in voice production, myriad other factors are known to influence voice. Some simple examples include age, exhaustion, smoking, which often makes the voice sound hoarse, or alcohol, which makes the voice sound slurred. The ensuing changes in the voice signal can be thought to be the "biomarkers" that give us information about the corresponding causative factors and allow us to infer their nature through voice analysis. Such relationships form the basis for artificial-intelligence (AI)-based voice profiling techniques that attempt to deduce a speaker's bio-relevant and environmentally related parameters from voice. However, virtually all research on voice profiling, diagnostics, and biometrics is currently predicated on clinically observed or statistically inferred relationships between changes in voice and the corresponding factors that are thought to cause them. The relationships that are chanced upon in this manner provide the basis for building predictive AI (machine-learning- or rule-based) mechanisms that can deduce the underlying factors that potentially influence voice through voice analysis.

For example, it is known that smoking affects voice. To establish this, a *human*observation-based approach would be: (a) an audiological one based on hearing the voices of smokers to determine if they deviate from those of non-smokers in an acoustic sense,



Citation: Singh, R. A Gene-Based Algorithm for Identifying Factors That May Affect a Speaker's Voice. *Entropy* **2023**, *25*, 897. https:// doi.org/10.3390/e25060897

Academic Editor: Deniz Gençağa

Received: 17 April 2023 Revised: 26 May 2023 Accepted: 28 May 2023 Published: 2 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and/or (b) a visual one where an analyst studies the spectrogram (or some other visual representation) of the speech signal to find patterns that distinguish one class of recordings from another. The spectrogram in this case is a "feature representation". A *statistical approach*, on the other hand, would gather examples of speech recordings from people who smoke and those who do not, extract feature representations from the recordings, and find significant differences in the statistics of these features obtained from the two sets of recordings. Alternatively, a classifier model may be trained to discriminate between voice samples from smokers and non-smokers. If high test accuracies are achieved in this task, the existence of a biomarker for smoking in voice is indicated. This is a purely data-driven approach for establishing the existence of biomarkers in voice.

The problem with these approaches is that neither is scalable. The number of factors that can influence the human persona is virtually infinite. Human observations are limited to the effects that are perceptually discernible in voice, and data-based discovery is confined and limited by the availability of representative data. This paper provides a more formal methodology for establishing the existence of biomarkers and for identifying which factors are likely to affect voice and which are not. The methodology is based on genomic considerations, as explained below.

Before we proceed, however, it must be noted that this is *not* an algorithm for computational biology research. Its use to aid biological discovery or establish biological facts in its current form has not been tested. It is *only* meant to establish the tentative links that are needed to justify computational profiling efforts.

1.1. A genomic-Based Approach to Detect the Existence of Biomarkers

The working hypothesis for this paper, and one that has also recently been proposed in the context of voice profiling [1], is that if a given factor exerts an influence on the speaker, and if pathways of biological effects can be traced from that influence to the speaker's voice production system, then voice *must* be affected (and must carry biomarkers for the factor). The methodology proposed herein is a literal test of this hypothesis in that it traces biological pathways from cause to effect to establish the existence of biomarkers.

For this, we begin with the genetic underpinnings of human vocal capabilities. In this context, it is important to differentiate between voice production and speech production. The former refers to the production of acoustic energy in any form within the vocal tract, and the latter refers to the modulation of the acoustic signals thus produced to form words and sentences in a language used for interpersonal communication. We will use the term "vocal production" to refer to both.

Vocal production in humans is a complex and multifaceted process and involves interactions between multiple genes and environmental factors. The genetic basis of vocal production is not fully understood. Nevertheless, a number of genes have been found to be involved in the process. Some that are now known to influence vocal production include:

- 1. **FOXP2**: This gene codes for a protein called "Forkhead Box P2", which is involved in the development and function of the brain, including the areas responsible for language and speech.
- 2. **TAFT**: This gene codes for a protein called TAFT1, which is involved in the development and function of the larynx, a structure in the throat that is involved in vocal production.
- 3. **OTOF**: This gene codes for a protein called Otoferlin, which is involved in the development and function of the auditory system, including the inner ear. Feedback from this system greatly influences vocal production.
- 4. **MYO15A**: This gene codes for a protein called Myosin XVa, which is involved in the development and function of the auditory system, including the hair cells in the inner ear.
- 5. **SEMA3A**: This gene codes for a protein called Semaphorin 3A, which is involved in the development and function of the auditory system, including the auditory nerve.

Recent efforts to identify and delineate the genes responsible for functional speech in humans have especially highlighted the importance of FOXP2, one of the protein-coding genes mentioned above. It is involved in a variety of biological pathways and cascades that are thought to regulate language development. It is autosomal dominant, and mutations in it cause speech and language disorders (OMIM: SPCH1). In this paper, we choose FOXP2 as an example gene to work with. This choice was only made for illustrative purposes. The methodology presented in its context is itself a generic one and can be applied to any of the genes listed above (and possibly others that exist) whose functions are relevant to the analysis at hand.

For illustrative purposes, we proceeded with the broad and simplifying assumption that any influence on speech and language is ultimately the phenotypic expression of FOXP2. The objective was thus to:

- Formalize the methodology to find a link between an influencing factor and this gene;
- Validate the methodology;
- Demonstrate its predictive potential.

To accomplish these goals, for simplicity, we chose the example of a category of medical conditions for which the underlying genetic causes are known; the effects of the conditions are observed and reported in medical literature, and the effects can involve problems with vocal production.

The medical conditions we choose are chromosomal microdeletion syndromes, which result from the deletion of specific genes in specific cytogenetic locations on human chromosomes. We propose an algorithm to find a link between the genes in the regions of microdeletions, to FOXP2. The connections in these links are derived using a path-search algorithm applied to a graph composed from known biological pathways that involve these (and other) genes. The "strength" of these connections is then defined in terms of the characteristics of the linkage discovered. In the validation stage, our goal is to show that there is a direct correlation between the strength of linkages found and the extent of vocal symptoms experienced by the affected individuals.

Before we proceed with the algorithm, in the paragraphs below, we first provide a working categorization of speech disorders, as reported in various medical literature. This is necessary for the clarity of the results presented later in this paper.

1.2. Anomalies in Speech Production

From a bio-mechanical perspective, human speech is the result of two complex processes that happen simultaneously: one that produces sound—the pressure wave that we sense as the voice signal—and another that modulates this signal (through articulator movements) to produce speech, thus altering the voice signal's frequency characteristics and shaping it into sounds with unique identities that are uttered sequentially to form words and sentences in a language. The overall process of voice and speech production is driven and controlled by neuromuscular and cognitive factors to different degrees. It is also moderated to different degrees by feedback obtained through auditory pathways. Generally, diseases that affect these functions, naturally also influence speech and alter the characteristics of the voice signal to proportional degrees. In most cases, when reporting such changes, references to "speech" implicitly include voice, as we see below.

Changes in speech are categorically described in terms of six major aspects of speech production: respiration, phonation, articulation, resonance, evolution, and prosody. In addition, terminology that relates to *voice quality* is often used to describe speech. Voice quality is, however, a subjective term and comprises many constituents, or sub-qualities (e.g., nasal, breathy, rough, twangy, etc.), that refer to the perceptual flavor of speech (or how a speaker's voice sounds to the listener). Physical anomalies that affect the shape and tissue structure of the vocal tract cause changes in all of these aspects. Speech delays and language difficulties result from cognitive and learning disabilities. These and other intellectual disabilities affect articulation, evolution, and prosody. Their effect on voice also manifests as changes in voice quality. Craniofacial anomalies affect the physical dimensions

of the vocal tract structures, often restricting the movement of the articulators as a result and causing speaking impairments. Motor problems affect articulation, phonation, and respiration. These cause speech aberrations and also affect voice quality. Hearing problems disturb the feedback mechanisms involved in controlling speech production and often lead to difficulties in prosody and articulation of speech.

In this paper, we do not focus explicitly on voice acoustic or quality characteristics, focusing instead on problems with speech (that subsume voice characteristics to some extent) as described under the OMIM (referring to the catalog Online Mendelian Inheritance in Man: https://www.omim.org/ (accessed on 21 September 2022)) category "Speech and Language disorders (SPCH1)". Even this category, however, is too broad and encompasses a wide range of speech problems, such as delays in acquisition of speech abilities, retardation in speech development with age, speech anomalies resulting from language delay, expression, and articulation.

For the purpose of this paper, it is necessary to make finer distinctions between these categories. The problem in doing so is that the language used in the literature to describe voice- and speech-related problems in the context of genetic syndromes is not standardized. For example, the terms "speech disorder", "speech disturbance", "speech anomalies", "speech aberrations", and "speech impairment" may each refer to a range of symptoms that may be overlapping to various degrees. For the purpose of this paper, it is therefore useful to map the broad range of speech problems into the following categories that are sufficiently discriminatory in terms of the different aspects of speech production mentioned above while still being limited in number:

- 1. **Absence of speech:** Phrases (in clinical/scientific literature) referring to (a) no development of speech capabilities, (b) no expressive speech, which is mostly limited to vocalizations, or (c) almost absent speech with a severely limited vocabulary (0–4 words).
- 2. **Apraxia:** Phrases referring to difficulty using language correctly while speaking, leading to speaking and communication difficulties.
- Delayed speech: Phrases referring to developmental delay, the retarded development of the ability to speak, or the retarded acquisition of language skills and communication skills (ability to use a vocabulary correctly to communicate in a cogent manner).
- 4. **Dysarthric speech or dysarthria**: Phrases referring to speaking problems resultant from damaged, paralyzed, or weakened muscles of the articulators caused by motor problems. Dysarthria results in slurred words, poor phonation, etc. The speaker uses vocabulary as in normal speech but finds it difficult to move the articulators (tongue, lips, jaw, etc.) correctly to form the proper sounds to utter the words.
- Idiosyncratic speech: Phrases referring to poor conformance to cogent language or incoherent language with articulation abnormalities.
- 6. **Impaired speech:** Phrases referring to poor articulation and phonation, as well as difficulties that result in sparse and disfluent speech.

2. Methodology

As mentioned earlier, we consider the example of syndromes resulting from chromosomal microdeletions and focus on their symptoms relating to speech abilities.

Chromosomal microdeletions are structural anomalies of chromosomes in which small sections of a chromosome are deleted or missing. The loss of the specific set of genes from the deleted section often results in phenotypic changes. An *implicated gene* is a gene in the deleted region of a chromosome that is known to cause much of the observed effects of the syndrome in affected individuals. These are identified through microarray and other studies. For most microdeletion syndromes, some such genes have been identified and reported in the medical literature. We use this information from the medical literature as-is in the sections below.

Although the human genome is very large in comparison to a typical chromosomal microdeletion region, microdeletions often cause serious problems. In fact, only a small

set of deletions are compatible with life or fetal survival. This set continues to expand with the addition of newly discovered deletions in surviving individuals who have the means to reach genetic testing facilities. However, it is still a very small set and can be exhaustively studied. Most known deletions are well documented in the literature, both from the genetic and medical perspectives. Information about the genes associated with them is readily available through well-curated publicly accessible repositories. Thus, they are good example cases for this paper.

The methodology proposed herein analyzes ensembles of biological pathway *chains*, each of which connects a specific gene in the cytogenetic region of chromosomal microdeletion to the FOXP2 gene. A "biological pathway" here is defined as in standard terminology, referring to a physiological process at the cellular level that is enabled by the action of multiple genes that perform specific functions within the process.

We define a *pathway chain* as the sequential linkage of pathways where links between pathways are shared genes (implicitly meaning that the molecules resultant from genes are shared—we will use the term "gene" with this implicit meaning in the context of pathways for generality going forward). For example, consider a pathway that signals for a cell to stop dividing when an injury to the nuclear DNA strand is being repaired. It may involve the coordinated chemical action of molecules that are formed by the transcription of multiple genes that perform different functions. It would *also* be connected to a repair pathway by necessity. Thus, the two pathways can be considered to be links in a single pathway *chain* (they must share some genes in a functional sense) that perform the function of relaying messages from one pathway to another. Such genes may also perform other functions that are essential to both pathways.

The hypothesis we make here is that for a gene, if a chain from its pathway(s) (i.e., from the biological pathways it contributes to directly) extends to pathways that influence voice production, then the phenotype resultant from the absence or aberrant functioning of the gene can be expected to include anomalies in speech production and voice characteristics.

2.1. Voice Chains

Our definition of a *voice chain* extends our definition of a pathway chain in that the head of the chain must now necessarily be a pathway that includes a gene that influences voice or speech production, while the termination of the chain is not necessarily a biological pathway but could include any given set of genes with a common characterization (such as a common cytogenetic location or function).

In this paper, the voice-related gene chosen is FOXP2, but in other analyses, voice chains could involve other genes (e.g., as in [2]) without loss of generality. The terminal link in the chain is taken to be a genetic microdeletion syndrome. A voice chain thus establishes a relationship between an influencing factor—a genetic microdeletion syndrome in this case—and a corresponding effect on voice/speech production. We refer to a voice chain that includes a sequence of α pathways from the microdeletion region to the FOXP2 gene as a *level-* α voice chain (Figure 1). The specific genes on the microdeletion that link it to the voice chain are referred to as "chainlink" genes. We represent the set of chainlink genes that connect a microdeletion to level- α voice chains as the chainlink set $\mathcal{V}_{N\alpha}$. Since we aim to analyze the genetic basis of the effect of microdeletions on voice, these chainlink sets are the focus of our analysis. Note that the subscript N in $\mathcal{V}_{N\alpha}$ denotes the *manner in which* overlaps or linkages between biological pathways are defined. This subscript is fixed for the purpose of this paper, for which the exact manner in which pathway overlaps are defined is described in Figures 1 and 2 below. We however leave the N in place to facilitate future differentiations and variations of the proposed algorithm based on how the pathway intersections (or unions) may be defined.



Figure 1. Voice chains of different levels. In this reversed perspective ideogram, a link depicts a pathway, the black dots on it are genes, and the chromosomes that they belong to are shown as rods where relevant. The lines connecting genes are only meant for visual clarity. Chains are formed with respect to genes on a chromosome (a microdeletion region in this case, shown shaded in yellow). In this ideogram, Gene-1 and FOXP2 lie on the same pathway, contributing to a level-1 voice chain. In a level-2 chain, FOXP2 and the microdeletion region are on different pathways, but the pathways share a set of genes. Gene 2 and Gene 3 have level-2 voice chains.

In order to trace the genetic links between a microdeletion syndrome and voice, we first attempt to identify voice chains of different lengths that link to the genes in the microdeletion region. For this, we must find voice chains that link the FOXP2 gene to the syndrome, and identify the specific genes from the syndrome through which they are linked. We do so using the graph-search algorithm described below.

From our perspective, a biological pathway \mathcal{B} is represented by the set of genes it involves: $\mathcal{B} := \{g : g \text{ is a gene in the specified pathway}\}$. Two pathways, \mathcal{B}_1 and \mathcal{B}_2 , are *linked* if there are genes that are common to both pathways, i.e., $\mathcal{B}_1 \cap \mathcal{B}_2 \neq \emptyset$. Thus, the set of all pathways can be represented as a graph where the nodes are biological pathways, and two nodes are linked only if the corresponding pathways have common genes, as illustrated in Figure 2a.

A pathway *chain* is any non-repeating sequence of pathways $\mathcal{B}_1 \mathcal{B}_2 \mathcal{B}_3 \cdots \mathcal{B}_N$ such that $\mathcal{B}_i \cap \mathcal{B}_{i+1} \neq \emptyset$ and $\mathcal{B}_i \neq \mathcal{B}_j$ for $i \neq j$, i.e., where every pair of adjacent pathways has common genes, and there are no closed loops in the chain. In terms of the graph (see Figure 2a), a pathway chain is any path between any two nodes in the graph. A *voice chain* \mathcal{V} is any chain $\mathcal{V} = \mathcal{B}_V \mathcal{B}_2 \mathcal{B}_3 \cdots \mathcal{B}_N \mathcal{S}$ where the *head* node \mathcal{B}_V (and the head node alone) is a pathway that includes the FOXP2 gene, i.e., $FOXP2 \in \mathcal{B}_V$, and the terminal node \mathcal{S} is a set of genes with common characterization, as mentioned earlier. The *length* of the chain $|\mathcal{V}|$ is the number of nodes α in the chain, not counting the terminal node \mathcal{S} . For the purpose of this paper, we will assume \mathcal{S} to be the set of genes in a microdeletion region associated with a syndrome. Thus, $\mathcal{S} = \{g : g \text{ is a gene in the microdeletion region}\}$. All voice chains of length α form the set of level- α voice chains, and the chainlink genes that connect \mathcal{S} to the level- α voice chains form the chainlink set $\mathcal{V}_{N\alpha}$.



Figure 2. Voice chains of different types. (**a**): Shows the different components of a graph comprising biological pathways. Each node in this graph represents a biological pathway. The genes g_i that contribute to each pathway are listed within the node. Nodes are linked to each other through edges that represent the set of shared genes. If two nodes have no gene in common, no edge exists between them. Genes that link pathways (nodes) are explicitly shown on the edges. (**b**): Explains what a chainlink gene means in the context of a microdeletion region and the target voice gene (FOXP2 in the example used in this paper). To compose voice chains, the set of genes of interest (e.g., from a microdeletion region) is added to the a graph as a node (shaded pink). The "chainlink" genes that link the set to the graph are also shown. The topmost (left) node represents a biological pathway that contains the FOXP2 gene (g_1 , highlighted in yellow). (**c**) A Level-1 voice chain is the edge shown in red. The "chainlink" gene g_2 that links the set (in pink) to the voice chain is also shown. (**d**) Exemplifies Level-2 chains. The colored edges represent Level-2 chains. The "chainlink" genes are also shown.

To find voice chains of the form $\mathcal{B}_V, \mathcal{B}_1, \dots, \mathcal{S}$ arising from the microdeletion region \mathcal{S} (which we will refer to as the "syndrome" for brevity), we introduce the microdeletion region in the pathway graph (Figure 2b). Voice chains are now the paths from \mathcal{B}_V to \mathcal{S} (Figure 2c,d). A breadth-first algorithm, described in Algorithm 1, is used to extract the chainlink sets $\mathcal{V}_{N_{\alpha}}$ for voice chains of multiple levels. The outcome of the algorithm is the

set of chainlink genes $\mathcal{V}_{N\alpha}[S]$ that connect each syndrome S to voice chains of level α , for $1 \leq \alpha \leq 2$. We restrict ourselves to chains of lengths of up to 2 since, at greater lengths, the chained influences cannot be disambiguated, as indicated by prior studies in the (highly related) context of protein–protein interactomes, e.g., [3]. Another reason for restricting ourselves to level 2 chains is that for the specific example chosen in this paper, there are not enough data that allow us to build deeper chains meaningfully (without resorting to self-loops, which may lead to incorrect conclusions).

Algorithm 1: Pseudocode for a breadth-first algorithm for computing the set of chainlink genes that form level 1 and level 2 voice chains for FOXP2.

	0
1 I	nput:
2	<i>Syndromes</i> := set of all syndromes in database;
3	Genes[P] := set of all genes in a pathway P (for all pathways);
4	<i>Pathways</i> [<i>g</i>] := set of all pathways involving gene <i>g</i> (for all genes);
5	GenesInSyndrome[S] := set of all genes in syndrome S (for all syndromes);
6	<i>FoxP2Pathway</i> := voice pathways that include FOXP2;
7 S	Set:
8	$\mathcal{V}_{N1}[\mathcal{S}] := \{\} \forall \mathcal{S}, \text{ set of all chainlink genes in level-1 voice chains from syndrome } \mathcal{S};$
9	$\mathcal{V}_{N2}[S] := \{\} \forall S, \text{ set of all chainlink genes in level-2 voice chains from syndrome } S$:
10	L1P := {} set of pathways connected at level-1 to a FOXP2 pathway through any gene:
11	<i>Level2Genes</i> := {} set of all chainlink genes in level-2 voice chains:
12 5	Start/Iterate:
13	/* Loop over all genes in ForP2Pathway *
14	for g in Genes [FoxP2Pathway] do
15	/* All syndromes linked to these genes have a level-1 pathway *
16	for S in Syndromes do
17	if (g in GenesInSyndrome[S]) $\mathcal{V}_{N1}[S] \leftarrow \mathcal{V}_{N1}[S] \cup \{g\}$
18	end
19	/* $L1P$ includes all pathways (except FoxP2Pathway) that include g *
20	$L1P \longleftarrow L1P \cup (Pathways[g] \setminus FoxP2Pathway)$
21	end
22	/* Construct the set of all level-2 chainlink genes *
23	for P in L1P do
24	$Level2Genes \longleftarrow Level2Genes \cup Genes[P]$
25	end
26	/* Construct the list of level-2 voice chains for all syndromes *
27	for S in Syndromes do
28	for g in Level2Genes do
29	$ \text{if } (g \text{ in } GenesInSyndrome}[S]) \mathcal{V}_{N2}[S] \longleftarrow \mathcal{V}_{N2}[S] \cup \{g\}$
30	end
31	end
32 (Dutput:
33	$V_{N1}[S] \forall S \in Syndromes;$
34	$V_{N2}[S] \forall S \in Syndromes;$

2.2. Ensemble Analysis

In the methodology we propose, for any microdeletion region S, we derive the set of chainlink genes within it for which α -level chains exist. The size and composition of this set can then be used in conjunction with the level of the voice chain to indicate the effect on

voice (in a later analysis). In general, we can work with any level- α voice chains in such an analysis; however, we restrict ourselves to $\alpha = 1$ and $\alpha = 2$.

3. Analysis

 $V_{N\alpha}$, where $\alpha = 1, 2$, were computed for a total of 82 microdeletion syndromes of chromosomes 1–20/22/X,Y. Genomic information, including gene names, was obtained from the HUGO Gene Nomenclature Committee's (HGNC) human genome database, comprising 42,764 gene symbols and names and 3245 gene families and sets as of the time of conducting this analysis. Information about the phenotypes and the specific genes implicated in a syndrome was obtained from a survey of the current literature on medical genetics and genomics and from the Online Mendelian Inheritance in Man (OMIM) repository for authoritative information about human genes and genetic phenotypes.

The FOXP2 gene chosen for this analysis has been strongly implicated in speech and language disorders [4,5], including monogenic speech disorders. The cytogenetic location (chromosome locus) of this gene is 7q31.1. Mutations in this gene are known to cause speech and language disorder Type 1, also called "Autosomal dominant speech and language disorder with orofacial dyspraxia". The phenotype description and known molecular basis for this disorder can be found under OMIM entry SPCH1:602081. The FOXP2 gene encodes for the protein "Forkhead Box Protein P2" [6]. This protein is a transcription factor; it controls the activity of other genes. It binds to the DNA of the genes that it controls through a region known as a Forkhead Domain. It thus plays a critical role in several protein-coding and other biological pathways and has been well studied [7]. A more detailed summary of this gene can be obtained from the Human Protein Atlas [8].

The ensemble of pathways used for this analysis was obtained from the Carcinogenic Potency Database (CPDB), described on its website as "a single standardized resource of the results of 45 years of chronic, long-term carcinogenesis bioassays". Its current database of human biological pathways contains 4319 pathways and their gene compositions. This database has been used extensively in the medical literature and was chosen in this case for illustrative purposes since there is (importantly) no inherent bias towards the speech phenotype in it. In this database, there is only one pathway that contains the gene FOXP2. This is the Adenoid Cystic Carcinoma (ACC) pathway, which contains 63 genes, listed below for reference:

Gene membership of the ACC pathway:

AKT1 ARID1A ARID4B ARID5B ATM ATRX BCOR BCORL1 BRCA1 BRD1 CEBPA CMTR2 CNTN6 CREBBP CTBP1 DTX4 EP300 ERBB2 ERBIN FBXW7 FGF16 FGFR4 FOXO3 FOXP2 H1-4 H2AC16 HRAS IL17RD INSRR JMJD1C KANSL1 KAT6A KDM6A KDM6B KMT2C MAGI1 MAGI2 MAML3 MAP2K2 MAX MGA MORF4L1 MYB MYBL1 MYC MYCBP MYCN NCOR1 NFIB NOTCH1 NSD1 PIK3CA PRKDC PTEN RAF1 SETD2 SMARCA2 SMARCE1 SMC1A SRCAP TLK1 TP53 UHRF1

Table A1, given in Appendix A, documents the voice chains found for a set of 75 documented microdeletion syndromes. This range excludes chromosome 21, for which sufficient documentation was not found in the literature. Only voice chains up to level 2 are shown in this table and used in the analysis presented in this paper. This is sufficient to demonstrate the viability of the methodology for the discovery of voice chains proposed in this paper. The entries in the rows and columns of this table are explained in detail in Appendix A.

Table 1 summarizes some of the information in Table A1 to help understand the analysis given in the next section. The information given in Table 1 includes, for each syndrome listed in it, the corresponding implicated genes that are *also* discovered to be chainlink genes by the algorithm proposed in this paper; the overall counts of level-1 and level-2 chainlink genes for each syndrome, along with the number of additional pathways they collectively connect to (in parentheses); and the corresponding phenotypic effects on speech that have been reported in the scientific literature.

Table 1. Information about level-1 and level-2 voice chain ensembles for 76 chromosomal microdeletion syndromes. For each syndrome, the corresponding implicated gene (culled from medical literature, with details given in the Appendix A) that was *also discovered to be a chainlink gene by the proposed algorithm* in either level-1 or level-2 chains is listed in the second column. The number of chainlink genes (i.e., genes that have chains that connect to the ACC biological pathway of FOXP2) is shown in the third and fourth columns for level-1 and level-2 chains, respectively. The total number of pathways that they collectively influence is indicated next to each count, in parentheses. The observed phenotypic effects on speech are given in the last column. **Del:** delayed speech; **Imp:** impaired speech; **Norm:** normal speech; **Abs:** absent speech; **Apr:** apraxia; **Dys:** dysarthria; **Idio:** idiosyncratic.

Syndrome	Implicated Genes from Microarray Studies That Also form Voice Chains	No. of Genes in $\{\mathcal{V}_{N1}\}$	No. of Genes in $\{\mathcal{V}_{N2}\}$	Reported Effects on Speech
1p36	SPEN	1 (10)	226 (3152)	Del [9], or Abs
1q21.1–q21.2	RBM8A, GJAS	-	41 (906)	Del [10] or Imp [11]
1q41–q42	DISP1, LEFTY1, LEFTY2, BPNT1	-	60 (1020)	Apr [12]
1q43–q44	AKT3	-	78 (903)	Del, Imp, or Abs
2p16.1-p15		-	17 (508)	Dys, Apr, or Imp [13]
2p21	SLC3A1	-	24 (557)	Apr or Idio [14]
2q23.1		-	2 (31)	Del or Imp
2q32–q33	COL3A1, COL5A2, GTF3C3, CASP8, CASP10	-	62 (1429)	Abs
2q37.3	HDAC4	-	38 (330)	Imp [15]
3p13	FOXP1	-	8 (48)	Del, Idio, Imp and Dys (all severe), Apr [16]
3q13.31	DRD3, GAP43, LSAMP	-	5 (56)	Imp [17] or Abs
3q29	PAK2, DLG1	-	24 (413)	Del
4p16.3	FGFR3	1 (25)	44 (495)	Del [18] or Abs
4q21		-	41 (696)	Del or Abs; Imp [19,20]
5p (5p15.2 and/or (5p15.3 or 5p15.33))	TERT, CTNND2	_	37 (500)	Del, Abs, and Apr [21]
5q14.3	MEF2C	_	7 (275)	Abs
5q33.1	RPS14	-	17 (153)	Dys [22]
5q35.3	NSD1	1 (8)	47 (653)	Norm [23] or Del
6pter-p24	FOXC1, GMDS	-	41 (416)	Del
6q25.3	ARID1B	-	21 (363)	Del, Apr, Dys [24]
7p21	TWIST1	-	18 (243)	Del [25]
7q11.23	ELN, LIMK1, GTF2IRD1, GTF2I	_	35 (536)	Norm or Del
8p23.1	GATA4	-	49 (337)	No significant anomaly reports
8q22.1	CCNE2	-	13 (152)	Del

Syndrome	Implicated Genes from Microarray Studies That Also form Voice Chains	No. of Genes in $\{\mathcal{V}_{N1}\}$	No. of Genes in $\{\mathcal{V}_{N2}\}$	Reported Effects on Speech
8q24.11-q24.13	EXT1	_	21 (211)	Del
9p24.3		1 (9)	3 (13)	Del, Dys, Apr [26]
9q34.3	EHMT1	1 (62)	46 (718)	Del [27], Apr [28] and Abs
10pter–p13 or 10p14–p15.1	GATA3	-	49 (796)	Sensorineural hearing loss
10q23	PTEN, BMPR1A	1 (107)	59 (1003)	Del or Abs; Imp [29]
10q26	DOCK1	-	48 (708)	Imp; Del [30]
11p11.2–p12	EXT2	-	46 (737)	Idio, Dys, Del, Apr [31]
11p13–p12	PAX6, SLC1A2, PRRG4	-	23 (342)	Imp [32]
11p15.5	IGF2	1 (347)	50 (1243)	Imp [33]
11q13.3	FGF4, FGF3, FADD	_	11 (608)	Del [34] Imp [35]
11q23		-	77 (1421)	Dysarthric, Abs [36]; Imp [37]
11q23.3-q25	FLI1, JAM3	_	119 (1522)	Imp, Del, Apr [38]
12q14.3	HMGA2	-	9 (108)	Abs, Imp [38] or Del [39]
13q12.3		_	6 (119)	Del [40]
13q14	RB1	-	46 (595)	Norm [41]
13q22.3	EDNRB	-	3 (37)	Norm [41]
13q33–q34	SOX1, ARHGEF7	-	30 (635)	Apr
14q11–q22	PAX9, SUPT16H, CHD8, RALGAPA1	_	173 (2021)	Del [42] or Abs
14q22.1-q23.1	PTGDR, BMP4	-	41 (530)	Imp [43]
14q32.2	DLK1, MEG3	_	23 (242)	Del; Idio [44]
15q11.2	NIPA1, NIPA2, CYFIP1, TUBGCP5	_	11 (72)	Del
15q11-q13	NDN, SNRPN	-	30 (230)	Del or Imp
15q11-q13	UBE3A		-Same as above-	Abs; Imp [45]
15q13.3	CHRNA7, OTUD7A	-	8 (47)	Imp or Idio
15q24	SIN3A	_	38 (623)	Del or Imp
16p11.2	SH2B1, TBX6, CORO1A	1 (1)	70 (1067)	Apr [46]; Dys [47], Del or Imp
16p12.2-p11.2	SH2B1	1 (1)	106 (1682)	Del or Imp
16p12.1		-	17 (157)	Del
16p13.11	MYH11	-	6 (116)	Del [48]
16p13.3	CREBBP, TRAP1	1 (140)	122 (1532)	Apr, Dys, Imp or Del [49]

Yq11

Syndrome	Implicated Genes from Microarray Studies That Also form Voice Chains	No. of Genes in $\{\mathcal{V}_{N1}\}$	No. of Genes in $\{\mathcal{V}_{N2}\}$	Reported Effects on Speech
16q22	CBFB	1 (2)	79 (830)	-Not available-
16q24.3–q24.2	CDH15, ZNF778, ZFPM1	-	28 (307)	Del or Imp [50]
17p11.2	LLGL1, UBB	-	36 (845)	Del; Dys [51]
17p13.1	KCNAB3, GUCY2D, TP53, TRAPPC1, MPDU1, FXR2, EFNB3	2 (214)	74 (1387)	Abs
17p13.3	PAFAH1B1, YWHAE	-	37 (487)	Del [52]
17q11.2	NF1	-	40 (620)	No significant issues
17q12	HNF1B, LHX1, CCL3L3	1 (124)	59 (890)	Del or Imp
17q21.31	KANSL1, MAPT, CRHR1	2 (68)	46 (740)	Del or Abs
17q23.1–q23.2		-	28 (307)	Del [53]
17q24.3–q24.2	ABCA5, MAP2K6, SOX9	_	22 (720)	Del [54]
18q	MBP	-	122 (1666)	Del [55] or Imp [56]
19p13.13		2 (263)	131 (2080)	Del; Imp [57]
19q13.11	UBA2, WTIP	1 (23)	22 (283)	Del or Abs; Imp [58]
20p12.3	BMP2	_	13 (415)	Del or Imp [59]
22q11.2	TBX1, COMT, TOP3B	_	66 (1294)	Apr, Dys, Del or Imp [60]
22q12.2	NF2	-	27 (253)	Del or Imp [61]
22q13.3	ARSA, SHANK3	1 (9)	45 (670)	Del or Abs
Xp11.3	RP2	1 (6)	20 (335)	Imp [62]
Xp21	GK, DMD, NR0B1	-	7 (121)	Del [63]
Xq28 (a)	ABCD1, BCAP31, SLC6A8	-	64 (1053)	Del [64]
Xq28 (b)	MECP2	_	-same as above-	Norm to Abs

Table 1. Cont.

4. Inferences

A wealth of conclusions can be drawn from Table 1 (and from its more detailed version, Table A1 in Appendix A). However, we focus only on those that help validate the usefulness of the proposed algorithm.

6 (40)

Norm

4.1. Voice Chains as Predictors of Speech Characteristics

_

Of the 76 syndromes in Table A1, voice chains were found to exist for all. By our hypothesis, this would imply that in all cases, there is a potential for voice to be affected. The syndromes 15q11–q13 and Xq28 have two versions each, divided in the medical literature based on symptoms, rather than gene composition of the microdeletion region. We can therefore combine them for analysis, leaving us with 74 syndromes to be analyzed. For the syndrome 16q22, no information about the speech issues was found in the medical literature. Only the remaining 73 syndromes are considered in the analysis below.

The incidence of speech pathologies (including *all* forms of pathologies) among the general population is reported to be about 5% [65], and between 2.3% and 24.6% among children [66]. Of the 73 syndromes, 17 syndromes had both level-1 and level-2 voice chains, while 56 had only level-2 chains. The occurrence of speech aberrations was reported for all 17 syndromes with level-1 chains and for all but 6 of the 56 syndromes with only level-2 voice chains. Thus, voice chains correlate highly with the existence of speech anomalies.

4.2. Voice Chains as Information-Carrying Entities

Let us study how voice chains correlate with the presence or absence of *specific voice problems*. Such correlations would show that voice chains carry information about how the voice may be affected. This information is expected to be coarse-grained since we only take the presence or absence of any gene into consideration and consider no other cytogenetic information related to it.

From Table A1, we observe the following.

Level-1 chains:

The number of level-1 chainlink genes is limited to 1 or 2 in all cases and is not amenable to statistical analysis. However, we make the following observations:

- 1. Level-1 voice chains co-occur with speech problems 100% of the time.
- 2. For all instances where level-1 voice chains are present, severe symptoms occur 100% of the time (impaired, delayed or absent speech).
- 3. For all instances of syndromes with no effect on speech (i.e., where normal is not just one of a range of other speech symptoms), level-1 chains are absent 100% of the the time.

Level-2 Chains:

Level-2 voice chains are present in all cases and co-occur with speech disorders in all but 6 cases; thus, in only 6 cases has speech been reported to be normal. Therefore, level-2 voice chains co-occur with speech problems 91.8% of the time.

We note that a syndrome may have level-2 voice chains through many chainlink genes, which could number in the tens or even hundreds. Each of the chainlink genes could, in turn, also be associated with multiple other pathways, in addition to the one connecting it to FOXP2. We refer to the total number of pathways that include the chainlink genes of a syndrome as its "chainlink connectivity".

Table 2 presents some statistics of syndromes, chainlink genes, and chainlink connectivity associated with speech disorders of different severity. The problems considered are: absent speech (the most severe symptom), impaired speech (a symptom that is less severe than absent), delayed speech (a cognitive symptom that is also less severe than absent and comparable in severity to impaired speech—a physical symptom), dysarthric speech (a symptom related to physical issues), and apraxic speech (due to CNS disorders; this is less severe compared to absent speech and often subsumes idiosyncratic speech). Each row of the table represents one type of speech problem and shows the number of syndromes associated with it, the mean and median of the counts of chainlink genes for the syndromes. From an inspection of Table 2, a distinct pattern emerges. Rank ordering the symptoms by ascending order of the means of the counts of chainlink genes, we see that their connectivities also fall in almost the same order:

normal
$$(32, 404) < apraxia(43, 675) < dysarthria(43, 725) < impaired(46, 734)$$
 (1)
< delayed(47, 682) < absent(60, 902)

This rank ordering is consistent with the rank ordering of symptom severity based on the descriptions in the medical literature. In general, statistically speaking, the *number* of chainlink genes and the chainlink connectivity both appear to relate monotonically to the severity of the speech disorder.

Smooth Turno	Count	Chainlink Genes		Chainlinl	Connectivity
Speech Type	Count	Mean	Median	Mean	Median
Normal	6	32	40	404	595
Apraxic	19	43	30	675	557
Dysarthric	11	43	36	725	737
Impaired	31	46	38	734	608
Delayed	51	47	37	682	500
Absent	19	60	46	902	718

Table 2. Row-wise: Statistics showing the number of syndromes (count) associated with each voice disorder, the number of chainlink genes associated with the corresponding set of syndromes, and the connectivity of chainlink genes for the set.

Figure 3 shows scatter plots for counts of chainlink genes, chainlink connectivity, and a scatter of chainlink gene counts vs. normalized (per-chainlink-gene) chainlink connectivity for different severities of voice problems. Once again, it is apparent from the figures that the distributions of chainlink counts and chainlink connectivity is predictive of the type of speech problem. In particular, as is evident from Figure 3c, the distribution for normal speech stands out distinctly, as does that for absent speech, although the latter is not as distinctive as the former. Among the other levels, the distributions for apraxic and dysarthric speech appear similar, and so also do those for impaired and delayed speech appear similar.

In order to quantify these differences, we modeled the distributions of chainlink counts and chainlink connectivities for the different severity levels. These distributions have the general characteristics of over-dispersed Poisson distributions and can be modelled as Conway–Maxwell–Poisson (CMP) distributions [67]. The CMP distribution is a twoparameter exponential-family PMF over non-negative integers that takes the form

$$P(n) = \frac{\lambda^n}{(n!)^{\nu}} \frac{1}{Z(\lambda, \nu)}$$
(2)

where λ , $\nu > 0$ are the parameters of the distribution. Given a set of integers, λ and ν can be obtained through a maximum likelihood estimator [68].

Figure 4 shows the maximum likelihood estimates of the probability distributions of chainlink counts (Figure 4a) and connectivities (Figure 4b) for syndromes associated with speech problems of different severity levels. These, too, follow the visible trends of Figure 3, where the distributions of both chainlink counts and chainlink connectivities for syndromes associated with the two extreme conditions, normal and absent speech, are distinct from those for other types of problems.

To quantify the differences in the distributions, we define the *code distance* between two sets of integers $C_i = \{n_1, \dots, n_i\}$ and $C_j = \{m_1, \dots, m_j\}$ as the excess number of bits required to encode them if each set is encoded using the optimal code for the other set rather than itself.

$$D(P_i, P_j) = \sum_{n \in C_i} \log_2\left(\frac{P_j(n)}{P_i(n)}\right) + \sum_{n \in C_j} \log_2\left(\frac{P_i(n)}{P_j(n)}\right)$$
(3)

where $P_i()$ and $P_j()$ are the estimates of the PMFs for C_i and C_j , respectively. In our case, we choose the maximum likelihood estimates of the CMP distributions for the sets to compute this metric.

Table 3a shows the code distances between the chainlink counts for different types of speech problems. Table 3b shows the same for their chainlink connectivities. In both cases, we observe that the distributions for normal speech stand clearly apart from those for the other types of speech problems. The distributions for fully absent speech, too, are distinctive from those for other problem types. Among apraxic, dysarthric, impaired, and delayed speech, the differences between the distributions of adjacent degrees of severity is minimal; however, the distances show a distinct increasing trend with increases in the degree of impairment.



Figure 3. (a) Chainlink counts: Scatter of chainlink gene counts for syndromes associated with voice problems of different severities. Each cross represents a single syndrome. The horizontal axis value indicates the number of chainlink genes for the syndrome. The *y*-axis is a dummy axis. (b) **Chainlink connectivity**: Scatter of chainlink connectivities for syndromes associated with voice problems of different severities. Each panel shows the aggregate no. of pathways that all chainlink genes connect to (*x*-axis) for each syndrome (denoted by a cross) that exhibits the labeled speech characteristic. The *y*-axis is a dummy axis. (c) **Count vs. normalized connectivity**: Chainlink gene count (*x*-axis) vs. normalized chainlink gene connectivity (*y*-axis). The normalized connectivity is the average connectivity per chainlink gene.



Figure 4. Conway–Maxwell–Poisson models for the distributions of chainlink counts and chainlink connectivities for different types of speech problems. (**a**) Distribution of chainlink counts. (**b**) Distribution of chainlink connectivities.

Table 3. (a) Code distance between the (sets of) chainlink counts for different types of speech disorders.(b) Code distance between the (sets of) chainlink connectivities for different types of speech disorders.

			(a)			
	Normal	Apraxic	Dysarthric	Impaired	Delayed	Absent
Normal	0	2.1	2.3	7.6	14.7	15.3
Apraxic		0	0.2	1.1	2.9	6.8
Dysarthric			0	0.2	1.2	3.7
Impaired				0	0.5	3.7
Delayed					0	4.6
Absent						0
			(b)			
	Normal	Apraxic	Dysarthric	Impaired	Delayed	Absent
Normal	0	6.3	7.0	17.4	22.5	22.2
Apraxic		0	0.7	1.5	1.9	5.9
Dysarthric			0	0.0	0.4	1.7
Impaired				0	0.9	2.4
Delayed					0	5.5
Absent						0

Overall, from the above analysis, the properties of the level-2 chains of a syndrome appear predictive of the degree of the speech problems associated with it. Our analysis has considered the chainlink counts and connectivities indpendently, and each of them shows this behavior. A joint analysis of both may show stronger dependencies.

Most importantly, note that in all of the analysis above, we have ignored the secondary effects of other issues, such as intellectual disability and craniofacial anomalies, a highly simplifying assumption. A more correct information measure that takes these into account is expected to show even stronger relationships between the level and degree of connectivity of a syndrome to the FOXP2 pathway and its effect on speech.

4.3. Why Are There No Instances of Missing Voice Chains?

Are voice chains redundant? The fact that there are no missing voice chains is easily explainable. The reason is linked to the size of the syndromic regions. To understand this, consider the following facts.

Our database comprises 4319 unique pathways. A total of 1205 of these pathways are linked to the pathway that carries the FOXP2 gene, and collectively, these include 11,746 genes. Thus, a randomly chosen gene from the entire human genome of 42.7k genes (as in the HGNC Human Genome database) has a 27.5% chance of being on a pathway that links to the FOXP2 containing pathway, i.e., of being a level-2 chainlink gene.

The shortest microdeletion considered (2q23.1) includes 9 genes, each of which has a 27% chance of being a level-2 chainlink gene. The syndrome itself then has a 94.44% probability of having a level-2 voice chain purely by chance. The second shortest pathway includes 26 genes and has a 99.98% probability of having a level-2 voice chain by chance. The remaining pathways are larger (in terms of the number of genes), and it is virtually impossible for them *not* to have a level-2 chain.

As a result, it is realistic to expect that, as a consequence of the density with which the FOXP2 containing pathway is linked to other pathways, *any syndrome arising from genetic aberrations that includes even a moderately sized set of genes will have an effect on voice*. It remains a plausible hypothesis that any factor that influences gene function has at least some chance to ultimately affect voice—for example, at least a 27% chance within the boundaries of the example presented in this paper.

The above argument assumes that the genes in a microdeletion region are randomly chosen. The mean of the fraction of genes in a microdeletion that appears in any voice chain is observed to be 28.79% with a variance of 0.014, indicating concordance with the assumption of randomness. A secondary implication is that the likelihood of adjacent genes in the same cytogenetic region to be chainlink genes is independent of one another.

4.4. Ancillary Observations

Some important ancillary observations emerge from this study, which may be important to note. These are mentioned briefly below.

- 1. For each syndrome, some genes have been identified as largely important—i.e., these are *implicated* largely for the syndrome's effect on the individual. Of the syndromes for which there is information about implicated genes, we see that in only 8 syndromes (2p16.1–p15, 2q23.1, 9p24.3, 11q23, 13q12.3, 17q23.1–q23.2, 19p13.13, and Yq11), none of the implicated genes appear in the two levels of voice chains shown. In all other cases, the implicated genes impact FOXP2 pathways and are likely to have a bearing on speech anomalies. We have noted earlier that FOXP2 is not the only gene known to be related to voice production. If we had chosen some other gene as an example in this paper (instead of FOXP2), it is likely that the implicated genes for the 8 exceptions mentioned above would appear as chainlink genes (while some others may not). This hypothesis can be easily tested in corresponding experiments.
- 2. **Identifying candidate genes for further investigation:** Using only chainlink genes that appear on level-1 chains as illustrative examples (see Table A1 in the Appendix A for reference), we see that voice chains can be useful in identifying candidate genes for further investigation in the context of speech issues. Some examples are given below. The likely candidates are written in parentheses, while the already implicated genes are indicated in bold:
 - 1p36 (ARID1A): Although not implicated for this syndrome in studies so far, ARID1A is located in 1p36.11, a region frequently deleted in human cancers [69]. Disruption in its function may lead to the co-occurrence of oncological and speech issues. This hypothesis is verifiable.
 - 5q35.3 (NSD1): The gene NSD1 appears in a level-1 chain and is also an implicated gene. Ideally, this should not be a candidate for further investigation. However, paradoxically, while effects on speech are expected, the literature reports normal speech for some subjects for this case. This may be a result of biased sampling (the more severe cases may not be conducive to life due to other concurrent severe symptoms, which is a common occurrence in microdeletion syndromes; in some cases, only mosaic individuals survive). This warrants some investigation.
 - 11p15.5 (HRAS): Although not implicated, and although two studies cited under OMIM: 130650 for this syndrome explicitly mention HRAS as *not* significant, HRAS has nevertheless been independently found to be extremely significant in RASopathy and cancer studies, e.g., [70]. Its role in this syndrome needs to

be re-evaluated given its influence on 347 biological pathways and its strong influence on speech.

- 16p11.2 and 16p12.2–16p11.2 (SRCAP): Although not implicated, it connects to only one other pathway in the ensemble, and that is the ACC pathway of FOXP2. The effects on speech are expected to be strong if this gene is aberrant. This gene may be implicated in further investigations.
- 17p13.1 (KDM6B): Speech is absent in this syndrome. The gene TP53 is implicated, which also appears at level-1 and is associated with 206 pathways. KDM6B is the only other gene in the level-1 voice chains and connects to only 8 other pathways. It is likely that this gene also plays a strong role in influencing speech and merits investigation.
- 17q12 (ERBB2): The gene ERBB2 is associated with 124 pathways. It is a wellknown oncogene [71], in that perturbations in its function have been observed to have deleterious effects. If it is also connected to FOXP2, then its appearance in the voice chain allows a surprising hypothesis—that biomarkers of some oncological conditions may also be present in voice.
- 19p13.3 (MAP2K2,UHRF1): MAP2K2 and URHF1 are not implicated. However their appearance as level-1 chainlink genes warrants investigation, especially for MAP2K2, which influences 257 pathways. Prompted by this, a literature search did reveal that MAP2K2 *has* been implicated in this syndrome recently [72], although this is not on the OMIM records, which were largely consulted for this study.
- 22q13.3 (BRD1): The gene BRD1 is not implicated and appears in 9 pathways only, but the effect on speech is severe in this syndrome. This warrants the investigation of BRD1 independently in relation to speech characteristics. A literature search reveals that BRD1 is indeed strongly associated with brain development and susceptibility to both schizophrenia and bipolar affective disorder [73], and consequent effects on speech are highly likely.
- Xp11.22 (SMC1A): Although SMC1A is not implicated, it appears in 33 pathways. The speech issues are severe and the gene warrants investigation for this effect. A recent report in the literature has implicated it in severe intellectual disability and therapy-resistant epilepsy in females [74]. The former is known to be associated with severe speech anomalies.
- Xp11.3 (KDM6A): Although not implicated, KDM6A warrants investigation. In the literature, it is independently known to be associated with delayed speech and psychomotor development [75].
- 3. **Expression of speech characteristics:** The observation that deletions of genes on *all chromosomes* ultimately results in the expression of speech anomalies carries significance. From a much broader perspective, this suggests that the effect on speech may be supported by the action of multiple concurrent biological pathways. There may be no single gene or genes (on select chromosomes) that may code for speech capabilities per se, and FOXP2 may be one of a few genes that may *consolidate and regulate* the speech- and language-related emergent effects. It may be that genes directly code for structural elements in the range of phenotypes, while other properties, such as speech and language abilities, are emergent from the coordination of these (and epigenetic) factors.

A more prosaic argument for this can also be presented. Within the ensemble of syndromes analyzed, there are three kinds of of cause-and-effect relationships: (a) syndromes with *physical* structures of the vocal tract (e.g., craniofacial anomalies that include cleft palate, changes in lip shape, etc.), which adversely affect the biomechanical aspects of voice and speech production, (b) syndromes in which auditory and motor functions are compromised, and (c) syndromes that affect the normal functions of the brain, causing cognitive, learning, memory, and other issues that are, in turn, likely to lead to speech problems. In no case do we see only speech aberrations in

isolation of these factors. The associations between speech and other expressed factors have, in fact, been ubiquitously observed, e.g., [48]. This may support the hypothesis that speech abilities are likely to be emergent from an ensemble of factors (including other phenotypes), rather than expressed directly by any "speech" gene.

5. Conclusions

The hypothesis that the existence of voice chains is correlated with speech characteristics is adequately validated by the statistical analysis presented in this paper. The analysis presented in this paper, in fact, also shows that the level of voice chains is correlated positively with the severity of speech problems. Based on this, a simple information measure has been suggested to rank-order the effects of specific sets of voice chains on speech. We also see how the methodology presented can potentially provide leads to specific genes that might be candidates for further investigation in the context of speech issues and microdeletion syndromes. While the example of chromosomal microdeletion syndromes used for this paper is very specific, the methodology itself may be easily generalized and extended to reveal the potential effects of other diseases with a genetic basis and of other factors that influence gene function in some manner on speech, voice and (in further refinements of the analysis) their specific qualities and characterisitcs. As a specific suggestion, one exercise that would allow for a more comprehensive analysis would be to explore the entire human genome database to identify which genes are connected via voice chains (and to what level), as well as whether or not there have been corresponding effects on voice reported in the biomedical literature. In cases where large amounts of data are available, one could also explore such connections in an entirely data-driven manner, using AI-based biomarker discovery mechanisms.

Funding: This research was funded in part by the U.S. Army Research Office and the U.S. Army Futures Command Grant No. W911NF-20-D-0002. Its content does not reflect the position or the policy of the U.S. Army and no official endorsement should be inferred.

Institutional Review Board Statement: Ethical review and approval were waived for this study since the data used for this research was obtained from public sources. No human subjects research was explicitly carried out for this work.

Data Availability Statement: The code required to reproduce the results in this paper is archived for public use at https://datadryad.org/ (accessed on 21 September 2022) under the title "Data for Connecting voice profiling to genomics".

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

This table lists the voice chains found for a set of 76 documented microdeletion syndromes. This range excludes chromosome 21, for which sufficient documentation was not found in the literature. The analysis was conducted for voice chains up to level 2. The format of each row in this table is as follows:

First column:

- In each row, the top left entry is the microdeletion syndrome.
- Below this, on the left, is the OMIM record for the cytogenetic region of the syndrome.
- Below the OMOM record, the common names by which the syndrome is referred to in the medical literature are listed.
- Below the list of common names, sets of chainlink genes that form level-1 and level-2 voice chains (denoted as {V_{N1}} and {V_{N2}}) respectively, found by the proposed algorithm are listed. These sets are listed only if they are present.
- For each voice chain listed, the first entry denotes the level of the voice chain. Following this is the list of chainlink genes that belong to the corresponding microdeletion region, which are linked to FOXP2 through the corresponding voice chain level.

- For each gene, the number of pathways that a gene connects to (in general, inclusive of connections to the ACC pathway of FOXP2) is written as its subscript.
- All chainlink genes are not named. For brevity, the first 10 genes with the greatest number of links are listed, and the total number of the rest of the genes (total chainlink gene count) is indicated. At the end, the genes with a single pathway link are explicitly listed. They are *included* in the total chainlink gene count mentioned above. Thus, for example, the level-2 voicechain for the syndrome 2p16.1–p15 has a total chainlink gene count of 17, *including* all genes that are explicitly listed in the table.
- In each row, the genes in the voice chains that have been implicated for the syndrome's phenotypic expression in prior studies are shown in parentheses on the top right. The genes that are also present as chainlink genes in the voicechains found for the syndrome are shown in bold font.

Second column:

• The second column in this table lists the corresponding speech characteristics, with references. Where no reference is cited, the information is found in the OMIM record for the syndrome (where possible, OMIM references are used for brevity).

Table A1. Chainlink genes for level-1 and level-2 voice chain ensembles for 76 chromosomal microdeletion syndromes. Each chain forms a link from the gene shown to the ACC biological pathway of FOXP2 and has been automatically derived. The total number of pathways that each gene influences independently is shown as a subscript to its name. Observed phenotypic effects on speech are given in the last column. When references are not cited, the information reflects that in the OMIM records for the syndrome.

Syndrome	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information		
Voice Chains and Their Member Chainlink Genes		
Chromosome 1 deletions		
1p36	(Multiple incl. RERE, SPEN) ^{†‡}	Delayed [9] or absent speech
OMIM: 607872; Cytogenetic location: 1p36; Genom	nic coordinates (GRCh38): 1:23,600,000–	
27,600,000		
Names: Chromosome 1p36 deletion sydrome; mono	osomy 1p36 syndrome	
$\{\mathcal{V}_{N1}\}$ ARID1A ₁₀		
$\{\mathcal{V}_{N2}\}$: PIK3CD ₁₆₈ MTOR ₁₄₆ GNB1 ₁₄₀ CDC42 ₁₄₀ C	ASP9 ₁₀₈ RPS6KA1 ₉₃ PRKCZ ₇₅ CNKSR1 ₅₇	
SFN ₅₅ DVL1 ₄₈ (216 more) SSU72 ₁ RBP7 ₁ PLEK	HN1 ₁ PLEKHM2 ₁ PEX10 ₁ NBL1 ₁ MTCO3P12 ₁	
MMP23B ₁ MMP23A ₁ HSPB7 ₁ HMGN2 ₁ GPR3 ₁ EX	OSC101 DISP31 CROCCP21 CAMTA11	
AHDC1 ₁ ACAP3 ₁		
1q21.1–q21.2	(Multiple incl. RBM8A ^{+‡} , GJAS ⁺)	Delayed [10] or Impaired [11]
OMIM: 612474/274000, Cytogenetic location: 1q21	.1 Genomic coordinates (GRCh38):	speech
1:143,200,000-147,500,000 Names: Thrombocytoper	ia with absent radius (TAR) syndrome (OMIM	
274000)		
$\{\mathcal{V}_{N2}\}$: H4C14 ₆₇ H4C15 ₆₆ H2BC21 ₆₃ PRKAB2 ₅₇ H3	C15 ₅₆ H3C14 ₅₆ H3C13 ₅₆ H2AC20 ₅₃ H2AC19 ₅₃	
H2AC18 ₅₃ (31 <i>more</i>)		
1q41–q42 (Multiple incl. DISP1 , HPE10	, LEFTY1 , LEFTY2 , WDR26, TSEN2, BPNT1) ^{†‡}	Apraxia [12]
OMIM: 612530; Cytogenetic location: 1q41-q42 Ge	nomic coordinates (GRCh38): 1:214,400,000-	
236,400,000		
Names: Chromosome 1q41–q42 deletion syndrome;	holoprosencephaly 10; HPE10	
$\{\mathcal{V}_{N2}\}$: NUP133 ₈₈ H2BU1 ₆₃ TGFB2 ₆₂ DUSP10 ₅₄ H3	3-3A ₅₀ WNT3A ₄₇ ARF1 ₄₆ PARP1 ₄₁ MIR3620 ₃₈	
$PSEN2_{36} (50 more \dots) DISP1_2 MIR215_1 MIR194-1_2$	DNAH14 ₁ CDC42BPA ₁	
1q43–q44	(AKT3 , ZBTB18) ‡	Delayed, impaired, or absent
OMIM: 612337		speech
Names: Mental retardation, autosomal dominant 2	2; MRD22; chromosome 1q43–q44 deletion	
syndrome (included); chromosome 1qter deletion sy	vndrome (included)	
$\{\mathcal{V}_{N2}\}$: AKT3 ₁₆₉ RYR2 ₇₄ ACTN2 ₆₉ CHRM3 ₄₂ MTR	₃₈ FH ₃₁ ADSS2 ₃₀ EXO1 ₂₆ RGS7 ₁₆ KMO ₁₅	
(68 more) $FMN2_1$		

21 of 33

Syndrome	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information		
Voice Chains and Their Member Chainlink Gene	95	
Chromosome 2 deletions		
2p16.1–p15 OMIM: 612513; Cytogenetic location: 2p16.1–p15	(Multiple incl. BCL11A) ⁺⁺ Genomic coordinates (GRCh38): 2:54,700,000–	Dysarthria, apraxia, or im- paired speech [13]
63,900,000 Names: Chromosome 2n16.1, n15 deletion syndrom	m 0	
$\{V_{M2}\}$: RPS27A ₂₀₂ XPO1 ₅₄ UGP2 ₂₂ MDH1 ₂₇ REL	22 CCT414 RTN412 B3GNT212 VRK210 USP347	
(7_{N2}) (12_{23}) $(12_{$		
2p21 OMIM: 606407; Cytogenetic location: 2p21 Genor 47,500,000	(Multiple incl. SLC3A1 , PREPL) nic coordinates (GRCh38): 2:41,500,000–	Apraxia or idiosyncratic speech [14]
Names: Hypotonia–cystinuria syndrome; cystinu 2p21 deletion syndrome	ria with mitochondrial disease; homozygous	
$\{\hat{\mathcal{V}}_{N2}\}$: CALM2 ₂₃₄ PRKCE ₇₃ SLC3A1 ₃₆ HAAO ₂₂	ATP6V1E2 ₁₉ ABCG5 ₁₈ MSH2 ₁₇ ABCG8 ₁₇	
EPAS1 ₁₆ COX7A2L ₁₅ $(14 \text{ more } \dots)$ SIX3 ₁ EML4 ₁		
2q23.1	(MBD5) ‡	Delayed or impaired speech
Names: Mental retardation, autosomal dominant 1 $\{V_{N2}\}$: ORC4 ₂₉ KIF5C ₂	; MRD1; chromosome 2q23.1 deletion syndrome	
2q32–q33 (SATB2) [‡] (HOXD cluster and regulator	y elements, COL3A1 COL5A2, GTF3C3, CASP8,	Absent speech
OMIM: 612313		
Names: Glass syndrome; GLASS; chromosome 2q	32–q33 deletion syndrome; SATB2-associated	
syndrome		
$\{\mathcal{V}_{N2}\}$: CREB1 ₂₀₃ STAT1 ₁₃₀ CASP8 ₁₀₃ NUP35 ₇₉ IT	CGAV ₆₄ CD28 ₆₀ SUMO1 ₅₂ AOX1 ₅₀ FZD5 ₄₃	
$FZD7_{37} (52 more) TMEFF2_1 KLF7_1 HSPE1_1 DT$	USP19 ₁ DNAH7 ₁	
2q37.3 OMIM: 600430 (2q37); Cytogenetic location: 2q37 242,193,529	(HDAC4) [‡] Genomic coordinates (GRCh38): 2:236,400,000–	Impaired speech [15]
Names: Chromosome 2q37 deletion syndrome, br Albright hereditary osteodystrophy-like syndrome ()2ya): HDAC4 22 GPC127 AGXT25 COL6A340 AC	achydactyly–intellectual disability syndrome; Type 3 KR344 NEU412 NDUEA1012 PRI H11 PER211	
$\frac{(7N_2)}{28 \text{ more } \dots)} \text{TWIST2}_1 \text{ ILKAP}_1$		
Chromosome 3 deletions		
3p13	(FOXP1) ^{†‡}	Delayed, idiosyncratic,
OMIM: 613670	and the second design of the factories	and impaired speech and dysarthria (all severe)
Names: Mental retardation with language impairin $\{\mathcal{V}_{M3}\}$: MITE ₁₀ PROK2 ₈ PPP4R2 ₅ GPR27 ₅ EIF4E3 ₅	FOXP12 GXYLT22 RYBP1	apraxia [16]
3q13.31	(DRD3, ZBTB20 [‡] , GAP43, LSAMP) ^{†‡}	Impaired [17] or absent
OMIM: 615433; Cytogenetic location: 3q13.31 Ger 117,600,000	nomic coordinates (GRCh38): 3:113,700,000-	speech
Names: Chromosome 3q13.31 deletion syndrome $\{\mathcal{V}_{N2}\}$: DRD3 ₂₂ ATP6V1A ₂₂ GAP43 ₆ QTRT2 ₃ LS A	AMP ₃	
3q29 OMIM: 609425; Cytogenetic location: 3q29 Genor	(Multiple incl. PAK2, DLG1) ^{†‡} nic coordinates (GRCh38): 3:192,600,000–	Delayed speech
Names: Chromosome 3q29 deletion syndrome; m	icrodeletion 3q29 syndrome; 3qter deletion syn-	
drome { V_{N2} }: PAK2 ₇₄ DLG1 ₇₁ NCBP2 ₅₃ HES1 ₃₉ RPL354	A22 TFRC21 RNF16818 BDH117 PCYT1A15	
MUC20 ₁₃ (14 <i>more</i>) FBXO45 ₁		

22 of 33

Syndrome (Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information	
Voice Chains and Their Member Chainlink Genes	
4p16.3 (Multiple incl. FGFR3 , MSX1) ⁺ ±	Delayed [18] or absent speech
OMIM: 194190; Cytogenetic location: 4p16.3 Genomic coordinates (GRCh38): 4:0–4,500,000	
Names: Wolf-Hirschhorn syndrome; Pitt-Rogers-Danks syndrome; Pitt syndrome; Wittwer syn-	
drome; Dillan 4p syndrome	
$\{V_{N1}\}$ CTBP1 ₂₅	
$\{V_{N2}\}$: FGFK3 ₁₀₉ INIF 2 ₃₂ NELFA ₂₇ NIK945 ₂₆ CIDF 1 ₂₅ ADKA2C ₂₂ N5D2 ₁₉ GKK4 ₁₅ DGKQ ₁₄	
$\frac{PDE0D_{13} \text{ HAUS5}_{13} \text{ SLC20A1}_{12} \text{ SLDP}_{12} \text{ AIP5}_{\text{ME12}} \text{ KNP4}_{11} \text{ CPLA1}_{11} \text{ KGS12}_{10} (27 \text{ more} \dots)$	
$\frac{2\Gamma\GammaVE2o_1\GammaCGFo_1NOG1_1NIAD4_1ADCA11F_1}{4o^{21}}$ (Multiple)	Delayed or abcent speech:
OMIM: 613509: Cytogenetic location: 4g21 Genomic coordinates (GRCh38): 4:86.000.000–	impaired speech [19,20]
87,100,000	
Names: Chromosome 4q21 deletion syndrome	
$\{\mathcal{V}_{N2}\}$: MAPK10 ₁₂₂ FGF5 ₁₁₈ NUP54 ₇₉ PAQR3 ₄₇ PRKG2 ₂₇ CXCL10 ₂₂ CXCL9 ₁₈ ABRAXAS1 ₁₈	
SEC31A ₁₆ CXCL11 ₁₆ (31 <i>more</i>) SHROOM3 ₁ SEPTIN11 ₁ HNRNPDL ₁ G3BP2 ₁ BMP2K ₁ AFF1 ₁	
Chromosome 5 deletions	
5p (5p15.2 and/or (5p15.3 or 5p15.33)) (Multiple incl. TERT, CTNND2) ⁺	Delayed or absent speech and
OMIM: 123450	apraxia [21]
syndrome: 5n monosomy: partial monosomy 5	
{V _{N2} }: ADCY2 ₁₁₆ SLC6A3 ₇₃ SDHA ₃₇ TERT ₂₄ TRIO ₂₃ MTRR ₁₉ LPCAT1 ₁₅ CEP72 ₁₅ SRD5A1 ₁₄	
SLC9A3 ₁₃ (27 more) OTULINL ₁ CTNND2 ₁ CLPTM1L ₁	
$\frac{5a14.3}{(MEF2C)^{\frac{1}{2}}}$	Absent speech
OMIM: 612881 (distal version 5q14.3–q15); Cytogenetic location: 5q14.3–q15 Genomic coordinates	nobelii opeeeli
(GRCh38): 5:83,500,000–98,900,000	
Names: Distal chromosome 5q14.3 deletion syndrome; periventricular heterotopia associated with	
chromosome 5q deletion; periventricular nodular heterotopia 5; PVNH5	
$\frac{\{V_{N2}\}: \text{KASA189 CCINH_{74} MEF2C_{70} FOLK3G_{17} COA/C_{15} \text{ MIR9-27 EDIL33}}{5\sigma^{23} 1}$ (PPS1 4) ^{†‡}	Ducarthria [22]
OMIM: 153550	
Names: Chromosome 5q deletion syndrome; 5q syndrome; refractory macrocytic anemia due to 5q	
deletion; MAR	
$\{\mathcal{V}_{N2}\}$: RPS14 ₂₉ GPX3 ₁₈ DCTN4 ₁₃ SPARC ₁₂ SLC36A1 ₁₂ NMUR2 ₁₁ CD74 ₁₀ NDST1 ₉ GM2A ₈	
TNIP1 ₆ (7 more) SYNPO ₁ IRGM ₁	
5q35.3 (NSD1) [‡]	Normal [23] or delayed
OMIM: 117550	speech
Names: Sotos syndrome 1; Sotos1; Nevo syndrome; cerebral gigantism, Nevo type; chromosome	
$\{\mathcal{V}_{N1}\}$ NSD18	
$\{\mathcal{V}_{N2}\}$: MAPK9 ₁₆₇ LTC4S ₅₃ F12 ₃₃ SQSTM1 ₃₀ MAML1 ₂₉ RACK1 ₂₅ FLT4 ₂₃ CANX ₂₂ GRK6 ₂₀	
$GRM6_{16}$ (37 more) $CLK4_1$	
Chromosome 6 deletions	
6pter–p24 (Multiple incl. FOXC1, GMDS) ^{†‡}	Delayed speech
OMIM: 612582; Cytogenetic location: 6pter-p24 Genomic coordinates (GRCh38): 6:0–13,400,000	
Names: Chromosome 6pter-p24 deletion syndrome	
$\{V_{N2}\}$: KITK183 EDIN143 FI3AI34 IFAF2A19 D5P18 BMP617 IKF415 IUBB2A14 GMDS13 ELOVL213	
$(31 more) PAK1IP1_1 MAK_1 FUXQI_1 DUSP22_1 C6orf201_1$	
6q25.3 (Multiple incl. ARID1B) [‡]	Delayed speech, apraxia,
	dysarthria [74]
OMIM: N.A. {V _{N2} }: SLC22A253 GTF2H550 ACAT240 SLC22A124 EZR24 SOD200 IGF2R20 SYNI210 SLC22A314	dysarthria [24]

23 of 33

Syndrome	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information		
Voice Chains and Their Member Chainlink Gen	es	
Chromosome 7 deletions		
/p21 OMIM: 101400	(TWISTI (/p21.1)) +	Delayed speech [25]
Names: Saethre–Chotzen syndrome: acrocephalos	syndactyly III: ACS3: ACS III: Chotzen syndrome:	
acrocephaly, skull asymmetry, and mild syndactyl	V	
$\{\mathcal{V}_{N2}\}$: RPA3 ₇₄ PRPS1L1 ₃₀ HDAC9 ₂₆ ITGB8 ₁₈ AI	HR ₁₈ POLR1F ₁₆ NDUFA4 ₁₅ DGKB ₁₄ TWIST1 ₉	
THSD7A ₄ (8 <i>more</i>) MIOS ₁ MEOX2 ₁ ARL4A ₁		
7q11.23 (N	Iultiple incl. ELN, LIMK1, GTF2IRD1, GTF2I) ^{†‡}	Normal or delayed speech
OMIM: 194050; Cytogenetic location: 7q11.23 Ge	nomic coordinates (GRCh38): 7:72,700,000–	
Names: Williams syndrome: WS: WMS: chromos	ome 7a11 23 deletion syndrome 1.5- to 1.8-MB	
Williams–Beuren syndrome; WBS		
$\{V_{N2}\}$: YWHAG ₆₄ POM121 ₅₉ RFC2 ₅₇ MDH2 ₄₂ L	IMK1 ₃₄ HSPB1 ₃₁ STX1A ₃₀ FZD9 ₃₀ NCF1 ₂₉	
POR ₁₅ (25 <i>more</i>) ABHD11 ₁		
Chromosome 8 deletions		
8p23.1	(GATA4)	No significant speech
OMIM: 179613 (not exclusive to syndrome)		anomaly reports
tion and 8nter-p23 1 deletion: San Luis Valley syn	drome	
$\{\mathcal{V}_{N2}\}$: FDFT1 ₄₄ CTSB ₁₇ BLK ₁₇ TNKS ₁₆ GATA4 ₁₁	5 AGPAT514 ANGPT212 MIR124-110 NEIL29	
$CLDN23_8$ (39 more) PINX1 ₁ MIR598 ₁	14 12 10 10	
8g22.1	(CCNE2, TMEM67, FAM92A1)	Delaved speech
OMIM: 608156; Cytogenetic location: 8q22.1 Gen	omic coordinates (GRCh38): 8:92,300,000-	
97,900,000		
Names: Nablus mask-like facial syndrome; NMLI	FS; chromosome 8q22.1 deletion syndrome	
$\{V_{N2}\}$: CCNE2 ₄₅ SDC2 ₃₄ TP53INP1 ₁₃ GDF6 ₁₂ UC	$QCRB_{11} PIDSSI_9 ESRPI_8 PDPI_5 NDUFAF6_5$	
CDH17 ₄ (3 more)	<u>r</u>	
8q24.11-q24.13	(TRPS1, EXT1) ⁺	Delayed speech
OMINI: 150230; Cytogenetic location: 8q24.11–q2 8.116 700 000_126 300 000	4.13 Genomic coordinates (GRCn38):	
Names: Langer–Giedion syndrome: LGS: chromo	some 8a24.1 deletion syndrome: trichorhinopha-	
langeal syndrome type II; TRPS2		
$\{\mathcal{V}_{N2}\}$: SQLE ₄₃ TAF2 ₂₀ RAD21 ₂₀ MIR3610 ₁₉ TNF	RSF11B ₁₅ DERL1 ₁₃ NDUFB9 ₁₂ EXT1 ₁₂	
SLC30A8 ₁₀ FBXO32 ₈ (11 <i>more</i>) WASHC5 ₁ FA	M91A1 ₁ DSCC1 ₁	
Chromosome 9 deletions		
9p24.3	(DMRT1, DMRT2) ⁺ ‡	Delayed speech, dysarthria,
OMIM: 154230; Cytogenetic location: 9p24.3 Gene	omic coordinates (GRCh38): 9:0–2,200,000	apraxia [26]
deletion: chromosome 9n24 3 deletion syndrome	ual dysgenesis, partial or complete, with 9p24.5	
$\{\mathcal{V}_{N1}\}$ SMARCA29		
$\{\mathcal{V}_{N2}\}$: SMARCA29 DOCK83 WASHC11		
9q34.3	(EHMT1) ‡	Delayed speech [27], Apraxia
OMIM: 610253	talana taia dalatian ana darara darara	[28] and Absent speech
Page 3 deletion syndrome; 9q syndrome; 9q sul	oleiometric deletion synarome; chromosome	
$\{\mathcal{V}_{N1}\}$ NOTCH162		
$\{\mathcal{V}_{N2}\}$: GRIN1 ₁₃₉ TRAF2 ₇₈ NOTCH1 ₆₂ PTGDS ₅₁	ANAPC237 TUBB4B28 NELFB27 ENTPD825	
CACNA1B ₂₀ AGPAT2 ₂₀ (36 more) UAP1L1 ₁	MIR602 ₁ FUT7 ₁ CYSRT1 ₁	

Syndrome	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information		
Voice Chains and Their Member Chainlink Ger	nes	
Chromosome 10 deletions		
10pter–p13 or 10p14–p15.1	(Multiple incl. GATA3)	Sensorineural hearing loss
UMIM: 146255	a concertingunal destinges and repaidie	
Names: barakat syndrome; hypoparathyroidish	i, sensorineural dearness, and renal dis-	
$\int V_{22} = 0$ II 2R Δ_{22} PRKCO- Δ KR1C3-2 IDI1 α CA	MI 5 ₂₆ AKR1C4 ₂₅ CALMI 3 ₂₂ CATA 3 ₂₇	
$\left(V_{N2} \right)$. ILZIC 183 TREEQ/T FRACEO/0 IDT40 CF		
10. 22 IAI 321 (39 more) ZMINDII1 CE.		
10q23 (10 $q22.2, q22.2$)	(PIEN, BMPKIA)	Delayed or absent speech;
Names: Chromosome $10q22.3 - q23.2$ deletion sur	dromo	impaned speech [27]
$\{\mathcal{V}_{3,1}\}$ PTFN ₁₀₇	laronie	
$\{\mathcal{V}_{N2}\}$: PTEN ₁₀₇ CYP2C8 ₁₀₀ NRG3 ₈₅ CYP2C9 ₇₈	FAS70 CYP2C1966 GLUD146 LIPA39 BMPR1A28	
$PLCE1_{25} (49 more) PDLIM1_1 PCGE5_1 MMR^2$	N21 IFIT21	
	(HMY3 DOCK1 C10OPE00) ^{††}	Impaired apoch. dologod
OMIM. 609625: Cytogenetic location: 10/226 Cer	(ПИЛО, DUCKI, CIUUKF9U) '*	speech [30]
133 797 422	lonne coordinates (Greenso). 10.120,000,000-	
Names: Chromosome 10g26 deletion syndrome:	terminal chromosome 10a26 deletion syndrome	
$\{\mathcal{V}_{N2}\}$: FGFR2 ₁₁₃ CYP2E1 ₉₁ ECHS1 ₈₈ ACADSB	1 DOCK133 BUB328 OAT21 RPL21P1620 EIF3A19	
$GRK5_{17}$ (38 more) RAB11FIP21 MKI671 EBE	$3_1 \text{ CUZD1}_1 \text{ CALY}_1$	
Chromosome 11 deletions		
11p11 2_p12	(FXT2 AI X4) ‡	Idiosyncratic speech
OMIM: 601224 (11p11.2): Cytogenetic location:	11p11.2 Genomic coordinates (GRCh38):	dysarthria, delayed speech,
11:43.400.000–48.800.000		apraxia [31]
Names: Potocki–Shaffer syndrome; PSS; chromo	osome 11p11.2 deletion syndrome: proximal 11p	
deletion syndrome; DEFECT11 syndrome		
$\{\mathcal{V}_{N2}\}$: TRAF6 ₁₃₄ NUP160 ₈₈ F2 ₈₅ PSMC3 ₃₀ NR1	H3 ₂₆ CREB3L1 ₂₆ CKAP5 ₂₅ DDB2 ₂₂ SPI1 ₁₉	
HSD17B12 ₁₉ (36 <i>more</i>)		
11p13-p12 (Multip	ble incl. WT1, PAX6 , BDNF) ^{†‡} (SLC1A2 , PRRG4) [*] ,	Impaired speech [32]
OMIM: 612469; Cytogenetic location: 11p13-p1	2 Genomic coordinates (GRCh38): 11:31,000,000–	1 1 1 1
43,400,000		
Names: Wilms tumor, aniridia, genitourinary and	omalies, mental retardation and obesity syndrome;	
WAGRO; WAGRO syndrome; WAGR syndrome	e with obesity; chromosome 11p13-p12 deletion	
syndrome		
$\{V_{N2}\}$: TRAF6 ₁₃₄ CAT ₃₇ CD44 ₂₇ CD59 ₁₉ CSTF3	$_{16}$ SLC1A2 ₁₄ EIF3M ₁₄ PDHX ₁₂ WTT ₁₀ APIP ₁₀	
(13 more) PRR5L ₁ ELF5 ₁		
11p15.5	(Multiple incl. CDKN1C, H19, IGF2) ^{†‡}	Impaired speech [33]
OMIM: 130650;		
Names: 130650: Beckwith–Wiedemann syndron	ne; BWS; exomphalos–macroglossia–gigantism	
syndrome; EMG syndrome; Wiedemann–Beckwi	th syndrome; WBS	
$\{V_{N1}\}$: HKA5 ₃₄₇		
$\{\nu_{N2}\}$: FIKA3347 IN3103 FOLK2L85 KCINQI69 D	$ \begin{array}{c} \textbf{U}_{\text{D}} \textbf{F}_{053} \text{ ININ1552} \text{ ININ1252} \text{ A} \textbf{F}_{2} \textbf{A} \textbf{F}_{4} \textbf{F}_{39} \text{ I} \textbf{H}_{30} \\ \textbf{F}_{05} \textbf{F}_{05} \textbf{F}_{15} $	
(40 more) KIC8A ₁ KASSF7 ₁ NLKP6 ₁ MIR48	3_1 KKIAP5- 4_1 KKIAP5- 1_1 DEAF1 $_1$ CKACR2B $_1$	
<u>DK5K21</u>		
11q13.3 OMIM: 166750	(FGF4, FGF3, FADD) '+	Delayed speech [34]; Im-
Names: Chromosome 11/213 deletion syndrome	ododental syndrome: otodental dysplasia and	paneu specch [JJ]
coloboma due to 11a13 3 microdeletion	, ououentai synuionie, otouentai uyspiasia allu	
$\{\mathcal{V}_{N2}\}$: FGF4 ₁₃₅ CCND1 ₁₂₂ FGF3 ₁₀₂ FGF19 ₀₅ FA	DD73 CPT1A38 CTTN22 PPFIA18 TPCN25	
SHANK2 ₄ ANO1 ₃	, , , , , , , , , , , , , , , , , , ,	
1 0		

Syndrome (Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information	
Voice Chains and Their Member Chainlink Genes	
11q23 (Multiple incl. FLI1) ^{†‡} OMIM: 188025; Cytogenetic location: 11q23 Genomic coordinates (GRCh38): 11:114,600,000– 121 300 000	Dysarthric speech, absent speech [36]; impaired speech [37]
Names: Chromosome 11g23 deletion syndrome: thrombocytopenia, Paris-Trousseau type: TCPT:	
Paris–Trousseau syndrome; 11g terminal deletion syndrome	
$\{\mathcal{V}_{N2}\}$: FXYD2 ₁₅₈ PPP2R1B ₁₄₅ CBL ₁₀₄ H2AX ₆₂ NCAM1 ₅₆ CD3G ₄₃ CD3D ₄₀ SC5D ₃₉ DLAT ₃₉	
APOA1 ₃₉ (67 more) ZPR1 ₁ TMPRSS4 ₁ TMPRSS13 ₁ TAGLN ₁ SIK3 ₁ POU2F3 ₁ MPZL2 ₁	
C11orf521	
11q23.3-q25 (FLI1, BSX, NRGN, FRA11B, JAM3) ⁺	Impaired speech, delayed
OMIM: 147791; Cytogenetic location: 11q23.3–q25 Genomic coordinates (GRCh38): 11:114,600,000–135,086,622	speech, apraxia [38]
Names: Jacobsen syndrome; JBS; Del(11)(qter); distal deletion of 11q; distal monosomy 11q; mono-	
somy 11qter	
$\{V_{N2}\}$: FXYD2 ₁₅₈ CBL ₁₀₄ KCNJ5 ₇₀ H2AX ₆₂ HSPA8 ₅₁ CHEK1 ₄₈ CD3G ₄₃ CD3D ₄₀ SC5D ₃₉ APOA1 ₃₉	
(109 more) ZPR1 ₁ VPS26B ₁ TMPRSS4 ₁ TMPRSS13 ₁ TAGLN ₁ SLC37A2 ₁ SIK3 ₁ POU2F3 ₁	
MPZL2 ₁ IGSF9B ₁ GRAMD1B ₁ EI24 ₁	
Chromosome 12 deletions	
12q14.3 (HMGA2) ⁺	Absent, impaired [38] or
UMIM: 618908 Names: Silver, Bussell sundrome 5: SPS5	delayed speech [39]
{V ₂₁₀ } IRAK340 WIF14F MIR65024F CNS4F MSRB340 I FMD340 CRIP10 HMCA27 CAND17	
Chromosome 13 deletions	
13g12.3 (POMP) ^{†‡}	Delayed speech [40]
OMIM: 601952	
Names: Keratosis linearis with ichthyosis congenita and sclerosing keratoderma syndrome; KLICK	
syndrome	
$\{\mathcal{V}_{N2}\}$: HMGB1 ₄₃ FLT1 ₃₆ ALOX5AP ₁₃ SLC7A1 ₁₁ HSPH1 ₁₁ B3GLCT ₅	
13q14 (RB1) [‡]	Normal speech [41]
OMIM: 613884; Cytogenetic location: 13q14 Genomic coordinates (GRCh38): 13:50,300,000–	
54,700,000 Namos: Chromosomo 13a14 dolation syndromo	
$\{V_{\lambda 12}\}$: RB1 ₀₄ FOXO1 ₀₁ GTF2F2 _{F1} KBTBD7 _{F0} SLC25A15 ₂₄ HTR2A ₂₂ TNFSF11 ₁₈ LPAR6 ₁₇	
$(V_{N2}) \cdot RD194 + O(O(191 O(101 D_{21}) RD1DD) = 0 O(22 O(100 D_{24}) O(101 D_{13}) O(101 D_{13})$	
(EDNIDE) ^{††}	Normal march [41]
OMIM: 277580	Normal speech [41]
Names: Waardenburg syndrome, Type 4A; WS4A; Waardenburg syndrome with Hirschsprung	
disease, Type 4A; Waardenburg–Shah syndrome; Shah–Waardenburg syndrome; WS4	
$\{\mathcal{V}_{N2}\}$: EDNRB ₂₆ FBXL3 ₁₀ MYCBP2 ₁	
13q33–q34 (SOX1, ARHGEF7) ⁺ ‡	Apraxia
OMIM: 619148; Cytogenetic location: 13q33–q34 Genomic coordinates (GRCh38): 13:106,400,000–	
114,364,328 Namosi Chromosomo 12,522, 524 deletion sun dreme	
Names: Chroniosome 15q55–q54 detention synarome $\{1\}_{2,2}$ (DC16ar COL 4A1a; COL 4A2aa ARHCEF7 a;	
(V_{N2}) . INS2[08 RASA552 11D1 148 1746 11045 CDC1037 COL4A136 COL4A233 ARTIGET73]	
Unromosome 14 deletions	
¹⁴ q11–q22 (Multiple incl. PAX9 , SUP116H , CHD8 , KALGAPA1] ⁺ OMIM: 613457; Cytogenetic location: 14q11–q22 Genomic coordinates (GRCh38): 14:18,200,000– 57 600 000	belayed [42] or Absent speech
Names: Chromosome 14a11-a22 deletion syndrome	
$\{\mathcal{V}_{N2}\}$: NFKBIA ₁₆₈ ADCY499 GNG2 ₈₉ SOS2 ₇₅ POLE2 ₄₇ PNP ₄₅ SLC7A7 ₄₀ SLC7A8 ₃₉ BMP4 ₃₃	
GZMB ₂₈ (163 more) ZNF219 ₁ WDHD1 ₁ TRD ₁ TEP1 ₁ RALGAPA1 ₁ PAX9 ₁ NDRG2 ₁	
MIR208A ₁ DLGAP5 ₁ CEBPE ₁ BAZ1A ₁ AKAP6 ₁	

Syndrome (Implicated Genes from Microarray Studies) **Reported Effects on Speech Other Information** Voice Chains and Their Member Chainlink Genes 14q22.1-q23.1 (Multiple incl. PTGDR, BMP4)[‡] Impaired speech [43] OMIM: 609640; Cytogenetic location: 14q22.1-q22.3 Genomic coordinates (GRCh38): 14:50,400,000-57.600.000 Names: Frias syndrome; Chromosome 14q22 deletion syndrome; growth deficiency, facial anomalies, and brachydactyly {V_{N2}}: GNG289 MNAT170 PRKCH35 BMP433 PSMA326 GNPNAT125 PELI224 PPM1A23 PYGL21 PTGER2₂₀ (31 more ...) WDHD1₁ SIX6₁ SIX1₁ DLGAP5₁ (DLK1, MEG3, RTL1) ‡ 14q32.2 Delayed speech; idiosyncratic OMIM: 608149 (paternal) / 616222 (maternal); Cytogenetic location: 14q32 Genomic coordinates speech [44] (GRCh38): 14:103,500,000-107,043,718 Names: 14q32.2 Kagami–Ogata syndrome; uniparental disomy, paternal, chromosome 14/14q32.2; Temple syndrome; uniparental disomy, maternal, chromosome 14 $\{\mathcal{V}_{N2}\}$: BDKRB2₃₂ CCNK₂₅ BDKRB1₂₄ YY1₂₁ MIR6764₂₀ CYP46A1₁₉ PAPOLA₁₈ DEGS2₁₃ EVL₁₂ WARS1₁₀ (12 more...) MIR345₁ MIR342₁ MEG3₁ **Chromosome 15 deletions** 15q11.2 (NIPA1, NIPA2, CYFIP1, TUBGCP5) ^{†‡} Delayed speech OMIM: 615656; Cytogenetic location: 15q11.2 Genomic coordinates (GRCh38): 15:20,500,000-25,500,000 Names: Burnside-Butler syndrome; 15q11.2 BP1-BP2 microdeletion $\{\mathcal{V}_{N2}\}$: CYFIP1₁₆ UBE3A₁₅ TUBGCP5₇ SNURF₇ OR4N4₆ SNRPN₅ OR4N4C₅ OR4M2₅ NIPA2₂ NIPA12 NDN2 15q11-q13 (NDN, SNRPN) ^{†‡} Delayed or Impaired speech OMIM: 176270 (for 5q11.2) Names: Prader-Willi syndrome; Prader-Lambhart-Willi syndrome; Labhart-Willi syndrome; Prader's syndrome; Prader-Labhart-Willi-Fanconi syndrome $\{\mathcal{V}_{N2}\}$: HERC2₂₂ TJP1₂₀ RYR3₁₆ GABRA5₁₆ CYFIP1₁₆ UBE3A₁₅ GABRB3₁₅ CHRNA7₁₄ GABRG3₁₃ TUBGCP5₇ (20 more ...) MTMR10₁ HMGN2P5₁ FMN1₁ (UBE3A) ^{+‡} 15q11-q13 Absent speech; impaired speech [45] OMIM: 105830 (for 15q11.2) Names: Angelman syndrome; happy puppet syndrome -Same as above-(CHRNA7, CHRFAM7A, OTUD7A)[†]‡ 15q13.3 Impaired or idiosyncratic speech OMIM: 612001; Cytogenetic location: 15q13.3 Genomic coordinates (GRCh38): 15:30,900,000-33.400.000 Names: Chromosome 15q13.3 microdeletion syndrome $\{\mathcal{V}_{N2}\}$: RYR3₁₆ **CHRNA7**₁₄ TRPM1₄ OTUD7A₄ FAN1₄ ARHGAP11A₃ MTMR10₁ FMN1₁ (Multiple incl. SIN3A) ^{+‡} 15q24 Delayed or impaired speech OMIM: 613406 Names: Witteveen-Kolk syndrome; WITKOS; chromosome 15q24 deletion syndrome (included) {*V*_{N2}}: CSK₈₉ NRG4₈₂ CYP1A2₇₉ HCN4₅₆ CYP1A1₄₈ PML₃₂ CYP11A1₃₂ SIN3A₂₅ COX5A₁₆ MPI₁₅ (28 more ...) SNX331 PTPN91 CLK31 **Chromosome 16 deletions** (SH2B1[‡], TBX6, CORO1A)^{†‡} 16p11.2 Apraxia [46]; dysarthria [47]; delayed or impaired speech OMIM: 611913; Cytogenetic location: 16p11.2 Genomic coordinates (GRCh38): 16:28,500,000-35,300,000 Names: Chromosome 16p11.2 deletion syndrome (593 kb) or 16p11.2 deletion syndrome (220 kb) $\{\mathcal{V}_{N1}\}$ SRCAP₁ {V_{N2}}: MAPK3447 ALDOA46 CD1938 ITGAM32 VKORC131 SULT1A129 ITGAL25 STX422 PYCARD₁₉ CDIPT₁₉ (60 more...) SRCAP₁ RNF40₁ PYDC1₁ ORAI3₁ MAZ₁ HIRIP3₁ AHSP₁

Syndrome (Implicated Genes from Microarray Studies) **Reported Effects on Speech Other Information** Voice Chains and Their Member Chainlink Genes 16p12.2-p11.2 (Multiple incl. SH2B1) * [76] Delayed or impaired speech OMIM: 613604; Cytogenetic location: 16p12.2-p11.2 Genomic coordinates (GRCh38): 16:21,200,000-35,300,000 Names: Chromosome 16p12.2-p11.2 deletion syndrome, 7.1 to 8.7 MB $\{\mathcal{V}_{N1}\}$ SRCAP₁ $\{V_{N2}\}$: MAPK3₄₄₇ PRKCB₁₈₄ PLK1₆₅ ALDOA₄₆ CD19₃₈ SCNN1G₃₅ SCNN1B₃₅ TNRC6A₃₃ ITGAM₃₂ VKORC1₃₁ (96 more...) XPO6₁ USP31₁ SRCAP₁ RNF40₁ PYDC1₁ ORAI3₁ MAZ₁ IGSF61 HIRIP31 CLN31 CHP21 AHSP1 (Multiple)[‡] 16p12.1 Delayed speech OMIM: 136570; Cytogenetic location: 16p12 Genomic coordinates (GRCh38): 16:24,200,000-28,500,000 Names: Chromosome 16p12.1 deletion syndrome, 520 KB $\{\mathcal{V}_{N2}\}: \texttt{TNRC6A}_{33} \texttt{IL4R}_{24} \texttt{CACNG3}_{20} \texttt{EIF3CL}_{14} \texttt{HS3ST4}_8 \texttt{GTF3C1}_8 \texttt{SLC5A11}_7 \texttt{IL27}_7 \texttt{AQP8}_7 \texttt{AQP8$ NSMCE1₅ (7 more...) XPO6₁ CLN3₁ 16p13.11 (Multiple incl. MYH11) ⁺ Delayed speech [48] OMIM: 619351 Names: 16p13.1 microdeletion predisposing to autism and/or ID; megacystis-microcolonintestinal hypoperistalsis syndrome 2; MMIHS2 $\{\mathcal{V}_{N2}\}$: ABCC1₄₁ NDE1₂₃ MIR484₂₃ MYH11₁₈ RRN3₇ ABCC6₄ (CREBBP, DNASE1, TRAP1) ^{†‡} 16p13.3 Apraxia, dysarthria, impaired or delayed speech [49] OMIM: 610543; Cytogenetic location: 16p13.3 Genomic coordinates (GRCh38): 16:0-7,800,000 Names: Chromosome 16p13.3 deletion syndrome, proximal; severe Rubinstein-Taybi syndrome (RTS); broad thumb-hallux syndrome; Rubinstein syndrome; Rubinstein-Taybi deletion syndrome; RSTS deletion syndrome $\{\mathcal{V}_{N1}\}$ **CREBBP**₁₄₀ $\{\mathcal{V}_{N2}\}$: PDPK1₁₇₃ **CREBBP**₁₄₀ ADCY9₁₀₂ TSC2₇₈ MLST8₆₉ GNG13₆₈ CACNA1H₆₀ AXIN1₄₉ UBE2I₄₄ ELOB₄₀ RPS2₃₅ (111 more...) WDR24₁ TFAP4₁ SOX8₁ SEPTIN12₁ RHBDL1₁ RBFOX1₁ PGP1 NPRL31 NAGPA1 MIR31761 HAGHL1 GNPTG1 GLIS21 FAHD11 E4F11 CHTF181 BAIAP31 16q22 (CBFB) ^{†‡} -Not available-OMIM: 614541; Cytogenetic location: 16q22 Genomic coordinates (GRCh38): 16:72,800,000-74,100,000 Names: Chromosome 16q22 deletion syndrome $\{\mathcal{V}_{N1}\}$: CMTR2₂ {V_{N2}}: CDH1₆₀ TRADD₅₀ SNTB2₅₀ SLC7A6₄₂ NFATC3₃₀ TAT₂₉ PARD6A₂₇ E2F4₂₆ NQO1₂₅ ST3GAL2₂₄ (83 more...) WWP2₁ PSKH1₁ PKD1L3₁ NUTF2₁ NOB1₁ CHTF8₁ CDH16₁ 16q24.3-q24.2 (CDH15, ZNF778, ANKRD11, ZFPM1) ** Delayed or impaired speech $\{\mathcal{V}_{N2}\}$: SLC7A5₄₈ MVD₄₂ CYBA₂₈ APRT₂₈ RPL13₂₂ CDT1₂₀ TUBB3₁₃ MC1R₁₁ DPEP1₁₀ [50]TRAPPC2L₉ (18 more...) SPIRE2₁ CHMP1A₁ **Chromosome 17 deletions** 17p11.2 (LLGL1, RAI1, UBB) ^{+‡} Delayed speech; dysarthria OMIM: 182290 Names: Smith-Magenis syndrome; chromosome 17p11.2 deletion syndrome [51] $\{\mathcal{V}_{N2}\}$: UBB₂₈₀ MAP2K3₁₀₈ MAPK7₇₂ ALDH3A1₅₀ SHMT1₄₉ ALDH3A2₄₆ SREBF1₂₇ TOP3A₂₅ MIR6778₂₀ KCNJ12₁₇ (26 more ...) (Multiple incl. KCNAB3, GUCY2D, TP53, TRAPPC1, MPDU1, CDG1F, FXR2, FMRP, 17p13.1 Absent speech EFNB3) ^{†‡} OMIM: 613776; Cytogenetic location: 17p13.1 Genomic coordinates (GRCh38): 17:6,500,000-10,800,000 Names: Chromosome 17p13.1 deletion syndrome $\{\mathcal{V}_{N1}\}$ **TP53**₂₀₆ KDM6B₈ $\{\mathcal{V}_{N2}\}$: TP53₂₀₆ ATP1B2₁₆₁ PIK3R5₁₀₄ DLG4₇₆ POLR2A₇₁ ALOX12₅₅ ALOX15B₅₁ ALOX12B₄₈ DVL245 SLC2A438 VAMP237 (63 more...) TNK11 SHBG1 MYH41 MYH131 MYH11 MIR4971 MIR3241 DNAH21

Syndrome Other Information	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Voice Chains and Their Member Chainlink Gene	S	
17p13.3 OMIM: 247200; Cytogenetic location: 17p13.3 Gen Names: Miller–Dieker lissencephaly syndrome; M (included) $\{V_{N2}\}$: CRK ₉₈ RPA1 ₈₆ YWHAE ₈₃ PAFAH1B1 ₃₄ IN	(LIS1, PAFAH1B1 , YWHAE) ^{†‡} omic coordinates (GRCh38): 17:0–3,400,000 IDLS; chromosome 17p13.3 deletion syndrome IPP5K ₁₆ ABR ₁₂ SERPINF2 ₁₀ SLC43A2 ₉ OR3A1 ₈	Delayed speech [52]
OR1G1 ₈ (27 <i>more</i>) SMYD4 ₁ SERPINF1 ₁		
17q11.2	(NF1)‡	No significant speech issues
OMIM: 162200 Names: Neurofibromatosis type I; NF1; Von Reckl eral type; Morbus–Recklinghausen	inghausen disease; neurofibromatosis, periph-	
$\{\mathcal{V}_{N2}\}$: SLC6A4 ₇₃ KSR1 ₆₆ NF1 ₆₂ NOS2 ₄₁ PSMD11 ALDOC ₁₉ (30 <i>more</i>) RAB11FIP4 ₁ MIR451A ₁ M	₃₀ VTN ₂₈ NLK ₂₇ CDK5R1 ₂₆ RPL23A ₂₂ 11R423 ₁	
17q12 OMIM: 614527; Cytogenetic location: 17q12 Geno 39,800,000	(Multiple incl. HNF1B , LHX1 , CCL3L3 , SNIP) ^{†‡} mic coordinates (GRCh38): 17:33,500,000–	Delayed or impaired speech
Names: Chromosome 17q12 deletion syndrome $\{V_{N1}\}$ ERBB2 ₁₂₄		
{ \mathcal{V}_{N2} }: CACNB1 ₁₂₉ ERBB2 ₁₂₄ AP2B1 ₄₉ PIP4K2B ₄₅ RPL23 ₂₃ (49 more) PEX12 ₁ MMP28 ₁ DUSP14 ₁	ACACA ₄₁ CCL5 ₃₀ CCL2 ₃₀ PSMB3 ₂₅ CCL4 ₂₄	
17q21.31	(KANSL1, MAPT, CRHR1) [‡]	Delayed or absent speech
OMIM: 610443		
Names: Koolen–De Vries syndrome; KDVS; chron	nosome 17q21.31 deletion syndrome; Microdele-	
tion 1/q21.31 syndrome { \mathcal{V}_{N1} }: KANSL1 ₄ BRCA1 ₆₄ { \mathcal{V}_{N2} }: ITGA2B ₉₅ BRCA1 ₆₄ MAP3K14 ₅₉ G6PC1 ₄₃	FZD2 ₃₈ WNT3 ₃₆ DUSP3 ₃₁ HDAC5 ₂₈ NSF ₂₇	
AOC3 ₂₄ (36 more) TMEM106A ₁ RND2 ₁ HEXI	M1 ₁	
17q23.1-q23.2 OMIM: 613355; Cytogenetic location: 17q23.1-q23 17:59,500,000–63,100,000	(TBX4) ^{†‡} 3.2 Genomic coordinates (GRCh38):	Delayed speech [53]
Names: Chromosome 17q23.1–q23.2 deletion synd $\{\mathcal{V}_{N2}\}$: RPS6KB1 ₉₁ CLTC ₅₄ BRIP1 ₂₄ MIR21 ₁₉ SKA	rome 2 ₁₂ MRC2 ₁₂ PTRH2 ₇ CA4 ₇ MED13 ₅ DHX40 ₄	
(5 more) BCASS ₁ AFFBF2 ₁	(Multiple incl. APCAE MADDVG COVO) ^{††}	Deleve demonstration [54]
OMIM: 135400; Cytogenetic location: 17q24.2–q24. 72,900,000	3 Genomic coordinates (GRCh38): 17:66,200,000–	Delayed speech [54]
Names: Hypertrichosis, congenital generalized, w fibromatosis, gingival, with hypertrichosis; chromo $\{\mathcal{V}_{N/2}\}$: PRKCA237 MAP2K6 131 PRKAR1A122 KCN	ith or without gingival hyperplasia; HTC3; psome 17q24.2–q24.3 deletion syndrome II271 PSMD1231 CACNG421 KPNA216 KCNI1613	
$(14 more) BPTF_1$		
Chromosome 18 deletions		
18q OMIM: 601808; Cytogenetic location: 18q Genom 80,373,285	(Multiple incl. MBP , TSHZ1) ^{†‡} ic coordinates (GRCh38): 18:18,500,000–	Delayed [55] or impaired [56] speech
Names: Chromosome 18q deletion syndrome; 18q	syndrome CA A240 NFATC1=2 PIK3C342 SMAD726	
LAMA3 ₃₅ SLC14A2 ₃₄ (112 more) SS18 ₁ SETB	$P1_1$ SERPINB4 ₁ SERPINB13 ₁ RIOK3 ₁ NETO1 ₁	
$\operatorname{MIR122}_{1}\operatorname{MBD1}_{1}\operatorname{LINC}\operatorname{ROR}_{1}\operatorname{GAREM1}_{1}\operatorname{CELF4}_{1}$		

Syndrome (Implicated Genes from Microarray Studies)	Reported Effects on Speech		
Other Information			
Voice Chains and Their Member Chainlink Genes			
Chromosome 19 deletions			
19p13.13 (Multiple incl. NFIX, MAST1, CALR) ⁺	Delayed speech; impaired		
OMIM: 613638; Cytogenetic location: 19p13.13 Genomic coordinates (GRCh38): 19:12,600,000–	speech [57]		
13,800,000			
Names: Chromosome 19p13.13 deletion syndrome			
$\{V_{N1}\}$ MAP2K2 ₂₅₇ UHKFI ₆			
$\{V_{N2}\}$: MAP2K2257 FGF22102 VAV1100 FOLK2E85 5HC281 GNG775 FIF5K1C58 G1F2F153 GNA1148			
$PSPN_{47} (121 \text{ more} \dots) ZBTB7A_1 TJP3_1 SEMA6B_1 REXO1_1 PLIN4_1 MYDGF_1 MIR7-3_1 DPP9_1$			
DAZAP1 ₁			
19q13.11 (Multiple incl. LSM14A, UBA2, WTIP, TSHZ3) ⁺	Delayed or absent speech;		
OMIM: 613026 (distal)/ 617219 (proximal); Cytogenetic location: 19q13.11 Genomic coordinates	impaired speech [58]		
(GRCh38): 19:31,900,000–35,100,000 Names: Chromosome 19q13.11 deletion syndrome, distal;			
chromosome 19q13.11 deletion syndrome, proximal			
$\{V_{N1}\}$ CEBPA ₂₃			
$\{V_{N2}\}$. SCIVID ₆₂ GF151 SEC/A739 I SIVC 4_{28} CEDIA ₂₃ SEC/A10 ₁₃ ODA2 ₈ CH3107 W H1 ₆			
KG59BP ₆ (12 more)			
Chromosome 20 deletions			
20p12.3 (Multiple incl. BMP2)	Delayed or Impaired speech		
$\{V_{N2}\}$: PLCB1 ₁₅₉ PCNA ₇₉ PLCB4 ₆₈ BMP2 ₃₃ MCM8 ₂₅ PROKR2 ₉ CRLS1 ₉ CDS2 ₉ GPCPD1 ₈	[59]		
KPS18P1 ₆ HAO1 ₆ 1KM16 ₃ LKKN4 ₁			
Chromosome 22 deletions			
22q11.2 (Multiple Incl. IBAL*, COMI*, INIT*, IOP3B*+) OMIM: 611867: Cytogenetic legation: 22g11 2 Companie goordinates (CBCh28): 22;22 100 000	Apraxia, dysarthria, delayed,		
OWING 611667; Cytogenetic location: 22q11.2 Genomic coordinates (GRCh56): 22:25,100,000–	of inipared speech [00]		
23,500,000 Names: Chromosome 22a11.2 deletion syndrome: distal chromosome 22a11.2 deletion syndrome			
{V _{MD} }: MAPK1460 CRKL68 GGT162 BIDEC ADORA2A2E COMT 22 GNAZ22 UPB124 ATP6V1E122			
$[GL_{22}] (56 more) PIWII 31 MIR1851$			
22212 2 (NE2) [±]	Delayed on immeriad an each		
$22q_{12.2}$ (INF2) + OMIM: 101000	[61]		
Name: Neurofibromatosis type II: neurofibromatosis central type: acoustic schwannomas hilat-			
eral: hilateral acoustic neurofibromatosis: BANE: acoustic neurinoma bilateral: ACN			
{Vyra}: LIE ₂ , PLA2C322 AP1B122 LIMK222 INPP5L2 OSM27 PISD24 SEI122 RNF18522 MTMR322			
(17 more) SEC1/L 2, PIK3IP1, MIR3200, DEPDC5.			
$(17 \text{ more}) \text{ SECTED 1 Roll 1 Windows (ABCA SHANK2)}^{\dagger}$	Deleveral an eleverat en elev		
22q13.5 (AK5A, SHAINK5) +	Delayed or absent speech		
OMINI: 000252 Names: Phalan McDermid sundrome: PHMDS: shromesome 22a13.3 deletion sundrome: telem			
eric 22a13 monosomy syndrome			
$\{\mathcal{V}_{N1}\}$ BRD10			
$\{\mathcal{V}_{N2}\}$: MAPK11 ₁₄₀ MAPK12 ₈₄ NUP50 ₇₉ PRR5 ₃₈ PPARA ₃₅ WNT7B ₂₁ TYMP ₂₃ HDAC10 ₁₉ CHKB ₁₉			
$ARSA_{10}$ (35 more) UPK3A1 PLXNB21 GRAMD41 CELSR11			
Chromosome X deletions			
$\begin{array}{c} \text{(Multiple incl. RP2 ZNF674)}^{++} \end{array}$	Impaired speech [62]		
OMIM: 300578; Cytogenetic location: Xp11.3 Genomic coordinates (GRCh38): X:42.500 000–	impuned speech [02]		
47,600,000			
Names: Chromosome Xp11.3 deletion syndrome; mental retardation, X-linked, with retinitis pig-			
mentosa			
$\{\mathcal{V}_{N1}\}$ KDM6A ₆			
$\{\mathcal{V}_{N2}\}$: ARAF ₁₀₆ MAOA ₇₂ MAOB ₃₇ UBA1 ₁₈ TIMP1 ₁₆ NDUFB11 ₁₀ MIR221 ₁₀ CHST7 ₁₀ MIR222 ₉			
USP11 ₈ (10 more)			

Table A1. Cont.		
Syndrome	(Implicated Genes from Microarray Studies)	Reported Effects on Speech
Other Information		1 1
Voice Chains and Their Member Chainlink Gen	es	
Xp21	(GK, DMD, NR0B1) ^{†‡}	Delayed speech [63]
OMIM: 300679; Cytogenetic location: Xp21 Geno	mic coordinates (GRCh38): X:31,500,000-	
37,800,000		
Names: Chromosome Xp21 deletion syndrome; C	omplex glycerol kinase deficiency	
$\{\mathcal{V}_{N2}\}$: TAB3 ₅₂ CYBB ₂₆ GK ₁₆ DMD ₁₂ NR0B1 ₈ XB	$K_5 ARX_2$	
Xq28 (a)	(ABCD1, BCAP31, SLC6A8) ^{†‡}	Delayed speech [64]
OMIM: 300475		
Names: Deafness, dystonia, and cerebral hypomy	elination; DDCH; contiguous ABCD1/DXS1375E	
deletion syndrome, included; CADDS, included		
$\{\mathcal{V}_{N2}\}$: IKBKG ₁₆₁ IRAK1 ₈₅ MIR718 ₅₄ DUSP9 ₅₂ F8	46 H2AB143 NSDHL35 G6PD33 FLNA28 IDH3G26	
$(54 \text{ more } \dots)$ MPP1 ₁ MIR224 ₁ MIR105-1 ₁ MAGE	A11 ₁	
Xq28 (b)	(MECP2) ^{†‡}	Normal to absent speech
OMIM: 312750		
Names: Rett syndrome; RTT; RTS autism, dement	ia, ataxia, and loss of purposeful hand use	
-same as above-		
Chromosome Y deletions		
Yq11	(USP9Y, BPY2, CDY1) ^{+‡}	Normal speech
OMIM: 415000		
Names: Spermatogenic failure, Y-linked, 2; SPGF	Y2; Sertoli-cell-only syndrome; Del Castillo syn-	
drome; germ cell aplasia, spermatogenic failure		
$\{\mathcal{V}_{N2}\}$: RPS4Y2 ₂₅ UTY ₄ NLGN4Y ₄ KDM5D ₃ CD2	4P4 ₃ TMSB4Y ₁	

Sources: †: [77], ‡: [78], †‡: From OMIM records, *: [32], **: [50].

References

- 1. Singh, R. Profiling Humans from Their Voice; Springer-Nature: Singapore, 2019.
- 2. Sataloff, R.T. Genetics of the voice. J. Voice 1995, 9, 16–19. [CrossRef] [PubMed]
- 3. Ganapathiraju, M.K.; Thahir, M.; Handen, A.; Sarkar, S.N.; Sweet, R.A.; Nimgaonkar, V.L.; Loscher, C.E.; Bauer, E.M.; Chaparala, S. Schizophrenia interactome with 504 novel protein–protein interactions. *NPJ Schizophr.* **2016**, *2*, 16012. [CrossRef] [PubMed]
- 4. Morgan, A.; Fisher, S.E.; Scheffer, I.; Hildebrand, M. FOXP2-Related Speech and Language Disorders. In *GeneReviews*[®]; University of Washington: Seattle, WA, USA, 2017.
- 5. Fisher, S.E.; Scharff, C. FOXP2 as a molecular window into speech and language. *Trends Genet.* **2009**, *25*, 166–177. [CrossRef] [PubMed]
- 6. FOXP2. Online Mendelian Inheritance in Man (OMIM) Entry 605317. Available online: https://omim.org/entry/605317 (accessed on 21 September 2022).
- Den Hoed, J.; Fisher, S.E. Genetic pathways involved in human speech disorders. *Curr. Opin. Genet. Dev.* 2020, 65, 103–111. [CrossRef]
- 8. FOXP2. The Human Protein Atlas. Available online: https://www.proteinatlas.org/ENSG00000128573-FOXP2 (accessed on 21 September 2022).
- 9. Bac, C. Investigation of Speech Delay in Individuals with 1p36 Deletion Syndrome. Ph.D. Thesis, University of Cincinnati, Cincinnati, OH, USA, 2015.
- 10. Pang, H.; Yu, X.; Kim, Y.M.; Wang, X.; Jinkins, J.K.; Yin, J.; Li, S.; Gu, H. Disorders Associated With Diverse, Recurrent Deletions and Duplications at 1q21.1. *Front. Genet.* **2020**, *11*, 577. [CrossRef]
- 11. Brazil, A.; Stanford, K.; Smolarek, T.; Hopkin, R. Delineating the phenotype of 1p36 deletion in adolescents and adults. *Am. J. Med. Genet. Part A* **2014**, *164*, 2496–2503. [CrossRef]
- 12. He, J.; Xie, Y.; Kong, S.; Qiu, W.; Wang, X.; Wang, D.; Sun, X.; Sun, D. Psychomotor retardation with a 1q42.11–q42.12 deletion. *Hereditas* **2017**, *154*, 6. [CrossRef]
- 13. Peter, B.; Matsushita, M.; Oda, K.; Raskind, W. De novo microdeletion of BCL11A is associated with severe speech sound disorder. *Am. J. Med. Genet. Part A* 2014, 164, 2091–2096. [CrossRef]
- 14. Eggermann, T.; Spengler, S.; Venghaus, A.; Denecke, B.; Zerres, K.; Baudis, M.; Ensenauer, R. 2p21 Deletions in hypotonia– cystinuria syndrome. *Eur. J. Med. Genet.* **2012**, *55*, 561–563. [CrossRef]
- 15. Chen, C.P.; Lin, S.P.; Chern, S.R.; Tsai, F.J.; Wu, P.C.; Lee, C.C.; Chen, L.F.; Lee, M.S.; Wang, W. Deletion 2q37.3 → qter and duplication 15q24.3 → qter characterized by array CGH in a girl with epilepsy and dysmorphic features. *Genet. Couns.* **2010**, 21, 263.

- Palumbo, O.; D'Agruma, L.; Minenna, A.F.; Palumbo, P.; Stallone, R.; Palladino, T.; Zelante, L.; Carella, M. 3p14.1 de novo microdeletion involving the FOXP1 gene in an adult patient with autism, severe speech delay and deficit of motor coordination. *Gene* 2013, 516, 107–113. [CrossRef]
- 17. Lowther, C.; Costain, G.; Melvin, R.; Stavropoulos, D.J.; Lionel, A.C.; Marshall, C.R.; Scherer, S.W.; Bassett, A.S. Adult expression of a 3q13.31 microdeletion. *Mol. Cytogenet.* **2014**, *7*, 23. [CrossRef]
- Van Borsel, J.; De Grande, S.; Van Buggenhout, G.; Fryns, J.P. Speech and language in Wolf-Hirschhorn syndrome: A case-study. J. Commun. Disord. 2004, 37, 21–33. [CrossRef]
- 19. Bonnet, C.; Andrieux, J.; Beri-Dexheimer, M.; Leheup, B.; Boute, O.; Manouvrier, S.; Delobel, B.; Copin, H.; Receveur, A.; Mathieu, M.; et al. Microdeletion at chromosome 4q21 defines a new emerging syndrome with marked growth restriction, mental retardation and absent or severely delayed speech. *J. Med. Genet.* **2010**, *47*, 377–384. [CrossRef]
- Tran, T.M.; Sherwood, J.K.; Doolittle, M.J.; Sathler, M.F.; Hofmann, F.; Stone-Roy, L.M.; Kim, S. Loss of cGMP-dependent protein kinase II alters ultrasonic vocalizations in mice, a model for speech impairment in human microdeletion 4q21 syndrome. *Neurosci. Lett.* 2021, 759, 136048. [CrossRef]
- Kristoffersen, K.E. Speech and language development in cri du chat syndrome: A critical review. *Clin. Linguist. Phon.* 2008, 22, 443–457. [CrossRef] [PubMed]
- Flax, J.F.; Hare, A.; Azaro, M.A.; Vieland, V.J.; Brzustowicz, L.M. Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci. J. Neurodev. Disord. 2010, 2, 210–223. [CrossRef]
- Rauch, A.; Beese, M.; Mayatepek, E.; Dörr, H.G.; Wenzel, D.; Reis, A.; Trautmann, U. A novel 5q35.3 subtelomeric deletion syndrome. *Am. J. Med. Genet. Part A* 2003, 121, 1–8. [CrossRef]
- 24. Peter, B.; Lancaster, H.; Vose, C.; Fares, A.; Schrauwen, I.; Huentelman, M. Two unrelated children with overlapping 6q25.3 deletions, motor speech disorders, and language delays. *Am. J. Med. Genet. Part A* 2017, 173, 2659–2669. [CrossRef]
- 25. Bianchi, E.; Aricŏ, M.; Podestă, A.F.; Grana, M.; Fiori, P.; Beluffi, G.; Opitz, J.M.; Reynolds, J.F. A family with the Saethre-Chotzen syndrome. *Am. J. Med. Genet.* **1985**, *22*, 649–658. [CrossRef]
- Vanzo, R.J.; Martin, M.M.; Sdano, M.R.; South, S.T. Familial KANK1 deletion that does not follow expected imprinting pattern. *Eur. J. Med. Genet.* 2013, 56, 256–259. [CrossRef] [PubMed]
- Yatsenko, S.; Cheung, S.; Scott, D.; Nowaczyk, M.; Tarnopolsky, M.; Naidu, S.; Bibat, G.; Patel, A.; Leroy, J.; Scaglia, F.; et al. Deletion 9q34.3 syndrome: Genotype-phenotype correlations and an extended deletion in a patient with features of Opitz C trigonocephaly. J. Med. Genet. 2005, 42, 328–335. [CrossRef] [PubMed]
- Samango-Sprouse, C.; Lawson, P.; Sprouse, C.; Stapleton, E.; Sadeghin, T.; Gropman, A. Expanding the phenotypic profile of Kleefstra syndrome: A female with low-average intelligence and childhood apraxia of speech. *Am. J. Med. Genet. Part A* 2016, 170, 1312–1316. [CrossRef] [PubMed]
- 29. Septer, S.; Zhang, L.; Lawson, C.E.; Cocjin, J.; Attard, T.; Ardinger, H.H. Aggressive juvenile polyposis in children with chromosome 10q23 deletion. *World J. Gastroenterol. WJG* 2013, 19, 2286. [CrossRef]
- Nishi, E.; Uehara, T.; Yanagi, K.; Hasegawa, Y.; Ueda, K.; Kaname, T.; Yamamoto, T.; Kosaki, K.; Okamoto, N. Clinical spectrum of individuals with de novo EBF3 variants or deletions. *Am. J. Med. Genet. Part A* 2021, 185, 2913–2921. [CrossRef]
- Kim, H.G.; Rosenfeld, J.A.; Scott, D.A.; Bénédicte, G.; Labonne, J.D.; Brown, J.; McGuire, M.; Mahida, S.; Naidu, S.; Gutierrez, J.; et al. Disruption of PHF21A causes syndromic intellectual disability with craniofacial anomalies, epilepsy, hypotonia, and neurobehavioral problems including autism. *Mol. Autism* 2019, 10, 35. [CrossRef]
- Xu, S.; Han, J.; Morales, A.; Menzie, C.; Williams, K.; Fan, Y.S. Characterization of 11p14-p12 deletion in WAGR syndrome by array CGH for identifying genes contributing to mental retardation and autism. *Cytogenet. Genome Res.* 2008, 122, 181–187. [CrossRef]
- Borsel, J.V.; Morlion, B.; Snick, K.V.; Leroy, J.S. Articulation in Beckwith-Wiedemann syndrome: Two case studies. Am. J. Speech-Lang. Pathol. 2000, 9, 202–213. [CrossRef]
- Kim, Y.S.; Kim, G.H.; Byeon, J.H.; Eun, S.H.; Eun, B.L. Chromosome 11q13 deletion syndrome. *Korean J. Pediatr.* 2016, 59, S10. [CrossRef]
- 35. Chilian, B.; Abdollahpour, H.; Bierhals, T.; Haltrich, I.; Fekete, G.; Nagel, I.; Rosenberger, G.; Kutsche, K. Dysfunction of SHANK2 and CHRNA7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci. *Clin. Genet.* **2013**, *84*, 560–565. [CrossRef]
- 36. Takahashi, I.; Takahashi, T.; Sawada, K.; Shimojima, K.; Yamamoto, T. Jacobsen syndrome due to an unbalanced translocation between 11q23 and 22q11.2 identified at age 40 years. *Am. J. Med. Genet. Part A* **2012**, 158, 220–223. [CrossRef]
- 37. Penny, L.A.; Dell'Aquila, M.; Jones, M.C.; Bergoffen, J.; Cunniff, C.; Fryns, J.P.; Grace, E.; Graham, J.M.; Kousseff, B.; Mattina, T.; et al. Clinical and molecular characterization of patients with distal 11q deletions. *Am. J. Hum. Genet.* **1995**, *56*, 676.
- Manolakos, E.; Orru, S.; Neroutsou, R.; Kefalas, K.; Louizou, E.; Papoulidis, I.; Thomaidis, L.; Peitsidis, P.; Sotiriou, S.; Kitsos, G.; et al. Detailed molecular and clinical investigation of a child with a partial deletion of chromosome 11 (Jacobsen syndrome). *Mol. Cytogenet.* 2009, 2, 26. [CrossRef]
- Lynch, S.A.; Foulds, N.; Thuresson, A.C.; Collins, A.L.; Annerén, G.; Hedberg, B.O.; Delaney, C.A.; Iremonger, J.; Murray, C.M.; Crolla, J.A.; et al. The 12q14 microdeletion syndrome: Six new cases confirming the role of HMGA2 in growth. *Eur. J. Hum. Genet.* 2011, 19, 534–539. [CrossRef]

- Bartholdi, D.; Stray-Pedersen, A.; Azzarello-Burri, S.; Kibaek, M.; Kirchhoff, M.; Oneda, B.; Rødningen, O.; Schmitt-Mechelke, T.; Rauch, A.; Kjaergaard, S. A newly recognized 13q12.3 microdeletion syndrome characterized by intellectual disability, microcephaly, and eczema/atopic dermatitis encompassing the HMGB1 and KATNAL1 genes. *Am. J. Med. Genet. Part A* 2014, 164, 1277–1283. [CrossRef]
- Tüysüz, B.; Collin, A.; Arapoğlu, M.; Suyugül, N. Clinical variability of Waardenburg–Shah syndrome in patients with proximal 13q deletion syndrome including the endothelin-B receptor locus. *Am. J. Med. Genet. Part A* 2009, 149, 2290–2295. [CrossRef]
- Fonseca, D.J.; Prada, C.F.; Siza, L.M.; Angel, D.; Gomez, Y.M.; Restrepo, C.M.; Douben, H.; Rivadeneira, F.; de Klein, A.; Laissue, P. A de novo 14q12q13.3 interstitial deletion in a patient affected by a severe neurodevelopmental disorder of unknown origin. *Am. J. Med. Genet. Part A* 2012, 158, 689–693. [CrossRef]
- Martínez-Fernández, M.L.; Bermejo-Sánchez, E.; Fernández, B.; MacDonald, A.; Fernández-Toral, J.; Martínez-Frías, M.L. Haploinsufficiency of BMP4 gene may be the underlying cause of Frias syndrome. *Am. J. Med. Genet. Part A* 2014, 164, 338–345. [CrossRef]
- Huang, H.; Mikami, Y.; Shigematsu, K.; Uemura, N.; Shinsaka, M.; Iwatani, A.; Miyake, F.; Kabe, K.; Takai, Y.; Saitoh, M.; et al. Kagami–Ogata syndrome in a fetus presenting with polyhydramnios, malformations, and preterm delivery: A case report. *J. Med. Case Rep.* 2019, 13, 340. [CrossRef]
- 45. Murthy, S.; Nygren, A.; El Shakankiry, H.; Schouten, J.; Al Khayat, A.; Ridha, A.; Al Ali, M. Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment. *Cytogenet. Genome Res.* 2007, 116, 135–140. [CrossRef]
- 46. Mei, C.; Fedorenko, E.; Amor, D.J.; Boys, A.; Hoeflin, C.; Carew, P.; Burgess, T.; Fisher, S.E.; Morgan, A.T. Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. *Eur. J. Hum. Genet.* **2018**, *26*, 676–686 [CrossRef] [PubMed]
- 47. Demopoulos, C.; Kothare, H.; Mizuiri, D.; Henderson-Sabes, J.; Fregeau, B.; Tjernagel, J.; Houde, J.F.; Sherr, E.H.; Nagarajan, S.S. Abnormal speech motor control in individuals with 16p11.2 deletions. *Sci. Rep.* **2018**, *8*, 1274. [CrossRef] [PubMed]
- Sahoo, T.; Theisen, A.; Rosenfeld, J.A.; Lamb, A.N.; Ravnan, J.B.; Schultz, R.A.; Torchia, B.S.; Neill, N.; Casci, I.; Bejjani, B.A.; et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet. Med.* 2011, 13, 868–880. [CrossRef] [PubMed]
- 49. Hennekam, R.C.; Baselier, A.C.; Beyaert, E.; Bos, A.; Blok, J.; Jansma, H.; Thorbecke-Nilsen, V.; Veerman, H. Psychological and speech studies in Rubinstein-Taybi syndrome. *Am. J. Ment. Retard.* **1992**, *96*, 645–660. [PubMed]
- Novara, F.; Rinaldi, B.; Sisodiya, S.M.; Coppola, A.; Giglio, S.; Stanzial, F.; Benedicenti, F.; Donaldson, A.; Andrieux, J.; Stapleton, R.; et al. Haploinsufficiency for ANKRD11-flanking genes makes the difference between KBG and 16q24.3 microdeletion syndromes: 12 new cases. *Eur. J. Hum. Genet.* 2017, 25, 694–701. [CrossRef]
- 51. Gropman, A.L.; Duncan, W.C.; Smith, A.C. Neurologic and developmental features of the Smith-Magenis syndrome (del 17p11.2). *Pediatr. Neurol.* **2006**, *34*, 337–350. [CrossRef]
- Schiff, M.; Delahaye, A.; Andrieux, J.; Sanlaville, D.; Vincent-Delorme, C.; Aboura, A.; Benzacken, B.; Bouquillon, S.; Elmaleh-Berges, M.; Labalme, A.; et al. Further delineation of the 17p13.3 microdeletion involving YWHAE but distal to PAFAH1B1: Four additional patients. *Eur. J. Med. Genet.* 2010, *53*, 303–308. [CrossRef]
- 53. Schönewolf-Greulich, B.; Ronan, A.; Ravn, K.; Baekgaard, P.; Lodahl, M.; Nielsen, K.; Rendtorff, N.D.; Tranebjaerg, L.; Brøndum-Nielsen, K.; Tümer, Z. Two new cases with microdeletion of 17q23.2 suggest presence of a candidate gene for sensorineural hearing loss within this region. *Am. J. Med. Genet. Part A* 2011, 155, 2964–2969. [CrossRef]
- 54. Vergult, S.; Dauber, A.; Delle Chiaie, B.; Van Oudenhove, E.; Simon, M.; Rihani, A.; Loeys, B.; Hirschhorn, J.; Pfotenhauer, J.; Phillips, J.A.; et al. 17q24.2 microdeletions: A new syndromal entity with intellectual disability, truncal obesity, mood swings and hallucinations. *Eur. J. Hum. Genet.* **2012**, *20*, 534–539. [CrossRef]
- 55. Cody, J.D.; Sebold, C.; Malik, A.; Heard, P.; Carter, E.; Crandall, A.; Soileau, B.; Semrud-Clikeman, M.; Cody, C.M.; Hardies, L.J.; et al. Recurrent interstitial deletions of proximal 18q: A new syndrome involving expressive speech delay. *Am. J. Med. Genet. Part* A 2007, 143, 1181–1190. [CrossRef]
- Marseglia, G.; Scordo, M.R.; Pescucci, C.; Nannetti, G.; Biagini, E.; Scandurra, V.; Gerundino, F.; Magi, A.; Benelli, M.; Torricelli, F. 372 kb Microdeletion in 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment. *Eur. J. Med. Genet.* 2012, 55, 216–221. [CrossRef]
- 57. Bonaglia, M.C.; Marelli, S.; Novara, F.; Commodaro, S.; Borgatti, R.; Minardo, G.; Memo, L.; Mangold, E.; Beri, S.; Zucca, C.; et al. Genotype–phenotype relationship in three cases with overlapping 19p13.12 microdeletions. *Eur. J. Hum. Genet.* 2010, *18*, 1302–1309. [CrossRef]
- 58. Melo, J.B.; Estevinho, A.; Saraiva, J.; Ramos, L.; Carreira, I.M. Cutis Aplasia as a clinical hallmark for the syndrome associated with 19q13.11 deletion: The possible role for UBA2 gene. *Mol. Cytogenet.* **2015**, *8*, 21. [CrossRef]
- Amasdl, S.; Natiq, A.; Sbiti, A.; Zerkaoui, M.; Lyahyai, J.; Amzazi, S.; Liehr, T.; Sefiani, A. 20p12.3 deletion is rare cause of syndromic cleft palate: Case report and review of literature. *BMC Res. Notes* 2016, 9, 5. [CrossRef]
- Solot, C.B.; Sell, D.; Mayne, A.; Baylis, A.L.; Persson, C.; Jackson, O.; McDonald-McGinn, D.M. Speech-language disorders in 22q11.2 deletion syndrome: Best practices for diagnosis and management. *Am. J. Speech-Lang. Pathol.* 2019, 28, 984–999. [CrossRef]

- 61. Davidson, T.B.; Sanchez-Lara, P.A.; Randolph, L.M.; Krieger, M.D.; Wu, S.Q.; Panigrahy, A.; Shimada, H.; Erdreich-Epstein, A. Microdeletion del (22)(q12.2) encompassing the facial development-associated gene, MN1 (meningioma 1) in a child with Pierre-Robin sequence (including cleft palate) and neurofibromatosis 2 (NF2): A case report and review of the literature. *BMC Med. Genet.* 2012, *13*, 19. [CrossRef]
- 62. Hayashi, S.; Mizuno, S.; Migita, O.; Okuyama, T.; Makita, Y.; Hata, A.; Imoto, I.; Inazawa, J. The CASK gene harbored in a deletion detected by array-CGH as a potential candidate for a gene causative of X-linked dominant mental retardation. *Am. J. Med. Genet. Part A* **2008**, *146*, 2145–2151. [CrossRef]
- 63. Fries, M.H.; Lebo, R.V.; Schonberg, S.A.; Golabi, M.; Seltzer, W.K.; Gitelman, S.E.; Golbus, M.S. Mental retardation locus in Xp21 chromosome microdeletion. *Am. J. Med. Genet.* **1993**, *46*, 363–368. [CrossRef]
- Gedeon, A.; Meinänen, M.; Ades, L.; Kääriäinen, H.; Gecz, J.; Baker, E.; Sutherland, G.; Mulley, J. Overlapping submicroscopic deletions in Xq28 in two unrelated boys with developmental disorders: Identification of a gene near FRAXE. *Am. J. Hum. Genet.* 1995, 56, 907.
- 65. National Institute on Deafness and Other Communication Disorders. Statistics on voice, speech, and language. *IEEE/ACM Trans. Audio, Speech Lang. Process.* **2008**, *25*, 2098–2111.
- 66. American Speech-Language-Hearing Association. Speech Sound Disorders: Articulation and Phonology. Practice Portal. 2017. Available online: www.asha.org/Practice-Portal/Clinical-Topics/Articulation-and-Phonology (accessed on 21 September 2022).
- 67. Shmueli, G.; Minka, T.P.; Kadane, J.B.; Borle, S.; Boatwright, P. A useful distribution for fitting discrete data: Revival of the Conway–Maxwell–Poisson distribution. J. R. Stat. Soc. Ser. C (Appl. Stat.) 2005, 54, 127–142. [CrossRef]
- 68. Sellers, K.F.; Borle, S.; Shmueli, G. The COM-Poisson model for count data: A survey of methods and applications. *Appl. Stoch. Model. Bus. Ind.* **2012**, *28*, 104–116. [CrossRef]
- 69. Huang, J.; Zhao, Y.L.; Li, Y.; Fletcher, J.A.; Xiao, S. Genomic and functional evidence for an ARID1A tumor suppressor role. *Genes Chromosom. Cancer* **2007**, *46*, 745–750. [CrossRef] [PubMed]
- Sheffels, E.; Sealover, N.E.; Theard, P.L.; Kortum, R.L. Anchorage-independent growth conditions reveal a differential SOS2 dependence for transformation and survival in RAS-mutant cancer cells. *Small GTPases* 2021, 12, 67–78. [CrossRef]
- Bertucci, F.; Borie, N.; Ginestier, C.; Groulet, A.; Charafe-Jauffret, E.; Adélaïde, J.; Geneix, J.; Bachelart, L.; Finetti, P.; Koki, A.; et al. Identification and validation of an ERBB2 gene expression signature in breast cancers. *Oncogene* 2004, 23, 2564–2575. [CrossRef]
- 72. Siggberg, L.; Olsén, P.; Näntö-Salonen, K.; Knuutila, S. 19p13.3 aberrations are associated with dysmorphic features and deviant psychomotor development. *Cytogenet. Genome Res.* 2011, 132, 8–15. [CrossRef]
- Severinsen, J.; Bjarkam, C.R.; Kiar-Larsen, S.; Olsen, I.M.; Nielsen, M.M.; Blechingberg, J.; Nielsen, A.L.; Holm, I.E.; Foldager, L.; Young, B.D.; et al. Evidence implicating BRD1 with brain development and susceptibility to both schizophrenia and bipolar affective disorder. *Mol. Psychiatry* 2006, 11, 1126–1138. [CrossRef]
- Jansen, S.; Kleefstra, T.; Willemsen, M.; De Vries, P.; Pfundt, R.; Hehir-Kwa, J.; Gilissen, C.; Veltman, J.; de Vries, B.; Vissers, L. De novo loss-of-function mutations in X-linked SMC1A cause severe ID and therapy-resistant epilepsy in females: Expanding the phenotypic spectrum. *Clin. Genet.* 2016, 90, 413–419. [CrossRef]
- 75. Porntaveetus, T.; Abid, M.F.; Theerapanon, T.; Srichomthong, C.; Ohazama, A.; Kawasaki, K.; Kawasaki, M.; Suphapeetiporn, K.; Sharpe, P.T.; Shotelersuk, V. Expanding the oro-dental and mutational spectra of Kabuki syndrome and expression of KMT2D and KDM6A in human tooth germs. *Int. J. Biol. Sci.* 2018, 14, 381. [CrossRef]
- 76. Sagi-Dain, L.; Maya, I.; Peleg, A.; Reches, A.; Banne, E.; Baris, H.N.; Tenne, T.; Singer, A.; Ben-Shachar, S. Microarray analysis in pregnancies with isolated unilateral kidney agenesis. *Pediatr. Res.* **2018**, *83*, 825–828. [CrossRef]
- Žilina, O.; Teek, R.; Tammur, P.; Kuuse, K.; Yakoreva, M.; Vaidla, E.; Mölter-Väär, T.; Reimand, T.; Kurg, A.; Õunap, K. Chromosomal microarray analysis as a first-tier clinical diagnostic test: Estonian experience. *Mol. Genet. Genom. Med.* 2014, 2,166–175. [CrossRef]
- 78. Hsu, F.; Kent, W.J.; Clawson, H.; Kuhn, R.M.; Diekhans, M.; Haussler, D. The UCSC known genes. *Bioinformatics* 2006, 22, 1036–1046. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.