Phil. Trans. R. Soc. B. **370** doi: 10.1098/rstb.2014.0092

Supplementary Material

Defining the biological bases of individual differences in musicality

Dr. Bruno Gingras, Department of Cognitive Biology, University of Vienna, Austria

Prof. Henkjan Honing, Amsterdam Brain and Cognition (ABC), Institute of Logic, Language and Computation (ILLC), University of Amsterdam, Netherlands

Prof. Isabelle Peretz, Faculty of Psychology, University of Montreal, Canada

Prof. Laurel J. Trainor, Faculty of Psychology, McMaster University, Canada

* Prof. Simon E. Fisher, Max Planck Institute for Psycholinguistics, and Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

* *Corresponding author*

A primer on genes and genomes

To appreciate how the human genome might impact on the formation and function of neural circuits that are relevant for musicality, it is essential to first have a basic understanding of what genes actually are. Genes and genomes encode biological information in the form of deoxyribonucleic acid (DNA) - a chain of adenosine (A), cytosine (C), guanine (G), and thymine (T) nucleotides. DNA is usually configured as a double helix, comprising two complementary chains, or strands, of nucleotides, with each nucleotide on one strand being paired with its complement nucleotide (A with T, and G with C) on the opposite strand. The linear ordering of nucleotides (sequence) encodes the genetic information, as a four-letter code. The human nuclear genome consists of more than 3 billion nucleotides, distributed across 22 different chromosomes as well as the sex chromosomes X and Y. Each human chromosome is effectively one long molecule of DNA, with multiple genes spanning its length; tight packaging with other molecules allows this material, containing a vast amount of genetic information, to be compactly stored in the nucleus of a cell.

Genes come in a variety of forms, but the ones that we know the most about are those that code for proteins, chains of amino acids that are responsible for the majority of functions in a cell (or multicellular organism). For a gene to be expressed, its DNA code needs to be transcribed into an intermediate molecule, known as messenger RNA, which is passed on to the protein-building machinery of the cell. The messenger RNA molecule undergoes splicing in which sections of sequence that do not code for protein (introns) are cut out, leaving only the coding parts (exons). Next, in the translation step, the exons are decoded to produce the amino acid sequence corresponding to a protein. There are twenty

different amino acids, with different properties, that can be used as building blocks in this process. Thus, the linear sequence of nucleotides in a gene is able to specify the linear sequence of amino acids in a particular protein. The function of a protein is determined by the 3-dimensional shape that it forms, and this shape is itself tightly dependent on the precise sequence of amino acids that it contains.

In this way, the DNA of the human genome carries the information for building more than 20,000 different proteins with a vast array of different sizes, shapes and functions; enzymes that catalyze reactions, structural proteins that are integral to the cell, molecules for sending and receiving signals, channels that allow ions to flow across cell membranes, and a great many more. (Interestingly, due to a certain degree of flexibility during the splicing process, the same gene can encode different versions of a particular protein, which substantially increases the range of proteins that the genome can specify.) The genome also contains information that specifies when and where particular genes should be switched on or off (i.e. whether or not their protein product should be made).

People harbour a range of different types of genetic variation. There can sometimes be large-scale rearrangements of chromosomes which are visible under a microscope. For example, part of one chromosome might be exchanged with part of another chromosome (translocation), or a section of it could be missing (deletion) or repeated (duplication). Submicroscopic rearrangements include copy number variants – extra or missing copies of one or more genes. Even changes to just a single nucleotide of DNA (single nucleotide variants) can have functional consequences, by changing the amino-acid sequence of an encoded protein (and hence altering its shape and function), or by affecting the

intermediate steps leading from the gene to the protein (transcription, splicing, translation, etc.). The impact of a genetic variant can be very severe if it seriously damages the way a crucial protein works, or leads to a complete loss of that protein. Many variants are expected to have more moderate effects, for instance by more subtly altering the efficiency of a protein's function, how much of it is made, and so on. Moreover, a large proportion of the variants existing in human populations are thought to be completely benign, with no functional effects. The population frequencies of individual gene variants can range from rare mutations (even ones that are unique to a particular individual or family) through to common polymorphisms that are present in a large proportion of people.