Segmental duplications as a source of innovation in brain development

Lodewijk, G.A.

Publication date
2021

Document Version
Other version

License
Other

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

Summary of experimental chapters
1: Introduction
The human brain is known for its rapid increase in size during recent evolution, which has been associated with the gain of higher cognitive abilities. However, the exact genetic determinants for the size and complexity of the human brain and how evolution has played on its basic developmental program remains largely elusive. In order for us to understand how our brains are built up in terms of complexity and how disease can impact its functioning, we first need a better understanding of how our brain evolved and what genetic changes are underlying that. In this thesis, we focus specifically on some of the most recent structural variation in our genome, which accounts for a major fraction of evolutionary new DNA sequences in the human genome. Using a variety of experimental techniques, I interrogated their role in normal human brain development in order to form hypotheses about their contributions to the evolutionary increase in size and complexity of the human brain. As a red line in this thesis, we repeatedly observed that the beneficial effect of structural genomic adaptations comes at a remarkable price: These beneficial evolutionary adaptations often lead to increased genomic instability, which can result in higher vulnerability to disease caused by genetic defects. In the examples described in this thesis, it is clear that we cannot have one without the other. For human evolution this means that this evolutionary trade-off between beneficial and detrimental effects of the same mutation, might have been central to how we evolved as a species, and how we will keep evolving in the millions of years ahead.

1.1: A brief history of humankind
In the past 50,000-70,000 years, anatomically modern humans migrated out of Africa, and rapidly spread throughout the continents (Oppenheimer 2012; López et al. 2015; Pagani et al. 2016; Posth et al. 2016; Schlebusch and Jakobsson 2018; Haber et al. 2019). As a result, virtually every corner of the world has been inhabited by humans today. However, it was not the first time human ancestors have migrated out of Africa. Fossils and tools indicative of human-ancestor presence have been found in the Middle East and Asia dating back 1-2 million years ago (Ascenzi et al. 1996; Larick et al. 2001; Carbonell et al. 2008; Ferring et al. 2011; Zaim et al. 2011; Zhu et al. 2018). This shows that ancestral human lineages, like the Homo erectus, migrated out of Africa before and already populated various parts of the world. These lineages were preceded by Ardiptithicus and Australopithecus species, that lived between 2-7 million years.
Chapter 1

A  ~3 - 2 million years ago: Migration of human ancestors out of Africa

B  ~500,000 - 400,000 years ago: Neanderthal, Denisova and early homo sapiens migration out of Africa

C  ~65,000 - 45,000 years ago: Later, major migration of homo sapiens out of Africa
Introduction

Figure 1. General estimations of migrations by human-ancestors, Neanderthals, Denisovans and modern humans. **A**) Earliest known migrations as indicated by fossil findings of *Homo* species or tools indicating their presence. Mya: million years ago. **B**) Fossil findings indicating dispersal of Neanderthals and Denisova throughout Eurasia, and early migration of modern humans. kya: thousand years ago. **C**) Later, major migrations of modern humans throughout the world. kya: thousand years ago. Colored arrows indicate approximate time and locations of human dispersal around the world as indicated in the legend.

ago in Africa, and are some of earliest known human-like ancestors (Barnabas 1990; Alemseged et al. 2006; Wynn et al. 2006; White et al. 2009; Brunet 2010; Haile-Selassie et al. 2016; Haile-Selassie et al. 2019). Around 0.6-0.8 million years ago, common ancestors in Africa split into lineages closely related to anatomically modern humans (Rightmire 1998; Hublin 2009; Tattersall and Schwartz 2009; Tattersall 2011). These include lineages like Homo heidelbergensis and rhodesiensis, which continued to live and evolve on the African continent. A small group of these populations, of which a subset was later identified as Neanderthals (Prüfer et al. 2014; Prüfer et al. 2017; Hajdinjak et al. 2018; Slon et al. 2018) and Denisovans (Meyer et al. 2012; Slon, Viola, et al. 2017), migrated out of Africa around 400k-500k years ago and occupied large regions of Europe, Middle East and Asia. Evidence for the first presence in Africa of anatomically modern humans, Homo sapiens, are dated to around 300k-400k years ago, with many fossils found in East Africa dating 100k-200k years ago (Stringer 2016). Interestingly, there is some evidence for early migrations of Homo sapiens, from fossils found dating back to before the major migration of humans, in Morocco (300kya) (Hublin et al. 2017), Israel (200kya) (Hershkovitz et al. 2018), Greece (200kya) (Harvati et al. 2019), China (100kya) (Liu et al. 2015), Sumatra (70kya) (Westaway et al. 2017) and Australia (65kya) (Clarkson et al. 2017). This hints at the possibility that human origins were not strictly East Africa based, but it is unknown if early humans at other geographical locations established stable populations or if these represent early pioneers that died out during their early migration. Further fossil findings and genetic studies will have to confirm the exact timing and origins of the complex human lineages suggested by various studies. Around 50,000-70,000 years ago, the later and major migration of Homo sapiens from Africa across the world began. Soon after, all other human-ancestor related lineages, including Neanderthals and Denisovans, that already lived in various continents for a long time, began to go extinct. For now, it can only be speculated what set these peculiar events in motion.
Introduction

Figure 2. Schematics of brain size and development in different species. A) Overview of brain size in primates. B) Simplified overview of neocortical development in mouse, monkey and man. VZ: Ventricular zone. SVZ: Subventricular zone. IZ: Intermediate zone. CP: Cortical plate. iSVZ: Inner subventricular zone. oSVZ: Outer subventricular zone.

1.2: Innovations in human brain development

The spread of human lineages and the capacity to thrive in many different environments has in part been attributed to cognitive advances. These have likely been a result of evolutionary changes in brain structures. For instance, the human brain has tripled in size when compared to modern great ape species, such as chimpanzees (Herculano-Houzel 2009; Hofman 2014) (Figure 2A). Interestingly, fossil evidence shows this increase happened in the past 2 million years, during the emergence of the Homo species (Holloway et al. 2004; Rightmire 2004). The increase in brain size is thought to be accommodated by increased self-renewal potential of neural stem cells and their delayed differentiation into neurons and glial cells (Rakic 1995; Lukaszewicz et al. 2005; Kriegstein et al. 2006; Betizeau et al. 2013; Nowakowski et al. 2016; Rayon et al. 2019). Therefore, the expanded progenitor pool has the capacity to generate more cells and larger tissues. Multiple subtypes of neural stem cells have been described. The most well-known and studied are radial glia cells. These reside with their cell bodies in the ventricular zone, and have processes extending to the basal and apical borders of the neuroepithelium. The tightly packed organization of radial glia cell bodies within the ventricular supports their self-renewing state via cell-cell contacts and multiple signaling pathways (Noctor et al. 2002; Lui et al. 2011; Martynoga et al. 2012). Through several rounds of divisions, radial glia cells provide the required amount of neural progenitor cells for proper brain development. In later stages of brain development, they also produce glial progenitor cells and astrocytes (Kriegstein and Alvarez-Buylla 2009). Recently, much information has been gathered on a subtype of progenitor cells, the outer radial glia (Smart et al. 2002; Hansen et al. 2010). Like typical radial glia, they have the capacity to proliferate and differentiate into different cell types of the brain. However, the outer radial glia cells have unique molecular profiles and their morphology is different, as they only have a process extending to the basal lamina (Fietz et al. 2010; Ostrem et al. 2014; Johnson et al. 2015; Pollen et al. 2015; Liu et al. 2017). These cells are largely found in a separate region in the developing brain, the outer subventricular zone, which is a specialized niche that supports the outer radial glia cells. Outer radial glia cells are
characterized by a unique gene expression profile, which is thought to support their self-renewal capacity outside of the ventricular zone. Interestingly, especially the outer subventricular zone underwent a great expansion in the primate lineage, indicating outer radial glia cells had a major contribution to evolutionary expanded brain volumes (Figure 2B) (Fish et al. 2008; Dehay et al. 2015; Bakken et al. 2016; Borrell 2019).

1.3: NOTCH signaling in the developing neocortex

The NOTCH signaling pathway is conserved across animals and is involved in cell fate decisions and tissue organization. In mammals, 4 NOTCH receptors (NOTCH1-4) and 5 canonical ligands (JAG1, JAG2, DLL1, DLL3 and DLL4) are present, each with slightly different features (Figure 3A) (Gordon et al. 2008; Andersson et al. 2011). Many non-canonical ligands have also been described that can influence NOTCH signaling activity (D'Souza et al. 2010). It is known for its relatively simple layout, where activation of the NOTCH receptor via trans-interaction with a ligand leads to cleavage of the receptor intracellular domain, which then translocates to the nucleus to activate expression of target genes (Figure 3B). In the developing neocortex, the NOTCH pathway is highly active in formation of the neuroepithelium and in the subsequently formed radial glia cells in the ventricular zone. Radial glia are tightly packed and have extensive cell-cell contacts, including NOTCH receptor - ligand trans-interactions (Imayoshi et al. 2010; Molnár et al. 2014; Long and Huttner 2019). Activation of the NOTCH pathway leads to expression of several core genes related to maintaining neural stem cell identity of radial glia cells. One of the most well-known NOTCH targets is the HES1 gene, which is a basic helix-loop-helix (bHLH) transcriptional repressor that binds to E-box sequences in gene promoters (Ohtsuka et al. 2001; Kageyama et al. 2008; Boareto et al. 2017). HES1 protein has an important role in repressing proneural genes, such as NGN2 and ASCL1 (Dennis et al. 2019), to inhibit differentiation of radial glia cells and promote self-renewal. HES1 also binds to its own gene promoter which creates an autoregulatory negative feedback loop (Takebayashi et al. 1994). This leads to an oscillatory expression of HES1 with a period of about 2-3 hours and, via this feedback circuit, also of some of its proneural target genes like NGN2 (Figure 3C) (Shimojo et al. 2011). When HES1 expression is high, NGN2 expression is low and vice versa. Only when sustained high NGN2 and low HES1 expression occurs, the steps towards differentiation are set in motion. Sustained high HES1 expression is associated with a dormant cell
Figure 3. NOTCH signaling dynamics. A) Structure of NOTCH receptors and canonical ligands. B) NOTCH-ligand trans-interaction lead to cleavage of NOTCH receptor and translocation of NOTCH intracellular domain to the nucleus. C) Feedback mechanisms between neighbouring cells cause oscillating activity of NOTCH genes and pro-neural genes. D) Dynamic cell states as a result of NOTCH oscillations and neighbour feedback.
state, which may be important in border regions within the brain (Baek et al. 2006). Neighbouring radial glia cells in the ventricular zone also regulate HES1 and NGN2 expression dynamics via NOTCH receptor-ligand trans-interactions. This process of lateral inhibition, combined with HES1 autoinhibition, results in continuously changing cell states of high and low NOTCH activity in radial glia cells (Figure 3D) (Shimojo et al. 2011). The mechanisms regarding cell states in the ventricular zone are well characterized, but this is less understood for the more recently discovered outer radial glia cells. Several signaling pathways have been identified as possible factors in maintenance of outer radial glia cell state, such as signaling via secreted factors like LIF, PTN, BMP, FGF (Pollen et al. 2015), PDGFD (Lui et al. 2014) and SHH (Wang et al. 2016). Also, local interactions via the extracellular matrix consisting of integrins, syndecans and Tenascin-C may contribute to this (Fietz et al. 2010; Pollen et al. 2015). Whereas NOTCH lateral inhibition and HES1 signaling oscillations are thought to be essential for radial glia cell maintenance in the ventricular zone, it is unknown how this is regulated in the different structure of the outer subventricular zone. Outer radial glia cells express HES1 and are negatively affected in self-renewal by inhibiting gamma-secretase, a protease required for cleavage of the NOTCH receptor upon ligand stimulation (Hansen et al. 2010). This suggests NOTCH signaling is active in these cells, but it is not known how this is regulated and if, similar to radial glia cells, NOTCH lateral inhibition and HES1 oscillations are present. As much of these fundamental processes have been studied in mouse models, in which generation of outer radial glia cells is much more restricted (Shitamukai et al. 2011; Wang et al. 2011; Wong et al. 2015; Vaid et al. 2018), there is little knowledge about NOTCH and HES1 dynamics in these cells and in the human background in general.

1.4: Segmental duplications as a source of evolutionary adaptation
While anatomical differences in brain development and morphology have been known for some time, the genetic changes underlying these differences have long remained elusive. Early comparisons of protein-coding DNA sequences between humans and chimpanzees, showed they were highly similar, but these studies only had limited genome data available (King and Wilson 1975; Lander et al. 2001; Collins et al. 2004; Kitano et al. 2004; Glazko et al. 2005; Mikkelsen et al. 2005; Varki and Altheide 2009). Increased efforts in genome sequencing led to the discovery of human-specific genetic adaptations, in both coding and non-coding genome sequences, although identification and functional testing of human-
specific genome variation remains a monumental task. Recently, previously unexplored regions of the genome have received attention. These may contain a rich source of evolutionary genome variation, but have proved difficult to study. They include, among others, retrotransposon derived sequences (Deininger et al. 2003; Cordaux and Batzer 2009; de Koning et al. 2011; Sundaram et al. 2014),
tandem / simple repeats (Sonay et al. 2015; Sulovari et al. 2019) and centromere repeats (Jain et al. 2018; Miga et al. 2019), structural genome variants (Feuk et al. 2006; Auton et al. 2015; Audano et al. 2019) and gene duplications (Marques-Bonet et al. 2009; Sudmant et al. 2010; Dennis and Eichler 2016). We focus on gene duplications as they are the major contributor to evolutionary novel sequences in the human genome and they are associated with rapid adaptation of biological features involving clusters of gene families (Ohno 1970; Holland et al. 1994; Lynch and Conery 2000; Crow and Wagner 2006; Han et al. 2009; Dennis and Eichler 2016; Lan and Pritchard 2016; Rodrigo and Fares 2018). Examples are expansions of Hox genes, related to body structure in early evolution of vertebrates (Wagner et al. 2003; Lynch et al. 2006; Soshnikova et al. 2013), opsins in the evolution of color vision (Carvalho et al. 2017), protocadherins in the octopus related to brain development (Albertin et al. 2015), and zinc-finger genes in primates, related to defense against recurrent retrotransposon invasions (Tadepally et al. 2008; Nowick et al. 2011; Thomas and Schneider 2011). Duplications occur most frequently in germ cells during meiosis (Baudat et al. 2010; Hunter 2015; Zickler and Kleckner 2015; Gray and Cohen 2016) or during mitosis in preimplantation embryos (Ambartsumyan and Clark 2008; Carbone and Chavez 2015; Bolton et al. 2016). During the process of cell division and chromosome segregation, cross-over between paralogous alleles can occur. DNA repair machinery untangles cross-over events, but this may be done erroneously, causing imbalanced inheritance of the involved genetic region between chromosome pairs. This results in a duplication of the involved locus on one allele, and a deletion on the other allele (Figure 4A). Newly duplicated genes often lose their function via mutations, as they are functionally redundant and not necessarily under evolutionary pressure. However, in some cases the newly duplicated gene is used for evolutionary adaptation, where the protein coding sequence is modified to gain new functions (Figure 4B) (Hurles 2004). Multiple rounds of duplications can cause expanded regions of homology within a genome, each containing the common ancestral sequence termed the core duplicon (Marques-Bonet and Eichler 2009). Because of their high homology to each other, these regions are highly unstable. This often leads to de novo genomic rearrangements resulting in a high level of genetic variation of these loci in the modern human population. Further complexity can arise from gene conversion, a balanced exchange between homologous loci (Figure 4C) (Chen et al. 2007). Because recurrent
rearrangements of several of these loci are associated to human diseases, much effort is being put into mapping the structural genome variation in the healthy and disorder-affected population (Shlien and Malkin 2009; Zhang et al. 2009; Coe et al. 2014; MacDonald et al. 2014; Sudmant et al. 2015; Zarrei et al. 2019). Several loci that contain evolutionary expanded duplicons are associated with brain disorders like autism. Example loci are 1q21.1 *(NOTCH2NL / NBPF)* (Brunetti-Pierri et al. 2008; Mefford et al. 2008; Dumas et al. 2012; Bernier et al. 2015; Dougherty et al. 2017; Fiddes et al. 2018), 15q13.3 *(GOLGA8)* (Antonacci et al. 2014) and 16p11.2 *(BOLA2)* (Nuttle et al. 2016; Giannuzzi et al. 2019), where past gene duplications may explain their present instability. This hints at an evolutionary trade-off, where genome instability is an important source of variation, but may also have negative effects in part of the population living today.

1.5: Duplications of the KRAB zinc-finger genes in primate evolution

Zinc finger (ZNF) proteins are the most common type eukaryotic transcription factors, involved in a multitude of genome regulatory processes. Examples are CTCF, involved in genome organization (Ong and Corces 2014), GATA genes, for specification of the early embryo (Tremblay et al. 2018) and BCL11B, a factor essential for gene regulation in brain development (Lennon et al. 2017). Their name is derived from the zinc ion that is required to stabilize the protein structure, of which the most common type is a domain organized by 2 cysteine and 2 histidine amino acids, the C2H2 ZNF domain *(Figure 5A)* (Lee et al. 1989). Several other types of ZNF domains exist, such as the RING, PHD and LIM types, characterized by different structural organization of cysteine and histidine residues around the zinc ion (Cassandri et al. 2017). One C2H2 ZNF domain can recognize a triplet of DNA via interactions with specific amino acids in the protein sequence 3 specific amino acids at positions 6, 3 and -1 interact with one nucleotide each (Wolfe et al. 2000; Persikov et al. 2009). It was also identified that additional factors may affect ZNF domain binding specificity, for instance an additional contact residue position 2 and interactions between neighboring ZNF domains (Wolfe et al. 2000; Persikov et al. 2009; Najafabadi et al. 2017). Most ZNF proteins contain multiple domains organized in arrays (Stubbs et al. 2011), which combined recognize a specific DNA sequence *(Figure 5B)*. However, not necessarily every ZNF domain in a ZNF array takes part in DNA binding. So, the DNA recognition motif of a ZNF protein is defined by the number of ZNF domains and variation in the amino acid contact
residues. These parameters make it possible to generate a highly diverse ZNF domain array - DNA target motif interactions (Najafabadi et al. 2017). Because of their specific modular organization, tandem ZNF arrays may be robust against a single amino acid substitutions in a single ZNF domain. Structural variants in ZNF arrays can have large effects by shifting binding to a target sequence, for instance ZNF domains are deleted or duplicated (Figure 5C). Similarly, a single nucleotide variant (SNV) may lead to binding loss of a single ZNF domain out of the complete array only. However, insertions or deletions in the target motif can disrupt binding of multiple ZNFs in the array, by causing a shift in the DNA sequence (Figure 5D). Such changes allow for rapid evolutionary adaptation of ZNF genome-wide targets. We focus part of our work on a subset of C2H2 ZNFs in human and primate lineages. These are the KRAB zinc finger proteins (KZNFs), which have an N-terminal Kruppel associated box (KRAB) domain. When bound to DNA, they are able to establish a repressive epigenetic state with several cofactors like KRAB-associated protein 1 (KAP1), leading to DNA methylation and histone H3 lysine 9 trimethylation (H3K9me3) (Urrutia 2003; Turelli et al. 2014). In humans there are over 300 KZNF genes present, of which ~170 have emerged since ancestral primate lineages around 60 million years ago up to now (Emerson and Thomas 2009). This gain is attributed to segmental duplications in specific loci harboring clusters of KZNF genes, of which many are located on human chromosome 19. These clusters are intrinsically unstable, caused by highly homologous zinc-finger sequences interspersed with satellite repeats, prone to recombination events leading to gene duplications or deletions (Eichler et al. 1998). In the past decade, several insights have been explaining the origin and function for newly duplicated KZNFs. There is substantial evidence that KZNF gene duplications are a response to repress retrotransposon invasions, maintaining host genome integrity (Thomas...
and Schneider 2011; Jacobs et al. 2014; Imbeault et al. 2017). Retrotransposons are remnants of viral DNA insertions in the germline genome, which are then carried on in the offspring. They also have the capability to copy and paste (type I) or cut and paste (type II) themselves within the host genome, proving a potential threat to host genome integrity and possible disturbances in gene developmental pathways. Recent efforts that mapped genome-wide KZNF binding sites found that many KZNFs bind to specific retrotransposons, providing a mechanism to silence and prevent harmful retrotransposon activity (Najafabadi et al. 2015; Schmitges et al. 2016; Imbeault et al. 2017). Via this repression and DNA sequence erosion over time, most retrotransposons are not able to spread throughout the human genome anymore today. However, KZNFs and retrotransposons appear to co-evolve in a predator-prey mechanism, the so-called evolutionary arms race, where the invasion of new primate-specific retrotransposons has been linked to emergence of KZNF gene duplications (Figure 5E). Active retrotransposons can escape KZNF-mediated repression by acquiring mutations or deletions in the specific binding motif within the retrotransposon DNA sequence. Subsequently, newly duplicated KZNFs can be modified to recognize mutated or new retrotransposons via altered ZNF arrays targeting a new motif (Jacobs et al. 2014; Imbeault et al. 2017). Unsilenced retrotransposon activity can have major effects on the host genome. For instance, the gibbon-specific LAVA element has caused large-scale chromosomal rearrangements across the gibbon genome (Carbone et al. 2014). There are also retrotransposons known for acting as gene regulatory sequences in the host genome. An example are MER41 elements that are enhancers of genes involved in the innate immune system (Chuong et al. 2016). This shows retrotransposon insertions may be domesticated to introduce genome variation for evolutionary adaptation. Similarly, it is hypothesized that KZNF genes may have acquired novel functions in host gene regulatory pathways in primate evolution (Nowick et al. 2011; Schmitges et al. 2016; Imbeault et al. 2017; Pontis et al. 2019; Farmiloe et al. 2020; Kauzlaric et al. 2020). As KZNFs essentially act as transcription factors, the large amount of new KZNFs introduced via segmental duplications may also provide a large source of evolutionary innovation able to modify existing gene regulatory pathways.

1.6: Organoids to investigate human brain development and disease

Much fundamental knowledge about brain development has been acquired in studies using model organisms such as mice (Leung and Jia 2016), zebrafish
Introduction

(Kalueff et al. 2014) and fruit flies (van der Voet et al. 2014), among many others. The application of genetic engineering techniques has enabled the study of particular genes and regulatory pathways, by modulating their expression and activity in various ways in the developing organism. Human-specific genes can be introduced in the mouse genome, to investigate their function. In addition, animal models are used to model human disease, by introducing a specific disease related mutation in the orthologous gene. An important disadvantage of using animal models is that human-specific factors are not present in these systems (Markou et al. 2009; Nestler and Hyman 2010; Kaiser and Feng 2015; Hodge et al. 2019).

Primary human tissues have been used as an alternative, and provided insights in basic structures of the human brain on the molecular level (Clowry et al. 2010). However, their availability is limited, which also hampers the setup of extensive controlled experimental conditions. In recent years, these studies have been complemented by a new technique, the derivation of brain structures in vitro from human embryonic stem cells, termed organoids (Eiraku et al. 2008; Fatehullah et al. 2016). Under the right conditions embryonic stem cells can differentiate into specific tissues, including brain tissue (Eiraku and Sasai 2012). Much effort has been put in establishing defined culture conditions to generate organoids that resemble various structures of the brain, such as the cortex (Lancaster et al. 2013; Camp et al. 2015; Otani et al. 2016; Field et al. 2019; Pollen et al. 2019), hippocampus (Sakaguchi et al. 2015), thalamus (Xiang et al. 2019), midbrain, cerebellum (Muguruma et al. 2015) and spinal cord (Ogura et al. 2018). A major advantage of the use of human stem cell-derived organoids is that experiments can now be done in the context of human brain development. The unlimited source of tissues opens up high throughput experimental setups while still providing the platform for traditional molecular and genetic experimental techniques. Recently, organoids resembling the developing human neocortex have been used widely in studies of brain evolution (Arlotta and Pașca 2019; Mostajo-Radji et al. 2020) and disease (Di Lullo and Kriegstein 2017). These cortical organoids accurately replicate early human cortical development in terms of gene expression and tissue structure. Of course, like any model system, also organoid-based model systems are not without limitations, especially in long-term cultures (Quadrato et al. 2016; Qian et al. 2019; Bhaduri et al. 2020; Lovett et al. 2020). To better understand the capabilities and limitations of organoid systems, comparative mapping of developmental trajectories using organoids and primary human tissues is being
Chapter 1

A. Directed differentiation

- Embryo
- Epiblast isolation
- ESCs
  + KLF4
  + OCT4
  + SOX2
  + MYC
- Somatic cells
- iPSCs
- Organoids i.e. brain

B. Interspecies comparisons

- Adult
- Somatic cells
- iPSCs
- ESCs / iPSCs

C. Disease modeling

- Control
- Patient

D. Functional genomics

- CRISPR-Cas9
- Wild-type
- Mutant
Figure 6. Common strategies to implement organoids in human biology research. A) Sources of pluripotent cells for directed differentiation towards organoids. B) Pluripotent cells isolated from different species can be used for evolutionary and inter-species development comparison. C) Induced pluripotent stem cells capture the genetic background of a specific individual, which can be used to model disease features. D) Targeted genetic engineering using CRISPR-Cas9 in isogenic cell lines to study gene regulatory functions and tissue development.

done (Pollen et al. 2019). Two other major technological advances have made the use of organoids even more widespread. First, the ability to generate stem cells via reprogramming of somatic cells, such as fibroblasts, which are called induced pluripotent stem cells (Takahashi and Yamanaka 2006; Shi et al. 2017; Rowe and Daley 2019). This captures the genetic background of individuals in a cell line, including any disease-related variants and any normal genetic variation they might carry in their genome. This allows, for example, for comparative studies of cell lines derived from individuals carrying a specific disease related allele to healthy controls. Second, the developments in genetic engineering using CRISPR-Cas9 has been used extensively for genome engineering in a multitude of ways (Cong et al. 2013; Adli 2018; Pickar-Oliver and Gersbach 2019). The ease of use and high efficiency of CRISPR-Cas9 mediated genetic engineering has made this an essential tool for investigating gene regulatory pathways in human embryonic stem cells and their derived organoid tissues.

1.7: The ancient DNA revolution
Classically, for reconstructing the evolutionary history of humans, the genomes of species living today are compared to infer genetic events that happened in the past. Between-species comparative genomics is used to estimate the time when two species diverged from a common ancestor, for instance based on their accumulated differences in genetic composition over time (Bromham and Penny 2003; Kumar 2005). With the introduction of next generation sequencing techniques this became a very successful approach, as the mapping of genomes from many different species provides a wealth of information for evolutionary genetic analyses. Still, because the split between humans and chimpanzees is estimated at 6 million years ago, reconstructing human-specific genetic changes that happened in the last few million years have remained very challenging or even impossible for a long time. This field was revolutionized when DNA was first
extracted from Neanderthal fossils (Krings et al. 1997), and later used in whole genome sequencing (Green et al. 2010). This made genomes available of ancient modern humans (Raghavan et al. 2013; Fu et al. 2014; Lazaridis et al. 2014; Slatkin and Racimo 2016), Neanderthals (Prüfer et al. 2014; Prüfer et al. 2017; Skoglund et al. 2017; Slon et al. 2018) and Denisovans (Meyer et al. 2014; Slon et al. 2017), who lived up to 30,000-150,000 years ago (Douka et al. 2019). These data provide a window back in evolutionary time, allowing direct analysis of ancestral genomes and any genetic information that was present then, and lost today. Since the availability of these ancient genome datasets, there has been much interest in identifying genetic regions associated with evolution of developmental features in Neanderthals, Denisovans and modern humans.

Physical properties of Neanderthals have been charted relatively well via discovery of skeletal remains in various locations. Shape of the skull and brain of Neanderthals appears more elongated compared to humans (Figure 7A) (Neubauer et al. 2018; Gunz et al. 2019). Denisovan fossils have been very scarce, some finger bones and teeth have been discovered and used for DNA extraction. More recently, a fraction of a Denisova skull and jawbone have been bound (Chen et al. 2019). Still, the overall limited knowledge about physical and physiological features makes it difficult to confidently make genotype - phenotype associations.
Another recent study used an intriguing approach to solve this, which predicted Denisova physical properties using genetic information (Gokhman et al. 2019). They investigated DNA methylation in the Denisova genome, and compared it to DNA methylation in chimpanzee genomes, modern human genomes and other ancient genomes. Differentially methylated regions were then linked to genotype-phenotype information in modern humans from the Human Phenotype Ontology database. The major limitation in ancient genome studies is the availability and quality of fossil DNA samples (Figure 7B). Over time, DNA degrades and fossils are contaminated by microorganisms, which is affected by various environmental factors (Allentoft et al. 2012). Certain fossilized materials preserve DNA better than others, as it was found that extracts from the petrous bone contains DNA of much higher quality and reduced contamination compared to other bones (Pinhasi et al. 2015). Several advances have been made to improve retrieval of endogenous ancient DNA in fossil bone samples, such as hypochlorite treatment (Kemp and Smith 2005; Korlević et al. 2015), uracil selection (Briggs et al. 2009; Gansauge and Meyer 2014; Rohland et al. 2015), or using probes to capture DNA of interest. There have even been successful efforts of retrieving ancient DNA from cave sediments (Slon et al. 2017). Also, well-preserved proteins extracted from tooth enamel has been used for proteomics on a 1.7 million years old rhinoceros fossil (Cappellini et al. 2019). As fossil samples are scarce and valuable, these technical advances help to increase the number of samples suitable for genomics analysis. In the human context, the genomic data retrieved from fossil samples has aided in more accurately estimating dispersion and branching of ancient human lineages (Schlebusch et al. 2017; Lipson et al. 2019). Furthermore, it revealed multiple events of human, Neanderthal and Denisova interbreeding and introgression (Racimo et al. 2017; Villanea and Schraiber 2019), which still affects our genetic traits today related to the immune system and metabolism (Huerta-Sánchez et al. 2014; Nédélec et al. 2016; Quach et al. 2016; Buckley et al. 2017; Enard and Petrov 2018). These data also hint at introgression of so-called “ghost” species, the presence of genetic variants that were passed on to humans via currently unknown extinct ancestral populations (Reich 2018). Using more recent human samples, it has provided major insights into the migratory routes of humans across the world, and human population dynamics in the last 50,000 -70,000 years (Pickrell and Reich 2014; Nielsen et al. 2017; Skoglund and Mathieson 2018).
1.8: Summary of experimental chapters

Chapter 2 describes the use of cortical organoids to study fundamental aspects of human brain disorders. A loss of function mutation in the \textit{CHMP1A} gene was identified as the cause of microcephaly with pontocerebellar hypoplasia and short stature in 3 families. Functional studies involving knockouts of the homologous gene in mice and zebrafish showed that \textit{CHMP1A} is involved in cell secretory mechanisms. Specifically, it is part of extracellular vesicles, which can transport various signalling proteins across the extracellular space. In the \textit{CHMP1A} knockout models, it was identified that decreased secretion of sonic hedgehog protein was one of the molecular features involved in the phenotype. To further investigate the microcephaly phenotype, we compared cortical organoids derived from human stem cells containing the \textit{CHMP1A} mutation introduced with CRISPR-Cas9, to control stem cells. Indeed, features of microcephaly were observed in the mutant organoids, when comparing expression profiles using RNA-seq. Exogenous activation of the sonic hedgehog pathway with a small molecule, hypothesized to rescue the phenotype, had widespread effects on gene expression in both mutant and control organoids. Interestingly, in the control organoids, many genes involving positive regulation of stem cell self-renewal were upregulated, while in the mutant this was much less apparent. This suggests the molecular profile and response to signaling factors may be altered in neural stem cells of the \textit{CHMP1A} knock-out organoids. Altogether, this strengthened the evidence that the \textit{CHMP1A} loss of function observed in patients causes defects in neural stem cell self-renewal, leading to underdeveloped brain structures.

In Chapter 3, we describe the identification of a human-specific gene cluster, the \textit{NOTCH2NL} genes. These emerged after various segmental duplication and recombination events, and functional copies of \textit{NOTCH2NL} genes are only found in the human lineage. 3 of the \textit{NOTCH2NL} genes reside in the 1q21 locus. Reciprocal copy-number variations of this locus have been associated with microcephaly and macrocephaly, among other phenotypes. Using CRISPR-Cas9, \textit{NOTCH2NL} genes were removed from the genomes in human embryonic stem cells. In essence, this converts the genome back to an ancestral state before the gain of \textit{NOTCH2NL} genes. Comparative profiling using cortical organoids we find that \textit{NOTCH2NL} genes contribute to self-renewal neural stem cells, where knock-out of \textit{NOTCH2NL} leads to premature depletion of the neural stem cell pool. In an extensive series of follow-up experiments, we show that the role of \textit{NOTCH2NL} in
neural stem cell self-renewal is likely mediated by enhancing the NOTCH pathway, a well-known gene-regulatory system involved in the maintenance and differentiation of progenitor cells. In patients that carry neurological disease associated 1q21 copy-number changes, we find evidence that the genomic breakpoints are located within NOTCH2NL genes. This hints at an evolutionary trade-off, where NOTCH2NL genes have benefited adaptation of brain development, but also cause increased genome instability linked to disease.

Chapter 4 explores recent evolution of NOTCH2NL genes, using DNA data from Neanderthals, Denisovans and ancient modern humans. We find that many Neanderthals have an extra copy of a NOTCH2NL gene. Also, Neanderthals and Denisovans have unique protein-coding variants, virtually absent in modern human populations today. These variants appear to modify the structure of function of the protein. The Denisova genome shows a very unusual configuration of NOTCH2NL genes, which we did not find in any modern humans we analyzed. This configuration is predicted to produce an extremely low dose of NOTCH2NL protein. Surprisingly, analysis of 279 whole-genome datasets and 49,599 exome datasets of modern humans revealed specific NOTCH2NL protein-coding variants which all act to reduce functional NOTCH2NL protein level. These variants appear to synergistically and cooperatively decrease the level of functional NOTCH2NL protein. The combined results from the study of ancient DNA and modern human genomes indicate that NOTCH2NL gene configuration was highly variable in recent evolution, and may still be under evolutionary pressure today. This also highlights a potential importance for analysis of 1q21 related neuronal disorders, which are associated with NOTCH2NL genes. From our findings, both copy-number variations and the presence of the specific gene variants have to be analyzed. This combined data will give a better insight in how NOTCH2NL dosage may affect various 1q21 related neurological disease phenotypes.

In Chapter 5, we study a specific member of the KRAB zinc-finger (KZNF) genes, ZNF675. During primate evolution, the KZNF gene family has expanded dramatically via segmental duplications. The human genome contains approximately 350 KZNF genes, of which 170 are uniquely present in primates. The canonical function of KZNFs is to bind retrotransposons via specific DNA-binding zinc-finger domain arrays, and to establish repressive epigenetic landscape mediated by the KRAB domain to silence retrotransposon activity. The
expansion of the KZNF family is seen as a response to repeated invasions of retrotransposon in primate evolution, also described as the evolutionary arms race between the host genome and retrotransposons. One region in the KZNF cluster on chromosome 19 was of particular interest, the 19p12 locus. Copy number variations of this locus, which contains only the two KZNF genes ZNF675 and ZNF681, result in a variety of neurodevelopmental disorders. We find that ZNF675 binds to a specific class of retrotransposons, following the evolutionary arms race model. To explore the disease mechanism, we deleted ZNF675 using CRISPR-Cas9 from human embryonic stem cells and generated cortical organoids. Surprisingly, the disease phenotype did not appear to be linked to dysregulation of retrotransposons. Instead, we found ZNF675 also binds a number of neuronal gene promoters. One of these was the promoter of the HES1 gene, an important factor of the NOTCH pathway, which was also differentially expressed in ZNF675 knock-out organoids. Following up on this effect, we validated in functional assays that ZNF675 alters HES1 expression dynamics, indicating that regulation of endogenous genes may have become an essential function of ZNF675. In a separate study, we further explored this for other KZNFs and find several examples that, like ZNF675, may have acquired a role in endogenous gene regulation (Farmiloe et al. 2020). This highlights the importance of KZNF genes in primate and human evolution.

References
Introduction


Chapter 1


Introduction


Introduction


Chapter 1


Introduction


Introduction


Chapter 1


Introduction


Chapter 1


Introduction


