1 Introduction

Learning to imitate sounds and the rules of grammar endows humans with the unique ability to communicate infinite meaning with a finite vocal repertoire using language. Although language is learned, a genetic bias towards this learning has already been proposed in Charles Darwin´s "Descent of Man" (Chapter III - Comparison of the Mental Powers of Man and the Lower Animals):

" […] language is an art, like brewing or baking; but writing would have been a better simile. It certainly is not a true instinct, for every language has to be learnt. It differs, however, widely from all ordinary arts, for man has an instinctive tendency to speak, as we see in the babble of our young children; whilst no child has an instinctive tendency to brew, bake, or write."

A modern account of the idea that learning language is not solely based upon experience was put forward by the Linguist Noam Chomsky. He developed the concept of a "universal grammar", which posits the existence of a universal set of rules common to all languages (Chomsky, 1957). This universal grammar shared by all languages suggests that some aspects of how language is learned are determined by intrinsic, genetically defined structural and functional characteristics of the human brain. The first example of a gene possibly contributing to such a genetic predisposition for language was provided by the discovery that disruptions of the FoxP2 gene cause developmental verbal dyspraxia (DVD). Individuals suffering from this speech and language disorder have severe difficulties with articulation and show impaired receptive and cognitive language skills. Although recent theoretical work also puts forward the idea that the universality of certain syntactic rules might just be the by-product of the scale-free network architecture of languages (i Cancho et al., 2005; Nowak et al., 2001), the case of FoxP2 obviously allows to take a closer look on the development and function of neural circuits associated with language from a molecular and cellular perspective.

1.1 FoxP2 and Developmental Verbal Dyspraxia

The causative link between FoxP2 and DVD was established when genomic alterations of FoxP2 were identified in all 16 affected members of the british KE-family and an unrelated

individual with a remarkably similar pathophenotype. Affected KE family members carry a substitution of arginine to histidine (R553H), which most likely renders the protein non-functional (Figure 1.1). This mutation is inherited in a dominant fashion and was found in KE DVD patients across three generations. In the unrelated individual FoxP2 is disrupted by a balanced translocation (Lai et al., 2001). The direct search for FoxP2 mutations in DVD patient panels meanwhile revealed more individuals with a disrupted FoxP2 allele (Feuk et al., 2006; MacDermot et al., 2005).

What is the common behavioral phenotype of individuals with DVD? Affected members of the KE family have severe difficulty in correctly articulating speech. In both word and the non-word repetition tests, where subjects have to repeat words (e.g. *killer*) and non-words (e.g. *rillek*) after hearing them, DVD patients score significantly worse than their unaffected family members (Watkins et al., 2002). The impairment increases gradually with the complexity of the words to be articulated. The DVD family members also have difficulties in the volitional control of skilled non-speech orofacial movements, a symptome called orofacial dyspraxia. Importantly, these difficulties cannot be attributed to a general impairment of motor control, since the patients´ limb praxis performance is indistinguishable from unaffected individuals (Watkins et al., 2002). The patients are also not impaired in their hearing ability. Interestingly, the DVD phenotype resembles that observed in patients with Broca´s aphasia (reviewed in (Damasio and Geschwind, 1984). However, there are important behavioral differences between the two pathologies. Aphasics perform better in the word than the non-word repetition test, whereas affected KE family perform equally bad in both tests. This could indicate that despite their actual problems with articulation, aphasics had learned to associate word articulation patterns with word meanings before the onset of the aphasia, which might help them finding the correct words. In contrast, affected KE family that never learned the correct word articulation patterns would fail in using word meaning to solve the word-repetition task.

In addition to the verbal and orofacial dyspraxia, KE family patients perform significantly worse than their unaffected relatives on tests that assess receptive and grammatical language. The deficit includes the inability to correctly inflect words (i.e. change tense or number) or to match sentences describing subtle relationships between objects with the corresponding pictures. Nevertheless, affected individuals score only slightly, but significantly lower on a non-verbal IQ-test than non-affected individuals and there is

considerable overlap between the groups (Alcock et al., 2000; Vargha-Khadem et al., 1998; Watkins et al., 2002). Taken together, these findings suggest that the primary deficit in the affected KE family members reflects a disruption of the sensorimotor mechanisms mediating the selection, control, and sequencing of learned fine movements of the mouth and face. An open question remains, if the receptive cognitive problems result from the primary articulation problem or if they constitute a second independent core deficit of the disorder. The first possibility would be consistent with the motor theory of speech perception (Liberman and Mattingly, 1985), which posits that decoding speech requires the brain circuitry involved in its production. Although recent human studies support this concept (Fadiga et al., 2002; Watkins et al., 2003), the possibility that aberrations of FoxP2 affect the development of grammatic skills independent of the articulation deficit cannot be ruled out.

First insights into the neural basis of the behavioral abnormalities shown by DVD patients came from the examination of affected and unaffected KE family members with structural and functional brain imaging techniques. Affected KE family members displayed bilateral structural deficits consisting of a reduction in the gray matter density of the caudate nucleus in the basal ganglia (Vargha-Khadem et al., 1998; Watkins et al., 2002) the ventral cerebellum (Belton et al., 2003) and Broca´s area. Abnormally high gray matter density was found in the putamen and Wernicke´s area. Interestingly, the volume of the caudate correlated well with the performance in the test of oral praxis (see above; (Watkins et al., 2002), indicating its involvement in the pathology. Given the well-established role of the basal ganglia in motor planning and sequencing (Graybiel, 1995), the structural abnormalities in the striatal regions of the basal ganglia (caudate and putamen) are generally consistent with an impaired control of orofacial motor function. However, it is less clear how they specifically compromise orofacial movements, without affecting other motor functions.

Functional imaging during the performance in covert (silent) and overt (spoken) tasks revealed lateralized disturbances in language-impaired subjects. In contrast to the typical left-dominant activation pattern involving Broca´s Area that is elicited by a verb generation test in unaffected KE family members, the signal distribution in affected individuals is more bilateral. Extensive bilateralization in the activation pattern was also observed for DVD subjects in the word repetition tasks described above. Consistent with the

morphological findings, an underactivation of Broca´s area and the putamen occurred in the affected family KE members (Liegeois et al., 2003). The observed overactivation of areas normally not involved in language has been interpreted to result from compensatory recruitment of additional brain areas, increased attention or a higher cognitive effort to solve the task. Taken together, the imaging work points to the frontostriatal and frontocerebellar networks as key circuitry affected in impaired KE family members.

1.2 FoxP2 Expression in the Brains of Mice and Men

Mapping the expression of FoxP2 in human and murine brains with *in situ* hybridization and immunohistochemistry has established where mammalian FoxP2 acts (Ferland et al., 2003; Lai et al., 2003; Takahashi et al., 2003). In adulthood, most prominent FoxP2 expression is found in the basal ganglia, in regions of the thalamus that receive input from the basal ganglia, in midbrain visual processing regions and in the inferior olive of the medulla. Further regions expressing high levels of FoxP2 include the cerebellar Purkinje cells, deep cerebellar nuclei, sensory auditory midbrain structures and layer VI neurons of the cerebral cortex (Ferland et al., 2003; Lai et al., 2003). Fetal FoxP2 expression is consistent with the adult expression pattern. In the rodent telencephalon, initial expression of FoxP2 is largely limited to the lateral ganglionic eminence [LGE (Ferland et al., 2003; Takahashi et al., 2003)], the mammalian subpallial germinal zone that gives rise to the striatal projection neurons of the basal ganglia and to the majority of cortical interneurons (Brazel et al., 2003). Within the LGE, FoxP2 is expressed in the subventricular zone and mantle region but not in the proliferative ventricular zone, suggesting that expression is initiated in postmitotic neurons. This interpretation is also compatible with the additional expression site in the non-proliferative cortical plate of the developing cortex (Ferland et al., 2003; Takahashi et al., 2003). Taken together, the FoxP2 expression pattern is consistent with the sites of pathology identified in affected KE family members by brain imaging techniques. However, the question whether the reduction of functional FoxP2 protein affects the function of speech-related neural circuits as a consequence of their improper development, or by means of disturbed neural transmission or both remains unanswered, due to the purely descriptive nature of gene expression mapping.

1.3 Molecular Function of FoxP2

From the molecular perspective, FoxP2 belongs to the large family of winged helix transcription factors that are characterized by a conserved Forkhead box (Fox) DNA-binding domain. The forkhead box binds to distinct sequences in promoter regions of a specific set of target genes, allowing their transcriptional regulation. Fox proteins affect cell fate and differentiation in various tissues, and mutations cause developmental disorders (Lehmann et al., 2003). The common feature in all individuals with speech abnormalities caused by genomic alteration of FoxP2 seems to be a reduction of functional FoxP2 protein by 50%. This haploinsuffiency results from the introduction of a premature stop codon in one patient (MacDermot et al., 2005), the disruption of the gene by a translocation in another patient or a substitution of arginine to histidine (R553H) in the DNA binding domain. All affected members of the KE family in which the speech phenotype was originally described (Lai et al., 2001) carry the R553H mutation. Homology modeling of the FoxP2 forkhead domain structure in conjunction with electrostatic charge calculations predict a net reduction in positive charge on the DNA-binding surface of the R553H mutation, sufficient to disrupt DNA-binding (Banerjee-Basu and Baxevanis, 2004).

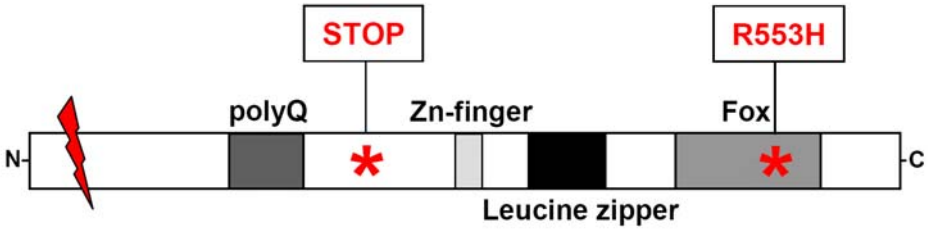

Figure 1.1 Functional domains of the FoxP2 protein. FoxP2 contains a glutamine-repeat region (polyQ), a C2H2 type zinc finger (Zn-finger), a leucine zipper and the forkhead box DNA-binding domain (Fox). All other FoxP family members (FoxP1, FoxP3 and FoxP4) have identical domain architecture with the exception of the polyQ region: in FoxP1, the polyQ stretch is shorter, varies in length among species and lies closer towards the N-terminus of the protein. FoxP3 and FoxP4 do not contain a polyQ region. The positions of the pathogenic alterations of the FoxP2 gene are indicated. In one patient, FoxP2 is disrupted by a balanced translocation (red flash). In another patient, a mutation introduces a stop codon (STOP). In the affected KE family members the mutation of arginine (R) to histidine (H) in position 553 of the amino acis sequence (*) disrupts the DNA-binding capacity of the Forkhead box (R553H).

Murine FoxP2 and the other three members of the FoxP family can act as transcriptional repressors, shown with reporter constructs in different cell lines (Li et al., 2004; Shu et al.,

2001). Thus in patients with FoxP2 mutations reduced levels of functional protein are expected to attenuate transcriptional repression of a specific set of target genes. Their identity is still unknown, in part because the exact DNA sequence to which FoxP2 binds has not been determined experimentally. However the sequence to which FoxP1, the closest homologue of FoxP2, binds is known (Shi et al., 2004; Wang et al., 2003). Interestingly, transcription reporter constructs containing the FoxP1 binding sequence also respond to FoxP2 (Shi et al., 2004), predicting a core motif to which both FoxP2 and FoxP1 can bind. This core motif is very similar to those of the two transcriptional activator families FoxO (Biggs et al., 2001) and FoxC (Saleem et al., 2003). These data suggest that Fox transcription factors are either functionally redundant or require additional protein interactions to specify target gene transcription.

For transcriptional repression to occur FoxP2 needs to dimerize either with itself, with FoxP1 or with FoxP4 (Li et al., 2004). This requirement distinguishes the FoxP family from the other Fox transcription factors. Dimerization depends on a conserved leucine zipper motif (Li et al., 2004). A C2H2 type zinc finger adjacent to the leucine zipper might modulate the specificity of the interaction between FoxP proteins, as reported for FoxP1 (Wang et al., 2003). FoxP1 and FoxP2 but not FoxP4 also interact with the transcriptional co-repressor C-terminal binding protein 1 (CtBP1). CtBP1 binding enhances, but is not essential, for transcriptional repression (Li et al., 2004). A plethora of FoxP2 isoforms including some that lack the forkhead box add further complexity to the system (Bruce and Margolis, 2002).

FoxP2 contains an N-terminal glutamine-repeat that could act as a polar zipper to join other transcription factors bound to separate DNA segments (Perutz et al., 1994), creating a multiprotein transcriptional unit. This hypothesis is consistent with the proximity of a binding site for FoxP1 to a number of other transcription factor binding sites in the *c-fms* promoter, a physiological target of FoxP1 (Shi et al., 2004). Regulation of *c-fms* expression by FoxP1 depends on the polyglutamine-repeat. Interestingly, the only neural sites of *c-fms* expression are the cerebellar Purkinje cells (Murase and Hayashi, 1998), which also strongly express FoxP2 (see below). The presence of a polyglutamine stretch in FoxP2 also prompted the search for pathogenic glutamine repeat extensions implicated in many neurodegenerative disorders (Zoghbi and Orr, 2000). However, the glutamine

region of FoxP2 is neither expanded in the DVD patients studied so far, nor in a set of 142 patients with progressive movement disorders (Bruce and Margolis, 2002).

The molecular factors that regulate FoxP2 expression and the neural target genes of FoxP2 are still unidentified. Analysis of signal transduction pathways relevant for the development of tissues in which FoxP2 is expressed and comparison with molecular interactions of other Fox genes converge on the sonic hedgehog (Shh) pathway as a candidate for interactions with FoxP2. FoxP2 is strongly expressed during lung morphogenesis (Shu et al., 2001), during which FoxA1 and FoxA2 regulate sonic hedgehog (Shh; (Wan et al., 2005). Knockout of FoxP2 (see below) and transgenic overexpression of FoxA2 in mice (Zhou et al., 2001) both disrupt cerebellar morphogenesis which also depends on Shh signaling (Dahmane and Ruiz-i-Altaba, 1999). FoxP2 could also lie downstream of Shh like FoxE1 (Eichberger et al., 2004), FoxM1 (Teh et al., 2002) and FoxF1 (Mahlapuu et al., 2001). In addition, the zinc finger of FoxP2 is highly homologous to those of the major Shh downstream transcriptional effectors Gli1, Gli2 and Gli3 (Shu et al., 2001).

Taken together, dimerization of FoxP proteins and their potential interaction with other transcription factors provide opportunity for complex patterns of target gene repression. In addition, the similarity of the predicted core DNA-motif to which both FoxP1 and FoxP2 bind raises the possibility that they can compensate for each other when co-expressed in the same cells.

1.4 FoxP2 Knockout Mouse

Whereas heart defects in FoxP1 knockout (KO) mice cause embryonic lethality (Wang et al., 2004), mice with disruption of both FoxP2 alleles live for three weeks after birth (Shu et al., 2005). They are developmentally delayed, and are impaired in tests that assay motor function. Heterozygous mice perform only moderately worse than wild-types and catch up by their second week of life. Adult heterozygous FoxP2 knockout mice show no deficits in the Morris water maze, which requires coordinated movement of the limbs and measures spatial learning abilities. Spatial learning depends on the hippocampus, which does not express FoxP2 in mice (Ferland et al., 2003; Lai et al., 2003) and would therefore not be expected to be strongly impaired in FoxP2 knockout mice.

Consistent with the conserved cerebellar FoxP2 expression (Ferland et al., 2003; Lai et al., 2003), FoxP2 knockout mice display cerebellar abnormalities. These include abnormal Bergmann glia and the delayed and incomplete postnatal resolution of the external granular layer, suggesting impaired cell migration. In addition, the molecular layer in heterozygous animals is thinner, the Purkinje cells have underdeveloped dendritic arbors and are misaligned. It is possible that the cerebellum is particularly vulnerable to the absence of FoxP2, because it lacks coexpression of FoxP1 (Tamura et al., 2004). FoxP1 might compensate for the absence of FoxP2 during development in regions that normally express both, e.g. the basal ganglia and the thalamus. The basal ganglia that strongly express FoxP2 and FoxP1 during development do not exhibit gross histological abnormalities in FoxP2 KO mice. Since KE family patients do have structural abnormalities of the basal ganglia (Watkins et al., 2002) it will be interesting to analyze the anatomy and behavioral function of the basal ganglia in FoxP2 KO mice in more detail.

Homozygous FoxP2 knockout pups vocalize less in the sonic range than heterozygous and wild-type animals when separated from their mothers. In the ultrasonic range, both homo- and heterozygous knockout animals utter fewer whistles. Interestingly, the acoustic structure of the vocalizations is preserved in FoxP2 KO pups indicating that the motor areas controlling acoustic features of sound production are intact. Ultrasound communication in adult homozygotes could not be tested because they die too early (Shu et al., 2005). Because FoxP2 is implicated in cellular differentiation of the developing lung, pneumatic function might be compromised in the knockout mice, which could affect vocalizations. In fact, hypoxia strongly decreases the rate of postnatal vocalizations (Blumberg and Alberts, 1991). Given the speech pathophysiology of patients with FoxP2 mutations, it is particularly interesting that vocal behavior in the KO mice is impaired. The recent finding, that adult male mice are capable of vocalizations with a previously unrecognized complexity that shares major characteristics of song (Holy and Guo, 2005), opens the possibility for a more detailed study of vocalizations in FoxP2 knockout mice. In light of the relative ease of genetic manipulation in mice and the large collection of mouse disease models this seems a particularly promising area of future research. However, it is important to bear in mind that whether mouse vocalization, like human speech, is learned has yet to be determined.

1.5 Human Speech and Birdsong

Although language is unique to humans, a few orders like bats (Esser, 1994), cetaceans, e.g. dolphins (Janik, 2000) and three orders of birds (Baptista and Schuchmann, 1990; Hall et al., 1997; Kroodsma and Baylis, 1982) are capable of learning to produce the vocal repertoire required for communicating with their conspecifics. This capacity of auditory-guided, imitative learning has been studied particularly well in the three avian orders: songbirds, parrots and hummingbirds. Collectively, these studies revealed many parallels between human speech and learned birdsong, which are briefly discussed in the following (for review see (Doupe and Kuhl, 1999). Although many of the parallels mentioned below also apply to parrots and hummingbirds, I will refer to songbirds if not stated otherwise to avoid false generalizations.

Birdsong consists of ordered strings of sound, separated by brief silent intervals. The smallest sound unit in the song is the note, that can be defined as a continuous marking on a sound spectrogram. Notes can be grouped together to form syllables. By definition syllables are separated by silent intervals. They can be seen as the basic processing unit of birdsong, as birds interrupted by a light flash or sound while singing still complete the entire syllable (Cynx, 1990). In human speech, syllables are similarly considered to be the phonological building blocks of words. Song syllables are usually assembled to form phrases or motifs, which can be a series of identical or different syllables. Many of the avian song learners sing several motifs in a fixed order. The timing and sequencing of syllables and phrases is not random, but usually follows a set of rules, called syntax. It is important to keep in mind though, that the term syntax in human language refers to the rules of grammar, which allow to create an infinite number of dependencies between words in a sentence. This is not to the case for the syntax of song. Avian vocal learners also do not seem to actively change the syntax of their vocalization to convey a different symbolic content. One exception may be the alarm calls of Black-capped Chickadees, which signal size and threat of a predator by adjusting the frequency of a particular syllable within their mobbing vocalization (Templeton et al., 2005). Although song syllables also lack abstract meaning, it is definitely not meaningless for a bird to vocalize. Birdsong advertises for mating, territorial ownership and fitness and communicates species and individual identity, including “neighbor” and “stranger” (Collins, 2004).

The similarities between birdsong and human speech are evident not only with regard to the acoustic structure of the vocalization and their importance for communication, but also the mechanisms of their perceptual learning. Both song learning and speech acquisition proceeds through several stages. In an initial sensory phase, babies and birds have to build the auditory memory for the sound characteristics of their vocal repertoire. Babies have to memorize the phonetic units and prosodic (pitch and intonation) characteristics that typify the mother tongue, birds store the specific notes, syllables, and prosodic characteristics that typify their species. In the subsequent motor phase, the production of sound is initiated. Babies and birds use the patterns stored in memory to guide motor production through the process of imitation. The motor phase of intensive rehearsing leads to speech in humans and to the adult, crystallized song in songbirds. This adult song is usually very similar to the tutors song. In some songbird species, like the zebra finch, and in humans the sensory and motor phases highly overlap. The ability to learn decreases with age in humans and birds, pointing to the existence of a "critical period" usually before reaching adulthood, where vocal learning is achieved best (Marler, 1970; Vargha-Khadem et al., 1997). The exception of the rule are the so-called "open-learner" species, which continue to modify their song throughout adulthood, e.g. canaries. Another well known example of an open-learner is the budgerigar, a member of the psittaciformes [parrots (Hall et al., 1997)].

For both speech acquisition and song learning auditory input is critical. The absence of exposure to other individuals leads to abnormal vocalizations in humans (Fromkin et al., 1974; Lane, 1976) and in songbirds (Thorpe, 1958). The existence of local dialects (Marler and Tamura, 1964), cross-fostering (Immelmann, 1969) and deafening experiments (Konishi, 1965) have further demonstrated the importance of auditory tutoring in songbirds. Interestingly once vocalizations are learned, both humans and songbirds depend less on hearing their own voice, even though deafness acquired in the adulthood deteriorates speech and song to some degree (Cowie and Douglas-Cowie, 1992; Nordeen and Nordeen, 1993). Another parallel between humans and songbirds exists with respect to the effect of altered or delayed auditory feedback. Experimental manipulation of the auditory feedback negatively influences the stability of the vocalization more severely than the absence of auditory feedback, suggesting that sensory input has some access to the adult vocal system (Howell and Archer, 1984; Leonardo and Konishi, 1999). Finally, since vocal communication is a social behavior, it is maybe not surprising that the social

context is an important component of both song- and speech-learning (Goldstein et al., 2003; Kuhl, 2003).

Parallels between human speech and birdsong not only exist on the behavioral level, but also on the level of the neural circuits mediating these behaviors. The neural pathways for vocal control in both humans and songbirds are hierarchically organized (for an anatomical overview of the songbird brain see Figure 1.2). At the periphery, brainstem and midbrain areas direct the movement of the vocal tract and the respiratory motor neurons (Figure 1.3). Whereas the function of these areas in sound production is not limited to vocally learning animals, higher-level cortical (in humans) and pallial (in birds) control of vocalizations has only been described in animals capable of auditory-guided vocal learning. In the songbird forebrain these telencephalic structures include the nuclei HVC and RA, which form the initial part of the motor pathway (Figure 1.3). HVC initiates a "central motor program" for the song (Vu et al., 1994). It projects to RA, which then connects to all the nuclei involved with vocal motor and respiratory control (Wild, 1997). HVC generates sequences of sparse bursts during song apparently encoding the temporal structure of the syllables (Hahnloser et al., 2002). Interestingly, the sparse bursting patterns are sometimes recapitulated during sleep (Hahnloser et al., 2006). This is reminiscent of earlier findings, that timing and structure of activity elicited by song playback during sleep matches the activity during daytime singing. (Dave and Margoliash, 2000). The songbird forebrain motor pathway has to be intact in order to produce normal song. Lesions in any of the two nuclei HVC and RA disrupt song production at all stages of life of the animal (Nottebohm et al., 1976).

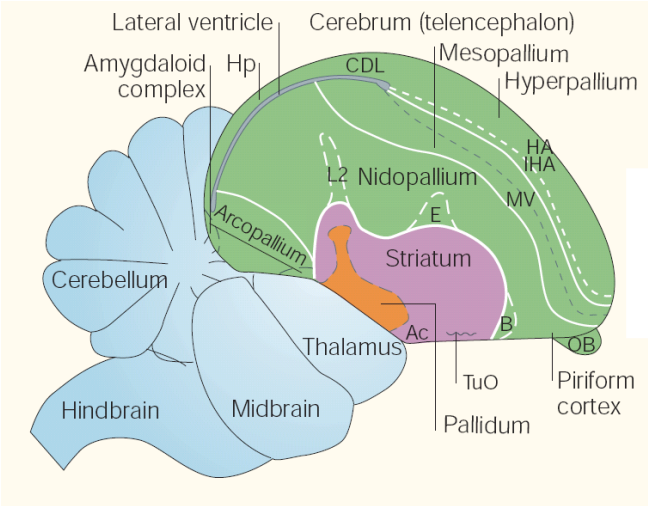

Figure 1.2 **Side view of the avian brain.** Solid white lines are lamina (cell-sparse zones separating brain subdivisions). Dashed grey lines divide regions that differ by cell density or cell size; dashed white lines separate primary sensory neuron populations from adjacent regions. Abbreviations: Ac, accumbens; CDL, dorsal lateral corticoid area; E, entopallium; B, basorostralis; HA, hyperpallium apicale; Hp, hippocampus; IHA, interstitial hyperpallium apicale; L2, field L2; MV, mesopallium ventrale; OB, olfactory bulb; TuO, olfactory tubercle (Figure from Jarvis et al., 2005).

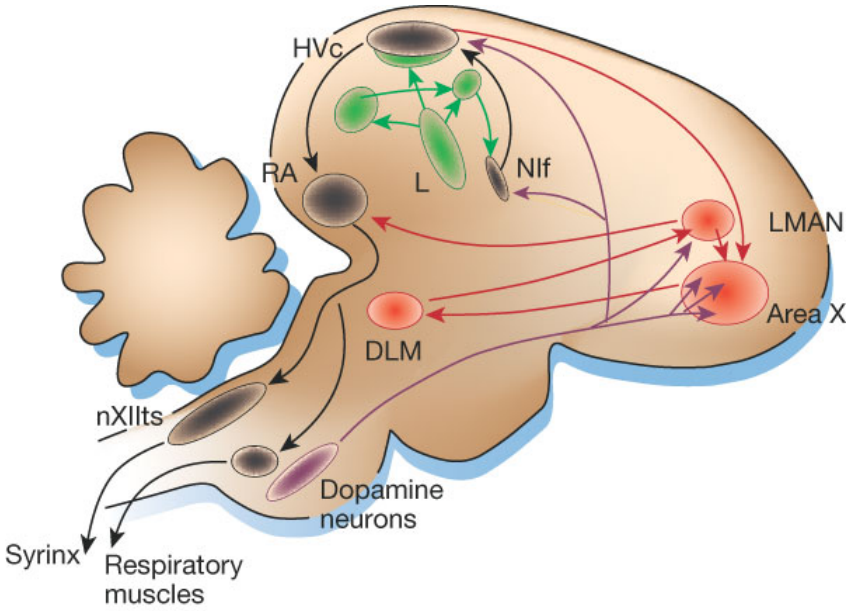

Figure 1.3 **The song system of the zebra finch.** The motor pathway (black) is necessary for normal song production throughout life, and includes HVC (abbreviation used as proper name) and the robust nucleus of the archistriatum (RA). RA projects to the tracheosyringeal portion of the hypoglossal nucleus (nXIIts), which controls the bird's vocal organ or syrinx, and to nuclei involved in control of respiration during song. Additional nuclei afferent to HVC, including the nucleus interfacialis (NIf), are likely to be part of the motor pathway, but their role is less clear. HVC sends a second projection to the anterior forebrain pathway (AFP, red). The AFP includes Area X, which is homologous to mammalian basal ganglia, the medial nucleus of the

dorsolateral thalamus (DLM), and the lateral magnocellular nucleus of the anterior neostriatum (LMAN; a frontal cortex-like nucleus). LMAN sends a projection back into to the motor pathway at the level of RA. Like basal ganglia in other vertebrates, Area X is the target of strong midbrain dopamine projections; LMAN, HVC and NIf also receive dopamine inputs (purple). The Field L complex is the avian primary forebrain auditory area and projects to a complex network of higher auditory areas (green), including the caudomedial neostriatum and caudal portion of the ventral hyperstriatum (not labelled). Auditory inputs likely enter the song system at the level of NIf and possibly HVC (Figure from Brainard and Doupe, 2002)

The higher complexity of the human neocortex and the layered, columnar organization make it more difficult to pinpoint the higher motor areas for speech in the human brain. According to the traditional view, Broca´s area in the posterior frontal inferior cortex is responsible for the production of speech, as patients with lesions in this part of the brain show expressive aphasia. However, investigation of brain activity with non-invasive techniques, like positron emission tomography and (PET), functional magnetic resonance imaging (fMRI), electro- and magnetoencelephalography (EEG, MEG) have revealed additional cortical areas active during speech generation i.e. the motor cortex, supplementary motor areas and the anterior cingulate. This suggests that there is not one single area for speech, but rather a parallel distribution of brain processes subserving different language functions in the brain (Ojemann, 1991).

On the subcortical level, three structures are involved with motor control of vocal output in humans: the cerebellum, the basal ganglia and the thalamus. These structures are generally implicated in the initiation of volitional movements and their modification on a minute-to-minute basis. The cerebellum is important for motor learning, the basal ganglia are critical for the ability to establish habits, procedures and stereotyped behaviors (for reviews see Boyden et al., 2004 and Packard and Knowlton, 2002). Cortical motor areas project to the basal ganglia and to the cerebellum. Both structures project back to the cortex, through the thalamus, building a cortico-basal-ganglia and a cortico-cerebellar loop, respectively. Whereas the output of the cerebellum is mainly excitatory, the output from the basal ganglia is mainly inhibitory. The balance between the two systems allows smooth coordinated movements and disturbances cause movements disorders. More specifically for speech and language, lesions in the cerebellum have been associated with articulatory deficits and slowed speech tempo (Ackermann et al., 1992), lesions in the caudate nucleus of the basal ganglia impair articulation and prosody, but interestingly also language comprehension (Damasio et al., 1982). Some evidence further indicates that parts of the

basal ganglia are activated when optimal performance in processing or production of speech is not (yet) achieved. A PET study revealed that dopamine requirement in the left striatum was negatively correlated with the accuracy and speed of phonological processing (Tettamanti et al., 2005). The left putamen has also been shown to be active when humans are generating words in a second language, but not when performing the same task in their native tongue (Klein et al., 1994). In another fMRI-study of bilingual brains, the left caudate has been identified to play a role in monitoring and controlling the language in use (Crinion et al., 2006).

Whereas the connectivity of the cerebellum within the song circuitry and its role for song production has yet to be established, the importance of the basal ganglia network for learned vocalizations in songbirds is well documented. The pallial nucleus HVC projects to the striatal nucleus Area X. Area X in turn projects to the thalamic nucleus DLM (medial nucleus of the dorsolateral thalamus). The pallial-basal-ganglia loop closes through a projection from DLM to the pallial nucleus lMAN (lateral magnocellular nucleus of the anterior nidopallium) which is connected with the motor pathway at the level of RA (Figure 1.3). The pallial-basal-ganglia loop of songbirds has been termed anterior forebrain pathway (AFP). In contrast to the songbird motor pathway, lesioning lMAN or Area X in most adult birds has no immediate consequence on song production. However, in Bengalese finches, a species in which adult animals rely more on auditory feedback (Okanoya and Yamaguchi, 1997) Area X lesions cause song deficits (Kobayashi et al., 2001). This suggests a role of the AFP in adult song maintenance. A common feature of AFP lesions in all songbirds is that song does not develop properly when they are performed in young birds (Bottjer et al., 1984; Scharff and Nottebohm, 1991). Electrical stimulation of lMAN can direct real-time changes in vocal output in young zebra finches (Brainard, 2004), which might suggest that the AFP corrects vocal output, whenever there is a mismatch between song heard and the song to be produced during phases of song learning or adult song plasticity. However the finding that firing patterns of neurons in lMAN are insensitive to abnormal auditory feedback, has rather promoted the idea that the AFP "injects" variability in the motor output during phases of learning (Leonardo, 2004). This variability is required to reinforce learning of the correct syllables by selecting for the appropriate vocal output. Consistent with this, pharmacological inactivation of lMAN in juvenile zebra finches reduces variability of syllable acoustic features and song syntax, which together with the fact that spiking patterns recorded in lMAN are highly variable

across song renditions, indicates that the AFP may initiate vocal experimentation (Olveczky et al., 2005).

The task of learning song in birds or speech in humans requires auditory input for two reasons. First, it is used for building an internal representation of the tutor song or speech, respectively. Second, auditory input is required to monitor self-produced vocalizations. Recognition of spoken language involves hierarchically organized cortical structures, that perform the acoustic-phonetic, phonological and lexical-semantical processing of language. The main areas involved in speech perception are the primary auditory cortex and additional auditory association areas, in particular the superior temporal gyri, including Wernicke´s area. The analogous structure to the human primary auditory cortex in the songbird brain is the field L, which is connected to a number of reciprocally connected secondary auditory nuclei, some of which process auditory information to HVC. The major source of auditory input to HVC is the nucleus interfacialis (NIf) afferent to HVC (Coleman and Mooney, 2004). Since NIf also shows premotor activity, it has been considered to be part of the motor pathway (McCasland, 1987). But bilateral lesions of NIf do not affect song production, demonstrating that at least in adult zebra finches, an intact NIf is dispensable for motor output (Cardin et al., 2005).

Another characteristic shared by human and songbird brains is the functional lateralization of the neural circuits for learned vocalization, although some important differences exist between the two systems. The avian syrinx is a bilateral vocal organ capable of producing two independent sounds from each syringeal half (Goller and Suthers, 1996). Each side of the syrinx receives input from one hemispheres via ipsilateral connections. In humans, contralateral projections from the two hemispheres converge on a single sound source. As is the case for speech processing in humans, there seems to be a bias towards the left hemisphere in song production in songbirds. Cutting the left tracheosyringeal nerve to the syringeal musculature in canaries results in more severe disruption of song, than when cutting the equivalent nerve on the right side (Nottebohm, 1970). Similarly, unilateral left HVC lesions are more detrimental to song than right HVC lesions (Nottebohm et al., 1976). Interestingly, in zebra finches, the right song system is dominant (Floody and Arnold, 1997).

Taken together, the many parallels between birdsong and humans speech on the functional, behavioral and neural circuit level emphasize the suitability and relevance of songbirds for the study of the basic principles of learned vocalizations, including speech. The songbird model, might also offer insights into the pathology of human speech abnormalities, like DVD. The striking overall consistency between FoxP2 expression pattern in the mammalian brain and the site of pathology in DVD patients, with areas involved with song learning draws the attention to a possible role of FoxP2 in the song system.

1.6 Genes for Vocal Learning ?

In both songbirds and humans, vocal learning is not solely dependant on acoustic cues, but there is evidence for innate mechanisms that govern aspects of how learning proceeds. Birds can discriminate between homo- and heterospecific (from another species) song very early in life (Dooling and Searcy, 1980; Nelson and Marler, 1993). They also show an initial innate predisposition for species-typical signals (Marler and Peters, 1977) and song learning in birds tutored with song from alien species, usually takes longer and is less accurate and less complete, than when tutored with conspecific song (Marler and Peters, 1977). Even in total absence of auditory input due to deafening, a few songbirds still produce some of the normal syntactical rules of the species song (Konishi, 1985). Crossing of two canaries strains with two distinct vocal repertoires yields offspring that develops a mixed repertoire when presented with songs from both repertoire types. Purebred birds tutored accordingly learned songs of their own genetic type (Mundinger, 1995). These data point to a genetic influence on the syllable catalog available to the canaries, suggesting the existence of innate mechanisms that restrict the vocal repertoire prior to any sensory input.

The situation in humans is less clear, in part because the “classical” bird experiments - deprivation of auditory input or tutoring with alien species song - cannot be carried out for obvious reasons. Nevertheless, there is evidence for an inborn perceptual bias for language. Humans can discriminate very early in life between different phonetic units (Kuhl, 1987). Babies born deaf start babbling normally, but their vocalization rapidly becomes distinguishable from hearing babies. Another subject of investigation has been the “spontaneous” development of sign language among deaf people from different cultures. The gestures of naturally evolving sign languages are assembled according to

rules that follow the general rules of human grammar (Goldin-Meadow and Mylander, 1998; Sandler et al., 2005; Senghas et al., 2004). Deaf babies who are exposed to sign language also babble using their hands (Petitto and Marentette, 1991). Moreover, simple "pidgin" languages can develop from a crude mixture of different languages into discrete, more complex languages in relatively short time, as children improve the grammar within every generation without external instruction (Pinker, 1994). If these examples are indicative of an innate predisposition specific to language or just reflect the generalized human capacity to learn to segment and group complex sensory inputs is still a matter of discussion (Fitch et al., 2005; Pinker and Jackendoff, 2005). Given that both aspects can be regarded as two sides of the same coin, this discussion however appears to be of rather semantic nature.

The innateness of certain aspects of learned vocalizations point to a genetically encoded neural circuitry that can later be shaped by perceptual learning in a species-specific way. Following this concept, it is important to point out that even though only a certain predisposition is genetically determined, the behavioral outcome of vocal learning is influenced by the action of genes at all levels - from building the brain, establishing the appropriate connections to their adjustment by experience. The genes involved in these different steps during the dynamic process of vocal learning are largely unknown. In songbirds, a few candidate genes have been identified, based on either their striking expression patterns in song nuclei or a known involvement in learning and memory in mammals (Scharff and White, 2004). Among those genes are IGFII, which is strongly expressed in the Area X-projecting neurons in HVC (Holzenberger et al., 1997), a gene for a yet-to-be-identified antigen which is expressed almost exclusively in RA (Akutagawa and Konishi, 2001) and α-synuclein, which is best known for its role in human Parkinson´s and Alzheimer disease, but is also differentially regulated during song learning (George et al., 1995). In addition, the immediate-early genes *c-Fos,* ZENK and Arc, are responsive to neural activity, and have provided much insight into the different activation patterns involved in song behavior (Jarvis et al., 1998; Kimpo and Doupe, 1997). Many glutamate receptor subtypes also show differential expression in songbird vocal nuclei (Wada et al., 2004) and haven been linked to forms of synaptic plasticity underlying learning and memory.

None of the above mentioned bird genes has directly been shown to be essential for vocal learning, in part because of the difficulty manipulating genes in avians (see below). Similarly, none of these genes have been shown to be specifically implicated in vocal learning, for some it is known they are not. In humans, while a number of genes have been found to impair cognitive abilities when disrupted (Ropers and Hamel, 2005), FoxP2 is the only gene known to be both essential and relatively specific for speech and language.

1.7 Analysis of Gene Function by Genetic Manipulation in Songbirds

The identification of FoxP2 as the cause of DVD, begs the question whether the parallels between human speech and birdsong also exist at the genetic level. In view of the FoxP2 haploinsufficiency in patients with DVD one could imagine generating homo- an heterozygous knockout-birds and subsequently assaying their capacity for song learning and -production. However, to date, no genetic modification of an avian vocal learner has been reported, mostly because of technical difficulties in the development of efficient methods for genetic modification of birds (Zajchowski and Etches, 2000). Recent success in the generation of transgenic chicken (McGrew et al., 2004) and quails (Scott and Lois, 2005) by use of lentiviral vectors have brought transgenic songbirds into close reach. Nevertheless, with these approaches it would still not be possible to target specific genes by homologous recombination, such that a gene can be "knocked out" or replaced with an expression reporter ("knock-in"). Another problem exists with respect to the temporal and spatial control of the genetic manipulation. To date no songbird promoters have been characterized, and although some already described mouse or chicken promoters might be of use, they would require intensive testing to confirm correct gene expression.

One method to circumvent both problems is to inject a lentiviral vector that induces RNA interference (RNAi) into defined brains areas at a defined time. RNAi is a mechanism of posttranscriptional gene silencing through sequence specific degradation of mRNA (Figure 1.4). In mammalians and chicken (Pekarik et al., 2003) it can be induced by double-stranded RNA of 21-23 nucleotide length (short interfering RNA or siRNA) that direct ribonucleases to homologous mRNA targets, thus leading to their cleavage (reviewed in Dykxhoorn et al., 2003). The triggering agent of RNAi, the siRNA, can also be expressed from vectors using promoters, originally derived from mammalian small nuclear RNA genes. These promoters are particularly suitable for the expression of small RNA´s

because they have a very precise transcription initiation start and recruit polymerase III (polIII) for transcription. In contrast to the more common polymerase II, which transcribes most mRNA´s, polIII does not add poly-A tails to the RNA transcripts. The "double-strandedness" of the expressed siRNA is achieved by designing the expression construct to encode the sense siRNA sequence, followed by a loop sequence and the antisense siRNA sequence, such that the linear transcript folds back to build a hairpin structure. Hairpin siRNA is also referred to as short hairpin RNA (shRNA). These hairpin structures have been shown to induce RNAi efficiently *in vitro* and *in vivo* (Krichevsky and Kosik, 2002; Rubinson et al., 2003). For the delivery of the shRNA, lentiviruses have become one of the most powerful genetic tools available. They readily infect non-dividing cells, escape transgene silencing effectively, usually integrate into transcriptionally active regions of the host genome and do not elicit an immune response in the host (Lois et al., 2002). The injection of a lentivirus encoding shRNA into brain regions of interest has been used successfully to alter neural gene expression and behavior in mice (Hovatta et al., 2005; Rumpel et al., 2005).

Figure 1.4 Theoretical model of RNAi induction by shRNA. Short hairpin RNA is processed by a protein complex including the RNAse III nuclease DICER. This generates double stranded RNA (dsRNA) molecules which are structurally similar to siRNA. The dsRNA mediates the recognition of the homologous mRNA target by the RNA-induced silencing complex (RISC). Argonaute, the catalytic component of the RISC then degrades the target mRNA by endonucleolytic cleaveage (after Dykxhoorn and Lieberman, 2005).

Given the rich knowledge about the neurobiology of birdsong, it seems desirable to adapt methods for the functional analysis of genes contributing to learning and production of song. A suitable songbird for establishing the above described method is the zebra finch, because is readily breeds in captivity and has already been studied extensively. The stereotypy of the zebra finch song as well as the availability of appropriate software for automated recording and quantitative analysis ease the investigation of song learning behavior with and without genetic manipulation (Tchernichovski et al., 2001).

1.8 The emergence of Vocal Learning and the Molecular Evolution of FoxP2

Why vocal learning evolved in some avian species and not others, is a matter of debate. What compensates for the cost of learning the vocal repertoire over using an innate, genetically encoded vocalization system? An advantage of learned song could be that unlike strict genetic transmission, socially transmitted behaviors can spread very quickly through a population (Freeberg, 2000). By creating mating boundaries in relative short time, vocal learning could thus have also increased speciation (Lachlan and Servedio, 2004). The developmental stress hypothesis relates the emergence of vocal learning in birds to its predictability of fitness. The quality of song learning success during juvenile life constitutes an honest trait which females can assess when choosing a mating partner (Nowicki and Searcy, 2004). Learned songs might also be used to identify individuals that are adapted to a particular habitat or social environment (Baker et al., 1981). Related to this, vocal learning could also have developed to maintain individual-specific bonds within changing social groupings. In cooperatively breeding birds, learned "calls" can function in a kin-recognition system, such that only the subset of kin within the population with whom the altruistic animal had direct association benefits from the cooperative behavior (Sharp et al., 2005). Learned songs could also be used to maximize outbreeding, by identifying the most distantly related mating partner. At least in Darwin's finches this does not seem to be the case (Millington and Price, 1984). Given the multitude of different vocally learning bird species, that differ in their ecological environment , their social structures etc. it seems unlikely, that a single, exclusive cause exists, to explain the emergence of avian vocal learning. It might have rather developed for many of the above described reasons, each contributing to varying degree, depending on the species studied.

Given the apparent selective advantage of the open-ended expressive power of modern language, it is surprising why it did not evolve in our closest relatives, the great apes, However, it is still a matter of debate, which selective advantage gave rise to the emergence of language. Suggestions have ranged from enhanced communication of information (Pinker and Bloom, 1990) to improved organization of internal thought (Dennett, 1995), sexual selection (Miller, 2000) and increased social cohesion (Dunbar, 2003). It is also not clear what came first - a means for the fine articulation of the vocal tract, a prerequisite for speech, or a means for combining individual communicative elements and coordinating them with meaning, a prerequisite for language. Alternatively

the two co-evolved (Liberman and Whalen, 2000). The origins of human language date to ~6 million years ago (MYA) and proficient language first appeared between 30,000 and 200,000 years ago in the species *Homo sapiens* concomitant with or subsequent to the emergence of anatomically modern humans (Klein, 1989; Wall and Przeworski, 2000). The invention of modern human language probably coincided with the explosive expansion of modern humans around the globe. The dynamic of this invention process was most likely gradual and involved morphological remodeling on different levels (MacWhinney, 2002). The evolution of bipedalism 7-10 MYA freed the hands, maybe allowing increased use of gestures. This could also have promoted the restructuring of the vocal tract, with the descent of the larynx as a result. The lower position of the larynx produces a larger pharyngeal cavity that is useful in making a wide variety of vowel sounds. On the level of the brain, a two to three-fold increase in size in the period between 2 MYA and 100,000 years ago probably increased cognitive abilities dramatically (MacWhinney, 2002). If he faculty of language inside these bigger brains emerged by a gradual extension of pre-existing communication schemes or by "high-jacking" already adapted systems like spatial or numerical reasoning remains unresolved (Fisher and Marcus, 2006; Hauser et al., 2002).

Since the disruption of FoxP2 impairs human speech, this gene might have constituted a genetic constraint during language evolution in humans. But, vertebrate species ranging from mice to chimpanzees also carry a FoxP2 gene in their genomes and all of these FoxP2 genes show an extraordinarily high degree of sequence conservation. This rather speaks for a general importance of FoxP2 for vertebrate fitness. However, the involvement of FoxP2 in speech and language is clearly unique to humans. This apparent discontinuity led to an analysis of the differences in the exact protein and genomic sequence of FoxP2 across mammalian species. A comparison of synonymous mutations (i.e. base substitutions that do not alter the amino acid sequence) and non-synonymous mutations (i.e. base substitutions that alter the amino acid sequence) in the FoxP2 sequences of mice, great apes and humans revealed that the gene must have been under selection pressure during recent human evolution (Clark et al., 2003; Enard et al., 2002). After divergence from the great apes, two non-synonymous but no synonymous substitutions occurred. However, one of the two previously presumed human-specific amino acids exists also in non-human carnivores (Zhang et al., 2002). The functional significance of the amino acid that remains unique to humans is unclear as it lies in an uncharacterized protein domain.

The pattern of FoxP2 sequence variation among humans further suggest that the human-specific allele was fixed in the population as a result of positive selection rather than relaxation of negative selection. Fixation is assumed to have occurred within the last 200,0000 years during which proficient language also appeared (see above). Taken together these findings indicate that FoxP2 might have been pivotal for the development of human language.

It is not known, which aspect of language might have been influenced through the evolution of the human unique FoxP2 allele, but the answer to this question is certainly intimately connected to FoxP2 function. If the FoxP2 allele proved to be necessary for the development of generative grammar, then also the selective sweep on FoxP2 should have been unique to human evolution. However, given the pathophenotype of affected KE family members, it rather seems plausible that the evolution of FoxP2 in humans improved their ability to learn and execute sequenced, orofacial motor behaviors. In this case, selection on FoxP2 might also have occurred in other species, particularly those capable of auditory-guided vocal imitation. The parallels between the neural circuits associated with the pathology of affected KE family subjects and the neural circuits involved in song learning in songbirds (Jarvis, 2004), emphasize the possibility that FoxP2 was under selection during the evolution of vocal learning in birds too. It has been proposed that vocal learning was gained three times independently in three distantly related groups of birds (hummingbirds, parrots, songbirds) during the evolution of the avian family, as a parallel loss of vocal learning from the last common ancestor in the 4 remaining groups of birds is considered rather unlikely (Sibley and Ahlquist, 1990). If this assumption is correct, evolution might also have left its mark on the FoxP2 sequences from songbirds, hummingbirds and parrots, as was the case for human FoxP2

1.9 Aims of this Study

It is well established, that the basic principles of acquired vocalization, including human speech can be studied in songbirds. Thus songbirds might also prove useful as a model for human speech pathologies, like DVD which is caused by mutations in the gene FoxP2. The overall consistency between the FoxP2 expression pattern in the mammalian brain, the site of pathology in DVD patients, and areas involved with song learning draws the attention to a possible role of FoxP2 in the song system. Therefore in the first part of this

study, I identify and characterize the zebra finch FoxP2 gene. Because of the functional interaction of the FoxP2 protein with its closest homolog the FoxP1 protein, the FoxP1 gene from the zebra finch is also cloned and sequenced. FoxP1 and FoxP2 expression patterns in the brains of songbirds are compared to the expression patterns in mammals. Is FoxP2 expressed in any of the well-characterized song nuclei of the zebra finch? Are FoxP2 expressing areas analogous to those involved in the human DVD pathophysiology? The strong conservation of the FoxP2 gene among vertebrates, most of which are not capable of auditory-guided vocal imitation leads to the question if there is something special to FoxP2 function in vocally learning species. Hence, FoxP2 brain expression is compared between birds that learn to vocalize, like the zebra finch and those that do not need to learn their vocalizations, like pigeons or doves. Taken together these experiments aim to answer the question if FoxP2 expression in the songbird brain is consistent with an involvement in learning and/or production of song.

To test if FoxP2 has a direct influence on song learning, I use a lentiviral expression system that induces gene knockdown by RNAi in the zebra finch brain *in vivo*. Stereotactic injection of pseudoviruses into defined brain areas of young zebra finches delivers expression constructs encoding shRNA in a temporally and spatially confined manner. Expressed shRNA target FoxP2 mRNA by RNAi, resulting in reduced FoxP2 levels in the brain. This is in analogy to the FoxP2 haploinsufficiency observed in KE family members. It is important to mention that the post-hatch genetic manipulation of FoxP2 allows to study FoxP2 function in the neural circuits for learning, isolated from its involvement in the development of the brain. All genetically manipulated animals are tutored and their songs recorded. Using software for the quantitative analysis of song the consequence of FoxP2 knockdown on song learning success was evaluated.

Analysis of the molecular evolution of human FoxP2 has revealed that the gene contains changes in amino-acid coding and a pattern of nucleotide polymorphisms which suggest it has been the target of selection during recent human evolution. This might indicate that FoxP2 was pivotal for the development of human language. If so, FoxP2 could also have been critical to the evolution of vocal learning in birds. To address the question of whether the FoxP2 gene has evolved differently in birds that learn their song from those whose song is not learned, I compare the FoxP2 sequences from avian vocal learners, non-learners and the evolutionary closest non-avian relative, the crocodile. The FoxP2 genes

from *Taeniopygia guttata* (zebra finch), *Glaucis hirsuta* (rufous-breasted hermit hummingbird), *Gallus gallus* (chicken), *Melospiza melodia* (song sparrow), *Sayornis phoebe* (phoebe), *Melopsittacus undulatus* (parrot, Budgerigar), *Archilochus colubris* (North Carolina hummingbird, Ruby-throated hummingbird), *Serinus canaria* (canary), *Columbia livia* (pigeon, rock dove), *Aphantochroa cirrhochloris* (sombre hummingbird) and *Alligator mississippiensis* (American alligator) are cloned and sequenced. The resulting DNA and amino acid sequences are analyzed for their phylogenetic relationship to test if a particular variant of FoxP2 segregates with the ability for vocal learning.