



UvA-DARE (Digital Academic Repository)

Segmental duplications as a source of innovation in brain development

Lodewijk, G.A.

Publication date

2021

Document Version

Other version

License

Other

[Link to publication](#)

Citation for published version (APA):

Lodewijk, G. A. (2021). *Segmental duplications as a source of innovation in brain development*. [Thesis, fully internal, Universiteit van Amsterdam].

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 6

Discussion

6: Discussion

Anatomical changes in human brain evolution have been known for a long time, but the genetic changes underlying these have remained unknown. In this thesis, we focussed on gene duplications and assessed their contributions to the evolution of the primate and human brain. In chapter 2, we used a cortical organoid model to assess consequences of *CHMP1A* loss-of-function, a proof of principle that cortical organoids may be used to successfully model features of the embryonic brain and related disorders (Coulter et al. 2018). In chapter 3 we describe the human-specific *NOTCH2NL* gene cluster and find evidence it plays a role in brain development and disease (Fiddes et al. 2018; Florio et al. 2018; Suzuki et al. 2018). *NOTCH2NL* genes join recent discoveries of other human and primate specific gene duplications involved in brain feature innovations, like *TBC1D3* (Ju et al. 2016), *ARHGAP11B* (Florio et al. 2015), *TMEM14B* (Liu et al. 2017) and *SRGAP2* (Charrier et al. 2012; Dennis et al. 2012). In chapter 4, we find that *NOTCH2NL* gene configurations are highly variable in the modern human population, suggesting they are still shaped by recent selective pressures. In chapter 5, we focus on one member of the KZNF gene family, *ZNF675*. Copy-number variations (CNVs) of this gene are associated with neurodevelopmental disorders and we aimed to establish underlying mechanisms. The findings presented in these chapters provide new insights in the genetic and molecular profile underlying human brain evolution.

6.1: Using organoids to model aspects of progenitor cells in brain development

To study the contribution of candidate genes in human-specific brain developmental features we used cortical organoids as one of our models. As a proof of principle, we collaborated on work regarding the *CHMP1A* gene (Coulter et al. 2018). Homozygous loss-of-function mutations in this gene are associated with developmental brain growth deficits, severely affecting the cerebellum and slightly affecting the neocortex (Mochida et al. 2012). We also found evidence for this in cortical organoids derived from a *CHMP1A* loss-of-function iPSC line compared to an isogenic iPSC line. This supports cortical organoids as a useful model to study neural stem cells, cell cycle and differentiation dynamics in the early human brain. In parallel, collaborators used a mouse knock-out model to study the *CHMP1A* loss-of-function phenotype observed in human patients described before. While generally showing similar phenotypes, the cerebellar hypoplasia appeared much more severe in human patients with the loss-of-function mutation compared to the mouse knock-out model. This suggests species-specific differences in the function of *CHMP1A* during brain development, which could be helpful in identifying human-specific molecular brain features in future studies. Following this line, treatment of *CHMP1A*-null cortical organoids and isogenic controls showed differences in response to smoothed agonist (SAG) treatment, a small molecule that activates the sonic-hedgehog (SHH) pathway. In this experiment the control group showed a much greater change in gene expression profiles associated with neural stem cell proliferation compared to the *CHMP1A*-null group. The relevant species-specific and brain region-specific differences involving *CHMP1A* or SHH pathway mechanisms could be studied by comparing cortical organoids and cerebellar organoids derived from human and mouse ESCs/iPSCs, supported by parallel experimental conditions including the *CHMP1A* loss-of-function and SAG treatment. Organoid models may therefore be a useful addition to other experimental setups, to identify mechanisms of human neurodevelopmental disorders and human-specific features of the brain in general, as also shown by other recent studies (Bershteyn et al. 2017; Field et al. 2019; Frega et al. 2019; Kanton et al. 2019; Pollen et al. 2019; Trujillo et al. 2019; Lovett et al. 2020). Future technical progress in organoid culture and tissue engineering protocols may allow modeling diverse aspects of brain development and functions beyond progenitor cells.

6.2: Evolution of *NOTCH2NL* genes and the origin human-specific brain features

We establish that *NOTCH2NL* is a human-specific gene cluster that emerged via a complex gene duplication and recombination steps involving the *NOTCH2* gene. Three *NOTCH2NL* paralogs are located near the *1q21.1* locus, which is linked to neurodevelopmental disorders caused by copy-number variations. *NOTCH2NL* protein potentiates *NOTCH* signalling, a process regulating neural stem cell self-renewal and inhibiting differentiation. This may have contributed to the evolutionary increase in human brain size via expansion of the progenitor cell pool. Notably, *NOTCH2NL* is expressed in outer radial glia cells, a specialized type of neural stem cell which is known for its role in the evolutionary expansion of the brain. Radial glia residing in the ventricular zone are known to be regulated by *NOTCH* signalling. The tightly organized structure in the ventricular zone allows regulation of cell fate via *NOTCH* lateral inhibition, inducing oscillatory gene expression patterns that are linked to maintaining the self-renewal state in radial glia. The outer radial cells in the outer subventricular zone do not have such an organized structure, which suggests there may be other ways to maintain self-renewal potential. Indeed, it was found that secreted factors and extracellular matrix components may establish a local niche supporting outer radial glia cells via several pathways (Fietz et al. 2010; Ostrem et al. 2014; Pollen et al. 2015; Wang et al. 2016). Moreover, these cells express *HES1*, indicating the *NOTCH* pathway is active despite the possible absence of lateral inhibition mechanisms observed in the ventricular zone (Hansen et al. 2010). As *NOTCH2NL* is both a secreted protein and an activator of the *NOTCH* pathway, these observations make it a likely candidate for contributing to *NOTCH*-dependent self-renewal mechanisms in the ventricular zone and outer subventricular zone of the human brain. From this perspective, it will be intriguing to see how *NOTCH* signalling is regulated in outer radial glia cells and how human-specific factors potentially modify their properties. In follow-up work, we find that multiple *NOTCH2NL* protein coding variants are present at high frequency, which act synergistically to reduce *NOTCH2NL* protein level. This suggests that recent selective pressures have fine-tuned *NOTCH2NL* dosage, which may still be on-going today in the human genome. While features of *1q21.1* CNV related disorders are well characterized, there are also many individual differences in disorder manifestation. Microcephaly (*1q21.1* deletion) or macrocephaly (*1q21.1* duplication) are not found in every individual (Brunetti-Pierri et al. 2008). This raises the question how individual genetic backgrounds may

influence such phenotypes. As *NOTCH2NL* genes are part of the *1q21.1* CNVs, the presence of specific *NOTCH2NL* coding variants could be modifiers in this particular case. Deletion or duplications of *NOTCH2NL* genes carrying the loss-of-function variants R113* and Exon2^{B-(Splice-mut)} may not have any consequences. However, an affected intact Exon1^{B-(High)} is expected to have a large NOTCH-dependent effect, based on its high contribution to NOTCH2NL protein levels. Further *in vitro* validation could be done by generating iPSC lines from individuals with different *NOTCH2NL* configurations followed by cortical organoid differentiation. Additionally, isogenic lines from H9 ESCs edited via CRISPR-Cas9 and homology directed repair, to revert Exon1^{A-(Low)} to Exon1^{B-(High)} configuration, can be used to study dose-dependent effects of NOTCH2NL expressed from endogenous loci. The advances in genome information and physiological properties available, such as provided by the UK Biobank, also provide interesting angles to study large scale genetic variation in humans using computational biology methods (Van Hout et al. 2019). The increasing amount of exome sequencing data, combined with measurements of anatomical, biological and physiological features, can provide evidence how NOTCH2NL dosage affects normal human variation. Finally, this can be complemented by doing the same analysis from an evolutionary perspective, by comparing modern human to ancient human datasets of fossil findings and ancient DNA sequencing (Neubauer et al. 2018; Gunz et al. 2019). These examinations of *NOTCH2NL* genotypes may improve our understanding of the clinical features of *1q21.1* CNVs, especially those involving brain size. New sequencing techniques using long read technologies may also aid in resolving complex rearrangements in *1q21.1* CNVs, to capture potential other sources of variation beyond *NOTCH2NL* in this locus. Collectively, these methods can be applied to study other recently duplicated genes in the human genome as well. Our findings highlight the importance to study such recently duplicated genes, which are often seen as pseudogenes and discarded from functional analysis (Cheetham et al. 2019). Instead, these newly introduced sequences may have been important factors in the evolution of our species.

6.3: KZNFs as new gene regulators in primate evolution

KRAB zinc finger genes (KZNFs) represent one of the largest gene families in the human genome, which emerged from gene duplications in the primate lineage. The instability of KZNF clusters provides an evolutionary strategy to rapidly generate

new and diverse KZNF genes. However, there may also be risk of affecting essential genes, where CNVs may lead to disease. KZNFs are recognized for their role in repressing retrotransposon activity. Besides this canonical function, it was hypothesized that KZNFs may have other roles as gene regulators, RNA metabolism and chromatin structure. Whereas ancient KZNFs show evidence for neofunctionalization to take part in these processes (Shin et al. 2011; Chauhan et al. 2013; Altomose et al. 2017; Ecco et al. 2017; Helleboid et al. 2019), it is less known if recently evolved KZNFs contributed to reshaping gene regulatory mechanisms. Several studies suggest that KZNFs may target gene promoters to act as traditional transcriptional regulators (Nowick et al. 2011; Schmitges et al. 2016; Imbeault et al. 2017; Pontis et al. 2019; Farmiloe et al. 2020; Kauzlaric et al. 2020). Mining existing genome-wide DNA binding datasets (ChIP-seq) shows that about 26% of all human genes may be under control of a primate-specific KZNF (Farmiloe et al. 2020). Since not all KZNFs have been profiled by ChIP-seq, this number may even be an underestimation. Here, we study the primate-specific gene *ZNF675* in detail, which was of particular interest because it is located in a locus where CNVs are associated with neurodevelopmental disorders. We find that the emergence and duplication of *ZNF675* is linked to the invasion of a specific retrotransposon family in early primates. However, the main function of *ZNF675* today may be regulating a small set of neuronal gene promoters. For instance, *ZNF675* binds to the *HES1* promoter and modifies its expression dynamics, suggesting it is required for regulating NOTCH signaling in the developing human brain. The study of *ZNF675* reveals there may be several constraints on the evolution of KZNF genes, following the consensus that they mainly target retrotransposons. For instance, The DNA-binding site location in retrotransposons may determine repression efficacy. Binding of the KZNF at important regulatory regions for the retrotransposon, such as long-terminal repeats (LTRs), may be very effective to repress their activity. However, at these sites sequences may be present that recruit endogenous transcription factors, that the transposon has hijacked for its own benefits. One example is the STAT1 motif in MER41 elements (Schmid and Bucher 2010; Chuong et al. 2016). As a consequence, KZNFs may target DNA motifs that resemble the binding site of endogenous transcription factors, which could be undesirable for the host genome. High expression of a KZNF may ensure efficient silencing of all transposons in the genome, but too high expression may cause more degenerate motifs to be bound as well, that are present in endogenous regulatory sequences like gene promoters. This hypothesis

suggests the possibility of non-reciprocal effects of loss or gain of specific KZNFs. Deletions may cause loss of target retrotransposon repression genome-wide, potentially having widespread effects on chromatin state and epigenetics. Duplications may lead to binding of endogenous regulatory sequences, such as promoters or enhancers containing similar DNA motifs, due to higher expression of the zinc-finger protein. This could disrupt normal function of gene regulatory networks. In short, in the evolution of KZNFs, the host genome needs to account and select for many factors, such as binding location within retrotransposons, DNA binding sequence, expression levels and potential cross-binding with endogenous transcription factor motifs. On the other hand, these parameters may be exploited to repurpose retrotransposons and KZNFs for adaptive evolution of gene regulation. This highlights the KZNF gene family as a new candidate for innovations in gene regulatory pathways during primate evolution. Further examination of KZNFs with gene-regulatory potential, as recently identified, will reveal how widespread their possible contributions are to human development.

6.4: The next era of identifying and understanding structural genetic variation

The identification and annotation of complex genomic regions has proven to be difficult with most currently available sequencing data. New sequencing techniques, like SMRT sequencing (Chaisson et al. 2015; Vollger et al. 2019), nanopore technology (Jain et al. 2016; Jain et al. 2018), linked reads (Dréau et al. 2019; Marks et al. 2019; Lu Zhang et al. 2019) and capture methods (Mercer et al. 2014; Warr et al. 2015; Dapprich et al. 2016), are rapidly resolving genome assembly gaps and complex structural variation. By combining high-quality datasets generated with different sequencing techniques, the first human chromosome has been assembled from end-to-end (Miga et al. 2019). This multi-platform analysis has also been used to extensively map large and small structural variation in a small set of genomes, providing gold standards for future analysis (Chaisson et al. 2019). These findings show that the last unmapped regions of the genome, including telomeres, centromeres, segmental duplications and repeat sequences, are now possible to analyze completely. Naturally, this gold standard analysis has only been used on a small selection of widely studied samples, but may be complemented by expanding this to genomes of various human ancestries. Already, considerable data has been generated that aims to map structural genome variation. Long read sequencing of 15 human genomes from different

ancestries revealed almost 100,000 new structural variants compared to the hg38 reference genome (Audano et al. 2019). In another study of 26 genomes, over 15,000 large variants (>2kb) were newly identified (Levy-Sakin et al. 2019). More specific instances were found in a Melanesian population, containing unique structural variations including gene duplications in the *8p21.3* and *16p11.2* loci (Hsieh et al. 2019). It was found these structural variants were the result of introgression with Neanderthals and Denisovans. The duplicated genes in these loci, identified as *TNFRSF10D1*, *TNFRSF10D2* and *NPIP16*, show evidence of positive selection based on relatively high frequency of non-synonymous variants compared to the ancestral genes. This highlights them as possible contributors to adaptations in this population. Another example is the identification of a new zinc finger in Dutch genome data, not annotated in the human reference genome, though the function of this particular zinc finger gene is not known (Hehir-Kwa et al. 2016). It does show the possible variability in zinc finger clusters encoded in human genomes from different ancestries. Further exploration of zinc finger clusters in genomes from different backgrounds may show how dynamic they are in present day humans. Similarly, we found genomes with previously unannotated additional copies of *NOTCH2NLR* or *NOTCH2NLC* in the Simons diversity data, for which detailed assemblies may also increase knowledge about sequence variation in the *1p12* and *1q21.1* loci. Given also the unique characteristics of *NOTCH2NL* loci in Neanderthal and Denisova genomes, the inclusion of ancient DNA datasets may provide additional clues on the emergence and trajectory of structural variants in human evolution. Finally, the improvements of reference genome assemblies and their diversification may aid in analysis of the ever expanding genome sequencing datasets. As population biases are still an important confounder in genetics, like genome-wide association studies, increased knowledge about genome diversity could improve accuracy of such studies by accounting for genetic background (Berg et al. 2019; Sohail et al. 2019). Structural variants are also linked to variation in gene expression regulation (Sonay et al. 2015; Chiang et al. 2017; Fotsing et al. 2019). Establishment of databases like the UK Biobank provide an unprecedented amount of genotype - phenotype information to answer questions in genetics research (Van Hout et al. 2019). About 50,000 exome sequencing datasets are now available in the UK Biobank, which will soon be increased to 150,000. The change of scale in analysis of this data, rising orders of magnitude compared to typical studies, has questioned the accuracy of a single reference assembly. To capture variation in alternate

sequences compared to the reference genome, creation of the pan-genome has been suggested, a set of all genomic sequences known in humans (Sherman and Salzberg 2020). Another recent development is the genome graph, which uses new methods for analysis of alternate sequences (haplotypes) and identification of genetic variants in the pan-genome (Paten et al. 2017; Rakocevic et al. 2019). The main advantage is to recover sequence information from reads otherwise lost, by the ability to map to non-reference haplotypes. The collective efforts to improve the reference assembly, reference genome diversity and the huge amount of datasets available provide the tools for the deepest study of human genetics yet. Time will tell if these efforts are enough to fully understand individual genome configurations and their link to specific biological traits.

6.5: Systems biology of evolutionary genomics

Genetic factors underlying human evolution are slowly being uncovered through comparative genomics analysis and functional assessment of human-specific DNA sequences. Especially functions related to the brain, immune system and metabolism are linked to recent human evolution (Charrier et al. 2012; Dennis et al. 2012; Dannemann et al. 2016; Ju et al. 2016; Nédélec et al. 2016; Pontzer et al. 2016; Quach et al. 2016; Reilly and Noonan 2016; Buckley et al. 2017; Liu et al. 2017; Llamas et al. 2017; Fiddes et al. 2018; Florio et al. 2018; Suzuki et al. 2018; Bitar et al. 2019; Perdomo-Sabogal and Nowick 2019; Won et al. 2019). One remaining challenge is to quantify the effect of individual factors into specific signaling pathways and to integrate this into the overarching biological systems (Ebisuya and Briscoe 2018). The development of organoid models, live imaging tools to quantify pathway activity and cell behaviour in culture, as well as new methods for genome-wide functional assays like CRISPR-screens (Dixit et al. 2016; Schmierer et al. 2017; Shifrut et al. 2018; Fei et al. 2019; Feldman et al. 2019; Mattioli et al. 2019), allow for matched comparisons of gene regulation and development between species. NOTCH signalling is one conserved pathway which has been studied extensively in quantitative biology to understand how NOTCH signalling and lateral inhibition is involved in tissue patterning, cell transcriptional states and cell fate decisions (Sprinzak et al. 2010; Sprinzak et al. 2011; Wang et al. 2011; Matsuda et al. 2015; Hunter et al. 2016; Shaya et al. 2017). The foundation of these models have been set by molecular biology work that revealed the basic mechanisms of this pathway (Bray 1998). The combination of these studies can have mutual benefits to better understand evolutionary differences in

signalling pathways. Models can be adjusted to incorporate species-specific information, such as the addition of a new factor in the case of *NOTCH2NL* in the NOTCH pathway. Recent discoveries of overall species-specific molecular properties are also interesting to include, such as increased protein half-life and cell cycle length in human cells compared to mouse cells (Rayon et al. 2019). One system that is studied in detail on many levels is the segmentation clock during somitogenesis (Masamizu et al. 2006; Kageyama et al. 2007; Baker et al. 2008; Morelli et al. 2009; Campanelli and Gedeon 2010; Uriu et al. 2010a; Uriu et al. 2010b; Hester et al. 2011; Ay et al. 2014). Here, oscillatory expression of NOTCH, WNT and FGF pathways are coupled to cyclic generation of somites (Dequéant and Pourquié 2008; Gibb et al. 2010). Different dynamics of these factors oscillation period are associated with differences in developmental timing of somitogenesis, like the size and amount of somites produced (Oates et al. 2012). For example, the human *HES7* oscillation period is twice as long as mouse *Hes7* oscillations, which coincides with the observation that development of human somites takes twice as long as mouse somites (Chu et al. 2019; Matsuda et al. 2019; Diaz-Cuadros et al. 2020). These oscillations are regulated by several parameters like synthesis and degradation rates of *HES7* mRNA and protein, intron delay, and *HES7* auto-inhibition. However, modifying individual parameters is not able to recapitulate species-specific changes. For instance, increasing *Hes7* protein stability or shortening intron delay in zebrafish does not lead to normal *Hes7* oscillation with altered properties, but create an out of balance system that leads to defective somite development (Hirata et al. 2004; Takashima et al. 2011; Harima et al. 2013). Another example is co-culture of human and rhesus monkey neural stem cells, which have different self-renewal and differentiation properties. Co-culture between cells of both species did not alter the output of their respective daughter cells, indicating they are not affected by extrinsic factors exerted by the cells from the other species. Instead, they generated cells at their usual species-specific rate, which may be regulated via intrinsic factors (Otani et al. 2016). An outstanding question is if oscillators like *HES1* and *HES5* expressed in neural stem cells show different dynamics between species, that may be linked to differences in brain development. These observations hint at a system where cell-intrinsic and extrinsic parameters that define tissue development may be tuned together in a species-specific way. Understanding these parameters may be required to explain effects of evolutionary innovations in development, such as regulation of brain size and structure in humans and other species.

6.6: Gene conversion and genetic variation in the earliest stages of life

Our observations regarding *NOTCH2NL* evolution raises questions about the rate of gene conversion in the human genome today. Rates of gene conversion have been reported in various studies, but have been highly variable depending on specific loci (Hurles 2001; Bosch et al. 2004; Padhukasahasram et al. 2004; Yin et al. 2009; Mansai et al. 2011; Glémin et al. 2015; Harpak et al. 2017; Pouyet et al. 2018). Gene conversion is also involved in several diseases, like spinal muscular atrophy for instance, via interlocus exchange of coding variants that disrupt the function of target genes (Chen et al. 2007). It is also known for causing variations in tandem repeat sequences after DNA damage in these regions (Pâques et al. 1998; Jakupciak and Wells 2000; Chuzhanova et al. 2009; Sharma et al. 2013; Massey and Jones 2018). Repeat expansions in specific genes are associated with severe neurological disorders, such as fragile X mental retardation syndrome (FMR) and Huntington's disease (HD) (La Spada and Taylor 2010; Banez-Coronel and Ranum 2019). More recently, a repeat expansion in *NOTCH2NLC* was linked to a neurodegenerative disorder (Deng et al. 2019; Ishiura et al. 2019; Okubo et al. 2019; Sone et al. 2019; Sun et al. 2019; Hayashi et al. 2020; Jiao et al. 2020). Following our findings regarding the instability of *NOTCH2NL* genes, it can be speculated repeat expansions may be caused by interlocus gene conversion. Structural variation in tandem repeat sequences were recently identified as a major component of genetic variation in humans (Audano et al. 2019), which also affects gene expression levels (Sonay et al. 2015; Chiang et al. 2017; Fotsing et al. 2019). Collectively, these findings suggest gene conversion may be an important mechanism shaping the human genome. While it is known that structural variation like copy-number variations and gene conversion are active during meiosis (Baudat et al. 2010; Hunter 2015; Zickler and Kleckner 2015; Gray and Cohen 2016), there is also evidence these can occur in early post-zygotic stages (Ambartsumyan and Clark 2008; Carbone and Chavez 2015; Bolton et al. 2016). During these phases chromosomal aberrations are very common, indicating extensive recombination events. Cells that contain chromosomal aberrations usually undergo apoptosis, and embryogenesis continues with the population of karyotypically normal cells. In the remaining cells, smaller structural variants or gene conversion may be present that are tolerated in formation. Early embryonic cell divisions are also associated with the introduction of new single nucleotide variants, linked to the short cell-cycle during this stage. So, in the early embryo, there may be widespread genetic variation between single cells. This leads to

somatic variation in the genome of progenitor cells, which creates mosaic genetic variations patterns in subsequently developing cells, tissues and organs. Somatic mutations are also involved in some neurological disorders, showing they can have drastic effects on development (Swami et al. 2009; Kurosawa and Ohta 2011; Ross 2011; Jamuar et al. 2014; Kraus-Perrotta and Lagalwar 2016; D’Gama et al. 2017; Lim et al. 2017; Nishioka et al. 2019). Extensive sequencing datasets and new technologies allow more detailed analysis of genome-wide gene conversion events. Especially trio genome sequencing analysis, where datasets are generated for parents and the offspring, is useful in deciphering gene conversion. Based on mendelian inheritance rules, predictions can be made for genetic configurations across the genome in the offspring, based on the genome of the parents. However, many sites in the offspring show mendelian inconsistent calls (MICs), which may be flagged as sequencing errors (Li et al. 2012; Patel et al. 2014; Toptas et al. 2018). Gene conversion could surpass mendelian inheritance patterns, which calls for re-examination of trio sequencing data and MICs to find potential associations with gene conversion. Some studies already suggests MICs are associated with copy number variations and repeat sequences (Eberle et al. 2017; Tatsumoto et al. 2017; Kothiyal et al. 2019; Yauy et al. 2019), which could be another hint that gene conversion is a factor at play. Additionally, monozygotic twin genomes may be used to establish somatic gene conversion rates. As they originate from the same zygote, any genetic variation is expected to occur from somatic changes. Moreover, some monozygotic twins are discordant for neurodevelopmental disorders like autism (Hallmayer et al. 2011; Wong et al. 2014; Castelbaum et al. 2019; Huang et al. 2019; Saffari et al. 2019), indicating that, if the underlying genetics are causing the discordance, significant genetic variation can occur in the early embryo. Single cell whole genome sequencing is a technique previously used to successfully identify somatic variation of single nucleotides and structural variants in neurons (Cai et al. 2014; Lodato et al. 2018). Although technically challenging, recent developments have improved accuracy of variant calling in single cell data (Luquette et al. 2019; Schnepf et al. 2019; Vu et al. 2019; Lei Zhang et al. 2019). Single cell genomes can provide new insights in gene conversion rates and hotspots across the genome. Of particular interest is single cell genome analysis of germ cells and early post zygotic embryonic cells, as gene conversion in these stages causes genetic variation in either all cells or a high percentage of cells in the organism. In vitro modelling of gene conversion mechanisms could be a supporting model to validate findings from genome

sequencing datasets. CRISPR-Cas9 can be used to induce DNA damage at specific loci of interest, followed by experiments to measure the DNA repair outcome and potential gene conversion events. The closest in vitro model representing early embryonic cells may be generation of naïve embryonic stem cells from primed embryonic stem cells (Collier and Rugg-Gunn 2018). Not much is known regarding genome stability in cultured naïve embryonic stem cells, but they do have epigenetic signatures resembling the more unstable genome of preimplantation cell types like genome-wide hypomethylation (Yagi et al. 2017). Further profiling of naïve ESCs are necessary to establish if they are a proper model to study somatic genetic variation. The datasets and tools to investigate gene conversion are becoming more and more numerous. These can be used to address the contribution of gene conversion in evolutionary adaptations in the human genome, as well as present-day genetic variation shaping human features in health and disease.

Concluding remarks

The questions revolving around our origins have been permeating from generation to generation of humankind. With advances in genome research at hand, there now is a much clearer picture of both human history and the function of recently evolved human genome sequences. Large scale structural variation, such as gene duplications, and small scale structural variation, such as tandem repeats and retrotransposon insertions, have greatly affected human genome evolution. The emergence of these sequences have been linked to unstable regions of the genome, which, due to their instability, rearranged and expanded in primate evolution. This allowed fast genomic variation by large scale alteration of genomic configurations, via duplications, deletions, insertions, etcetera. Inherently to this strategy of genome adaptation, the resulting sequences that contributed to human evolution are still prone to rearrangements today, creating variation in the human genome and sometimes causing disease. It appears there is a balance between genome stability and genome evolution, a double edged sword in evolutionary adaptation. Whereas the first milestones in identifying structural variation in the human genome have been set, the challenge remains to completely map and assemble all chromosomes. With the ever increasing amounts of sequencing data available, we are slowly uncovering the full spectrum of individual genome configurations and how they affect human phenotypes. Already, more and more structural variants are being discovered by using more advanced sequencing

techniques, and simply by the increased number of genomes analyzed. Model systems like organoids allow basic investigation of human development. Complementing techniques, like iPSC generation and CRISPR-Cas9 gene editing, have provided tools to investigate the function of human-specific genome sequences in health and disease. Together, these studies have synthesized into our understanding of human evolution so far. Without a doubt, it has become clear that structural genome variation has played a major role in human evolution. We are only beginning to see its impact on human biology in the present day.

References

- Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, Ross Chapman J, Radu Aricescu A, Myers SR. 2017. A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *Elife* 6:1–46.
- Ambartsumyan G, Clark AT. 2008. Aneuploidy and early human embryo development. *Hum. Mol. Genet.* 17:10–15.
- Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AME, Dougherty ML, Nelson BJ, Shah A, Dutcher SK, et al. 2019. Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* 176:663-675.e19.
- Ay A, Holland J, Sperlea A, Devakanmalai GS, Knierer S, Sangervasi S, Stevenson A, Özbudak EM. 2014. Spatial gradients of protein-level time delays set the pace of the traveling segmentation clock waves. *Dev.* 141:4158–4167.
- Baker RE, Schnell S, Maini PK. 2008. Mathematical Models for Somite Formation. *Curr. Top. Dev. Biol.* 81:183–203.
- Banez-Coronel M, Ranum LPW. 2019. Repeat-associated non-AUG (RAN) translation: insights from pathology. *Lab. Investig.* 99:929–942.
- Baudat F, Buard J, Grey C, Fedel-Alon A, Ober C, Przeworski M, Coop G, De Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* (80-.). 327:836–840.
- Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, Boyle EA, Zhang X, Racimo F, Pritchard JK, et al. 2019. Reduced signal for polygenic adaptation of height in UK biobank. *Elife* 8:1–47.
- Bershteyn M, Nowakowski TJ, Pollen AA, Di Lullo E, Nene A, Wynshaw-Boris A, Kriegstein AR. 2017. Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. *Cell Stem Cell* 20:435-449.e4.
- Bitar M, Kuiper S, O'Brien EA, Barry G. 2019. Genes with human-specific features are primarily involved with brain, immune and metabolic evolution. *BMC Bioinformatics* 20:1–12.
- Bolton H, Graham SJL, Van Der Aa N, Kumar P, Theunis K, Fernandez Gallardo E, Voet T, Zernicka-Goetz M. 2016. Mouse model of chromosome mosaicism reveals lineage-specific depletion of aneuploid cells and normal developmental potential. *Nat. Commun.* 7:1–12.

- Bosch E, Hurles ME, Navarro A, Jobling MA. 2004. Dynamics of a human interparalog gene conversion hotspot. *Genome Res.* 14:835–844.
- Bray S. 1998. Notch signalling in *Drosophila*: Three ways to use a pathway. *Semin. Cell Dev. Biol.* 9:591–597.
- Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, Lalani SR, Graham B, Lee B, Shinawi M, et al. 2008. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.* 40:1466–1471.
- Buckley MT, Racimo F, Allentoft ME, Jensen MK, Jonsson A, Huang H, Hormozdiari F, Sikora M, Marnetto D, Eskin E, et al. 2017. Selection in Europeans on fatty acid desaturases associated with dietary changes. *Mol. Biol. Evol.* 34:1307–1318.
- Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A, Walsh CA. 2014. Single-Cell, Genome-wide Sequencing Identifies Clonal Somatic Copy-Number Variation in the Human Brain. *Cell Rep.* 8:1280–1289.
- Campanelli M, Gedeon T. 2010. Somitogenesis clock-wave initiation requires differential decay and multiple binding sites for clock protein. *PLoS Comput. Biol.* 6.
- Carbone L, Chavez SL. 2015. Mammalian pre-implantation chromosomal instability: Species comparison, evolutionary considerations, and pathological correlations. *Syst. Biol. Reprod. Med.* 61:321–335.
- Castelbaum L, Sylvester CM, Zhang Y, Yu Q, Constantino JN. 2019. On the Nature of Monozygotic Twin Concordance and Discordance for Autistic Trait Severity: A Quantitative Analysis. *Behav. Genet.* .
- Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, et al. 2015. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517:608–611.
- Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, et al. 2019. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* 10:1–16.
- Charrier C, Joshi K, Coutinho-Budd J, Kim JE, Lambert N, De Marchena J, Jin WL, Vanderhaeghen P, Ghosh A, Sassa T, et al. 2012. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* 149:923–935.
- Chauhan Santosh, Goodwin JG, Chauhan Swati, Manyam G, Wang J, Kamat AM, Boyd DD. 2013. ZKSCAN3 Is a Master Transcriptional Repressor of Autophagy. *Mol. Cell* 50:16–28.
- Cheetham SW, Faulkner GJ, Dinger ME. 2019. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat. Rev. Genet.* .
- Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: Mechanisms, evolution and human disease. *Nat. Rev. Genet.* 8:762–775.
- Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, Hadzic T, Damani FN, Ganel L, Montgomery SB, et al. 2017. The impact of structural variation on human gene expression. *Nat. Genet.* 49:692–699.

- Chu LF, Mamott D, Ni Z, Bacher R, Liu C, Swanson S, Kendziorski C, Stewart R, Thomson JA. 2019. An In Vitro Human Segmentation Clock Model Derived from Embryonic Stem Cells. *Cell Rep.* 28:2247-2255.e5.
- Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351:1083–1087.
- Chuzhanova N, Chen JM, Bacolla A, Patrinos GP, Férec C, Wells RD, Cooper DN. 2009. Gene conversion causing human inherited disease: Evidence for involvement of non-B-DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. *Hum. Mutat.* 30:1189–1198.
- Collier AJ, Rugg-Gunn PJ. 2018. Identifying Human Naïve Pluripotent Stem Cells – Evaluating State-Specific Reporter Lines and Cell-Surface Markers. *BioEssays* 40:1–12.
- Coulter ME, Dorobantu CM, Lodewijk GA, Delalande F, Cianferani S, Ganesh VS, Smith RS, Lim ET, Xu CS, Pang S, et al. 2018. The ESCRT-III Protein CHMP1A Mediates Secretion of Sonic Hedgehog on a Distinctive Subtype of Extracellular Vesicles. *Cell Rep.* 24:973-986.e8.
- D’Gama AM, Woodworth MB, Hossain AA, Bizzotto S, Hatem NE, LaCoursiere CM, Najm I, Ying Z, Yang E, Barkovich AJ, et al. 2017. Somatic Mutations Activating the mTOR Pathway in Dorsal Telencephalic Progenitors Cause a Continuum of Cortical Dysplasias. *Cell Rep.* 21:3754–3766.
- Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* 98:22–33.
- Dapprich J, Ferriola D, Mackiewicz K, Clark PM, Rappaport E, D’Arcy M, Sasson A, Gai X, Schug J, Kaestner KH, et al. 2016. The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics* 17:1–14.
- Deng J, Gu M, Miao Y, Yao S, Zhu M, Fang P, Yu X, Li P, Su Y, Huang J, et al. 2019. Long-read sequencing identified repeat expansions in the 5’UTR of the NOTCH2NLC gene from Chinese patients with neuronal intranuclear inclusion disease. *J. Med. Genet.*:1–7.
- Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian S, Malig M, Kotkiewicz H, et al. 2012. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* 149:912–922.
- Dequéant M-L, Pourquié O. 2008. Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.* 9:370–382.
- Diaz-Cuadros M, Wagner DE, Budjan C, Hubaud A, Tarazona OA, Donnelly S, Michaut A, Al Tanoury Z, Yoshioka-Kobayashi K, Niino Y, et al. 2020. In vitro characterization of the human segmentation clock. *Nature*.
- Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167:1853-1866.e17.
- Dréau A, Venu V, Avdievich E, Gaspar L, Jones FC. 2019. Genome-wide recombination map construction from single individuals using linked-read sequencing. *Nat. Commun.* 10.

Chapter 6

Eberle MA, Fritzilas E, Krusche P, Källberg M, Moore BL, Bekritsky MA, Iqbal Z, Chuang HY, Humphray SJ, Halpern AL, et al. 2017. A reference dataset of 5 . 4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree Supplementary Information Content 1 . Identifying and Assessing Pedigree Consistent Variants 2 . Analysis of. *Genome Res.* 27:1–31.

Ebisuya M, Briscoe J. 2018. What does time mean in development? *Dev.* 145:0–3.

Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Dev.* 144:2719–2729.

Farmiloe G, Lodewijk GA, Robben SF, van Bree EJ, Jacobs FMJ. 2020. Widespread correlation of KRAB zinc finger protein binding with brain-developmental gene expression patterns. *Philos. Trans. R. Soc. B.*

Fei T, Li W, Peng J, Xiao T, Chen CH, Wu A, Huang J, Zang C, Liu XS, Brown M. 2019. Deciphering essential cistromes using genome-wide CRISPR screens. *Proc. Natl. Acad. Sci. U. S. A.* 116:25186–25195.

Feldman D, Singh A, Schmid-Burgk JL, Carlson RJ, Mezger A, Garrity AJ, Zhang F, Blainey PC. 2019. Optical Pooled Screens in Human Cells. *Cell* 179:787-799.e17.

Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, et al. 2018. Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* 173:1356-1369.e22.

Field AR, Jacobs FMJ, Fiddes IT, Phillips APR, Reyes-Ortiz AM, LaMontagne E, Whitehead L, Meng V, Rosenkrantz JL, Olsen M, et al. 2019. Structurally Conserved Primate LncRNAs Are Transiently Expressed during Human Cortical Differentiation and Influence Cell-Type-Specific Genes. *Stem Cell Reports* 12:245–257.

Fietz SA, Kelava I, Vogt J, Wilsch-Bräuninger M, Stenzel D, Fish JL, Corbeil D, Riehn A, Distler W, Nitsch R, et al. 2010. OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* 13:690–699.

Florio M, Albert M, Taverna E, Namba T, Brandl H, Lewitus E, Haffner C, Sykes A, Wong FK, Peters J, et al. 2015. Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. *Science* (80-.). 347:1465 LP – 1470.

Florio M, Heide M, Pinson A, Brandl H, Albert M, Winkler S, Wimberger P, Huttner WB, Hiller M. 2018. Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *Elife* 7:1–37.

Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M. 2019. The impact of short tandem repeat variation on gene expression. *Nat. Genet.* 51:1652–1659.

Frega M, Linda K, Keller JM, Gümüş-Akay G, Mossink B, van Rhijn JR, Negwer M, Klein Gunnewiek T, Foreman K, Kompier N, et al. 2019. Neuronal network dysfunction in a model for Kleefstra syndrome mediated by enhanced NMDAR signaling. *Nat. Commun.* 10:1–15.

Gibb S, Maroto M, Dale JK. 2010. The segmentation clock mechanism moves up a notch. *Trends Cell Biol.* 20:593–600.

Glémin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L. 2015. Quantification of GC-biased gene conversion in the human genome. *Genome Res.* 25:1215–1228.

- Gray S, Cohen PE. 2016. Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annu. Rev. Genet.* 50:175–210.
- Gunz P, Tilot AK, Wittfeld K, Teumer A, Shapland CY, van Erp TGM, Dannemann M, Vernot B, Neubauer S, Guadalupe T, et al. 2019. Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity. *Curr. Biol.* 29:120-127.e5.
- Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, Miller J, Fedele A, Collins J, Smith K, et al. 2011. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* 68:1095–1102.
- Hansen D V., Lui JH, Parker PRL, Kriegstein AR. 2010. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* 464:554–561.
- Harima Y, Takashima Y, Ueda Y, Ohtsuka T, Kageyama R. 2013. Accelerating the Tempo of the Segmentation Clock by Reducing the Number of Introns in the *Hes7* Gene. *Cell Rep.* 3:1–7.
- Harpak A, Lan X, Gao Z, Pritchard JK. 2017. Frequent nonallelic gene conversion on the human lineage and its effect on the divergence of gene duplicates. *Proc. Natl. Acad. Sci. U. S. A.* 114:12779–12784.
- Hayashi T, Katagiri S, Mizobuchi K, Yoshitake K, Kameya S, Matsuura T, Iwata T, Nakano T. 2020. Heterozygous GGC repeat expansion of *NOTCH2NLC* in a patient with neuronal intranuclear inclusion disease and progressive retinal dystrophy. *Ophthalmic Genet.* 00:1–3.
- Hehir-Kwa JY, Marschall T, Kloosterman WP, Francioli LC, Baaijens JA, Dijkstra LJ, Abdellaoui A, Koval V, Thung DT, Wardenaar R, et al. 2016. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat. Commun.* 7:1–10.
- Helleboid P, Heusel M, Duc J, Piot C, Thorball CW, Coluccio A, Pontis J, Imbeault M, Turelli P, Aebersold R, et al. 2019. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. *EMBO J.* 38:1–16.
- Hester SD, Belmonte JM, Gens JS, Clendenon SG, Glazier JA. 2011. A multi-cell, multi-scale model of vertebrate segmentation and somite formation. *PLoS Comput. Biol.* 7.
- Hirata H, Bessho Y, Kokubu H, Masamizu Y, Yamada S, Lewis J, Kageyama R. 2004. Instability of *Hes7* protein is crucial for the somite segmentation clock. *Nat. Genet.* 36:750–754.
- Van Hout CV, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, Gonzaga-Jauregui C, Khalid S, Liu D, Banerjee N, et al. 2019. Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv* :572347.
- Hsieh P, Vollger MR, Dang V, Porubsky D, Baker C, Cantsilieris S, Hoekzema K, Lewis AP, Munson KM, Sorensen M, et al. 2019. Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes. *Science* (80-.). 366:eaax2083.
- Huang Y, Zhao Y, Ren Y, Yi Y, Li X, Gao Z, Zhan X, Yu J, Wang D, Liang S, et al. 2019. Identifying Genomic Variations in Monozygotic Twins Discordant for Autism Spectrum Disorder Using Whole-Genome Sequencing. *Mol. Ther. - Nucleic Acids* 14:204–211.

- Hunter GL, Hadjivasiliou Z, Bonin H, He L, Perrimon N, Charras G, Baum B. 2016. Coordinated control of notch/delta signalling and cell cycle progression drives lateral inhibition-mediated tissue patterning. *Dev.* 143:2305–2310.
- Hunter N. 2015. Meiotic recombination: The essence of heredity. *Cold Spring Harb. Perspect. Biol.* 7.
- Hurles ME. 2001. Gene conversion homogenizes the CMT1A paralogous repeats. *BMC Genomics* 2:1–9.
- Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543:550–554.
- Ishiura H, Shibata S, Yoshimura J, Suzuki Y, Qu W, Doi K, Almansour MA, Kikuchi JK, Taira M, Mitsui J, et al. 2019. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* 51:1222–1232.
- Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17:1–11.
- Jain M, Olsen HE, Turner DJ, Stoddart D, Bulazel K V., Paten B, Haussler D, Willard HF, Akeson M, Miga KH. 2018. Linear assembly of a human centromere on the y chromosome. *Nat. Biotechnol.* 36:321–323.
- Jakupciak JP, Wells RD. 2000. Gene conversion (recombination) mediates expansions of CTG-CAG repeats. *J. Biol. Chem.* 275:40003–40013.
- Jamuar SS, Lam ATN, Kircher M, D’Gama AM, Wang J, Barry BJ, Zhang X, Hill RS, Partlow JN, Rozzo A, et al. 2014. Somatic mutations in cerebral cortical malformations. *N. Engl. J. Med.* 371:733–743.
- Jiao B, Zhou Lu, Zhou Y, Weng L, Liao X, Tian Y, Guo L, Liu X, Yuan Z, Xiao X, et al. 2020. Identification of expanded repeats in NOTCH2NLC in neurodegenerative dementias. *Neurobiol. Aging* .
- Ju X-C, Hou Q-Q, Sheng A-L, Wu K-Y, Zhou Y, Jin Y, Wen T, Yang Z, Wang X, Luo Z-G. 2016. The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. *Gleeson JG, editor. Elife* 5:e18197.
- Kageyama R, Masamizu Y, Niwa Y. 2007. Oscillator mechanism of notch pathway in the segmentation clock. *Dev. Dyn.* 236:1403–1409.
- Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchís-Calleja F, Gujjarro P, Sidow L, Fleck JS, Han D, et al. 2019. Organoid single-cell genomic atlas uncovers human-specific features of brain development.
- Kauzlaric A, Jang SM, Morchikh M, Cassano M, Planet E, Benkirane M, Trono D. 2020. KAP1 targets actively transcribed genomic loci to exert pleomorphic effects on RNA polymerase II activity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 375:20190334.
- Kothiyal P, Wong WSW, Bodian DL, Niederhuber JE. 2019. Mendelian Inconsistent Signatures from 1314 Ancestrally Diverse Family Trios Distinguish Biological Variation from Sequencing Error. *J. Comput. Biol.* 26:405–419.

- Kraus-Perrotta C, Lagalwar S. 2016. Expansion, mosaicism and interruption: mechanisms of the CAG repeat mutation in spinocerebellar ataxia type 1. *Cerebellum & Ataxias* 3:1–11.
- Kurosawa K, Ohta K. 2011. Genetic diversification by somatic gene conversion. *Genes (Basel)*. 2:48–58.
- Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AKY, McCaffrey J, Young E, Lam ET, Hastie AR, Wong KHY, et al. 2019. Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nat. Commun.* 10:1–14.
- Li B, Chen W, Zhan X, Busonero F, Sanna S, Sidore C, Cucca F, Kang HM, Abecasis GR. 2012. A Likelihood-Based Framework for Variant Calling and De Novo Mutation Detection in Families. *PLoS Genet.* 8.
- Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, D’Gama AM, Kim SN, Hill RS, Goldberg AP, et al. 2017. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* 20:1217–1224.
- Liu J, Liu W, Yang L, Wu Q, Zhang H, Fang A, Li L, Xu X, Sun L, Zhang J, et al. 2017. The Primate-Specific Gene TMEM14B Marks Outer Radial Glia Cells and Promotes Cortical Expansion and Folding. *Cell Stem Cell* 21:635–649.e8.
- Llamas B, Willerslev E, Orlando L. 2017. Human evolution: A tale from ancient genomes. *Philos. Trans. R. Soc. B Biol. Sci.* 372.
- Lodato MA, Rodin RE, Bohrsen CL, Coulter ME, Barton AR, Kwon M, Sherman MA, Vitzthum CM, Luquette LJ, Yandava CN, et al. 2018. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science (80-.)*. 359:555–559.
- Lovett ML, Nieland TJF, Dingle YL, Kaplan DL. 2020. Innovations in 3D Tissue Models of Human Brain Physiology and Diseases. 1909146:1–27.
- Luquette LJ, Bohrsen CL, Sherman MA, Park PJ. 2019. Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat. Commun.* 10.
- Mansai SP, Kado T, Innan H. 2011. The rate and tract length of gene conversion between duplicated genes. *Genes (Basel)*. 2:313–331.
- Marks P, Garcia S, Barrio AM, Belhocine K, Bernate J, Bharadwaj R, Bjornson K, Catalanotti C, Delaney J, Fehr A, et al. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Res.* 29:635–645.
- Masamizu Y, Ohtsuka T, Takashima Y, Nagahara H, Takenaka Y, Yoshikawa K, Okamura H, Kageyama R. 2006. Real-time imaging of the somite segmentation clock: Revelation of unstable oscillators in the individual presomitic mesoderm cells. *Proc. Natl. Acad. Sci. U. S. A.* 103:1313–1318.
- Massey TH, Jones L. 2018. The central role of DNA damage and repair in CAG repeat diseases. *DMM Dis. Model. Mech.* 11.
- Matsuda M, Hayashi H, Garcia-Ojalvo J, Yoshioka-Kobayashi K, Kageyama R, Yamanaka Y, Ikeya M, Toguchida J, Alev C, Ebisuya M. 2019. Species-specific oscillation periods of human and mouse

Chapter 6

segmentation clocks are due to cell autonomous differences in biochemical reaction parameters. *bioRxiv* :650648.

Matsuda M, Koga M, Woltjen K, Nishida E, Ebisuya M. 2015. Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat. Commun.* 6:1–12.

Mattioli K, Oliveros W, Gerhardinger C, Andergassen D, Maass PG, Rinn JL, Melé M. 2019. Analysis of regulatory element evolution between human and mouse reveals a lack of cis-trans compensation. *bioRxiv* :847491.

Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS. 2014. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* 9:989–1009.

Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon GA, et al. 2019. Telomere-to-telomere assembly of a complete human X chromosome. *bioRxiv*:735928.

Mochida GH, Ganesh VS, De Michelena MI, Dias H, Atabay KD, Kathrein KL, Huang HT, Sean Hill R, Felie JM, Rakiec D, et al. 2012. CHMP1A encodes an essential regulator of BMI1-INK4A in cerebellar development. *Nat. Genet.* 44:1260–1264.

Morelli LG, Ares S, Herrgen L, Schröter C, Jülicher F, Oates AC. 2009. Delayed coupling theory of vertebrate segmentation. *HFSP J.* 3:55–66.

Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, et al. 2016. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* 167:657-669.e21.

Neubauer S, Hublin J-J, Gunz P. 2018. The evolution of modern human brain shape. *Sci. Adv.* 4:eaa05961.

Nishioka M, Bundo M, Iwamoto K, Kato T. 2019. Somatic mutations in the human brain: implications for psychiatric research. *Mol. Psychiatry* 24:839–856.

Nowick K, Fields C, Gernat T, Caetano-Anolles D, Kholina N, Stubbs L. 2011. Gain, loss and divergence in primate Zinc-Finger genes: A rich resource for evolution of gene regulatory differences between species. *PLoS One* 6.

Oates AC, Morelli LG, Ares S. 2012. Patterning embryos with oscillations: Structure, function and dynamics of the vertebrate segmentation clock. *Development* 139:625–639.

Okubo M, Doi H, Fukai R, Fujita A, Mitsuhashi S, Hashiguchi S, Kishida H, Ueda N, Morihara K, Ogasawara A, et al. 2019. GGC Repeat Expansion of NOTCH2NLC in Adult Patients with Leukoencephalopathy. *Ann. Neurol.* 86:962–968.

Ostrem BEL, Lui JH, Gertz CC, Kriegstein AR. 2014. Control of outer radial glial stem cell mitosis in the human brain. *Cell Rep.* 8:656–664.

Otani T, Marchetto MC, Gage FH, Simons BD, Livesey FJ. 2016. 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. *Cell Stem Cell* 18:467–480.

- Padhukasahasram B, Marjoram P, Nordborg M. 2004. Estimating the rate of gene conversion on human chromosome 21. *Am. J. Hum. Genet.* 75:386–397.
- Pâques F, Leung W-Y, Haber JE. 1998. Expansions and Contractions in a Tandem Repeat Induced by Double-Strand Break Repair. *Mol. Cell. Biol.* 18:2045–2054.
- Patel ZH, Kottyan LC, Lazaro S, Williams MS, Ledbetter DH, Tromp G, Rupert A, Kohram M, Wagner M, Husami A, et al. 2014. The struggle to find reliable results in exome sequencing data: Filtering out Mendelian errors. *Front. Genet.* 5:1–13.
- Paten B, Novak AM, Eizenga JM, Garrison E. 2017. Genome graphs and the evolution of genome inference. *Genome Res.* 27:665–676.
- Perdomo-Sabogal Á, Nowick K. 2019. Genetic Variation in Human Gene Regulatory Factors Uncovers Regulatory Roles in Local Adaptation and Disease. *Genome Biol. Evol.* 11:2178–2193.
- Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, Di Lullo E, Alvarado B, Bedolli M, Dougherty ML, et al. 2019. Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* 176:743-756.e17.
- Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, et al. 2015. Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell* 163:55–67.
- Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* 24:724-735.e5.
- Pontzer H, Brown MH, Raichlen DA, Dunsworth H, Hare B, Walker K, Luke A, Dugas LR, Durazo-Arvizu R, Schoeller D, et al. 2016. Metabolic acceleration and the evolution of human brain size and life history. *Nature* 533:390–392.
- Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* 7:1–21.
- Quach H, Rotival M, Pothlichet J, Loh YHE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, et al. 2016. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* 167:643-656.e17.
- Rakocevic G, Semenyuk V, Lee WP, Spencer J, Browning J, Johnson IJ, Arsenijevic V, Nadj J, Ghose K, Suci MC, et al. 2019. Fast and accurate genomic analyses using genome graphs. *Nat. Genet.* 51:354–362.
- Rayon T, Stamatakis D, Perez-Carrasco R, Garcia-Perez L, Barrington C, Melchionda M, Exelby K, Tybulewicz V, Fisher EMC, Briscoe J. 2019. Species-specific developmental timing is associated with global differences in protein stability in mouse and human. *bioRxiv* :2019.12.29.889543.
- Reilly SK, Noonan JP. 2016. Evolution of Gene Regulation in Humans. *Annu. Rev. Genomics Hum. Genet.* 17:45–67.

Ross KA. 2011. Evidence for somatic gene conversion and deletion in bipolar disorder, Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis, type-1 diabetes, and type-2 diabetes. *BMC Med.* 9:1–29.

Saffari A, Arno M, Nasser E, Ronald A, Wong CCY, Schalkwyk LC, Mill J, Dudbridge F, Meaburn EL. 2019. RNA sequencing of identical twins discordant for autism reveals blood-based signatures implicating immune and transcriptional dysregulation. *Mol. Autism* 10:1–15.

Schmid CD, Bucher P. 2010. MER41 repeat sequences contain inducible STAT1 binding sites. *PLoS One* 5.

Schmierer B, Botla SK, Zhang J, Turunen M, Kivioja T, Taipale J. 2017. CRISPR/Cas9 screening using unique molecular identifiers. *Mol. Syst. Biol.* 13:945.

Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, Yin Y, Jolma A, Zhong G, Guo H, Kanagalingam T, et al. 2016. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.* 26:1742–1752.

Schnepp PM, Chen M, Keller ET, Zhou X. 2019. SNV identification from single-cell RNA sequencing data. *Hum. Mol. Genet.* 28:3569–3583.

Sharma A, Wolfgruber TK, Presting GG. 2013. Tandem repeats derived from centromeric retrotransposons. *BMC Genomics* 14.

Shaya O, Binshtok U, Hersch M, Rivkin D, Weinreb S, Amir-Zilberstein L, Khamaisi B, Oppenheim O, Desai RA, Goodyear RJ, et al. 2017. Cell-Cell Contact Area Affects Notch Signaling and Notch-Dependent Patterning. *Dev. Cell* 40:505-511.e6.

Sherman RM, Salzberg SL. 2020. Pan-genomics in the human genome era. *Nat. Rev. Genet.* .

Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, Bui CT, Li PJ, Diolaiti ME, Ashworth A, Marson A. 2018. Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* 175:1958-1971.e15.

Shin JH, Ko HS, Kang H, Lee Yunjong, Lee Yun Il, Pletinkova O, Troconso JC, Dawson VL, Dawson TM. 2011. PARIS (ZNF746) repression of PGC-1 α contributes to neurodegeneration in parkinson's disease. *Cell* 144:689–702.

Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CWK, Hirschhorn J, Daly MJ, Patterson N, et al. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. Nordborg M, McCarthy MI, Nordborg M, Barton NH, Hermisson J, editors. *Elife* 8:e39702.

Sonay TB, Carvalho T, Robinson MD, Greminger MP, Krützen M, Comas D, Highnam G, Mittelman D, Sharp A, Marques-Bonet T, et al. 2015. Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. *Genome Res.* 25:1591–1599.

Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, et al. 2019. Long-read sequencing identifies GGC repeat expansions in NOTCH2NL associated with neuronal intranuclear inclusion disease. *Nat. Genet.* 51:1215–1221.

La Spada AR, Taylor JP. 2010. Repeat expansion disease: Progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* 11:247–258.

- Sprinzak D, Lakhanpal A, LeBon L, Garcia-Ojalvo J, Elowitz MB. 2011. Mutual inactivation of Notch receptors and ligands facilitates developmental patterning. *PLoS Comput. Biol.* 7.
- Sprinzak D, Lakhanpal A, Lebon L, Santat LA, Fontes ME, Anderson GA, Garcia-Ojalvo J, Elowitz MB. 2010. Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* 465:86–90.
- Sun QY, Xu Q, Tian Y, Hu ZM, Qin LX, Yang JX, Huang W, Xue J, Li JC, Zeng S, et al. 2019. Expansion of GGC repeat in the human-specific NOTCH2NLC gene is associated with essential tremor. *Brain* 143:222–233.
- Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, et al. 2018. Human-Specific NOTCH2NLC Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* 173:1370–1384.e16.
- Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, Wheeler VC. 2009. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* 18:3039–3047.
- Takashima Y, Ohtsuka T, González A, Miyachi H, Kageyama R. 2011. Intronic delay is essential for oscillatory expression in the segmentation clock. *Proc. Natl. Acad. Sci. U. S. A.* 108:3300–3305.
- Tatsumoto S, Go Y, Fukuta K, Noguchi H, Hayakawa T, Tomonaga M, Hirai H, Matsuzawa T, Agata K, Fujiyama A. 2017. Direct estimation of de novo mutation rates in a chimpanzee parent-offspring trio by ultra-deep whole genome sequencing. *Sci. Rep.* 7:1–12.
- Toptas B, Rakocevic G, Kómar P, Kural D. 2018. Comparing complex variants in family trios. *Bioinformatics* 34:4241–4247.
- Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, Wang A, Wu W, Haddad GG, Chaim IA, et al. 2019. Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* 25:558–569.e7.
- Uriu K, Morishita Y, Iwasa Y. 2010a. Random cell movement promotes synchronization of the segmentation clock. *Proc. Natl. Acad. Sci. U. S. A.* 107:4979–4984.
- Uriu K, Morishita Y, Iwasa Y. 2010b. Synchronized oscillation of the segmentation clock gene in vertebrate development. *J. Math. Biol.* 61:207–229.
- Vollger MR, Dishuck PC, Sorensen M, Welch AE, Dang V, Dougherty ML, Graves-Lindsay TA, Wilson RK, Chaisson MJP, Eichler EE. 2019. Long-read sequence and assembly of segmental duplications. *Nat. Methods* 16:88–94.
- Vu TN, Nguyen HN, Calza S, Kalari KR, Wang L, Pawitan Y, Berger B. 2019. Cell-level somatic mutation detection from single-cell RNA sequencing. *Bioinformatics* 35:4679–4687.
- Wang L, Hou S, Han YG. 2016. Hedgehog signaling promotes basal progenitor expansion and the growth and folding of the neocortex. *Nat. Neurosci.* 19:888–896.
- Wang R, Liu K, Chen L, Aihara K. 2011. Neural fate decisions mediated by trans-activation and cis-inhibition in Notch signaling. *Bioinformatics* 27:3158–3165.
- Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. 2015. Exome sequencing: Current and future perspectives. *G3 Genes, Genomes, Genet.* 5:1543–1550.

Chapter 6

Won H, Huang J, Opland CK, Hartl CL, Geschwind DH. 2019. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat. Commun.* 10:1–11.

Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, Plomin R, Mill J. 2014. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol. Psychiatry* 19:495–503.

Yagi M, Yamanaka S, Yamada Y. 2017. Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. *Lab. Investig.* 97:1133–1141.

Yauy K, de Leeuw N, Yntema HG, Pfundt R, Gilissen C. 2019. Accurate detection of clinically relevant uniparental disomy from exome sequencing data. *Genet. Med.* 0:1–6.

Yin J, Jordan MI, Song YS. 2009. Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. *Bioinformatics* 25:231–239.

Zhang Lei, Dong X, Lee M, Maslov AY, Wang T, Vijg J. 2019. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl. Acad. Sci. U. S. A.* 116:9014–9019.

Zhang Lu, Zhou X, Weng Z, Sidow A. 2019. Assessment of human diploid genome assembly with 10x Linked-Reads data. *Gigascience* 8:1–11.

Zickler D, Kleckner N. 2015. Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harb. Perspect. Biol.* :1,2.