Connecting human voice profiling to genomics: A predictive algorithm for linking speech phenotypes to genetic microdeletion syndromes

Rita Singh

School of Computer Science, Carnegie Mellon University, Pittsburgh, USA

Abstract

Changes in vocal acoustic patterns are known to correlate with the occurrence of several diseases and syndromes, many of which do not directly affect the structures or processes that control voice production. In such cases, it is difficult to support the existence of correlated changes in voice. This paper presents a methodology for identifying potential genomic bases for such correlations, by finding links between specific genes involved in the conditions under study, and those involved in voice, speech or language generation. Syndromes associated with chromosomal microdeletions are examined as an illustrative case, with focus on their linkage to the FOXP2 gene which has been strongly implicated in speech and language disorders. A novel path-finding graph algorithm to detect pathway chains that connect the the former to the latter is proposed. Statistical analysis of ensembles of "voice" chains detected by this algorithm indicates that they are predictive of speech phenotypes for the syndromes. Algorithmic findings are validated against clinical findings in the literature pertaining to the actual speech phenotypes that have been found to be associated with these syndromes. This methodology may also potentially be used to predict the existence of voice biomarkers in naïve cases where the existence of voice biomarkers has not already been established.

I. Introduction

Aside from diseases that affect the biological structures and processes involved in voice production, myriad other factors are known to influence voice. The ensuing changes in the voice signal are the biomarkers that give us information about the causal 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 environmental parameters from voice. However, virtually all research on voice profiling, diagnostics and biometrics is predicated on clinically observed or statistically inferred relationships between changes in voice, and the corresponding factors that are thought to be causal. 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, 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 don't, 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, explained below.

I-A. *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.

The question of whether human voice and speech production capabilities have a genetic basis has been widely explored. In the quest for identifying and delineating the genes responsible for functional speech in humans, recent studies have highlighted the importance of FOXP2, a protein-coding gene, that is thought to be involved in a variety of biological pathways and cascades that may regulate language development. It is autosomal dominant, and mutations in it cause speech and language disorders (OMIM: SPCH1). In this paper, we use FOXP2 as an exemplary case, noting that the methodology presented in its context is itself a generic one, and can be applied to any set of genes whose functions are relevant to the analysis at hand. For illustrative purposes, we proceed with the broad and simplifying assumption that any influence on speech and language is ultimately the phenotypic expression of FOXP2. The objective is now to:

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

We accomplish these goals by reconciling the genomic findings for specific syndromes resultant from genetic aberrations, to clinical observations of the speech phenotype in each case. We choose to work with chromosomal microdeletions, and suggest a method to link the genes involved in the cytogenetic locations of the microdeletions to FOXP2, as a specific example. The connections are derived using a path-search algorithm applied to a graph composed from known biological pathways and the linkages between them.

I-B. 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 – 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 on voice acoustic or quality characteristics separately. Instead, we focus exclusively on problems with speech (subsuming voice to some extent) as described under the OMIM (referring to the catalog Online Mendelian Inheritance in Man: https://www.omim.org/) category "Speech and Language disorders (SPCH1)." This category, however, is too broad and encompasses a wide range of speech problems (or speech phenotypes), such as delay 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 speech phenotypes. 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 phenotypes 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 – mostly limited to vocalizations, c) almost absent speech, with severely limited vocabulary (0-4 words).
2) **Apraxia:** Phrases referring to difficulty using language correctly while speaking, leading to speaking and communication difficulties.
3) **Delayed speech:** Phrases referring to developmental delay, or the retarded development of the ability to speak, 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 it.
5) **Idiosyncratic speech:** Phrases referring to poor conformance to cogent language, or incoherent language with articulation abnormalities.
6) **Impaired speech:** Phrases referring to poor articulation, phonation, and difficulties that result in sparse and disfluent speech.

II. Methodology

We study the syndromes resulting from chromosomal microdeletions, focusing on the phenotypic expressions of 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 correlate well to the phenotypic expression of the syndrome represented by the deletion, as supported by microarray and other studies. Although the human genome is large in comparison to a typical chromosomal microdeletion region, only a specific set of deletions are compatible with life, or fetal survival. This set is small, and continues to expand with the addition of newly discovered deletions in surviving individuals who have the means to reach genetic testing facilities. Most known deletions are thus well documented in the literature, both from the genetic and medical perspectives. The genes associated with them are largely identified through microarray techniques. The choice of microdeletion syndromes for this study is driven by the fact that deletions represent simple but selective elimination of genes from chromosomes, which are expected to be easier to interpret for their phenotypic effect. They allow us to explore a binary presence/absence causation of phenotypes with respect to genes.

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 as defined 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 brevity, going forward). For example,a pathway that signals for a cell to stop dividing when an injury to the nuclear DNA strand is being repaired may involve the co-ordinated chemical action of molecules that are formed by the transcription of multiple genes which perform different functions, and 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 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 assumption we make here is that if, for a gene, a chain from its pathway(s) (i.e., the biological pathways it contributes to) 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.

II-A. *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.
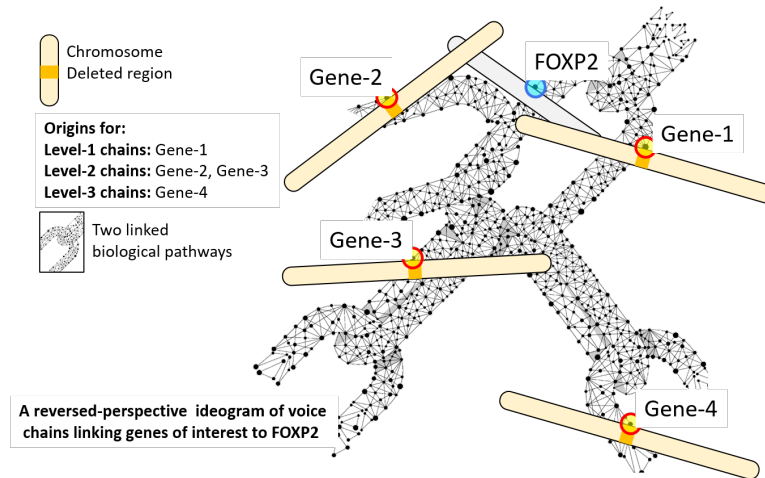
In this reversed perspective ideogram, a link depicts a pathway, the black dots on it are genes. and the chromosomes that they belong to are shown as rods where relevant. The lines connecting genes are only meant for visual clarity. Chains are formed with respect to genes on a chromosome (a microdeletion 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.

Fig. 1: Voice chains of different levels

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: $\mathcal{B} := \{g : g$ 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 $\mathcal{S}$. We represent the set of voice chains of length α as $\mathcal{V}_{N\alpha}$ (with the subscript N indicating that it is a set and not a single chain). $\mathcal{V}_{N\alpha}$ is thus the set of "level-α" chains. For the purpose of this paper, we will assume $\mathcal{S}$ to be set of genes in a microdeletion region associated with a syndrome. Thus $\mathcal{S} = \{g : g$ is a gene in the microdeletion region$\}$.

To find voice chains of the form $\mathcal{B}_V, \mathcal{B}_1, \cdots \mathcal{S}$ arising from the microdeletion region $\mathcal{S}$ (which we will refer to as "syndrome" for brevity) we introduce it in the pathway graph (Figure 2b). Voice chains are now the paths from $\mathcal{B}_V$ to $\mathcal{S}$ (Figures 2c, 2d). A

a) 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) To find voice chains, the set of genes of interest (*e.g.* from a microdeletion region) is added to the a graph as a node (shaded red). 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 yellow).

c) A Level-1 voice chain is the edge shown in red. The "chainlink" gene g_2 that links the set (red) to the voice chain is also shown.

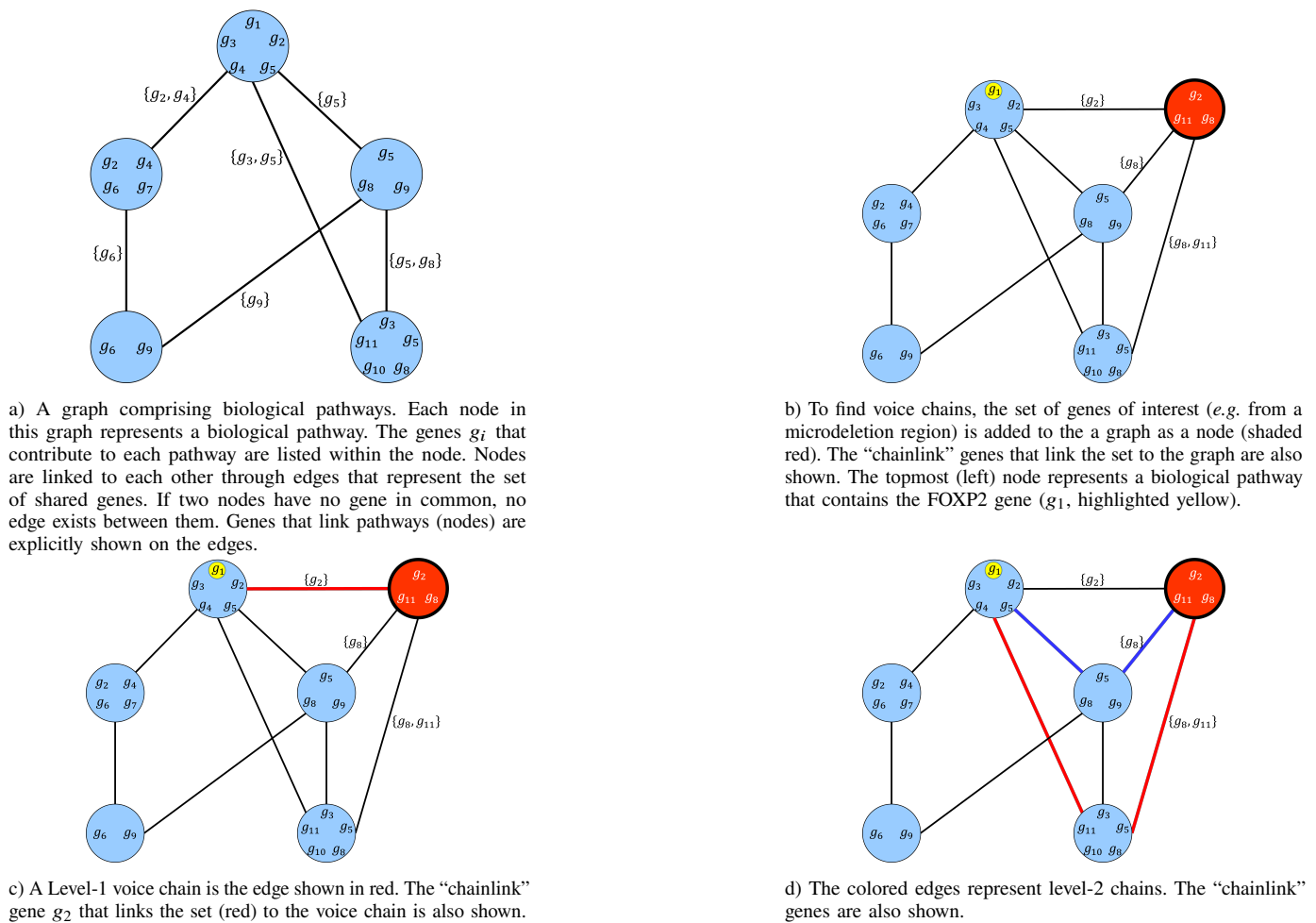d) The colored edges represent level-2 chains. The "chainlink" genes are also shown.

Fig. 2: Voice chains of different types

breadth-first algorithm described in Algorithm 1 is used to extract both, the voice chains and the genes through which the chains link to the syndrome. The outcome of the algorithm is the set of level-α voice chains $\mathcal{V}_{N\alpha}$ for $1 \le \alpha \le 2$, as well as the sets $\mathcal{G}[\mathcal{S}][\alpha]$ comprising the sets of chainlink genes from each syndrome $\mathcal{S}$ that links to voice chains of lengths α. We restrict ourselves to chains of length up to 2, since at greater lengths the chained influences cannot be disambiguated as indicated by prior studies in the (highly correlated) context of protein-protein interactomes, *e.g.* [3].

II-B. *Ensemble analysis*

In the methodology we propose, for any microdeletion region $\mathcal{S}$, we derive the set of chainlink genes within it for which level-α 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 phenotype (in 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 $\alpha = 2$.

III.   Analysis

$\mathcal{V}_{N\alpha}$, $\alpha = 1, 2$ were computed for a total of 82 microdeletion syndromes of chromosomes 1-20/22/X,Y. Genomic information including gene names were obtained from the HUGO Gene Nomenclature Committee's (HGNC) human genome database, comprising 42764 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 current literature in medical genetics and genomics, and from the Online Mendelian Inheritance in Man (OMIM) repository for authoritative information about human genes and genetic phenotypes.

Algorithm 1: Pseudocode for a breadth-first algorithm for computing the set of chainlink genes that form various levels of voice chains $\{\mathcal{V}_{N\alpha}\}$ for FOXP2.

---

1   **Input:**
2     $AllPathways$:= set of all pathways in database;
3     $AllVoicePathways$:= set of all pathways in database that include the target voice-coding gene, e.g. $FOXP2$;
4     $Syndromes$:= set of all syndromes to be analyzed;
5   N_{max} := Maximum depth of voice chains to analyze;
6   **Set:**
7     $Allchains = \{\}$ /* First-in first-out queue of pathway chains   */
     /* List of chainlink genes from syndrome S that are involved in a length l voice chain   */
8     $\mathcal{G}[S][\alpha] = \{\}$ $\forall\ S$ in $Syndromes,\ 1 \le \alpha \le N_{max}$;
9     $\mathcal{V}_{N\alpha} = \{\}$ $1 \le \alpha \le N_{max}$ /* Set of voice chains of length α */
     /* Function that returns the "neighbors" of a pathway e, i.e. the set of pathways that share genes with e */
10   **function** Neighbors(e) : {return $\{P$ in $AllPathways$ such that $e \cap P \ne \emptyset\}\}$;
11   **Start/Iterate:**
12   **for** f in $AllVoicePathways$ **do**
13        $Allchains = \{f\}$ /* Initialize queue with a chain that comprises only f itself   */
14   **while** $Not\_Empty(Allchains)$ **do**
15           $p = $ Pop($Allchains$) /* Extract earliest entry from the queue   */
16           $\alpha = $ Length(p) /* Number of pathways in pathway chain p */
17   if ($\alpha > N_{max}$) continue;
18           $e = $ Terminal_Node(p) /* The final pathway in pathway chain p */
19   **for** S in $Syndromes$ **do**
              /* Add any chainlink genes that link e and S to $L[S][l]$ */
20   **if** ($e \cap S \ne \emptyset$) **then**
21                 $\mathcal{G}[S][\alpha] = \mathcal{G}[S][\alpha] \cup (e \cap S)$;
22                 $\mathcal{V}_{N\alpha} = \mathcal{V}_{N\alpha} \cup Reverse(p)$;
23   **end**
24   **end**
          /* All neighbors of e excluding voice pathways and pathways already in p */
25           $expansions = Neighbors(e) \setminus AllVoicePathways \setminus Pathways\_in(p)$;
26   if ($expansions == \emptyset$) continue;
27   **for** n in $expansions$ **do**
28              $extension = p \odot n$ /* Concatenation   */
29   Push($Allchains,extension$) /* Enter $extension$ into the queue   */
30   **end**
31   **end**
32   **end**
33   **Output:**
34     $\mathcal{G}[S][\alpha]\forall S \in Syndromes,\ 1 \le \alpha \le N_{max}$;
35     $\mathcal{V}_{N\alpha}\ \ 1 \le \alpha \le N_{max}$;

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 I 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. The analysis was conducted for voice chains up to length 2 (Level-2). 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. The second column in Table I lists the corresponding speech phenotypes. References to the phenotypic description are cited. Where no reference is cited, the information reflects that found in the OMIM record for the syndrome. Where possible, OMIM references are used for brevity. The genes in the voice chains that have been implicated for the syndrome's phenotypic expression in prior studies are shown in bold font.

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
Chromosome 1 deletions		
1p36 — [Multiple incl. RERE, **SPEN**][†‡] OMIM: 607872; Cytogenetic location: 1p36; Genomic coordinates (GRCh38): 1:23,600,000-27,600,000 Names: Chromosome 1p36 deletion sydrome; Monosomy 1p36 syndrome $\{\mathcal{V}_{N1}\}$ ARID1A$_{10}$ $\{\mathcal{V}_{N2}\}$: PIK3CD$_{168}$ MTOR$_{146}$ GNB1$_{140}$ CDC42$_{140}$ CASP9$_{108}$ RPS6KA1$_{93}$ PRKCZ$_{75}$ CNKSR1$_{57}$ SFN$_{55}$ DVL1$_{48}$ (216 more...) SSU72$_1$ RBP7$_1$ PLEKHN1$_1$ PLEKHM2$_1$ PEX10$_1$ NBL1$_1$ MTCO3P12$_1$ MMP23B$_1$ MMP23A$_1$ HSPB7$_1$ HMGN2$_1$ GPR3$_1$ EXOSC10$_1$ DISP3$_1$ CROCCP2$_1$ CAMTA1$_1$ AHDC1$_1$ ACAP3$_1$		Delayed [9], or Absent speech
1q21.1–q21.2 — [Multiple incl. **RBM8A**[†‡], **GJAS**[†]] OMIM: 612474/274000, Cytogenetic location: 1q21.1 Genomic coordinates (GRCh38): 1:143,200,000-147,500,000 Names: Thrombocytopenia with absent radius (TAR) syndrome (OMIM 274000) $\{\mathcal{V}_{N2}\}$: H4C14$_{67}$ H4C15$_{66}$ H2BC21$_{63}$ PRKAB2$_{57}$ H3C15$_{56}$ H3C14$_{56}$ H3C13$_{56}$ H2AC20$_{53}$ H2AC19$_{53}$ H2AC18$_{53}$ (31 more...)		Delayed [10] or Impaired [11] speech
1q41–q42 — [Multiple incl. **DISP1**, HPE10, **LEFTY1**, **LEFTY2**, WDR26, TSEN2, **BPNT1**][†‡] OMIM: 612530; Cytogenetic location: 1q41-q42 Genomic 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$_{54}$ H3-3A$_{50}$ WNT3A$_{47}$ ARF1$_{46}$ PARP1$_{41}$ MIR3620$_{38}$ PSEN2$_{36}$ (50 more...) **DISP1**$_2$ MIR215$_1$ MIR194-1$_1$ DNAH14$_1$ CDC42BPA$_1$		Apraxia [12]
★1q43–q44 — [**AKT3**, ZBTB18][‡] OMIM: 612337 Names: Mental retardation, autosomal dominant 22; MRD22; Chromosome 1q43-q44 deletion syndrome (included); Chromosome 1qter deletion syndrome (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$_1$		Delayed, Impaired, or Absent speech
Chromosome 2 deletions		
★2p16.1–p15 — [Multiple incl. BCL11A][†‡] OMIM: 612513; Cytogenetic location: 2p16.1-p15 Genomic coordinates (GRCh38): 2:54,700,000-63,900,000 Names: Chromosome 2p16.1-p15 deletion syndrome $\{\mathcal{V}_{N2}\}$: RPS27A$_{293}$ XPO1$_{56}$ UGP2$_{33}$ MDH1$_{27}$ REL$_{23}$ CCT4$_{14}$ RTN4$_{13}$ B3GNT2$_{12}$ VRK2$_{10}$ USP34$_7$ (7 more...) OTX1$_1$ CCDC88A$_1$		Dysarthria, Apraxia, or Impaired speech [13]
★2p21 — [Multiple incl. **SLC3A1**, PREPL] OMIM: 606407; Cytogenetic location: 2p21 Genomic coordinates (GRCh38): 2:41,500,000-47,500,000 Names: Hypotonia-cystinuria syndrome; Cystinuria with mitochondrial disease; Homozygous 2p21 deletion syndrome $\{\mathcal{V}_{N2}\}$: CALM2$_{234}$ PRKCE$_{73}$ **SLC3A1**$_{36}$ HAAO$_{22}$ ATP6V1E2$_{19}$ ABCG5$_{18}$ MSH2$_{17}$ ABCG8$_{17}$ EPAS1$_{16}$ COX7A2L$_{15}$ (14 more...) SIX3$_1$ EML4$_1$		Apraxia or Idiosyncratic speech [14]
★2q23.1 — [MBD5][‡] OMIM: 156200 Names: Mental retardation, autosomal dominant 1; MRD1; Chromosome 2q23.1 deletion syndrome $\{\mathcal{V}_{N2}\}$: ORC4$_{29}$ KIF5C$_2$		Delayed or Impaired speech

Syndrome / Other information / Voice chains and their member Chainlink genes [Implicated genes from microarray studies]	Reported effects on speech
2q32–q33 [SATB2]‡[HOXD cluster + regulatory elements, **COL3A1 COL5A2**, **GTF3C3**, **CASP8**, **CASP10**]‡‡ **OMIM:** 612313 **Names:** Glass syndrome; GLASS; Chromosome 2q32-q33 deletion syndrome; SATB2-associated syndrome $\{\mathcal{V}_{N2}\}$: CREB1$_{203}$ STAT1$_{130}$ **CASP8**$_{103}$ NUP35$_{79}$ ITGAV$_{64}$ CD28$_{60}$ SUMO1$_{52}$ AOX1$_{50}$ FZD5$_{43}$ FZD7$_{37}$ (52 *more…*) TMEFF2$_1$ KLF7$_1$ HSPE1$_1$ DUSP19$_1$ DNAH7$_1$	Absent speech
⋆2q37.3 [**HDAC4**]‡ **OMIM:** 600430 (2q37); Cytogenetic location: 2q37 Genomic coordinates (GRCh38): 2:236,400,000-242,193,529 **Names:** Chromosome 2q37 deletion syndrome, Brachydactyly-intellectual disability syndrome; Albright hereditary osteodystrophy-like syndrome Type 3 $\{\mathcal{V}_{N2}\}$: **HDAC4**$_{33}$ GPC1$_{27}$ AGXT$_{25}$ COL6A3$_{19}$ ACKR3$_{14}$ NEU4$_{13}$ NDUFA10$_{12}$ PRLH$_{11}$ PER2$_{11}$ DTYMK$_{11}$ (28 *more…*) TWIST2$_1$ ILKAP$_1$	Impaired speech [15]
Chromosome 3 deletions	
3p13 [**FOXP1**]†‡ **OMIM:** 613670 **Names:** Mental retardation with language impairment and with or without autistic features Level$_{N2}$: MITF$_{19}$ PROK2$_8$ PPP4R2$_5$ GPR27$_5$ EIF4E3$_5$ **FOXP1**$_3$ GXYLT2$_2$ RYBP$_1$	Delayed, Idiosyncratic, Impaired speech and Dysarthria (all severe)
3q13.31 [**DRD3**, ZBTB20‡, **GAP43**, **LSAMP**]†‡ **OMIM:** 615433; Cytogenetic location: 3q13.31 Genomic coordinates (GRCh38): 3:113,700,000-117,600,000 **Names:** Chromosome 3q13.31 deletion syndrome $\{\mathcal{V}_{N2}\}$: **DRD3**$_{22}$ ATP6V1A$_{22}$ **GAP43**$_6$ QTRT2$_3$ **LSAMP**$_3$	Impaired [16] or Absent speech
3q29 [Multiple incl. **PAK2**, **DLG1**]†‡ **OMIM:** 609425; Cytogenetic location: 3q29 Genomic coordinates (GRCh38): 3:192,600,000-198,295,559 **Names:** Chromosome 3q29 deletion syndrome; Microdeletion 3q29 syndrome; 3qter deletion syndrome $\{\mathcal{V}_{N2}\}$: **PAK2**$_{74}$ **DLG1**$_{71}$ NCBP2$_{53}$ HES1$_{39}$ RPL35A$_{22}$ TFRC$_{21}$ RNF168$_{18}$ BDH1$_{17}$ PCYT1A$_{15}$ MUC20$_{13}$ (14 *more…*) FBXO45$_1$	Delayed speech
Chromosome 4 deletions	
4p16.3 [Multiple incl. **FGFR3**, MSX1]†‡ **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 syndrome; Dillan 4p syndrome $\{\mathcal{V}_{N1}\}$ CTBP1$_{25}$ $\{\mathcal{V}_{N2}\}$: **FGFR3**$_{109}$ TNIP2$_{32}$ NELFA$_{27}$ MIR943$_{26}$ CTBP1$_{25}$ ADRA2C$_{22}$ NSD2$_{19}$ GRK4$_{15}$ DGKQ$_{14}$ PDE6B$_{13}$ HAUS3$_{13}$ SLC26A1$_{12}$ SLBP$_{12}$ ATP5ME$_{12}$ RNF4$_{11}$ CPLX1$_{11}$ RGS12$_{10}$ (27 *more…*) ZFYVE28$_1$ PCGF3$_1$ NSG1$_1$ MXD4$_1$ ABCA11P$_1$	Delayed [17] or Absent speech
⋆4q21 [Multiple] **OMIM:** 613509; Cytogenetic location: 4q21 Genomic coordinates (GRCh38): 4:86,000,000-87,100,000 **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 *more…*) SHROOM3$_1$ SEPTIN11$_1$ HNRNPDL$_1$ G3BP2$_1$ BMP2K$_1$ AFF1$_1$	Delayed or Absent speech
Chromosome 5 deletions	
5p (5p15.2 and/or (5p15.3 or 5p15.33)) [Multiple incl. **TERT**, **CTNND2**]† **OMIM:** 123450 **Names:** Chromosome 5p deletion syndrome; Cri-du-chat syndrome; Cat cry syndrome; Lejeune syndrome; 5p monosomy; partial monosomy 5 $\{\mathcal{V}_{N2}\}$: 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$_1$	Delayed, Absent speech, and Apraxia [18]
⋆5q14.3 [**MEF2C**]‡ **OMIM:** 612881 (distal version 5q14.3-q15); Cytogenetic location: 5q14.3-q15 Genomic coordinates (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 $\{\mathcal{V}_{N2}\}$: RASA1$_{89}$ CCNH$_{74}$ **MEF2C**$_{70}$ POLR3G$_{17}$ COX7C$_{15}$ MIR9-2$_7$ EDIL3$_3$	Absent speech
5q33.1 [**RPS14**]†‡ **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$_1$	Dysarthria [19]

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
5q35.3 [NSD1][‡] **OMIM:** 117550 **Names:** Sotos syndrome 1; Sotos1; Nevo syndrome; Cerebral gigantism, Nevo type; Chromosome 5q35 deletion syndrome $\{\mathcal{V}_{N1}\}$ **NSD1** $\{\mathcal{V}_{N2}\}$: $MAPK9_{167}$ $LTC4S_{53}$ $F12_{33}$ $SQSTM1_{30}$ $MAML1_{29}$ $RACK1_{25}$ $FLT4_{23}$ $CANX_{22}$ $GRK6_{20}$ $GRM6_{16}$ (37 *more . . .*) $CLK4_1$		Normal [20] or Delayed speech
Chromosome 6 deletions		
6pter–p24 [Multiple incl. **FOXC1**, **GMDS**][†‡] **OMIM:** 612582; Cytogenetic location: 6pter-p24 Genomic coordinates (GRCh38): 6:0-13,400,000 **Names:** Chromosome 6pter-p24 deletion syndrome $\{\mathcal{V}_{N2}\}$: $RIPK1_{83}$ $EDN1_{43}$ $F13A1_{34}$ $TFAP2A_{19}$ DSP_{18} $BMP6_{17}$ $IRF4_{15}$ $TUBB2A_{14}$ **GMDS**$_{13}$ $ELOVL2_{13}$ (31 *more . . .*) $PAK1IP1_1$ MAK_1 $FOXQ1_1$ $DUSP22_1$ $C6orf201_1$		Delayed speech
⋆6q25.3 [Multiple incl. **ARID1B**][‡] **OMIM:** N.A. $\{\mathcal{V}_{N2}\}$: $SLC22A2_{53}$ $GTF2H5_{50}$ $ACAT2_{49}$ $SLC22A1_{24}$ EZR_{24} $SOD2_{20}$ $IGF2R_{20}$ $SYNJ2_{18}$ $SLC22A3_{16}$ $TCP1_{13}$ (11 *more . . .*)		Delayed speech, Apraxia, Dysarthria [21]
Chromosome 7 deletions		
⋆7p21 [**TWIST1** (7p21.1)][‡] **OMIM:** 101400 **Names:** Saethre-Chotzen syndrome; Acrocephalosyndactyly III; ACS3; ACS III; Chotzen syndrome; Acrocephaly, skull asymmetry, and mild syndactyly $\{\mathcal{V}_{N2}\}$: $RPA3_{74}$ $PRPS1L1_{30}$ $HDAC9_{26}$ $ITGB8_{18}$ AHR_{18} $POLR1F_{16}$ $NDUFA4_{15}$ $DGKB_{14}$ **TWIST1**$_9$ $THSD7A_4$ (8 *more . . .*) $MIOS_1$ $MEOX2_1$ $ARL4A_1$		Delayed speech [22]
7q11.23 [Multiple incl. **ELN**, **LIMK1**, **GTF2IRD1**, **GTF2I**][†‡] **OMIM:** 194050; Cytogenetic location: 7q11.23 Genomic coordinates (GRCh38): 7:72,700,000-77,900,000 **Names:** Williams syndrome; WS; WMS; Chromosome 7q11.23 deletion syndrome, 1.5- TO 1.8-MB; Williams-Beuren syndrome; WBS $\{\mathcal{V}_{N2}\}$: $YWHAG_{64}$ $POM121_{59}$ $RFC2_{57}$ $MDH2_{42}$ **LIMK1**$_{34}$ $HSPB1_{31}$ $STX1A_{30}$ $FZD9_{30}$ $NCF1_{29}$ POR_{15} (25 *more . . .*) $ABHD11_1$		Normal or Delayed speech
Chromosome 8 deletions		
8p23.1 [GATA4] **OMIM:** 179613 (not exclusive to syndrome) **Names:** Recombinant chromosome 8 syndrome; REC8 syndrome; Chromosome 8q22.1-qter duplication and 8pter-p23.1 deletion; San Luis Valley syndrome $\{\mathcal{V}_{N2}\}$: $FDFT1_{44}$ $CTSB_{17}$ BLK_{17} $TNKS_{16}$ **GATA4**$_{15}$ $AGPAT5_{14}$ $ANGPT2_{12}$ $MIR124\text{-}1_{10}$ $NEIL2_9$ $CLDN23_8$ (39 *more . . .*) $PINX1_1$ $MIR598_1$		No significant speech anomaly reports
⋆8q22.1 [**CCNE2**, TMEM67, FAM92A1] **OMIM:** 608156; Cytogenetic location: 8q22.1 Genomic coordinates (GRCh38): 8:92,300,000-97,900,000 **Names:** Nablus mask-like facial syndrome; NMLFS; Chromosome 8q22.1 deletion syndrome $\{\mathcal{V}_{N2}\}$: **CCNE2**$_{45}$ $SDC2_{34}$ $TP53INP1_{13}$ $GDF6_{12}$ $UQCRB_{11}$ $PTDSS1_9$ $ESRP1_8$ $PDP1_5$ $NDUFAF6_5$ $CDH17_4$ (3 *more . . .*)		Delayed speech
⋆8q24.11-q24.13 [TRPS1, **EXT1**][‡] **OMIM:** 150230; Cytogenetic location: 8q24.11-q24.13 Genomic coordinates (GRCh38): 8:116,700,000-126,300,000 **Names:** Langer-Giedion syndrome ; LGS; Chromosome 8q24.1 deletion syndrome; Trichorhinophalangeal syndrome type II; TRPS2 $\{\mathcal{V}_{N2}\}$: $SQLE_{43}$ $TAF2_{20}$ $RAD21_{20}$ $MIR3610_{19}$ $TNFRSF11B_{15}$ $DERL1_{13}$ $NDUFB9_{12}$ **EXT1**$_{12}$ $SLC30A8_{10}$ $FBXO32_8$ (11 *more . . .*) $WASHC5_1$ $FAM91A1_1$ $DSCC1_1$		Delayed speech
Chromosome 9 deletions		
9p24.3 [DMRT1, DMRT2][†‡] **OMIM:** 154230; Cytogenetic location: 9p24.3 Genomic coordinates (GRCh38): 9:0-2,200,000 **Names:** 46,XY sex reversal 4; SRXY4; 46,XY gonadal dysgenesis, partial or complete, with 9p24.3 deletion; Chromosome 9p24.3 deletion syndrome $\{\mathcal{V}_{N1}\}$ $SMARCA2_9$ $\{\mathcal{V}_{N2}\}$: $SMARCA2_9$ $DOCK8_3$ $WASHC1_1$		Delayed speech, Dysarthria, Apraxia [23]

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
9q34.3 OMIM: 610253 Names: Kleefstra syndrome; 9q- syndrome; 9q subtelometric deletion syndrome; Chromosome 9q34.3 deletion syndrome $\{\mathcal{V}_{N1}\}$ $NOTCH1_{62}$ $\{\mathcal{V}_{N2}\}$: $GRIN1_{139}$ $TRAF2_{78}$ $NOTCH1_{62}$ $PTGDS_{51}$ $ANAPC2_{37}$ $TUBB4B_{28}$ $NELFB_{27}$ $ENTPD8_{25}$ $CACNA1B_{20}$ $AGPAT2_{20}$ (36 *more...*) $UAP1L1_1$ $MIR602_1$ $FUT7_1$ $CYSRT1_1$	[**EHMT1**][‡]	Absent speech
Chromosome 10 deletions		
10pter–p13 or 10p14–p15.1 **OMIM:** 146255 **Names:** Barakat syndrome; Hypoparathyroidism, sensorineural deafness, and renal disease/dysplasia syndrome; HDRS; Nephrosis, nerve deafness and hypoparathyroidism $\{\mathcal{V}_{N2}\}$: $IL2RA_{83}$ $PRKCQ_{71}$ $AKR1C3_{70}$ $IDI1_{40}$ $CALML5_{36}$ $AKR1C4_{35}$ $CALML3_{33}$ **$GATA3_{27}$** $NUDT5_{22}$ $TAF3_{21}$ (39 *more...*) $ZMYND11_1$ $CELF2_1$	[Multiple incl. **GATA3**]	Sensorineural hearing loss
10q23 OMIM: 612242 (10q22.3-q23.2) Names: Chromosome 10q22.3-q23.2 deletion syndrome $\{\mathcal{V}_{N1}\}$ **$PTEN_{107}$** $\{\mathcal{V}_{N2}\}$: **$PTEN_{107}$** $CYP2C8_{100}$ $NRG3_{85}$ $CYP2C9_{78}$ FAS_{70} $CYP2C19_{66}$ $GLUD1_{46}$ $LIPA_{39}$ **$BMPR1A_{28}$** $PLCE1_{25}$ (49 *more...*) $PDLIM1_1$ $PCGF5_1$ $MMRN2_1$ $IFIT2_1$	[**PTEN**, **BMPR1A**]	Delayed or Absent speech
10q26 **OMIM:** 609625; Cytogenetic location: 10q26 Genomic coordinates (GRCh38): 10:128,800,000-133,797,422 **Names:** Chromosome 10q26 deletion syndrome; Terminal chromosome 10q26 deletion syndrome $\{\mathcal{V}_{N2}\}$: $FGFR2_{113}$ $CYP2E1_{91}$ $ECHS1_{88}$ $ACADSB_{41}$ **$DOCK1_{33}$** $BUB3_{28}$ OAT_{21} $RPL21P16_{20}$ $EIF3A_{19}$ $GRK5_{17}$ (38 *more...*) $RAB11FIP2_1$ $MKI67_1$ $EBF3_1$ $CUZD1_1$ $CALY_1$	[HMX3, **DOCK1**, C10ORF90][†‡]	Impaired speech
Chromosome 11 deletions		
11p11.2–p12 **OMIM:** 601224 (11p11.2); Cytogenetic location: 11p11.2 Genomic coordinates (GRCh38): 11:43,400,000-48,800,000 **Names:** Potocki-Shaffer syndrome; PSS; Chromosome 11p11.2 deletion syndrome: Proximal 11p deletion syndrome; DEFECT11 syndrome $\{\mathcal{V}_{N2}\}$: $TRAF6_{134}$ $NUP160_{88}$ $F2_{85}$ $PSMC3_{30}$ $NR1H3_{26}$ $CREB3L1_{26}$ $CKAP5_{25}$ $DDB2_{22}$ $SPI1_{19}$ $HSD17B12_{19}$ (36 *more...*)	[**EXT2**, ALX4][‡]	Idiosyncratic speech, Dysarthria, Delayed speech, Apraxia [24]
11p13-p12 **OMIM:** 612469; Cytogenetic location: 11p13-p12 Genomic coordinates (GRCh38): 11:31,000,000-43,400,000 **Names:** Wilms tumor, aniridia, genitourinary anomalies, mental retardation and obesity syndrome; WAGRO; WAGRO syndrome; WAGR syndrome with obesity; Chromosome 11p13-p12 deletion syndrome $\{\mathcal{V}_{N2}\}$: $TRAF6_{134}$ CAT_{37} $CD44_{27}$ $CD59_{19}$ $CSTF3_{16}$ **$SLC1A2_{14}$** $EIF3M_{14}$ $PDHX_{12}$ $WT1_{10}$ $APIP_{10}$ (13 *more...*) $PRR5L_1$ $ELF5_1$	[Multiple incl. WT1, **PAX6**, BDNF][†‡] [**SLC1A2**, **PRRG4**][⋆],	Impaired speech [25]
11p15.5 **OMIM:** 130650; **Names:** 130650: Beckwith-Wiedemann syndrome; BWS; Exomphalos-macroglossia-gigantism syndrome; EMG syndrome; Wiedemann-Beckwith syndrome; WBS $\{\mathcal{V}_{N1}\}$: $HRAS_{347}$ $\{\mathcal{V}_{N2}\}$: $HRAS_{347}$ INS_{103} $POLR2L_{85}$ $KCNQ1_{69}$ $DUSP8_{53}$ $TNNT3_{52}$ $TNNI2_{52}$ $AP2A2_{47}$ $IRF7_{39}$ TH_{30} (40 *more...*) $RIC8A_1$ $RASSF7_1$ $NLRP6_1$ $MIR483_1$ $KRTAP5-4_1$ $KRTAP5-1_1$ $DEAF1_1$ $CRACR2B_1$ $BRSK2_1$	[Multiple incl. CDKN1C, H19, **IGF2**][†‡]	Impaired speech [26]
11q13.3 **OMIM:** 166750 **Names:** Chromosome 11q13 deletion syndrome; Ododental syndrome; Otodental dysplasia and coloboma due to 11q13.3 microdeletion $\{\mathcal{V}_{N2}\}$: **$FGF4_{135}$** $CCND1_{122}$ **$FGF3_{103}$** $FGF19_{95}$ **$FADD_{73}$** $CPT1A_{38}$ $CTTN_{22}$ $PPFIA1_8$ $TPCN2_5$ $SHANK2_4$ $ANO1_3$	[**FGF4**, **FGF3**, **FADD**][†‡]	Delayed speech [27] Impaired speech [28]
11q23 **OMIM:** 188025; Cytogenetic location: 11q23 Genomic coordinates (GRCh38): 11:114,600,000-121,300,000 **Names:** Chromosome 11q23 deletion syndrome; Thrombocytopenia, Paris-Trousseau type; TCPT; Paris-Trousseau syndrome; 11q 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...*) $ZPR1_1$ $TMPRSS4_1$ $TMPRSS13_1$ $TAGLN_1$ $SIK3_1$ $POU2F3_1$ $MPZL2_1$ $C11orf52_1$	[Multiple incl. FLI1][†‡]	Reports of speech phenotype N.A.
11q23.3-q25 **OMIM:** 147791; Cytogenetic location: 11q23.3-q25 Genomic coordinates (GRCh38): 11:114,600,000-135,086,622 **Names:** Jacobsen syndrome; JBS; Del(11)(qter); Distal deletion 11q; Distal monosomy 11q; Monosomy 11qter $\{\mathcal{V}_{N2}\}$: $FXYD2_{158}$ CBL_{104} $KCNJ5_{70}$ $H2AX_{62}$ $HSPA8_{51}$ $CHEK1_{48}$ $CD3G_{43}$ $CD3D_{40}$ $SC5D_{39}$ $APOA1_{39}$ (109 *more...*) $ZPR1_1$ $VPS26B_1$ $TMPRSS4_1$ $TMPRSS13_1$ $TAGLN_1$ $SLC37A2_1$ $SIK3_1$ $POU2F3_1$ $MPZL2_1$ $IGSF9B_1$ $GRAMD1B_1$ $EI24_1$	[**FLI1**, BSX, NRGN, FRA11B, **JAM3**][†‡]	Impaired speech, Delayed speech, Apraxia [29]

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
Chromosome 12 deletions		
⋆12q14.3 **OMIM:** 618908 **Names:** Silver-Russell syndrome 5; SRS5 $\{\mathcal{V}_{N2}\}$: IRAK3$_{18}$ WIF1$_{15}$ MIR6502$_{15}$ GNS$_{15}$ MSRB3$_{12}$ LEMD3$_{10}$ GRIP1$_9$ **HMGA2**$_7$ CAND1$_7$	[**HMGA2**]†‡	Absent, Impaired [29] or Delayed speech [30]
Chromosome 13 deletions		
⋆13q12.3 **OMIM:** 601952 **Names:** Keratosis linearis with ichthyosis congenita and sclerosing keratoderma syndrome; KLICK syndrom $\{\mathcal{V}_{N2}\}$: HMGB1$_{43}$ FLT1$_{36}$ ALOX5AP$_{13}$ SLC7A1$_{11}$ HSPH1$_{11}$ B3GLCT$_5$	[POMP]†‡	Delayed speech [31]
13q14 OMIM: 613884; Cytogenetic location: 13q14 Genomic coordinates (GRCh38): 13:50,300,000-54,700,000 Names: Chromosome 13q14 deletion syndrome $\{\mathcal{V}_{N2}\}$: **RB1**$_{94}$ FOXO1$_{91}$ GTF2F2$_{51}$ KBTBD7$_{50}$ SLC25A15$_{24}$ HTR2A$_{23}$ TNFSF11$_{18}$ LPAR6$_{16}$ CYSLTR2$_{15}$ DGKH$_{14}$ (36 more...) WDFY2$_1$ ITM2B$_1$ DLEU2$_1$ DLEU1$_1$ AKAP11$_1$	[**RB1**]‡	–No information–
13q22.3 **OMIM:** 277580; **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$_1$	[**EDNRB**]†‡	Normal speech [32]
13q33–q34 OMIM: 619148; Cytogenetic location: 13q33-q34 Genomic coordinates (GRCh38): 13:106,400,000-114,364,328 Names: Chromosome 13q33-q34 deletion syndrome $\{\mathcal{V}_{N2}\}$: IRS2$_{108}$ RASA3$_{52}$ TFDP1$_{48}$ F7$_{46}$ F10$_{45}$ CDC16$_{37}$ COL4A1$_{36}$ COL4A2$_{33}$ **ARHGEF7**$_{31}$ ATP4B$_{27}$ (20 more...) ING1$_1$	[**SOX1**, **ARHGEF7**]†‡	Apraxia
Chromosome 14 deletions		
14q11–q22 OMIM: 613457; Cytogenetic location: 14q11-q22 Genomic coordinates (GRCh38): 14:18,200,000-57,600,000 **Names:** Chromosome 14q11-q22 deletion syndrome $\{\mathcal{V}_{N2}\}$: NFKBIA$_{168}$ ADCY4$_{99}$ 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$_1$	[Multiple incl. **PAX9**, **SUPT16H**, **CHD8**, **RALGAPA1**]†‡	Delayed [33] or Absent speech
14q22.1–q23.1 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, facila anomalies, and brachydactyly $\{\mathcal{V}_{N2}\}$: GNG2$_{89}$ MNAT1$_{70}$ PRKCH$_{35}$ **BMP4**$_{33}$ PSMA3$_{26}$ GNPNAT1$_{25}$ PELI2$_{24}$ PPM1A$_{23}$ PYGL$_{21}$ PTGER2$_{20}$ (31 more...) WDHD1$_1$ SIX6$_1$ SIX1$_1$ DLGAP5$_1$	[Multiple incl. **PTGDR**, **BMP4**]‡	Impaired speech [34]
14q32.2 OMIM: 608149 (paternal)/ 616222 (maternal); Cytogenetic location: 14q32 Genomic coordinates (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 $\{\mathcal{V}_{N2}\}$: BDKRB2$_{32}$ CCNK$_{25}$ BDKRB1$_{24}$ YY1$_{21}$ MIR6764$_{20}$ CYP46A1$_{19}$ PAPOLA$_{18}$ DEGS2$_{13}$ EVL$_{12}$ WARS1$_{10}$ (12 more...) MIR345$_1$ MIR342$_1$ **MEG3**$_1$	[**DLK1**, **MEG3**, RTL1]‡	Delayed speech
Chromosome 15 deletions		
⋆15q11.2 **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 $\{\mathcal{V}_{N2}\}$: **CYFIP1**$_{16}$ UBE3A$_{15}$ **TUBGCP5**$_7$ SNURF$_7$ OR4N4$_6$ SNRPN$_5$ OR4N4C$_5$ OR4M2$_5$ **NIPA2**$_2$ **NIPA1**$_2$ NDN$_2$	[**NIPA1**, **NIPA2**, **CYFIP1**, **TUBGCP5**]†‡	Delayed speech
15q11–q13 **OMIM:** 176270 (for 5q11.2) **Names:** Prader-Willi syndrome; Prader-Lambhart-Willi syndrome; Labhart-Willi syndrome; Prader's syndrome; Prader-Labhart-Willi-Fanconi syndrome $\{\mathcal{V}_{N2}\}$: HERC2$_{22}$ TJP1$_{20}$ RYR3$_{16}$ GABRA5$_{16}$ CYFIP1$_{16}$ **UBE3A**$_{15}$ GABRB3$_{15}$ CHRNA7$_{14}$ GABRG3$_{13}$ TUBGCP5$_7$ (20 more...) MTMR10$_1$ HMGN2P5$_1$ FMN1$_1$	[**NDN**, **SNRPN**]†‡	Delayed or Impaired speech
15q11–q13 **OMIM:** 105830 (for 15q11.2) **Names:** Angelman syndrome; Happy puppet syndrome –Same as above–	[**UBE3A**]†‡	Absent speech

12

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
15q13.3 [**CHRNA7**, CHRFAM7A, **OTUD7A**]†‡ OMIM: 612001; Cytogenetic location: 15q13.3 Genomic coordinates (GRCh38): 15:30,900,000-33,400,000 **Names:** Chromosome 15q13.3 microdeletion syndrome $\{\mathcal{V}_{N2}\}$: RYR3$_{16}$ **CHRNA7**$_{14}$ TRPM1$_4$ OTUD7A$_4$ FAN1$_4$ ARHGAP11A$_3$ MTMR10$_1$ FMN1$_1$		Impaired or Idiosyncratic speech
15q24 [Multiple incl. **SIN3A**]†‡ OMIM: 613406 **Names:** Witteveen-Kolk syndrome; WITKOS; Chromosome 15q24 deletion syndrome (included) $\{\mathcal{V}_{N2}\}$: CSK$_{89}$ NRG4$_{82}$ CYP1A2$_{79}$ HCN4$_{56}$ CYP1A1$_{48}$ PML$_{32}$ CYP11A1$_{32}$ **SIN3A**$_{25}$ COX5A$_{16}$ MPI$_{15}$ (28 *more...*) SNX33$_1$ PTPN9$_1$ CLK3$_1$		Delayed or Impaired speech
Chromosome 16 deletions		
16p11.2 [**SH2B1**‡, **TBX6**, **CORO1A**]†‡ 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) $\{\mathcal{V}_{N1}\}$ SRCAP$_1$ $\{\mathcal{V}_{N2}\}$: MAPK3$_{447}$ ALDOA$_{46}$ CD19$_{38}$ ITGAM$_{32}$ VKORC1$_{31}$ SULT1A1$_{29}$ ITGAL$_{25}$ STX4$_{22}$ PYCARD$_{19}$ CDIPT$_{19}$ (60 *more...*) SRCAP$_1$ RNF40$_1$ PYDC1$_1$ ORAI3$_1$ MAZ$_1$ HIRIP3$_1$ AHSP$_1$		Apraxia [35]; Dysarthria [36], Delayed or Impaired speech
16p12.2–p11.2 [Multiple incl. **SH2B1**]$^{*[37]}$ 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 $\{\mathcal{V}_{N1}\}$ SRCAP$_1$ $\{\mathcal{V}_{N2}\}$: MAPK3$_{447}$ PRKCB$_{184}$ PLK1$_{65}$ ALDOA$_{46}$ CD19$_{38}$ XPO6$_1$ USP31$_1$ SRCAP$_1$ RNF40$_1$ PYDC1$_1$ ORAI3$_1$ MAZ$_1$ IGSF6$_1$ HIRIP3$_1$ CLN3$_1$ CHP2$_1$ AHSP$_1$		Delayed or Impaired speech
★16p12.1 [Multiple]‡ OMIM: 136570; Cytogenetic location: 16p12 Genomic coordinates (GRCh38): 16:24,200,000-28,500,000 **Names:** Chromosome 16p12.1 deletion syndrome, 520-KB $\{\mathcal{V}_{N2}\}$: TNRC6A$_{33}$ IL4R$_{24}$ CACNG3$_{20}$ EIF3CL$_{14}$ HS3ST4$_8$ GTF3C1$_8$ SLC5A11$_7$ IL27$_7$ AQP8$_7$ NSMCE1$_5$ (7 *more...*) XPO6$_1$ CLN3$_1$		Delayed speech
★16p13.11 [Multiple incl. **MYH11**]† OMIM: 619351 **Names:** 16p13.1 microdeletion predisposing to autism and/or ID ; Megacystis-microcolon-intestinal hypoperistalsis syndrome 2; MMIHS2 $\{\mathcal{V}_{N2}\}$: ABCC1$_{41}$ NDE1$_{23}$ MIR484$_{23}$ **MYH11**$_{18}$ RRN3$_7$ ABCC6$_4$		Delayed speech [38]
16p13.3 [**CREBBP**, DNASE1, **TRAP1**]†‡ 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**$_{140}$ $\{\mathcal{V}_{N2}\}$: PDPK1$_{173}$ **CREBBP**$_{140}$ ADCY9$_{102}$ TSC2$_{78}$ MLST8$_{69}$ GNG13$_{68}$ CACNA1H$_{60}$ AXIN1$_{49}$ UBE2I$_{44}$ ELOB$_{40}$ RPS2$_{35}$ (111 *more...*) WDR24$_1$ TFAP4$_1$ SOX8$_1$ SEPTIN12$_1$ RHBDL1$_1$ RBFOX1$_1$ PGP$_1$ NPRL3$_1$ NAGPA$_1$ MIR3176$_1$ HAGHL$_1$ GNPTG$_1$ GLIS2$_1$ FAHD1$_1$ E4F1$_1$ CHTF18$_1$ BAIAP3$_1$		Apraxia, Dysarthria, Impaired or Delayed speech [39]
16q22 [**CBFB**]†‡ OMIM: 614541; Cytogenetic location: 16q22 Genomic coordinates (GRCh38): 16:72,800,000-74,100,000 **Names:** Chromosome 16q22 deletion syndrome $\{\mathcal{V}_{N1}\}$: CMTR2$_2$ $\{\mathcal{V}_{N2}\}$: CDH1$_{60}$ TRADD$_{50}$ SNTB2$_{50}$ SLC7A6$_{42}$ NFATC3$_{30}$ TAT$_{29}$ PARD6A$_{27}$ E2F4$_{26}$ NQO1$_{25}$ ST3GAL2$_{24}$ (83 *more...*)		–Not available–
16q24.3–q24.2 [**CDH15**, **ZNF778**, ANKRD11, **ZFPM1**]** $\{\mathcal{V}_{N2}\}$: SLC7A5$_{48}$ MVD$_{42}$ CYBA$_{28}$ APRT$_{28}$ RPL13$_{22}$ CDT1$_{20}$ TUBB3$_{13}$ MC1R$_{11}$ DPEP1$_{10}$ TRAPPC2L$_9$ (18 *more...*) SPIRE2$_1$ CHMP1A$_1$		Delayed or Impaired speech [40]
Chromosome 17 deletions		
17p11.2 [**LLGL1**, RAI1, **UBB**]†‡ OMIM: 182290 **Names:** Smith-Magenis syndrome; Chromosome 17p11.2 deletion syndrome $\{\mathcal{V}_{N2}\}$: **UBB**$_{280}$ MAP2K3$_{108}$ MAPK7$_{72}$ ALDH3A1$_{50}$ SHMT1$_{49}$ ALDH3A2$_{46}$ SREBF1$_{27}$ TOP3A$_{25}$ MIR6778$_{20}$ KCNJ12$_{17}$ (26 *more...*)		Delayed speech
17p13.1 [Multiple incl. **KCNAB3**, **GUCY2D**, **TP53**, **TRAPPC1**, **MPDU1**, CDG1F, **FXR2**, FMRP, **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**$_{206}$ KDM6B$_8$ $\{\mathcal{V}_{N2}\}$: **TP53**$_{206}$ ATP1B2$_{161}$ PIK3R5$_{104}$ DLG4$_{76}$ POLR2A$_{71}$ ALOX12$_{55}$ ALOX15B$_{51}$ ALOX12B$_{48}$ DVL2$_{45}$ SLC2A4$_{38}$ VAMP2$_{37}$ (63 *more...*) TNK1$_1$ SHBG$_1$ MYH4$_1$ MYH13$_1$ MYH1$_1$ MIR497$_1$ MIR324$_1$ DNAH2$_1$		Absent speech

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
17p13.3 [LIS1, **PAFAH1B1**, **YWHAE**]^{†‡} **OMIM:** 247200; Cytogenetic location: 17p13.3 Genomic coordinates (GRCh38): 17:0-3,400,000 **Names:** Miller-Dieker lissencephaly syndrome; MDLS; Chromosome 17p13.3 deletion syndrome (included) $\{\mathcal{V}_{N2}\}$: CRK_{98} $RPA1_{86}$ **$YWHAE_{83}$** **$PAFAH1B1_{34}$** $INPP5K_{16}$ ABR_{12} $SERPINF2_{10}$ $SLC43A2_9$ $OR3A1_8$ $OR1G1_8$ (27 more...) $SMYD4_1$ $SERPINF1_1$		Delayed speech [41]
17q11.2 [NF1][‡] **OMIM:** 162200 **Names:** Neurofibromatosis Type I; NF1; Von Recklinghausen disease; Neurofibromatosis, peripheral type; Morbus Recklinghausen $\{\mathcal{V}_{N2}\}$: $SLC6A4_{73}$ $KSR1_{66}$ **$NF1_{62}$** $NOS2_{41}$ $PSMD11_{30}$ VTN_{28} NLK_{27} $CDK5R1_{26}$ $RPL23A_{22}$ $ALDOC_{19}$ (30 more...) $RAB11FIP4_1$ $MIR451A_1$ $MIR423_1$		No significant speech issues
17q12 [Multiple incl. **HNF1B**, **LHX1**, **CCL3L3**, SNIP]^{†‡} **OMIM:** 614527; Cytogenetic location: 17q12 Genomic coordinates (GRCh38): 17:33,500,000-39,800,000 **Names:** Chromosome 17q12 deletion syndrome $\{\mathcal{V}_{N1}\}$ $ERBB2_{124}$ $\{\mathcal{V}_{N2}\}$: $CACNB1_{129}$ $ERBB2_{124}$ $AP2B1_{49}$ $PIP4K2B_{43}$ $ACACA_{41}$ $CCL5_{30}$ $CCL2_{30}$ $PSMB3_{25}$ $CCL4_{24}$ $RPL23_{23}$ (49 more...) $PEX12_1$ $MMP28_1$ $DUSP14_1$		Delayed or Impaired speech
17q21.31 [**KANSL1**, **MAPT**, **CRHR1**][‡] **OMIM:** 610443 **Names:** Koolen-De Vries syndrome; KDVS; Chromosome 17q21.31 deletion syndrome; Microdeletion 17q21.31 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...) $TMEM106A_1$ $RND2_1$ $HEXIM1_1$		Delayed or Absent speech
17q23.1-q23.2 [ABCA3, TBX4]^{†‡} **OMIM:** 613355; Cytogenetic location: 17q23.1-q23.2 Genomic coordinates (GRCh38): 17:59,500,000-63,100,000 **Names:** Chromosome 17q23.1-q23.2 deletion syndrome $\{\mathcal{V}_{N2}\}$: $SLC7A5_{48}$ MVD_{42} $CYBA_{28}$ $APRT_{28}$ $RPL13_{22}$ $CDT1_{20}$ $TUBB3_{13}$ $MC1R_{11}$ $DPEP1_{10}$ $TRAPPC2L_9$ (18 more...) $SPIRE2_1$ $CHMP1A_1$		Delayed speech [42]
17q24.3–q24.2 [Multiple incl. **ABCA5**, **MAP2K6**, **SOX9**]^{†‡} OMIM: 135400; Cytogenetic location: 17q24.2-q24.3 Genomic coordinates (GRCh38): 17:66,200,000-72,900,000 Names: Hypertrichosis, Congenital generalized, with or without gingival hyperplasia; HTC3; Fibromatosis, gingival, with hypertrichosis; Chromosome 17q24.2-q24.3 deletion syndrome $\{\mathcal{V}_{N2}\}$: $PRKCA_{237}$ **$MAP2K6_{131}$** $PRKAR1A_{122}$ $KCNJ2_{71}$ $PSMD12_{31}$ $CACNG4_{21}$ $KPNA2_{16}$ $KCNJ16_{13}$ (14 more...) $BPTF_1$		Delayed speech [43]
Chromosome 18 deletions		
18q [Multiple incl. **MBP**, TSHZ1]^{†‡} **OMIM:** 601808; Cytogenetic location: 18q Genomic coordinates (GRCh38): 18:18,500,000-80,373,285 **Names:** Chromosome 18q deletion syndrome; 18q- syndrome $\{\mathcal{V}_{N2}\}$: $BCL2_{100}$ $SMAD4_{79}$ $ROCK1_{78}$ $SMAD2_{66}$ $ACAA2_{60}$ $NFATC1_{52}$ $PIK3C3_{43}$ $SMAD7_{36}$ $LAMA3_{35}$ $SLC14A2_{34}$ (112 more...) $SS18_1$ $SETBP1_1$ $SERPINB4_1$ $SERPINB13_1$ $RIOK3_1$ $NETO1_1$ $MIR122_1$ $MBD1_1$ $LINC-ROR_1$ $GAREM1_1$ $CELF4_1$		Delayed [44] or Impaired [45] speech
Chromosome 19 deletions		
19q13.11 [Multiple incl. LSM14A, **UBA2**, **WTIP**, TSHZ3]^{†‡} **OMIM:** 613026 (distal)/ 617219 (proximal); Cytogenetic location: 19q13.11 Genomic coordinates (GRCh38): 19:31,900,000-35,100,000 **Names:** Chromosome 19q13.11 deletion syndrome, distal; Chromosome 19q13.11 deletion syndrome, proximal $\{\mathcal{V}_{N1}\}$ $CEBPA_{23}$ $\{\mathcal{V}_{N2}\}$: $SCN1B_{62}$ GPI_{51} $SLC7A9_{39}$ $PSMC4_{28}$ $CEBPA_{23}$ $SLC7A10_{13}$ **$UBA2_8$** $CHST8_7$ **$WTIP_6$** $RGS9BP_6$ (12 more...)		Delayed or Absent speech
19p13.13 [Multiple incl. NFIX, MAST1, CALR]^{†‡} **OMIM:** 613638; Cytogenetic location: 19p13.13 Genomic coordinates (GRCh38): 19:12,600,000-13,800,000 **Names:** Chromosome 19p13.13 deletion syndrome $\{\mathcal{V}_{N1}\}$ $MAP2K2_{257}$ $UHRF1_6$ $\{\mathcal{V}_{N2}\}$: $MAP2K2_{257}$ $FGF22_{102}$ $VAV1_{100}$ $POLR2E_{85}$ $SHC2_{81}$ $GNG7_{75}$ $PIP5K1C_{58}$ $GTF2F1_{53}$ $GNA11_{48}$ $PSPN_{47}$ (121 more...) $ZBTB7A_1$ $TJP3_1$ $SEMA6B_1$ $REXO1_1$ $PLIN4_1$ $MYDGF_1$ $MIR7-3_1$ $DPP9_1$ $DAZAP1_1$		Delayed speech
Chromosome 20 deletions		
⋆20p12.3 [Multiple incl. **BMP2**] $\{\mathcal{V}_{N2}\}$: $PLCB1_{159}$ $PCNA_{79}$ $PLCB4_{68}$ **$BMP2_{33}$** $MCM8_{25}$ $PROKR2_9$ $CRLS1_9$ $CDS2_9$ $GPCPD1_8$ $RPS18P1_6$ $HAO1_6$ $TRMT6_3$ $LRRN4_1$		Delayed or Impaired speech [46]
Chromosome 22 deletions		

Syndrome Other information Voice chains and their member Chainlink genes	[Implicated genes from microarray studies]	Reported effects on speech
22q11.2 [Multiple incl. **TBX1**[†], **COMT**[†], INI1[†‡], **TOP3B**[†‡]] **OMIM:** 611867; Cytogenetic location: 22q11.2 Genomic coordinates (GRCh38): 22:23,100,000-25,500,000 **Names:** Chromosome 22q11.2 deletion syndrome; Distal chromosome 22q11.2 deletion syndrome $\{\mathcal{V}_{N2}\}$: MAPK1$_{469}$ CRKL$_{68}$ GGT1$_{62}$ BID$_{56}$ ADORA2A$_{35}$ **COMT**$_{33}$ GNAZ$_{32}$ UPB1$_{24}$ ATP6V1E1$_{23}$ IGL$_{22}$ (56 *more* ...) PIWIL3$_1$ MIR185$_1$		Apraxia, Dysarthria, Delayed or Impaired speech [47]
22q12.2 [**NF2**][‡] **OMIM:** 101000 **Names:** Neurofibromatosis Type II; Neurofibromatosis, central type; Acoustic schwannomas, bilateral; Bilateral acoustic neurofibromatosis; BANF; Acoustic neurinoma, bilateral; ACN $\{\mathcal{V}_{N2}\}$: LIF$_{26}$ PLA2G3$_{20}$ AP1B1$_{20}$ LIMK2$_{19}$ INPP5J$_{18}$ OSM$_{15}$ PISD$_{14}$ SFI1$_{13}$ RNF185$_{13}$ MTMR3$_{12}$ (17 *more* ...) SEC14L2$_1$ PIK3IP1$_1$ MIR3200$_1$ DEPDC5$_1$		Delayed or Impaired speech [48]
22q13.3 [**ARSA**, **SHANK3**][‡] **OMIM:** 606232 **Names:** Phelan-McDermid syndrome; PHMDS; Chromosome 22q13.3 deletion syndrome; Telomeric 22q13 monosomy syndrome $\{\mathcal{V}_{N1}\}$ BRD1$_9$ $\{\mathcal{V}_{N2}\}$ MAPK11$_{140}$ MAPK12$_{84}$ NUP50$_{79}$ PRR5$_{38}$ PPARA$_{35}$ WNT7B$_{31}$ TYMP$_{23}$ HDAC10$_{19}$ CHKB$_{19}$ **ARSA**$_{19}$ (35 *more* ...) UPK3A$_1$ PLXNB2$_1$ GRAMD4$_1$ CELSR1$_1$		Delayed or Absent speech
Chromosome X deletions		
Xp11.3 [Multiple incl. **RP2**, ZNF674][†‡] **OMIM:** 300578; Cytogenetic location: Xp11.3 Genomic coordinates (GRCh38): X:42,500,000-47,600,000 **Names:** Chromosome Xp11.3 deletion syndrome; Mental retardation, X-linked, with retinitis pigmentosa $\{\mathcal{V}_{N1}\}$ 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* ...)		Impaired speech [49]
Xp21 [**GK**, **DMD**, **NR0B1**][†‡] **OMIM:** 300679; Cytogenetic location: Xp21 Genomic coordinates (GRCh38): X:31,500,000-37,800,000 **Names:** Chromosome Xp21 deletion syndrome; Complex glycerol kinase deficiency $\{\mathcal{V}_{N2}\}$: TAB3$_{52}$ CYBB$_{26}$ **GK**$_{16}$ **DMD**$_{12}$ **NR0B1**$_8$ XK$_5$ ARX$_2$		Delayed speech [50]
Xq28 (a) [**ABCD1**, **BCAP31**, **SLC6A8**][†‡] **OMIM:** 300475 **Names:** Deafness, dystonia, and cerebral hypomyelination; DDCH; Contiguous ABCD1/DXS1375E deletion syndrome, included; CADDS, included $\{\mathcal{V}_{N2}\}$: IKBKG$_{161}$ IRAK1$_{85}$ MIR718$_{54}$ DUSP9$_{52}$ F8$_{46}$ H2AB1$_{43}$ NSDHL$_{35}$ G6PD$_{33}$ FLNA$_{28}$ IDH3G$_{26}$ (54 *more* ...) MPP1$_1$ MIR224$_1$ MIR105-1$_1$ MAGEA11$_1$		Delayed speech [51]
Xq28 (b) [**MECP2**][†‡] **OMIM:** 312750 **Names:** Rett syndrome; RTT; RTS autism, dementia, ataxia, and loss of purposeful hand u se –same as above–		Normal to Absent speech
Chromosome Y deletions		
★Yq11 [USP9Y, BPY2, CDY1][†‡] OMIM: 415000 **Names:** Spermatogenic failure, Y-linked, 2; SPGFY2; Sertoli cell-only syndrome; Del Castillo syndrome; Germ cell aplasia, Spermatogenic failure $\{\mathcal{V}_{N2}\}$: RPS4Y2$_{25}$ UTY$_4$ NLGN4Y$_4$ KDM5D$_3$ CD24P4$_3$ TMSB4Y$_1$		Normal speech

TABLE I: Chainlink genes for Level-1 and Level-2 voice chain ensembles for 75 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.

Sources: †: [52], ‡: [53], †‡: From OMIM records, ★: [25], ∗∗: [40]

IV. Inferences

A wealth of conclusions can be drawn from Table I.

1) **Voice chains as predictors of speech phenotype:** Of the 75 syndromes analyzed, 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. In three of these

(11q23, 13q114 and 16q22) no information about the speech phenotype was found. Only the remaining 72 syndromes are considered in the analysis below.

a) The incidence of speech pathologies (including *all* forms of pathologies) among the general population is reported to be about 5% [54], and between 2.3% and 24.6% among children [55]. Of the 72 syndromes, 18 had both level-1 and level-2 voice chains, while 54 had only level-2 chains. Occurrence of speech aberrations was reported for all 18 syndromes with level-1 chains, and for all but three (13q22.2, 17q11.2, and Yq11) of the 54 syndromes with only level-2 voice chains. Thus voice chains correlate highly with the existence of speech anomalies.

In addition, we note that there were four syndromes (5q35.3, 7q11.23, 13q22.3 and X28(b)) for which studies have reported a range of phenotypes ranging from normal to completely absent speech. Only one of these syndromes (5q35.3) has a level-1 chain. However, given the inconclusiveness of the phenotypic reports, definitive conclusions cannot be made based on these syndromes. We consider these to be associated with aberrant speech for our analysis.

b) For each syndrome, studies so far have implicated some genes as being especially important, *i.e. implicated*. We find that of the 69 syndromes with reported speech anomalies, 2 do not have a specific implicated gene. These are the entries that simply say "Multiple" (4q21, 16p12.1). Of the 67 remaining that *do* have specific implicated genes, we note that in only 8 syndromes (2p16.1–p15, 2q23.1, 9p24.3, 11q23, 13q12.3, 17q23.1-q23.2, 19p13.13, Yq11), none of the implicated genes appear in the two levels of voice chains shown.

Thus, in 59 of a total of 67 cases (where implicated genes are present), the implicated genes impact FOXP2 pathways, and the corresponding phenotypes show speech anomalies. However, in 2 of the three syndromes for which no aberrant speech has been reported, implicated genes are indeed found on the voice chains. The mere presence of an implicated gene on a voice chain may not be predictive of speech anomalies by itself. This may support the hypothesis that there may be factors other than the implicated gene that influence the phenotype for the syndrome.

2) **Why are there no instances of missing voice chains?** Are voice chains redundant? The fact that there are no missing voice chains is linked in part to the size of the syndromic regions. Our database comprises 4319 unique pathways, 1205 of which are linked to the pathway that carries the FOXP2 gene (labelled 1422 in the ACC pathways database referenced earlier), and include 11746 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 1422 pathway. The shortest microdeletion considered (2q23.1) includes 9 genes, and thus 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 1422 (FOXP2) 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. More generally, any factor that influences gene function may be expected to ultimately affect voice.

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 belong to a voice chain is independent of one another.

A second hypothesis is that the *strength* of the influence could potentially be related to the *level* of the voice chain – level-1 voice chains may lead to more severe effects on speech. This is examined below.

3) **Level of voice chain and degree of influence:** 18 of the 72 syndromes considered have Level-1 voice chains, which means genes from these microdeletion regions also feature in the pathway on which FOXP2 also appears. With this, the speech phenotype is expected to be relatively more severe than for those syndromes for which Level-1 voice chains do not exist. Here we ignore the secondary effects of other phenotypes such as intellectual disability and craniofacial anomalies, a highly simplifying assumption. The syndromes with Level-1 voice chains, their corresponding genes and observed speech phenotypes are listed below:

1p36:ARID1A:Absent+Delayed

5q35.3:**NSD1**:Delayed+Normal

9q34.3:NOTCH1:Absent

11p15.5:HRAS:Impaired

16p12.2–16p11.2:SRCAP:Delayed+Impaired

16q22:CMTR2:—-

17q12:ERBB2:Delayed+Impaired

19q13.11:CEBPA:Absent+Delayed

22q13.3:BRD1:Absent+Delayed

4p16.3:CTBP1:Absent+Delayed

9p24.3:SMARCA2:Apraxia+Delayed+Dysarthria

10q23:**PTEN**:Absent+Delayed

16p11.2:SRCAP:Aprxia+Delayed+Dysarthria+Impaired

16p13.3:**CREBBP**:Apraxia+Delayed+Dysarthria+Impaired

17p13.1:KDM6B,**TP53**:Absent

17q21.31:BRCA1,**KANSL1**:Absent+Delayed

19p13.13:MAP2K2,UHRF1:Delayed

Xp11.3:KDM6A:Impaired

If we rank order the phenotypes by occurrence count, we get the following order:

Normal(1) < Dysarthria(3) = Apraxia(3) < Impaired(6) < Absent(8) < Delayed(13)

We see that together, the absence of speech, delay in speech development, and impairment of speech – all relatively stronger phenotypes – constitute the dominant set. This is consistent with our expectation that shorter voice chains to FOXP2 may have more intense effects.

Analysis of level-2 chains is more complicated. While a syndrome can only have one level-1 voice chain, it can potentially have a large number of level-2 chains. Consequently, the number of chainlink genes may be expected to be much larger for level-2 chains than for level-1 chains. The average number of chainlink genes for level-1 pathways is only 1.166, whereas that number for level-2 pathways is 45.04. As a result, although the increased length of the level-2 chains may be expected to reduce the direct influence of any gene, both, the increased number of genes and the increased number of level-2 paths through which any individual gene may connect to FOXP2 can compound this influence.

To evaluate this hypothesis, we assigned a 0-4 rating to the speech phenotypes by increasing degree of severity as follows: Normal:0, Dysarthria:1, Apraxia:2, Idiosyncratic:2, Impaired:2, Delayed:3 and Absent:4.

Each syndrome was assigned a speech score, comprising an average of the ratings of the phenotypes reported for a syndrome. We note that all 18 syndromes with level-1 chains also included level-2 chains. The remaining 54 (of 72) syndromes only had level-2 chains.

The average phenotype score for syndromes with both level-1 and level-2 chains was found to be 2.85 with a variance of 0.64. The average score for syndromes with only level-2 chains was found to be 2.40 with a variance of 0.75. A t-test shows the difference to be significant with $p < 0.05$, showing that in spite of the increased connectivity, level-2 chains have significantly lower influence than level-1 chains. Nonetheless, the increased variance of phenotypes for level-2 chains can result in serious impairment phenotypes. For instance, syndromes 14q11-q22, with 185 genes in level-2 chains and 4q21 with 47 genes in level-2 chains both are associated with absent speech. On the other hand, all of the syndromes that have no reported influence on speech also only have level-2 chains, but no level-1 chains.

However, even such an analysis would be naïve, since they do not consider the secondary effects of other phenotypes, such as intellectual disability, craniofacial anomalies, hearing disabilities, psychomotor retardation (all of which affect speech), which must be taken into account. Our current analysis is agnostic of the co-occurrence and secondary effects of other factors, which are likely to make the prediction of voice anomalies more certain, and the predicted effects stronger.

4) **Identifying candidates genes for further investigation:** Using only level-1 chains as illustrative examples, we see that voice chains can be useful in identifying candidate genes for further investigation in the context of speech phenotypes. As evidence of this, we consider the testimony of the genes that have been implicated for the syndromes through independent studies.

- 1p36 (Likely candidate (lc): ARID1A): Although not implicated for this syndrome in studies so far, ARID1A is located in 1p36.11, a region frequently deleted in human cancers [56]. Disruption in its function may lead to co-occurrence of oncological and speech phenotypes. This hypothesis is verifiable.

- 5q35.3 (lc:**NSD1**) The gene NSD1 appears in the 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 phenotype. The reason for this is not clear yet and potentially warrants investigation.

- 11p15.5 (lc: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. [57]. Its role in this syndrome needs to be re-evaluated, given its influence on 347 biological pathways and the strong speech phenotype.

- 16p11.2 and 16p12.2–16p11.2 (lc: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 (lc:KDM6B) Speech is absent in this syndrome. The gene TP53 is implicated which also appears at Level-1 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 the speech phenotype and merits investigation.

- 17q12 (lc:ERBB2) The gene ERBB2 is associated with 124 pathways. It is a well-known oncogene [58], 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 (lc:MAP2K2,UHRF1) MAP2K2 and URHF1 are not implicated. However their presence in the level-1 chain warrants investigation, especially for MAP2K2 which influences 257 pathways. Prompted by this, a literature search reveals that MAP2K2 *has* been implicated in this syndrome recently [59], although this is not on the OMIM records, which were largely consulted for this study.

- 22q13.3 (lc:BRD1) The gene BRD1 is not implicated, and appears in 9 pathways only, but the speech phenotype is

severe in this syndrome. This warrants the investigation of BRD1 independently in speech phenotypes. A literature search reveals that BRD1 is indeed strongly associated with brain development and susceptibility to both schizophrenia and bipolar affective disorder [60], and consequent effects on speech are highly likely.

- Xp11.22 (lc:SMC1A) Although SMC1A is not implicated, It appears in 33 pathways. The speech phenotypes are severe and the gene warrants investigation for this phenotype. A recent report in the literature has implicated it in severe intellectual disability and therapy-resistant epilepsy in females [61]. The former is known to be associated with severe speech anomalies.

- Xp11.3 (lc:KDM6A) Although not implicated, KDM6A warrants investigation. In the literature, it is independently known to be associated with delayed speech, and psychomotor development [62].

5) **Expression of the speech phenotype:** 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 overwhelmingly suggests that the speech phenotype may be an emergent capability, regulated and 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 co-ordination of these (and epigenetic) factors.

A more prosaic argument in favor of a phenotype being an emergent effect 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 phenotypes. The associations between speech and other phenotypes have in fact been ubiquitously observed , e.g. [38]. This supports the categorization of speech abilities as an emergent property of an ensemble of factors (including other phenotypes), rather than a primary phenotype that is likely to be expressed directly by a gene.

The emergence hypothesis is further supported by the fact that both, monosomies *and* trisomies in the case of autosomal aneuploidies have deleterious effects. This suggests the existence of an unstable equilibrium state where "normalcy" of phenotype is emergent from the correct functioning of all pathways, and disrupted by a disruption in any. Perturbation of this state from equilibrium in any sense seems to cause not only anomalies, but almost exclusively deleterious ones.

The implications of an emergence genesis for some phenotypes are important to consider: since emergent properties are not as tightly bound to the direct expression of genes, it may be possible to restore aberrant functions in these by adjustment of the overall properties of the pathway ensemble involved. In simpler words, the "cure" for a genetic condition may not necessarily only lie in restoring gene function, but may possibly also be in restoring some quantifiable equilibrium.

V. Conclusions

The hypothesis that the existence of voice chains is correlated with speech phenotypes is adequately validated by the statistical analysis presented in this paper. We also illustrate how the methodology presented can potentially provide leads to specific genes that might be candidates for further investigation in the context of speech phenotypes and microdeletion syndromes. The methodology itself may potentially be generalized and extended to reveal the potential effects of other diseases with genetic basis, and of other factors that influence gene function in some manner, on speech.

VI. Acknowledgment

References

[1] R. Singh, *Profiling humans from their voice*. Springer, 2019.

[2] R. T. Sataloff, "Genetics of the voice," *Journal of Voice*, vol. 9, no. 1, pp. 16–19, 1995.

[3] M. K. Ganapathiraju, M. Thahir, A. Handen, S. N. Sarkar, R. A. Sweet, V. L. Nimgaonkar, C. E. Loscher, E. M. Bauer, and S. Chaparala, "Schizophrenia interactome with 504 novel protein–protein interactions," *NPJ schizophrenia*, vol. 2, no. 1, pp. 1–10, 2016.

[4] A. Morgan, S. E. Fisher, I. Scheffer, and M. Hildebrand, "Foxp2-related speech and language disorders," 2017.

[5] S. E. Fisher and C. Scharff, "Foxp2 as a molecular window into speech and language," *Trends in Genetics*, vol. 25, no. 4, pp. 166–177, 2009.

[6] F. B. P. FOXP2, "Online mendelian inheritance in man (omim) entry 605317." https://omim.org/entry/605317, –.

[7] J. Den Hoed and S. E. Fisher, "Genetic pathways involved in human speech disorders," *Current Opinion in Genetics & Development*, vol. 65, pp. 103–111, 2020.

[8] FOXP2, "The human protein atlas." https://www.proteinatlas.org/ENSG00000128573-FOXP2.

[9] C. Bac, *Investigation of Speech Delay in Individuals with 1p36 Deletion Syndrome*. PhD thesis, University of Cincinnati, 2015.

[10] H. Pang, X. Yu, Y. M. Kim, X. Wang, J. K. Jinkins, J. Yin, S. Li, and H. Gu, "Disorders associated with diverse, recurrent deletions and duplications at 1q21. 1," *Frontiers in genetics*, vol. 11, p. 577, 2020.

[11] A. Brazil, K. Stanford, T. Smolarek, and R. Hopkin, "Delineating the phenotype of 1p36 deletion in adolescents and adults," *American Journal of Medical Genetics Part A*, vol. 164, no. 10, pp. 2496–2503, 2014.

[12] J. He, Y. Xie, S. Kong, W. Qiu, X. Wang, D. Wang, X. Sun, and D. Sun, "Psychomotor retardation with a 1q42. 11–q42. 12 deletion," *Hereditas*, vol. 154, no. 1, pp. 1–7, 2017.

[13] B. Peter, M. Matsushita, K. Oda, and W. Raskind, "De novo microdeletion of bcl11a is associated with severe speech sound disorder," *American journal of medical genetics Part A*, vol. 164, no. 8, pp. 2091–2096, 2014.

[14] T. Eggermann, S. Spengler, A. Venghaus, B. Denecke, K. Zerres, M. Baudis, and R. Ensenauer, "2p21 deletions in hypotonia–cystinuria syndrome," *European journal of medical genetics*, vol. 55, no. 10, pp. 561–563, 2012.

[15] C. P. Chen, S.-P. Lin, S.-R. Chern, F.-J. Tsai, P.-C. Wu, C.-C. Lee, L.-F. Chen, M.-S. Lee, and W. Wang, "Deletion 2q37. 3 [arrow right] qter and duplication 15q24. 3 [arrow right] qter characterized by array cgh in a girl with epilepsy and dysmorphic features," *Genetic Counseling*, vol. 21, no. 2, p. 263, 2010.

[16] C. Lowther, G. Costain, R. Melvin, D. J. Stavropoulos, A. C. Lionel, C. R. Marshall, S. W. Scherer, and A. S. Bassett, "Adult expression of a 3q13. 31 microdeletion," *Molecular cytogenetics*, vol. 7, no. 1, pp. 1–7, 2014.

[17] J. Van Borsel, S. De Grande, G. Van Buggenhout, and J.-P. Fryns, "Speech and language in wolf-hirschhorn syndrome: a case-study," *Journal of communication disorders*, vol. 37, no. 1, pp. 21–33, 2004.

[18] K. E. Kristoffersen, "Speech and language development in cri du chat syndrome: a critical review," *Clinical linguistics & phonetics*, vol. 22, no. 6, pp. 443–457, 2008.

[19] J. F. Flax, A. Hare, M. A. Azaro, V. J. Vieland, and L. M. Brzustowicz, "Combined linkage and linkage disequilibrium analysis of a motor speech phenotype within families ascertained for autism risk loci," *Journal of neurodevelopmental disorders*, vol. 2, no. 4, pp. 210–223, 2010.

[20] A. Rauch, M. Beese, E. Mayatepek, H.-G. Dörr, D. Wenzel, A. Reis, and U. Trautmann, "A novel 5q35. 3 subtelomeric deletion syndrome," *American Journal of Medical Genetics Part A*, vol. 121, no. 1, pp. 1–8, 2003.

[21] B. Peter, H. Lancaster, C. Vose, A. Fares, I. Schrauwen, and M. Huentelman, "Two unrelated children with overlapping 6q25. 3 deletions, motor speech disorders, and language delays," *American Journal of Medical Genetics Part A*, vol. 173, no. 10, pp. 2659–2669, 2017.

[22] E. Bianchi, M. Aric?, A. F. Podest?, M. Grana, P. Fiori, G. Beluffi, J. M. Opitz, and J. F. Reynolds, "A family with the saethre-chotzen syndrome," *American journal of medical genetics*, vol. 22, no. 4, pp. 649–658, 1985.

[23] R. J. Vanzo, M. M. Martin, M. R. Sdano, and S. T. South, "Familial kank1 deletion that does not follow expected imprinting pattern," *European journal of medical genetics*, vol. 56, no. 5, pp. 256–259, 2013.

[24] H.-G. Kim, J. A. Rosenfeld, D. A. Scott, G. Bénédicte, J. D. Labonne, J. Brown, M. McGuire, S. Mahida, S. Naidu, J. Gutierrez, *et al.*, "Disruption of phf21a causes syndromic intellectual disability with craniofacial anomalies, epilepsy, hypotonia, and neurobehavioral problems including autism," *Molecular autism*, vol. 10, no. 1, pp. 1–15, 2019.

[25] S. Xu, J. Han, A. Morales, C. Menzie, K. Williams, and Y.-S. Fan, "Characterization of 11p14-p12 deletion in wagr syndrome by array cgh for identifying genes contributing to mental retardation and autism," *Cytogenetic and genome research*, vol. 122, no. 2, pp. 181–187, 2008.

[26] J. V. Borsel, B. Morlion, K. V. Snick, and J. S. Leroy, "Articulation in beckwith-wiedemann syndrome: Two case studies," *American Journal of Speech-Language Pathology*, vol. 9, no. 3, pp. 202–213, 2000.

[27] Y.-S. Kim, G.-H. Kim, J. H. Byeon, S.-H. Eun, and B.-L. Eun, "Chromosome 11q13 deletion syndrome," *Korean journal of pediatrics*, vol. 59, no. Suppl 1, p. S10, 2016.

[28] B. Chilian, H. Abdollahpour, T. Bierhals, I. Haltrich, G. Fekete, I. Nagel, G. Rosenberger, and K. Kutsche, "Dysfunction of shank2 and chrna7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci," *Clinical genetics*, vol. 84, no. 6, pp. 560–565, 2013.

[29] E. Manolakos, S. Orru, R. Neroutsou, K. Kefalas, E. Louizou, I. Papoulidis, L. Thomaidis, P. Peitsidis, S. Sotiriou, G. Kitsos, *et al.*, "Detailed molecular and clinical investigation of a child with a partial deletion of chromosome 11 (jacobsen syndrome)," *Molecular cytogenetics*, vol. 2, no. 1, pp. 1–5, 2009.

[30] S. A. Lynch, N. Foulds, A.-C. Thuresson, A. L. Collins, G. Annerén, B.-O. Hedberg, C. A. Delaney, J. Iremonger, C. M. Murray, J. A. Crolla, *et al.*, "The 12q14 microdeletion syndrome: six new cases confirming the role of hmga2 in growth," *European Journal of Human Genetics*, vol. 19, no. 5, pp. 534–539, 2011.

[31] D. Bartholdi, A. Stray-Pedersen, S. Azzarello-Burri, M. Kibaek, M. Kirchhoff, B. Oneda, O. Rødningen, T. Schmitt-Mechelke, A. Rauch, and S. Kjaergaard, "A newly recognized 13q12. 3 microdeletion syndrome characterized by intellectual disability, microcephaly, and eczema/atopic dermatitis encompassing the hmgb1 and katnal1 genes," *American Journal of Medical Genetics Part A*, vol. 164, no. 5, pp. 1277–1283, 2014.

[32] B. Tüysüz, A. Collin, M. Arapoğlu, and N. Suyugül, "Clinical variability of waardenburg–shah syndrome in patients with proximal 13q deletion syndrome including the endothelin-b receptor locus," *American Journal of Medical Genetics Part A*, vol. 149, no. 10, pp. 2290–2295, 2009.

[33] D. J. Fonseca, C. F. Prada, L. M. Siza, D. Angel, Y. M. Gomez, C. M. Restrepo, H. Douben, F. Rivadeneira, A. de Klein, and P. Laissue, "A de novo 14q12q13. 3 interstitial deletion in a patient affected by a severe neurodevelopmental disorder of unknown origin," *American Journal of Medical Genetics Part A*, vol. 158, no. 3, pp. 689–693, 2012.

[34] M. L. Martínez-Fernández, E. Bermejo-Sánchez, B. Fernández, A. MacDonald, J. Fernández-Toral, and M. L. Martínez-Frías, "Haploinsufficiency of bmp4 gene may be the underlying cause of frias syndrome," *American Journal of Medical Genetics Part A*, vol. 164, no. 2, pp. 338–345, 2014.

[35] C. Mei, E. Fedorenko, D. J. Amor, A. Boys, C. Hoeflin, P. Carew, T. Burgess, S. E. Fisher, and A. T. Morgan, "Deep phenotyping of speech and language skills in individuals with 16p11. 2 deletion," *European Journal of Human Genetics*, vol. 26, no. 5, pp. 676–686, 2018.

[36] C. Demopoulos, H. Kothare, D. Mizuiri, J. Henderson-Sabes, B. Fregeau, J. Tjernagel, J. F. Houde, E. H. Sherr, and S. S. Nagarajan, "Abnormal speech motor control in individuals with 16p11. 2 deletions," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[37] L. Sagi-Dain, I. Maya, A. Peleg, A. Reches, E. Banne, H. N. Baris, T. Tenne, A. Singer, and S. Ben-Shachar, "Microarray analysis in pregnancies with isolated unilateral kidney agenesis," *Pediatric research*, vol. 83, no. 4, pp. 825–828, 2018.

[38] T. Sahoo, A. Theisen, J. A. Rosenfeld, A. N. Lamb, J. B. Ravnan, R. A. Schultz, B. S. Torchia, N. Neill, I. Casci, B. A. Bejjani, *et al.*, "Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems," *Genetics in Medicine*, vol. 13, no. 10, pp. 868–880, 2011.

[39] R. C. Hennekam, A. C. Baselier, E. Beyaert, A. Bos, J. Blok, H. Jansma, V. Thorbecke-Nilsen, and H. Veerman, "Psychological and speech studies in rubinstein-taybi syndrome.," *American journal on mental retardation*, 1992.

[40] F. Novara, B. Rinaldi, S. M. Sisodiya, A. Coppola, S. Giglio, F. Stanzial, F. Benedicenti, A. Donaldson, J. Andrieux, R. Stapleton, *et al.*, "Haploinsufficiency for ankrd11-flanking genes makes the difference between kbg and 16q24. 3 microdeletion syndromes: 12 new cases," *European Journal of Human Genetics*, vol. 25, no. 6, pp. 694–701, 2017.

[41] M. Schiff, A. Delahaye, J. Andrieux, D. Sanlaville, C. Vincent-Delorme, A. Aboura, B. Benzacken, S. Bouquillon, M. Elmaleh-Berges, A. Labalme, *et al.*, "Further delineation of the 17p13. 3 microdeletion involving ywhae but distal to pafah1b1: four additional patients," *European journal of medical genetics*, vol. 53, no. 5, pp. 303–308, 2010.

[42] B. Schönewolf-Greulich, A. Ronan, K. Ravn, P. Baekgaard, M. Lodahl, K. Nielsen, N. D. Rendtorff, L. Tranebjaerg, K. Brøndum-Nielsen, and Z. Tümer, "Two new cases with microdeletion of 17q23. 2 suggest presence of a candidate gene for sensorineural hearing loss within this region," *American Journal of Medical Genetics Part A*, vol. 155, no. 12, pp. 2964–2969, 2011.

[43] S. Vergult, A. Dauber, B. Delle Chiaie, E. Van Oudenhove, M. Simon, A. Rihani, B. Loeys, J. Hirschhorn, J. Pfotenhauer, J. A. Phillips, *et al.*, "17q24. 2 microdeletions: a new syndromal entity with intellectual disability, truncal obesity, mood swings and hallucinations," *European Journal of Human Genetics*, vol. 20, no. 5, pp. 534–539, 2012.

[44] J. D. Cody, C. Sebold, A. Malik, P. Heard, E. Carter, A. Crandall, B. Soileau, M. Semrud-Clikeman, C. M. Cody, L. J. Hardies, *et al.*, "Recurrent interstitial deletions of proximal 18q: a new syndrome involving expressive speech delay," *American Journal of Medical Genetics Part A*, vol. 143, no. 11, pp. 1181–1190, 2007.

[45] G. Marseglia, M. R. Scordo, C. Pescucci, G. Nannetti, E. Biagini, V. Scandurra, F. Gerundino, A. Magi, M. Benelli, and F. Torricelli, "372 kb microdeletion in 18q12. 3 causing setbp1 haploinsufficiency associated with mild mental retardation and expressive speech impairment," *European journal of medical genetics*, vol. 55, no. 3, pp. 216–221, 2012.

[46] S. Amasdl, A. Natiq, A. Sbiti, M. Zerkaoui, J. Lyahyai, S. Amzazi, T. Liehr, and A. Sefiani, "20p12. 3 deletion is rare cause of syndromic cleft palate: case report and review of literature," *BMC research notes*, vol. 9, no. 1, pp. 1–4, 2016.

[47] C. B. Solot, D. Sell, A. Mayne, A. L. Baylis, C. Persson, O. Jackson, and D. M. McDonald-McGinn, "Speech-language disorders in 22q11. 2 deletion syndrome: Best practices for diagnosis and management," *American journal of speech-language pathology*, vol. 28, no. 3, pp. 984–999, 2019.

[48] T. B. Davidson, P. A. Sanchez-Lara, L. M. Randolph, M. D. Krieger, S.-Q. Wu, A. Panigrahy, H. Shimada, and A. Erdreich-Epstein, "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 medical genetics*, vol. 13, no. 1, pp. 1–8, 2012.

[49] S. Hayashi, S. Mizuno, O. Migita, T. Okuyama, Y. Makita, A. Hata, I. Imoto, and J. Inazawa, "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," *American Journal of Medical Genetics Part A*, vol. 146, no. 16, pp. 2145–2151, 2008.

[50] M. H. Fries, R. V. Lebo, S. A. Schonberg, M. Golabi, W. K. Seltzer, S. E. Gitelman, and M. S. Golbus, "Mental retardation locus in xp21 chromosome microdeletion," *American journal of medical genetics*, vol. 46, no. 4, pp. 363–368, 1993.

[51] A. Gedeon, M. Meinänen, L. Ades, H. Kääriäinen, J. Gecz, E. Baker, G. Sutherland, and J. Mulley, "Overlapping submicroscopic deletions in xq28 in two unrelated boys with developmental disorders: identification of a gene near fraxe.," *American journal of human genetics*, vol. 56, no. 4, p. 907, 1995.

[52] O. Žilina, R. Teek, P. Tammur, M. Kuuse, M. Yakoreva, E. Vaidla, T. Mölter-Väär, T. Reimand, A. Kurg, and K. Õunap, "Chromosomal microarray analysis as a first-tier clinical diagnostic test: E stonian experience," *Molecular genetics & genomic medicine*, vol. 2, no. 2, pp. 166–175, 2014.

[53] F. Hsu, W. J. Kent, H. Clawson, R. M. Kuhn, M. Diekhans, and D. Haussler, "The ucsc known genes," *Bioinformatics*, vol. 22, no. 9, pp. 1036–1046, 2006.

[54] N. I. on Deafness and O. C. Disorders, "Statistics on voice, speech, and language," 2008.

[55] A. S.-L.-H. Association *et al.*, "Speech sound disorders: articulation and phonology," *Practice Portal). Retrieved from www. asha. org/Practice-Portal/Clinical-Topics/Articulation-and-Phonology*, 2017.

[56] J. Huang, Y.-L. Zhao, Y. Li, J. A. Fletcher, and S. Xiao, "Genomic and functional evidence for an arid1a tumor suppressor role," *Genes, Chromosomes and Cancer*, vol. 46, no. 8, pp. 745–750, 2007.

[57] E. Sheffels, N. E. Sealover, P. L. Theard, and R. L. Kortum, "Anchorage-independent growth conditions reveal a differential sos2 dependence for transformation and survival in ras-mutant cancer cells," *Small GTPases*, vol. 12, no. 1, pp. 67–78, 2021.

[58] F. Bertucci, N. Borie, C. Ginestier, A. Groulet, E. Charafe-Jauffret, J. Adélaïde, J. Geneix, L. Bachelart, P. Finetti, A. Koki, *et al.*, "Identification and validation of an erbb2 gene expression signature in breast cancers," *Oncogene*, vol. 23, no. 14, pp. 2564–2575, 2004.

[59] L. Siggberg, P. Olsén, K. Näntö-Salonen, and S. Knuutila, "19p13. 3 aberrations are associated with dysmorphic features and deviant psychomotor development," *Cytogenetic and genome research*, vol. 132, no. 1-2, pp. 8–15, 2011.

[60] J. Severinsen, C. R. Bjarkam, S. Kiar-Larsen, I. M. Olsen, M. M. Nielsen, J. Blechingberg, A. L. Nielsen, I. E. Holm, L. Foldager, B. D. Young, *et al.*, "Evidence implicating brd1 with brain development and susceptibility to both schizophrenia and bipolar affective disorder," *Molecular psychiatry*, vol. 11, no. 12, pp. 1126–1138, 2006.

[61] S. Jansen, T. Kleefstra, M. Willemsen, P. De Vries, R. Pfundt, J. Hehir-Kwa, C. Gilissen, J. Veltman, B. de Vries, and L. Vissers, "De novo loss-of-function mutations in x-linked smc1a cause severe id and therapy-resistant epilepsy in females: expanding the phenotypic spectrum," *Clinical genetics*, vol. 90, no. 5, pp. 413–419, 2016.

[62] T. Porntaveetus, M. F. Abid, T. Theerapanon, C. Srichomthong, A. Ohazama, K. Kawasaki, M. Kawasaki, K. Suphapeetiporn, P. T. Sharpe, and V. Shotelersuk, "Expanding the oro-dental and mutational spectra of kabuki syndrome and expression of kmt2d and kdm6a in human tooth germs," *International journal of biological sciences*, vol. 14, no. 4, p. 381, 2018.