Autism-associated Genes, Circuits & Behaviour in Zebrafish

Marcus Ghosh

Department of Cell & Developmental Biology
University College London

2018

A thesis presented in partial fulfilment of the degree of Doctor of Philosophy, University College London
Declaration

I, Marcus Ghosh confirm that the work presented herein is my own. When information has been derived from other sources, I confirm that this has been acknowledged.
“Ideas are like fish.

If you want to catch little fish, you can stay in the shallow water.

But if you want to catch the big fish, you’ve got to go deeper.

Down deep, the fish are more powerful and more pure.

They’re huge and abstract.

And they’re very beautiful.”

David Lynch, Catching the Big Fish.
Abstract

Autism spectrum disorder (autism) is a debilitating and highly prevalent disorder, thought to affect approximately one percent of the population. Clinically autism is defined by deficits in social interaction and abnormal repetitive behaviours, though the rate of comorbidities, including epilepsy and sleep disorders, is high. Recent genetic studies have linked hundreds of genes with diverse functions to autism. However, the mechanisms by which these genetic lesions lead to autism remain unclear and consequently there are currently no effective treatments available. To uncover these mechanisms, I developed a generalisable framework for mapping the behavioural and neuronal effects of autism-associated mutations in zebrafish, an ideal model system for this approach given their genetic tractability, vertebrate brain plan and amenability to pharmacological screening. At a behavioural level I used a high-throughput set-up to monitor the activity of hundreds of zebrafish larvae across multiple days and nights and developed machine learning tools to describe behaviour in this assay. This approach provided insight into the organisation of zebrafish behaviour across the day/night cycle as well as the effects of drug exposure. Extending this methodology to mutants in two autism-associated genes, chd8 and bckdk, identified behavioural abnormalities in both, demonstrating the power of this approach. At a circuit level I used a registration algorithm to align whole-brain ‘snapshots’ of neuronal activity in three dimensions to identify areas with differential activity in mutant animals. Mapping the chd8 behavioural phenotype to circuits revealed increased neuronal activity in the right, ventral tectum stratum periventriculare in chd8 mutants, providing a starting point for dissecting how this mutation impacts the brain and behaviour. Ultimately, the work described herein establishes a generalisable framework for uncovering the mechanisms by which genetic mutations lead to disease as well as for identifying potential pharmacological therapeutics for such disorders.
Impact Statement

Recently, genetic studies have linked hundreds of genes to autism risk, providing a starting point for uncovering the mechanisms underlying autism as well as for developing treatments. These are pressing questions, given autism's high prevalence and the lack of effective therapeutics, which together generate an annual societal cost in the United Kingdom higher than the combined costs of cancer, chronic heart disease and stroke. One approach to addressing these questions is to use animal models in which the effects of autism-associated mutations can be studied in ways not possible in humans. Zebrafish are particularly well suited to this purpose given their genetic tractability, vertebrate brain plan and amenability to pharmacological screening.

To study the behavioural effects of autism-associated mutations in zebrafish I developed a machine learning approach, suited to the study of large scale datasets. This approach, which will be made publicly available through publication, is a substantial improvement over previous methodologies and will likely find general use in the zebrafish and other animal model communities. In particular this approach is well suited to addressing questions that require either long timescale recordings, such as studies of sleep or ageing, or high-throughput such as pharmacological or genetic screens.

To uncover the causes of abnormal mutant behaviour, I set-up tools to compare whole-brain activity between typical and mutant animals. The general utility of this approach is demonstrated by the uptake of this method by colleagues at UCL to address questions including the effects of drug exposure and injection of the Alzheimer’s related amyloid beta peptide upon neuronal activity. I anticipate that it will continue to provide insight into a variety of such questions in future.

The application of these tools to a range of datasets revealed insights that could not have been gained through previous methods, providing starting points for uncovering several pharmacological and genetic mechanisms. Ultimately, the work described herein establishes a generalisable framework for uncovering the mechanisms linking genetic mutations to disease and is poised to facilitate future drug discovery efforts. Given the rapid pace at which genes are being linked to disorders including depression, schizophrenia and autism-spectrum disorder, such a framework is key.
Acknowledgements

A PhD involves a symbiotic relationship between student and research, the nature of which is complicated, fluctuating between competition (defined as mutual harm), amensalism (where one is harmed, with no loss or gain to the other) and even parasitism (where one benefits at the others expense). Mostly, however, this relationship can be described as mutualisistic, with both student and research benefiting as they learn to successfully co-exist. Like many organisms this symbiont, requires care to develop properly. Below is a non-exhaustive list of those who have taken the time to nurture me (and my research), to whom I’m thankful.

To Jason Rihel, for taking me under your fin as first an IBSc and later MBPhD student. To you no experiment is too ambitious and every experiment is exciting. Your contagious passion carried me through my PhD and I’m sure will carry me on to other adventures in future.

To my parents, who have always inspired me with their caring and selfless nature’s.

To Judith Gardom and Bruce Kinsey, for introducing me first to philosophy and then neuroscience, thank you for sharing your love of knowledge with me.

To everyone on the first floor, thank you for all of the help, support and cake.

To Isaac, thank you for taking the time mentor me.

To Pedro, Eirinn and Ricardo, thank you for looking after me like brothers.

To Sumi, for looking after my experiments, my fish and myself.

To Swarali, Aisling and Winnie, for being great students.

To staff at UCL, for all of the support. Particularly, the team in stores, the fish facility, the IT and Remote Computing Teams, and the Burrito man.

To Daniel Bendor and Michael Orger for my stimulating viva.

To the MBPhD program and MRC, for this opportunity.

Penultimately, to friends both old and new, though for life: Hassan, Al, Tom, Connor, Jazz, Helen, Kirandeep, Ali and Marshal.

Finally, to Ida who makes me a better person everyday.
To friends and family,
# Table of Contents

Declaration .......................................................... 2  
Abstract .................................................................. 4  
Impact Statement .................................................. 5  
Acknowledgements ............................................... 6  
Abbreviations ....................................................... 13  

Chapter One: Introduction .................................... 14  
From Genes to Behaviour ........................................ 14  
Autism Spectrum Disorder ..................................... 15  
  Autism’s Aetiology ............................................. 15  
  Common Variants ............................................ 16  
  Rare Inherited Alleles ..................................... 16  
  De novo Mutations ......................................... 17  
From Genes to Mechanisms ................................... 17  
  Network Approaches .................................... 18  
  Single Model Approaches ............................... 18  
  CNTNAP2 .................................................... 19  
  CHD8. ......................................................... 19  
  Predictive Validity in Autism Models ................. 20  
Zebrafish ............................................................. 21  
  Single Gene Models ..................................... 22  
  Screening Genetic Models ......................... 23  
Aims & Approaches ............................................. 24  

Chapter Two: Behaviour ....................................... 25  
Quantitative Descriptions of Behaviour .................. 25  
Behavioural Analysis Pipeline ............................... 27  
  Improved Behavioural Resolution .................... 28  
  Identifying Behavioural Modules ..................... 32  
  Hierarchical Compression ............................... 36  
  Context Specific Behavioural Motifs ............... 42  
Chd8 and bckdk Mutant Behaviour .......................... 47  
Discussion .......................................................... 51  
  Low-dimensional Representations of Behaviour ...... 51  
  Modular Descriptions of Behaviour ................. 52
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised Motif Selection</td>
<td>91</td>
</tr>
<tr>
<td>Genotyping</td>
<td>92</td>
</tr>
<tr>
<td>DNA Extraction</td>
<td>92</td>
</tr>
<tr>
<td>PCR</td>
<td>92</td>
</tr>
<tr>
<td>KASP</td>
<td>92</td>
</tr>
<tr>
<td>Iconography</td>
<td>93</td>
</tr>
<tr>
<td>Immunohistochemistry</td>
<td>93</td>
</tr>
<tr>
<td>Fluorescent In-situ Hybridisation</td>
<td>93</td>
</tr>
<tr>
<td>pERK/tERK Immunohistochemistry</td>
<td>94</td>
</tr>
<tr>
<td>Microscopy</td>
<td>95</td>
</tr>
<tr>
<td>Pharmacology</td>
<td>95</td>
</tr>
<tr>
<td>Bibliography</td>
<td>97</td>
</tr>
<tr>
<td>Appendix of Figures</td>
<td>109</td>
</tr>
<tr>
<td>Appendix of Tables</td>
<td>122</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Long Time-scale Behavioural Tracking ........................................... 30
Figure 2. Behaviour Varies With Behavioural Context ................................. 31
Figure 3. Unsupervised Learning Identifies Behavioural Modules .. .......................... 34
Figure 4. Module Usage Varies Across the Day/Night ................................. 35
Figure 5. Hierarchical Compression Reveals Structure in Behaviour . ............. 39
Figure 6. Compressibility Varies With Time ............................. .......................... 40
Figure 7. Compression Identifies Motifs in Zebrafish Behaviour .......... ................. 41
Figure 8. Supervised Learning Identifies Day/Night Motifs .......................... 45
Figure 9. Pharmacologically-relevant Behavioural Motifs ......................... 46
Figure 10. ASD Mutant Inactive Module Phenotypes ................................. 49
Figure 11. Compression of ASD Mutant Behaviour ..................................... 50
Figure 12. Establishing Registration Tools ................................................... 62
Figure 13. Whole-Brain Activity Mapping in chd8 Mutants ......................... 66
Figure 14. Neuronal Activity Across Genotypes, Time and Hemisphere ......... 67
Figure 15. Phenotypic Sub-typing in a Prospective ASD Atlas ........... .......................... 75

Appendix Figure 1. Behavioural Set-up & Analysis ..................................... 110
Appendix Figure 2. Bout Features .............................................................. 111
Appendix Figure 3. Module Details ............................................................. 113
Appendix Figure 4. Evidence Accumulation Based Clustering ................. .......................... 114
Appendix Figure 5. Melatonin, PTZ and hcrtr Module Probabilities .... .......................... 115
Appendix Figure 6. Hierarchical Compression Metrics ............................... 116
Appendix Figure 7. Motif Classifier Performance ..................................... 117
Appendix Figure 8. ASD Mutant Bout Features ........................................ 118
Appendix Figure 9. ASD Mutant Module Probabilities ............................. 119
Appendix Figure 10. Behaviour Analysis Pipeline .................................. 120
Appendix Figure 11. Imaging Analysis Pipeline ........................................ 121
List of Tables

Appendix Table 1. Wild Type Motif Classifier Performance . . . . . . . 123
Appendix Table 2. hcrtr and Pharmacological Classifiers . . . . . . . . 124
Appendix Table 3. ASD Classifier Performance . . . . . . . . . . . . 125
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Autism spectrum disorder</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BCAA</td>
<td>Branched chain amino acids</td>
</tr>
<tr>
<td>CaMPARI</td>
<td>Calcium modulated photoactivatable ratiometric integrator</td>
</tr>
<tr>
<td>ChIP-sequencing</td>
<td>Chromatin immunoprecipitation sequencing</td>
</tr>
<tr>
<td>CMTK</td>
<td>Computational morphometry toolkit</td>
</tr>
<tr>
<td>Cv</td>
<td>Cross validated</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>dpf</td>
<td>Days post-fertilisation</td>
</tr>
<tr>
<td>Het</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>Hom.</td>
<td>Homozygous</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>KASP</td>
<td>Kompetitive allele specific PCR</td>
</tr>
<tr>
<td>Mc</td>
<td>Majority class</td>
</tr>
<tr>
<td>mGluR5</td>
<td>Metabotropic glutamate receptor 5</td>
</tr>
<tr>
<td>mRMR</td>
<td>Minimal-redundancy-maximal-relevance</td>
</tr>
<tr>
<td>NMDA</td>
<td>N-methyl-D-aspartate receptor</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pERK</td>
<td>Phosphorylated-extracellular signal-regulated kinase</td>
</tr>
<tr>
<td>PTZ</td>
<td>Pentylenetetrazol</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>STD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>tERK</td>
<td>Total-extracellular signal-regulated kinase</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>Z-Brain</td>
<td>Zebrafish Brain Atlas</td>
</tr>
</tbody>
</table>
Chapter One: Introduction

“I think prime numbers are like life. They are very logical but you could never work out the rules, even if you spent all your time thinking about them”

Mark Haddon, The Curious Incident of the Dog in the Night-Time

From Genes to Behaviour

As with physical traits, behavioural traits can be genetically determined. For example, circadian behaviour, is controlled by a genetically programmed transcription-translation-based feedback loop (Takahashi, 2017). Genes are, however, prone to mutation, and consequently even within a species there is an abundance of genetic variation. The human exome (the protein-coding region of the genome) for example contains an average of 85 heterozygous and 35 homozygous protein truncating variants per person (Lek et al., 2016). This variation likely leads to the exploration of behaviour at species level, and through a process of natural selection will over time lead to an accumulation of genetic variants coding for advantageous behavioural traits. Conversely, however, genetic variation can also lead to behavioural detriments. Indeed, recent genetic studies have linked genes to both specific behavioural abnormalities, such sleep disruption (Hammerschlag et al., 2017; Lane et al., 2017), as well as multi-faceted behavioural disorders including depression (Wray et al., 2018), schizophrenia (Ripke et al., 2014) and autism spectrum disorder (Poelmans et al., 2013; Krumm et al., 2014; Grove et al., 2017). These genetic associations provide a starting point for uncovering the mechanisms linking genes to disease, a potentially crucial step towards the development of treatments.

Progress along this vector has however been limited, for two predominant reasons. Firstly, most neuropsychiatric diseases have been linked to hundreds of genes. Secondly, even with these genes in hand, it remains unclear how to uncover their disease-causing mechanisms, or how such knowledge could be leveraged to develop treatments. Autism spectrum disorder (autism) exemplifies these challenges. To date, 831 genes have been positively linked to autism (gene.sfari.org), and while a variety of approaches have been taken to uncover the mechanisms linking these genes to autism, these results have not yet proven clinically fruitful. Consequently, a new framework for uncovering such mechanisms and for leveraging this knowledge to identify potential therapeutics is required.

In this chapter autism and its aetiology are discussed, as well as the approaches that have been employed to gain mechanistic understanding of the disorder and attempts to translate these findings to the clinic. Finally, a case is made highlighting the advantages of using zebrafish as a model system to tackle these questions and the overarching aim of this thesis, to use zebrafish as a model to study the impact of autism-associated mutations upon neuronal circuits and behaviour, is broken down into three sequential aims that are addressed in the following chapters.
Autism Spectrum Disorder

Autism spectrum disorder (autism) is a highly prevalent and debilitating disorder, thought to affect approximately 1% of the population in the United Kingdom (Brugha et al., 2012) and 1.7% in the United States (Baio et al., 2018), at a ratio of 4:1 affected males:females (Fombonne, 2009). Clinically, autism describes a set of heterogeneous neurodevelopmental disorders defined by a dyad of deficits in social interaction and repetitive patterns of behaviour (American Psychiatric Association, 2013). The disorder’s heterogeneity arises from both the severity of these core phenotypes and the high prevalence of comorbidities including epilepsy (8-30%), anxiety (42-56%), and sleep disorders (50-80%) (Lai, Lombardo and Baron-Cohen, 2014).

In the United Kingdom the median age of diagnosis is 4 years and 7 months (Brett et al., 2016), and the average life expectancy of autism cases is shorter by a mean of 16 years and 4 months than gender-, age- and country of residence-matched controls, though no single cause of premature death is particularly evident (Hirvikoski et al., 2016). To date there is no ‘cure’ for autism per se, nor strong evidence for either behavioural (Weitlauf et al., 2014) or medical interventions (McPheeters et al., 2011). Consequently, no pharmacological treatments are recommended for routine use in treating the core symptoms of autism (Howes et al., 2018). Psychotropic drugs are, however, often prescribed, and poly-pharmacy in autistic cases is common, (Murray et al., 2014). Notably most of these prescriptions are targeted at treatable comorbidities, such as sleep disorders, the amelioration of which can improve core symptoms (Herrmann, 2015; Singh and Zimmerman, 2015) but only to a certain extent.

Together autism’s high prevalence, debilitating nature and lifelong impact result in total annual societal costs, including care and productivity loss, of approximately £32.1 billion in the United Kingdom (Buescher et al., 2014). The cost of autism is thus significantly higher than the costs of dementia (£23 billion), cancer (£12 billion), chronic heart disease (£8 billion), stroke (£5 billion) and even the combined costs of the latter three (Luengo-Fernandez, Leal and Gray, 2012).

Autism’s Aetiology

Both environmental and genetic factors have been linked to autism’s aetiology. However, while environmental factors are thought to contribute to autism risk, strong evidence for specific factors is lacking (Modabbernia, Velthorst and Reichenberg, 2017). A notable exception, however, is prenatal exposure to valproic acid, a drug commonly used in the treatment of epilepsy and other neuropsychological disorders, which significantly increases autism risk (Christensen et al., 2013).

In contrast, genetic factors are strongly implicated in autism (Ramaswami and Geschwind, 2018). Autism’s heritability is estimated to be 50% and an infant’s risk of autism is increased 2 or 10-fold if a cousin or full sibling has the diagnosis (Sandin et al., 2014). The genetic architecture of autism is, however, incredibly diverse. At the time of writing for example, 831
genes have been positively linked to autism risk (gene.sfari.org), albeit with varying degrees of confidence. Indeed, only 86 are currently considered high confidence or strong candidates; defined generally as genes identified from case-control comparisons in multiple studies. Autism-associated genetic aberrations can be broadly divided into three classes: common variants, rare inherited alleles, and de novo mutations.

Common Variants
The additive effects of common variants (i.e. genetic polymorphisms present in at least 5% of the population) are estimated to account for 50% of genetic risk in autism (Gaugler et al., 2014), making them the single most common genetic cause. Putative common variants have been identified from genome wide association studies (Poelmans et al., 2013). To date, however, these studies remain underpowered often sampling only a few thousand cases (Ramaswami and Geschwind, 2018). Thus, future work in this direction will require far larger sample sizes, akin to the 30,000 cases necessary for the discovery of common variants in schizophrenia (Ripke et al., 2014). Consequently, research has predominantly focussed upon the effects of rare inherited alleles and de novo mutations, despite the fact that no single gene currently accounts for more than 2% of autism cases (Ramaswami and Geschwind, 2018).

Rare Inherited Alleles
Rare inherited alleles are thought to explain 3% of genetic risk in autism (Gaugler et al., 2014). Many of these alleles define rare Mendelian syndromes which are associated with autism (Ramaswami and Geschwind, 2018). Fragile X syndrome for example is comorbid with autism in approximately 50% of cases, though the extent to which autism in Fragile X is representative of autism generally remains debated (Abbeduto, McDuffie and Thurman, 2014). Other inherited alleles have been identified in consanguineous families through linkage analysis (Strauss et al., 2006), homozygosity mapping (Novarino et al., 2012) and whole exome sequencing (Yu et al., 2013). Two examples of such mutations are homozygous loss of function mutations in the genes contactin-associated protein-like 2 (CNTNAP2) (Strauss et al., 2006) and branched chain ketoacid dehydrogenase kinase (BCKDK) (Novarino et al., 2012), both of which were identified in consanguineous families with autism and epilepsy. More recently exome sequencing in ‘quad’ families (families in which both parents, an affected child, and an unaffected sibling are all sequenced) has also identified rare, inherited copy number variants (regions of duplicated or deleted DNA than can encompass many genes) as a heritable autism risk factor (Krumm et al., 2015).
De novo Mutations
The final class of autism-associated genetic aberrations are single de novo mutations, which currently account for 3% of autism-risk (Gaugler et al., 2014), though it is estimated that this number will rise as future association studies are completed (Ronemus et al., 2014). Such mutations include copy number variants (Pinto et al., 2010), gene disrupting or loss of function mutations (Krumm et al., 2014) and even mutations in putative non-coding regulatory DNA (Turner et al., 2016). These mutations have predominantly been identified by either whole exome or whole genome sequencing in quads and trios (Krumm et al., 2014; Turner et al., 2016).

The best example of a de novo mutation are heterozygous mutations in chromodomain helicase DNA binding protein 8 (CHD8), an ATP-dependent chromatin remodelling factor (Thompson et al., 2008). To date, CHD8 exhibits the strongest association with autism of any single variant (gene.sfari.org), with de novo truncating mutations observed almost solely in autism cases in multiple studies (O’Roak, Vives, Fu, et al., 2012; O’Roak, Vives, Girirajan, et al., 2012; Talkowski et al., 2012; De Rubeis et al., 2014; Iossifov et al., 2014, 2015). Consequently, a well characterised cohort of individuals harbouring CHD8 mutations exists who exhibit a distinct autism syndrome defined by macrocephaly (in 80% of cases), gastrointestinal problems (80%) and sleep disruption (67%) (Bernier et al., 2014; Barnard, Pomaville and O’Roak, 2015; Merner et al., 2016). Further evidence supporting the specificity of the link between CHD8 and autism comes from exome sequencing of more than 60,000 humans (Lek et al., 2016), in which only 3 predicted heterozygous and no homozygous loss of function CHD8 variants were observed, while based upon the length of the CHD8 gene and its sequence mutability 87 of such variants were expected. This finding demonstrates that CHD8 is one of the most loss-of-function intolerant genes in the entire human genome (exac.broadinstitute.org). Despite the strength of this association, however, it should be kept in mind that currently CHD8 mutations only explain a very small fraction (0.21%) of autism cases (Ramaswami and Geschwind, 2018). However, though this fraction is low, it is similar to that of other de novo mutations associated with autism, such as SCN2A (0.13%) and ADNP (0.10%).

From Genes to Mechanisms
Collectively, the large number of implicated genes and heterogeneous nature of their mutations paint a complex picture of autism’s aetiology. On the one hand, all autism-associated mutations converge phenotypically on social interaction deficits and abnormal repetitive behaviours, suggesting that on some level mechanisms may be shared across genetic lesions. On the other hand, one interpretation of these results is that as a diagnosis, autism in fact encompasses a plethora of distinct genetic disorders or perhaps subtypes of disorders. This interpretation is appealing as it allows for the phenotypic heterogeneity in autism to be understood as arising from diverse mutations with distinct phenotype-causing mechanisms. Were this the case, then it may be possible to identify phenotypic sub-types of mutations, who may benefit from sub-type specific treatments. To date, there have been two
main approaches to translating autism-associated genetic findings to mechanisms: network approaches, which I will define as comparisons of multiple autism-associated genes aimed at identifying common mechanisms, and single model approaches which have focused on mapping the effects of single genetic mutations to autism.

Network Approaches
Network approaches have so far sought commonalities between autism-associated genes at three distinct levels: protein function, developmental timing and neuroanatomical. On a protein function level, one group used protein-protein interaction networks to demonstrate that autism-associated genes identified from whole exome sequencing studies converge upon three functional pathways: chromatin remodelling, Wnt signalling during development and synaptic function (Krumm et al., 2014). In terms of timing another group leveraged an atlas of the human brain transcriptome during development, and found a key point of convergence to be mid-foetal layer 5/6 cortical projection neurons (Willsey et al., 2013). Finally, on a neuroanatomical level a magnetic resonance imaging study of 26 mouse mutants in autism-associated genes, identified three clusters of neuroanatomical abnormalities (Ellegood et al., 2014).

Collectively, these results suggest that groups of autism risk genes converge on specific protein functions, developmental timepoints and neuroanatomical structures, implying that mechanistic subtypes of autism may exist. Generally, however, the diversity between genetic lesions in these studies is more notable than the identified commonalities. Furthermore, such methods offer only hints as to the underlying mechanisms, while currently, the mechanisms mapping even single genes to autism remain unclear.

Single Model Approaches
Animal models for any human disorder are generally conceptualised in terms of face, construct and predictive validity (Willner, 1984), which describe the similarity of the cause, symptoms and responses to interventions, between cases and the model. Notably in such efforts a full recapitulation of the human disorder, i.e. an autistic mouse, is neither expected or necessary. What is desirable is that the model recapitulates aspects of the disorder, the mechanisms of which can be uncovered and potentially leveraged to develop treatments (Servadio, Vanderschuren and Trezza, 2015; Sukoff Rizzo and Crawley, 2017). For example, studying an autism model with no social deficits, but abnormal repetitive behaviour could still be informative and could hold predictive value for this aspect of autism.

The majority of such work has so far been carried out in rodents, with autism modelled either by exposure to valproic acid (Nicolini and Fahnestock, 2018) or by inducing mutations in autism associated genes, specifically rare inherited alleles or de novo mutations (Lázaro and Golshani, 2015). Below I discus what has been learned from the study of a single gene from
each of those two classes of genetic aberrations: **CNTNAP2** and **CHD8**.

**CNTNAP2**
In humans, homozygous *CNTNAP2* mutations were first identified in a consanguineous family presenting with autism and with additional comorbidities including epilepsy and hyperactivity (Strauss *et al.*, 2006). In several of these cases resective surgery was carried out, which failed to alleviate seizures, though did reveal neuronal migration abnormalities (Strauss *et al.*, 2006). *Cntnap2* mutant mice recapitulate many of these phenotypes, including: reduced social preference, increased grooming behaviour, spontaneous seizures, hyperactivity and similar neuronal migration abnormalities (Peñagarikano *et al.*, 2011). Additionally, study of *Cntnap2* mutant mice revealed a reduced numbers of interneurons, predominantly of the parvalbumin subtype and reduced cortical neuronal synchrony (Peñagarikano *et al.*, 2011). Together these results suggested that altered excitatory:inhibitory balance could underlie the observed behavioural deficits. This hypothesis was supported by recent work that demonstrated an acute rescue of social deficits and hyperactivity in *Cntnap2* mutant mice by optogenetically increasing the excitability of parvalbumin neurons in the medical prefrontal cortex thereby increasing inhibitory tone (Selimbeyoglu *et al.*, 2017). Collectively, these results demonstrate the face validity of the *Cntnap2* mouse model and highlight the insight it has provided into the role of excitatory:inhibitory balance in *Cntnap2* mutant behavioural phenotypes.

**CHD8**
*CHD8* is an ATP-dependent chromatin remodelling factor (Thompson *et al.*, 2008), thought to regulate the expression of hundreds of genes by both directly binding transcriptional start sites and indirectly through secondary gene network effects (Sugathan *et al.*, 2014; Barnard, Pomaville and O’Roak, 2015). De novo heterozygous *CHD8* mutations have one of the strongest associations with autism of any gene and are characterised clinically by a distinct syndrome defined by: macrocephaly, gastrointestinal problems and sleep disorders (Bernier *et al.*, 2014). Given the strength of this association, *Chd8* has been the focus of much recent work in mice.

In general, heterozygous *Chd8* mutant mice recapitulate *CHD8* comorbid phenotypes with macrocephaly described in 4 out of 4 studies (Katayama *et al.*, 2016; Gompers *et al.*, 2017; Platt *et al.*, 2017; Suetterlin *et al.*, 2018) and gastrointestinal problems tested and identified in one study (Katayama *et al.*, 2016). *Chd8* mutant mice do not however recapitulate core autism-related behavioural deficits. In terms of social behaviour 1 study found no difference at all in social behaviour (Gompers *et al.*, 2017), and 3 have demonstrated increased duration of social contacts in *Chd8* mutant mice (Katayama *et al.*, 2016; Platt *et al.*, 2017; Suetterlin *et al.*, 2018), the opposite phenotype of what may be expected in an autism model. In terms of repetitive behaviours all 4 studies found no difference, assayed by time spent grooming or marble burying (Katayama *et al.*, 2016; Gompers *et al.*, 2017; Platt *et al.*, 2017; Suetterlin *et
These discrepancies between human and rodent CHD8 mutant behavioural phenotypes raise two points regarding the study of rodent models that merit further discussion. The first is how similar human and rodent social interactions can be considered to be, particularly as human social interactions are primarily visually driven, whilst rodent social interactions are predominantly olfaction based (Chen and Hong, 2018). Indeed, one group identified an increased interest in socially significant, though not other, odours in Chd8 mutant mice (Suetterlin et al., 2018). This finding could explain the mutant’s increased duration of social contacts and even the divergence of social phenotypes between human CHD8 cases and Chd8 mutant mice. The second point for discussion is the potential for anthropomorphism in animal studies. In terms of repetitive behaviours, excessive grooming can certainly be thought of a repetitive behaviour and perhaps marble burying too, but these are only examples of specific behaviours that could show alternations, not a generalised test for repetitive behaviours, or indeed the motor stereotypies that constitute repetitive behaviours in humans (American Psychiatric Association, 2013). Furthermore, these tests of repetitive behaviour are often brief (=10 minutes), which may simply be too short a time to detect subtle differences in behaviour. Collectively, the phenotypic results from Chd8 mutant mice suggest the validity of this model’s use in studying CHD8 co-morbidities, which it recapitulates, though perhaps not autism’s core behavioural deficits, for which it lacks face validity.

To identify the mechanisms linking Chd8 to phenotypes many rodent studies have used chromatin immunoprecipitation and RNA sequencing (given CHD8’s function as a chromatin remodelling factor) to compare mutant and wild-type brains. In general, these assays have detected an enrichment of Chd8 binding at transcriptional start sites in both mutant and wild-type brains, and board, small magnitude gene expression differences between genotypes, particularly in genes with neurodevelopmental functions (Katayama et al., 2016; Gompers et al., 2017; Platt et al., 2017; Suetterlin et al., 2018). Together, these results suggest that the phenotypic effects of mutations in Chd8 may be attributable to the synergistic effects of many developmental gene’s altered expression, rather than prominent changes in a single gene or subset of genes. However, no follow-up experiments have tested this hypothesis.

Overall results from Cntnap2, Chd8 and other mouse models demonstrate their face validity for modelling particular aspects of autism, as well as the mechanistic insights than can be gained from focussing upon these. What remains to be discussed is how such knowledge can be translated to clinical therapeutics.

**Predictive Validity in Autism Models**

The predictive validity of animal models can be considered in two directions: testing human treatments in animal models, and using animal models to identify novel therapeutics. In autism, testing the first is challenging as there are no effective treatments for autism per se (Howes et al., 2018). It is, however, possible to test compounds targeted at specific behaviours; risperidone, for example, is often prescribed to treat repetitive behaviours in autism, despite
weak evidence for its efficacy (Jesner, Aref-Adib and Coren, 2007). Interestingly, in Cntnap2 mutant mice, risperidone ameliorated repetitive behaviours (Peñagarikano et al., 2011). This result suggests that the Cntnap2 model could have predictive value, and furthermore that while risperidone may be of questionable efficacy in autism generally, it could be an effective targeted treatment for repetitive behaviours in humans with CNTNAP2 mutations.

Testing predictive value from animal models to humans is also challenging, as exemplified by the attempted treatment of Fragile X, a genetic disorder associated with autism, with metabotropic glutamate receptor 5 (mGluR5) antagonists; which failed in multiple double-blind placebo-controlled trials, despite extensive pre-clinical evidence from both mice and Drosophila (Erickson et al., 2017). Conversely, promising findings have recently been made in relation to both Cntnap2 and Bckdk mutant mice. In Cntnap2 mutant mice it was shown that administration of a selective melanocortin receptor 4 agonist, which causes endogenous oxytocin release, could acutely rescue social deficits (Peñagarikano et al., 2015). In Bckdk mutant mice a diet rich in branched chain amino acids ameliorated neurological abnormalities, demonstrating their reversibility (Novarino et al., 2012). It remains to be seen if either of these treatments will translate to humans.

Collectively, these examples illustrate how mechanistic insights gained from rodent models could be leveraged to develop treatments. They also, however, highlight the low-throughput of rodent models which is a significant weakness in the study of autism-associated genes for two reasons. Firstly, most rodent studies focus on only a single mutation, meaning that comparisons across genes to identify potentially convergent mechanisms must be made based upon data from independent studies often using different methodologies. Secondly, it necessitates hypothesis driven, targeted drug testing, which is inherently challenging, as illustrated by the failure of mGluR5 antagonists in treating Fragile X (Erickson et al., 2017). Thus, a preferable system would be one in which mutants in many autism-associated genes could be tested in standardised assays, and in which pharmacological compounds could be quickly screened at scale in a hypothesis free manner.

Zebrafish
Zebrafish are rapidly emerging as a powerful model organism for the study of psychiatric genetics (Haesemeyer and Schier, 2015), including both autism- (Hoffman et al., 2016) and schizophrenia-associated genes (Thyme et al., 2018). This is due to five key factors. Firstly, approximately 70% of human genes have at least one clear zebrafish orthologue (Howe et al., 2013) and a variety of gene editing technologies are available in zebrafish, making it easy to produce zebrafish mutants in human disease-associated genes. Secondly, zebrafish develop rapidly ex utero, from a single cell to a free-swimming larva in just four days (Kimmel et al., 1995), meaning that their development can be easily studied. Thirdly, even larval zebrafish display a variety of behaviours including a distinct locomotor repertoire (Marques et al., 2018), a diurnal pattern of activity (Zhdanova et al., 2001; Prober et al., 2006) and by three weeks of age, social preference (Dreosti et al., 2015). Fourthly, zebrafish have a vertebrate
brain plan and at larval stages are small and translucent, enabling cellular-resolution whole-brain imaging in both fixed tissue (Randlett et al., 2015) and live animals (Naumann et al., 2010; Ahrens et al., 2013; Kim et al., 2017). Finally, their small size facilitates high-throughput genetic (Chiu et al., 2016) and pharmacological screens (Kokel et al., 2010; Rihel et al., 2010) in which thousands of manipulations can be tested.

Together these five features make zebrafish an ideal model organism for the study of disease-associated genes. To date, the majority of work in this area has employed morpholinos to achieve knock-down of gene transcripts (Shams et al., 2018), results which should be interpreted with caution as morpholino-induced phenotypes often differ from mutant phenotypes due to their off-target effects (Kok et al., 2015). Two notable studies have used mutants to study neurological disease-associated genes in zebrafish, each with a unique approach: the first focussed upon detailed study of a \textit{cntnap2} mutant (Hoffman et al., 2016), the second studied a library of mutants in all schizophrenia-associated genes (Thyme et al., 2018).

**Single Gene Models**

Of relevance to autism, zebrafish mutants in the gene \textit{cntnap2} (Hoffman et al., 2016) recapitulate behavioural phenotypes observed in both \textit{Cntnap2} mutant mice (Peñaagarikano et al., 2011) and human cases (Strauss et al., 2006), including seizures and hyperactivity. On a mechanistic level \textit{cntnap2} mutant zebrafish also show interneuron deficits, similar to the mouse model, demonstrating a conversation of this mechanism across these species. Furthermore, in terms of predictive value, risperidone ameliorated \textit{cntnap2} mutant zebrafish hyperactivity (Hoffman et al., 2016), akin to its improvement of repetitive behaviours in \textit{Cntnap2} mutant mice (Peñaagarikano et al., 2011). Together, these results demonstrate the validity of the \textit{cntnap2} zebrafish model and its similarities with the mouse model.

This study, however, also provided unique insights into \textit{cntnap2} mutations that would have challenging to garner in other model systems. Specifically, a ‘predictive pharmacology’ approach was used to compare the \textit{cntnap2} phenotype (night-time hyperactivity) to a library of pharmacologically-induced behavioural phenotypes (Rihel, Prober and Schier, 2010). Two comparisons were made: correlation -- to identify compounds that copy the mutant phenotype (induce night-time hyperactivity), and anti-correlation -- to identify compounds that elicit the opposite phenotype (decrease night-time activity).

Correlation of the mutant to the pharmacological fingerprints revealed that NMDA receptor antagonists elicit night-time hyperactivity in wild-type fish. Subsequent testing of these compounds demonstrated that \textit{cntnap2} mutants were more sensitive to these molecules than wild type animals. A result which supports a model of excitatory:inhibitory imbalance in \textit{CNTNAP2} cases (Peñaagarikano et al., 2011; Selimbeyoglu et al., 2017), though implies a role for glutamatergic signalling, despite an absence of obvious anatomical glutamatergic deficits.
Anti-correlation of the mutant behaviour to the panel of small molecules identified a set of compounds, enriched for estrogenic agonists, which specifically decrease night-time activity. Applying these compounds to the mutants demonstrated not only that estrogenic compounds could suppress the mutant phenotype but also that they did so better than risperidone (Hoffman et al., 2016). This results implicates abnormal estrogenic signalling in cntnap2 mutants, and furthermore demonstrates the ability of such predictive pharmacology approaches to pre-select phenotypic suppressors.

Together, these results demonstrate how the zebrafish’s amenability to pharmacological screening can be exploited to identify both disrupted signalling pathways of interest and potential small molecules of predictive value, an approach that has so far only been applied to this mutant.

**Screening Genetic Models**

Schizophrenia-associated genes have predominantly been identified through genome-wide association studies, which rather than pinpointing single genes, identify genetic regions of interest that can contain many genes (Ripke et al., 2014). Consequently, it remains unclear which of the genes within these regions are disease causing. To tackle this question Thyme et al., (2018) generated zebrafish mutants in all 132 schizophrenia-associated genes. These mutants demonstrated a variety of behavioural abnormalities, which mirrored data from both mice and humans, demonstrating their face validity as a model. Leveraging the capacity to image whole-brains in zebrafish, this study additionally mapped alterations in brain morphology and neuronal activity in all of the mutants. Together, this approach generated an atlas of behavioural and neuronal phenotypes for the mutants, from which 30 genes could be prioritised for future work based upon their behavioural and neuroanatomical effects in zebrafish (Thyme et al., 2018). This study is an excellent demonstration of how the advantages of the zebrafish system, particularly it’s high-throughput, can be leveraged to gain insight into human disease.

Collectively, these results from zebrafish mutants in autism- and schizophrenia-associated genes highlight the advantages of this model system for the study of these disorders, particularly for comparisons across genetic lesions and for drug discovery. However, each of these studies utilised relatively simple behavioural measures, precluding the assessment of many potentially disease-relevant behavioural phenotypes, such as repetitive behaviours in autism mutants. If such phenotypes could be assessed in zebrafish, particularly in a manner suited to genetic and pharmacological screening, then the strengths of this model could be utilised to uncover their underlying mechanisms, and to identify potential therapeutics. Thus, in my thesis I aimed to develop a behavioural framework, suited to high-throughput genetic and pharmacological screening, that could simultaneously capture a plethora of autism-relevant behavioural phenotypes, and to demonstrate how such phenotypes provide a starting point for uncovering the mechanisms linking autism-associated genes to aberrant behaviour.
Aims & Approaches

The overarching aim of the work described herein was to use zebrafish as a model to study how mutations in autism-associated genes affect neuronal circuits and behaviour. This aim can be described in terms of three sequential steps:

1. Develop tools suited to high-throughput screening for quantifying zebrafish behaviour
2. Identify behavioural phenotypes in zebrafish with mutations in autism-associated genes
3. Map mutant behavioural abnormalities to neuronal circuits

While this body of work is focussed on only two autism-associated genes, \textit{CHD8} and \textit{BCKDK}, the longer-term goal of this project is to extend this approach to study libraries of autism-mutants and small molecules. Together, this prospective dataset will provide insight into the common or divergent mechanisms by which genetic aberrations translate to behavioural phenotypes in autism and could even facilitate the discovery of novel therapeutics.
Chapter Two: Behaviour

“What we observe is not nature itself, but nature exposed to our method of questioning”

Annie, quoting Heisenberg, to Cooper in Twin Peaks season two, episode 20.

Quantitative Descriptions of Behaviour

To survive, animals must coordinate patterns of action and inaction in response to their environment. For example, larval zebrafish swim by executing sequences of forward as well as left and right turns, resulting in a slaloming swim pattern and an efficient search of their environment (Dunn et al., 2016). These actions and inactions, which together I will define as behaviour, result from some function incorporating internal (e.g. transcriptional, hormonal or neuronal activity) and external (e.g. time of day or temperature) state. Thus, behavioural observations provide insight into the underlying mechanisms that control behaviour and are a necessary step in understanding these systems (Krakauer et al., 2017). As Marr (1982) writes “trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: it just cannot be done”. Therefore, my starting point for dissecting how mutations in autism-associated genes affect neuronal circuits was detailed behavioural analysis. Firstly, to gain insight into the behaviour of typical zebrafish and then secondly, to assess mutants in autism-associated genes for behavioural abnormalities.

This approach, best summarized as “Neuroscience Needs Behaviour” (Krakauer et al., 2017), is explicated by Marr’s framework for understanding how systems execute complex information-processing tasks, in which three levels of understanding are necessary (1982): Computational Theory – understanding the logic of how the task is executed; Algorithm and Representation – understanding how the systems inputs can be transformed to outputs; and Hardware Implementation – understanding how the algorithm could be physically realized. As applied to animal behaviour these levels correspond to: understanding a given behaviour through detailed analysis, generating theoretical models of how the behaviour could be controlled and finally searching for the physical correlates of this control. An elegant example of this approach is Naumann et al., (2016) in which the zebrafish optomotor response was studied through detailed behavioural analysis, modelling of potential neuronal strategies that could control this behaviour, and finally whole-brain imaging to identify circuits that implement these strategies. The importance of quantitative behavioural descriptions is further supported by recent studies that demonstrate how they can provide biological insights even in the absence of explicit algorithmic models; for example by correlating behaviour with genetic mutations (Brown et al., 2013) or the activity of neuronal circuits (Vogelstein et al., 2014; Lovett-Barron et al., 2017; Robie et al., 2017).
Animal behaviour, however, typically has many degrees of freedom and evolves over multiple timescales in parallel from milliseconds (Wiltschko et al., 2015) to days (Fulcher and Jones, 2017) and even across an animal’s entire lifespan (Jordan et al., 2013; Stern, Kirst and Bargmann, 2017). As such, quantitatively describing animal behaviour remains both conceptually and technically challenging (Berman, 2017; Brown and de Bivort, 2017).

Conceptually the problem is well described by Berman (2017) “what numbers should one put on the movements an animal is performing (or not performing)?”. Inspired by early ideas from ethology (Lashley, 1951; Tinbergen, 1963) recent work in behavioural neuroscience has sought to describe behaviour in terms of modules arranged into motifs. Behavioural modules are often defined from postural data, as stereotyped movements, such as walking and grooming in Drosophila (Berman et al., 2014; Vogelstein et al., 2014; Robie et al., 2017) and mice (Wiltschko et al., 2015), or different types of swim bouts in zebrafish (Girdhar, Gruebele and Chemla, 2015; Marques et al., 2018). Behavioural motifs are defined as sequences of modules, that capture the patterns inherent to animal behaviour, such as the slaloming swimming of zebrafish larvae (Dunn et al., 2016). Many of these approaches are underpinned by unsupervised learning algorithms (Todd, Kain and de Bivort, 2017) that unbiasedly identify behavioural modules, thus, eliminating the inter-rater reliability problems inherent to the manual scoring of behaviour. Advantageously, unsupervised learning also eliminates the potential for anthropomorphism, which in the context of animal models for neurological disorders is key. What do authors mean when they manually score and observe “autism-like behaviours” in a mouse? Furthermore, how should we decide which behaviours to study in these models in the first place?

I reasoned that a good approach to the study of autism-models would be to employ a hypothesis-free behavioural assay that could detect a broad range of phenotypes spanning the gamut of autism’s comorbidities. I decided upon the locomotor assay described by Rihel, Prober and Schier (2010), as it offered two major advantages. Firstly, the assay can accommodate up to 96 animals at a time, making it ideally suited to studying mutant behaviour as many animals of each genotype can be simultaneously studied, potentially allowing for the detection of subtle phenotypes. Secondly, the assay allows for long-timescale recordings, even over multiple days and nights, enabling the identification of phenotypes that may only manifest at certain times of the day/night cycle or even subtle phenotypes that may occur only infrequently. The trade-off for these advantages is that the assay, like many high-throughput or long-timescale assays (Churgin et al., 2017; Stern, Kirst and Bargmann, 2017), records only a one dimensional metric of activity over time for each animal. Simple behavioural representations such as this clearly result in a loss of information, for example direction of movement or position in the environment. However, in many such assays, this problem is compounded by further simplifications of the data, such as averaging the data into a smaller number of time periods.

I hypothesised that a hybrid approach coupling this assays simple behavioural representation with machine learning tools could be sufficient to describe zebrafish behaviour, and even to provide behavioural insights that may not have been apparent from other assays due to small
sample sizes or short recording times. To test this idea, I developed a behavioural analysis pipeline to analyse the behaviour of hundreds of zebrafish larvae recorded over multiple days and nights. Specifically, I used unsupervised learning to identify typical types of movements and pauses (modules), a hierarchical compression algorithm to identify sequences of modules (motifs) of a range of lengths, and finally a supervised learning approach to reveal motifs that were used in specific behavioural contexts. To develop this pipeline, I gathered data from wild-type larvae as well as those subject to pharmacological or genetic manipulations. The results from this analysis, as well as the subsequent application of this pipeline to mutants in two autism-associated genes are presented below.

**Behavioural Analysis Pipeline**

To design and evaluate my proposed behavioural analysis pipeline, I gathered and analysed a test dataset using the set-up described by Rihel, Prober and Schier (2010) (see Materials & Methods, Behavioural Set-up for more information, page 84). Briefly, the format of this set-up (Appendix Figure 1a, page 110) is a 96-square well plate with a single zebrafish larva housed in each well. From below the plate, white lights provide a 14:10 light:dark cycle (lights on from 09:00 a.m. to 23:00 p.m.) and infrared lights provide constant illumination allowing larvae to be recorded in the dark. A camera monitors the plate and a frame by frame subtraction method is applied per well to determine the number of pixels that change intensity each frame. This metric is termed Δ pixels and is recorded as a proxy for movement over time.

To explore the space of possible behaviours in this assay, I included conditions in the test dataset which previous data suggested would elicit a variety of behavioural phenotypes. Here, I define these conditions as behavioural contexts. Included in the test dataset were three experiments in which wild-type larvae were tracked across multiple days and nights, single experiments in which larvae were dosed with a range of concentrations of the sedative melatonin (Zhdanova *et al.*, 2001) or the epileptogenic pentylenetetrazol (PTZ) (Baraban *et al.*, 2005), and finally both heterozygous and homozygous hypocretin receptor mutants (*hcrtr*) from duplicate experiments, for which only subtle behavioural phenotypes have previously been described (Yokogawa *et al.*, 2007). Overall the test dataset thus consisted of data from 452 animals from 7 independent experiments. The results from the analysis of this test dataset are presented below, divided into the four main analytical steps of the behavioural analysis pipeline. A flow diagram depicting these steps and the associated code for their execution is shown in Appendix Figure 10 (page 120).
Improved Behavioural Resolution

Larval zebrafish behaviour consists of an alternating sequence of movements and pauses, termed bouts, structured at a sub-second timescale. Conventional analysis of data from this assay sums the data into non-overlapping one-minute blocks, resulting in a loss of this structure. Thus, as a first step I built upon previous work in the field by analysing Δ pixels on a frame by frame basis (see Materials & Methods, Computing, Behavioural Data Analysis, Δ pixels, page 88).

When recorded in this way, Δ pixels data is an alternating sequence of peaks of values greater than zero, and strings of zeros (Figure 1a). The peaks represent movements, and larger movements will result in larger Δ pixels values. I defined active bouts as any single or consecutive frames with non-zero Δ pixels values and described each bout using the following features: length, mean, standard deviation, total, minimum and maximum (Figure 1a). The strings of zeros record periods of inactivity of varying lengths. I defined inactive bouts as any single or consecutive frames with zero Δ pixels values, and described each inactive bout using its length (Figure 1a).

From the test dataset I extracted a total of 61,800,436 bouts from all 452 animals. As expected, wild-type larvae had more bouts during the day than the night (Figure 1b). Furthermore, I found that several bout features differed between the day and the night (Appendix figure 2a, page 111). For example, inactive bouts lasting around 0.4s long were more likely during the day, suggesting that these lengths are the typical pauses between swim bouts. Conversely, longer inactive bouts, which could constitute rest or even sleep (Prober et al., 2006), were more likely during the night (Figure 1c). Together these behavioural differences result in a diurnal pattern of activity (Figure 2a).

Compared to the old approach, this new methodology resulted in 1,500 times the data per animal, and a corresponding increase in what I will term behavioural resolution, the ability to distinguish between behaviours. This point is illustrated in Appendix Figure 1b (page 110), which shows how the previous analysis method would assign two distinct behavioural patterns the same value, and furthermore how the previous method likely underestimates the amount of inactivity as long inactive periods falling into two one-minute blocks will be lost. To attempt to quantify the information gained from the new method, I processed the wild-type data through both approaches and compared each animal’s data between the two. Specifically, I compared the number of inactive bouts equal to or longer than 60s, a threshold traditionally used to define sleep in zebrafish (Prober et al., 2006). As anticipated the one-minute block approach misses on average 1/3rd of each animal’s inactive bouts 60s or longer (Appendix Figure 1c, page 110), quantitatively demonstrating the information gained through this new approach to the data analysis.

In line with previous work (Zhdanova et al., 2001) melatonin dose dependently decreased larval activity (Figure 2b) and induced longer inactive bouts. Interestingly, however, melatonin also reduced active bout length and amplitude (Appendix Figure 2b, page 111). Similarly, PTZ (Figure 2c) altered both active and inactive bout parameters, eliciting on average longer,
lower amplitude active bouts and longer inactive bouts during the day (Appendix Figure 2c, page 112). hcrtr/ had subtle differences in active bout structure. Specifically shorter mean active bout length, and lower active bout standard deviation and total, compared to both hcrtr/+ and hcrtr/+ (Appendix Figure 2d, page 112). hcrtr/+ did not differ from hcrtr/+ by any metrics, demonstrating that hcrtr is haplosufficient for behaviour in this assay.

Collectively these results demonstrate how assessing Δ pixels data on a frame by frame basis substantially improves the behavioural resolution of this assay, and furthermore illustrates how this improvement can provide insight into both the behaviour of wild-type animals across the day/night cycle as well as those subject to pharmacological or genetic manipulations.
Figure 1. Long Time-scale Behavioural Tracking

a. Top panel: five consecutive frames from an individual well of a 96-well plate as a 6dpf zebrafish larva performs a swim bout. Blue highlights pixels that change intensity between frames (Δ pixels), numbering in seconds corresponds to the trace below. Lower panel: a Δ pixels trace from the larva above, highlighted are the features that describe each active and inactive bout.

b. The average number of bouts recorded from individual larvae at 5 and 6dpf during the day (light blue) and the night (dark blue). Each circle is 1 of 124 wild-type fish and the orange pluses mark the population means.

c. The probability of observing different lengths of inactivity during the day (light blue) or the night (dark blue) at 5 and 6dpf. Each fish’s data was fit by a probability density function (pdf), shown is a mean pdf (bold line) and standard deviation (shaded surround) with a log scale on the x-axis cropped to 10 seconds. Insert: the total probability per animal, of inactive bout lengths being longer than 10 seconds. Iconography as in b.
Figure 2. Behaviour Varies With Behavioural Context

a. The average activity of 124 wild-type fish from 4–7dpf, days (white background) and nights (grey background). Data for each larva was summed into seconds and then smoothed with a 15 minute running average. Shown is a mean summed and smoothed trace (bold line) and standard error of the mean (shaded surround).

b. Average activity across one day (white background) and night (dark background) for larvae dosed with either DMSO (control) or a range of doses of melatonin immediately prior to tracking at 6dpf. Data was summed and smoothed as in a. n denotes the number of animals per condition.

c. As in b, for a range of doses of PTZ or H₂O (control).
Identifying Behavioural Modules

Recent work has demonstrated that larval movements can be classified through an unsupervised learning approach into 13 distinct bout types (Marques et al., 2018). A full description of larval behaviour, however, requires quantification of both the movements and pauses that they execute. Thus, as a second step I sought to determine if distinct active or inactive bout types, which I will term modules, were identifiable from the bout data and if so, if module usage depended upon behavioural context. To address these questions, I separately clustered the active and inactive bouts using an evidence accumulation-based clustering algorithm (see Materials & Methods, Computing, Behavioural Data Analysis, Clustering, page 89). Briefly, 200 Gaussian Mixture Models were built for each data set, from samples each independent in terms of which and how many bouts were sampled. Then the results of these models were combined to generate an aggregate solution.

This clustering method identified 5 active and 5 inactive modules (Figure 3a-b and Appendix Figure 3-4, page 113), which I separately labelled from 1-5 from the shortest to longest mean length. The active modules corresponded to different shapes, in terms of amplitude and length, of Δ pixels changes (Figure 3a), while the inactive modules consisted of different lengths of inactivity (Figure 3b). The shortest inactive module (module 1) had a mean length of 0.06s and ranged from a minimum of 0.04s (the sampling limit) to a maximum of 0.12s. In contrast the longest inactive module (module 5) had a mean length of 96s and covered a huge range of values from a minimum of 20s to a maximum of 8.8 hours. Having classified each bout into a distinct module, I could represent each larva’s behaviour as an alternating sequence of active and inactive modules across time (Figure 3c).

In the wild-type data, active and inactive module usage varied with both time of day and development (Figure 4). For example, the probability of observing inactive module 2, which consists of typical day pause lengths (Figure 3b), was on average 0.6 during the day and only 0.24 during the night, as inactive modules 1, 4 and 5 became more likely (Figure 4a). To reveal finer-grain temporal dynamics, I examined each module’s mean frequency over time (Figure 4b). In general, the active and short inactive modules have high frequency during the day and peak at the light/dark transition as the fish respond to the sudden change in illumination. In contrast, the only module with a peak at dawn is inactive module 4, which also increases its frequency in the evening. Together these results reveal that zebrafish employ different bout types, in a time of day/night dependent manner.

Larvae dosed with melatonin showed a shift towards longer inactive modules and shorter active modules (Appendix Figure 5a, page 115). In PTZ dosed larvae there were shifts in active module probability, though no single active module clearly captured behavioural seizures (Appendix Figure 5b, page 115). One particularly notable difference, however, was that 27 of 28 (96.4%) PTZ dosed larvae did not use active module 1 at all during either the day or the night, while controls used this module with 0.12 probability during the day and 0.22 during the night. Consistent with the long inactive periods that have been described between seizures in zebrafish (Baraban et al., 2005), PTZ also increased the probability of longer inactive modules (Appendix Figure 5b, page 115). Finally, in hctr mutants there were
no differences in either active nor inactive module probabilities (Appendix Figure 5c, page 115), demonstrating that bout type usage is similar between these mutants and wild-type animals across both the day/night cycle and developmental time.

Collectively, these results reveal that zebrafish behaviour in this assay can be described by 5 types of active and 5 types of inactive modules. Furthermore, they demonstrate that module usage varies with behavioural context, including time of day and with drug exposure. Interestingly, in many contexts both active and inactive module probabilities were shifted, suggesting that the two may covary, perhaps by being arranged into sequences.
Figure 3. Unsupervised Learning Identifies Behavioural Modules

a. Average Δ pixels changes for each active module. Shown is the mean (bold line) and standard error of the mean (shaded surround) of 100 bouts randomly sampled from each module, from one representative fish.

b. Probability density functions showing the distribution of inactive bout lengths in seconds, on a log x-axis cropped to 60s, in each inactive module. Modules are numbered and coloured from shortest to longest mean length (see legend).

c. Matrices showing the active (left) or inactive (right) module assignment of every frame (x-axis) across days/nights 5 and 6 post-fertilisation, for each of 124 wild-type fish (y-axis). Each module’s corresponding colour is shown in the adjacent colormap.
Figure 4. Module Usage Varies Across the Day/Night

a. Average active (upper) and inactive (lower) module probability during day (light blue) and night (dark blue) 5 and 6 of development. Each of 124 wild-type animals is shown as a circle and orange crosses mark the population means. Active modules are sorted by average day probability from highest to lowest (left to right). Inactive modules are sorted by average length from shortest to longest (left to right). The blobs correspond to the colour used for each module in b and other figures.

b. The average frequency of each active (left) and inactive (right) module across days 5 and 6 of development. Shown is a mean smoothed with a 15 minute running average, rescaled to 0-1. Days are shown with a white background, nights with a dark background. Modules are sorted from shortest to longest (lower to upper panels).
Hierarchical Compression

I sought to determine if there are sequences inherent to zebrafish behaviour and to quantify how patterned their behaviour is. To address these questions I used, as in Gomez-Marin, Stephens and Brown (2016), a dictionary based compression algorithm (Nevill-Manning and Witten, 2000) (see Materials & Methods, Computing, Behavioural Data Analysis, Hierarchical Compression for more information, page 90). This algorithm shortens a given symbolic sequence by iteratively removing motifs (i.e. short sub-sequences of symbols), from the sequence and replacing them with new symbols until no more motifs can be identified and the sequence is maximally compressed. Figure 5a illustrates this process in the context of my analysis. Initially each animal’s behaviour across time is represented by a Δ pixels trace. Through an unsupervised learning approach stereotyped active and inactive modules are identified. Thus, each animal’s behaviour can be reduced to an alternating sequence of active and inactive modules across time. This symbolic, module sequence is input to the compression algorithm, which iteratively executes the following steps:

Hierarchical compression steps

1. Identify the most compressive motif in the sequence
2. Replace all instances of this motif with a new symbol
3. Add this symbol and its module sequence to the motif library

The algorithm terminates when no more motifs can be found. At this point the sequence is defined as being maximally compressed. Note that due to the replacements the algorithm makes, it can identify hierarchically organised motifs, i.e. a long motif can be made up of two shorter motifs joined together (Figure 5a). Ultimately, compression produces two outputs of interest: compressibility, a measure of how much the input sequence can be reduced by, and a library of identified motifs.

One can measure how compressive each sequence is, as follows. At each iteration (i) the most compressive motif is defined as the motif which makes the most saving, a balance between the length of the motif (W) and the number of times it occurs in the sequence (N), which also considers the combined cost of adding a new motif to the dictionary (W + 1) and of replacing the occurrences of this motif in the sequence with a new symbol (+N) (equation 2).

Equation 2:

\[ Savings_i = WN - (W + 1 + N) \]

The overall compressibility of a given input sequence can be calculated by summing these savings across all iterations and dividing this total by the length of the input sequence (in modules). This process results in a compressibility metric that ranges from 0-1 (low-high compressibility). Sequences with few repeats will have low compressibility, while more repetitive sequences (with more frequent repeats) will have higher compressibility. Consequently, compressibility can be used to measure the repetitiveness of sequences of
animal behaviour (Gomez-Marín, Stephens and Brown, 2016).

Initially I compressed every animal’s full module sequence. To determine if these sequences were more or less compressive than would be expected based upon the distribution of modules and the active to inactive module transition structure, I generated 10 sets of paired shuffle data per animal. All wild-type larvae were more compressive than their paired shuffled data, demonstrating that their behaviour is more repetitive than expected based on module probabilities alone (Appendix Figure 6a, page 116). Compressibility will, however, vary non-linearly with input sequence length, as longer sequences will be more likely to contain repeats (Appendix Figure 6b, page 116). Thus, to enable comparisons between samples (each with a different number of modules) I compressed non-overlapping blocks 500 modules long. Comparing compressibility through this metric revealed several insights. I found that compressibility was consistently higher during the day than the night, and increased between days/nights 5 and 6 of development (Figure 5b and Figure 6a). Focussing on the lighting transitions, I used a running block 500 modules long to compress 1,500 modules centred around each transition. This revealed that light to dark transitions in the evening led to an increase in compressibility then a decrease towards the lower night baseline (Figure 6b). While dark to light transitions in the morning led to an initial decrease and then an increase towards the elevated daytime baseline (Figure 6b).

Turing to the pharmacological and genetic data, PTZ increased compressibility to a constant day/night value (Figure 5b), capturing the repetitive structure of seizure cycles (Baraban et al., 2005). In contrast, melatonin decreased compressibility to the control night-time level, demonstrating that melatonin induces night-like levels of behavioural repetitiveness (Figure 5b). These drug-induced changes in compressibility, and other compression results, do not however simply reflect changes in overall activity levels. For example, PTZ exposed larvae are less active than controls during the day and more active during the night (Appendix Figure 2c, page 112) but have higher levels of compressibility during both the day and the night (Figure 5b); thus, supporting the concept that compressibility describes repetitive structure in larval behaviour. In hctr mutants I found no differences in compressibility, suggesting that their behaviour is as repetitive as wild-type animals (Figure 5b).

The algorithm’s second output are the motifs identified from the input sequences. From compressing every fish’s full module sequence and then merging the motifs identified across all animals, I generated a library of 46,554 unique behavioural motifs (Figure 7a). In terms of raw Δ pixels data each motif represents an approximately repeated pattern of movements and pauses of varying length (Figure 7b). The motifs ranged from 2-20 modules long with a median length of 8 modules. Based upon the average module lengths, these motifs span timescales from 0.1s-11.3 minutes with a median length of 3.84s, revealing the varied timescales at which zebrafish behaviour is organised. Interestingly, motifs of different module lengths used distinct sub-sets of modules (Figure 7a); for example, longer motifs had a lower probability of using long inactive modules.
This finding suggested that there are rules governing how modules are arranged into motifs. To draw an analogy, one could consider the modules as letters, motifs as words and these rules as the grammar. To dissect these grammatical rules, it was necessary to quantify how often each motif was observed in a behavioural context, independent of changes in module probability between contexts. To quantify this, I compared the number of times each motif was observed in a given animal and time window (e.g. day or night) to 10 sets of paired shuffled data, as well as comparing the shuffles to each other. By this metric, which I will term motif enrichment/constraint, positive and negative values, respectively, indicate motifs observed more or less than expected based upon module probabilities alone. Through this approach I found that enrichment/constraint scores from the real data were more prone to extreme positive and negative values than the shuffled data (Appendix Figure 6c, page 116), suggesting that a minority of behavioural motifs are either enriched or constrained.

Overall these results reveal the presence of structure, motifs spanning sub-second to minute timescales, in larval behaviour and demonstrate that the amount of structure varies with both time and drug exposure. Furthermore, these results indicate that motifs can be enriched and constrained, leading me to hypothesise that motif enrichment/constraint may be dependent upon behavioural context.
Figure 5. Hierarchical Compression Reveals Structure in Zebrafish Behaviour

a. Hierarchical compression approach explained using fictive data. From top to bottom: Δ pixels data (black trace) was discretised by clustering, the circles below show each fictive frame’s module assignment. Compression iteratively identifies motifs (shown as boxes) from modular sequences by replacing them with new symbols, until no more motifs can be identified, and the sequence is maximally compressed.

b. Each panel shows how compressibility, calculated from 500 module blocks, varies in different behavioural contexts. Each pale line shows an individual fishes average compressibility during the day and the night. The darker overlay shows a population day and night mean and standard deviation. In the wild-type data compressibility is higher during the day than the night \((p < 10^{-15})\) and increases from day/night 5 to 6 \((p < 10^{-4})\), findings consistent across triplicate experiments. There is no effect of \(hcrt\) genotype on compressibility. Melatonin decreases \((p < 10^{-10})\), while PTZ increases compressibility \((p < 10^{-9})\). Statistics are two or four-way ANOVA, comparing group, day/night, and for wild-type and \(hcrt\) data, development, experimental repeat, and all interactions.
Figure 6. Compressibility Varies With Time

a. Compressibility of 500 module blocks for each wild-type fish, averaged into 1-hour time points. Each pale blue line shows 1 of 124 larvae, the darker blue overlay shows the mean and standard deviation of this data every hour. Shown are days (white background) and nights (dark background) 5 and 6 of development.

b. Compressibility, per 500 modules, at the light transitions. Data is aligned such that each transition is at module 0 (dashed orange line). The dashed black lines indicate where the blocks start (-500, left) and stop (+500, right) including the transition. Data is shown as a mean and standard error of the mean from 124 wild-type larvae for four separate transitions (see legend). D:L - dark to light (morning), L:D - light to dark (evening).
Figure 7. Compression Identifies Motifs in Zebrafish Behaviour

a. All 46,554 unique motifs (y-axis) identified by compressing data from all animals. Each motif’s module sequence is shown, with the modules coloured according to the colormap on the right. Motifs are sorted by length and then sequentially by module. Motifs range in length from 2-20 modules long. Insert: the probability of observing each module in motifs of each observed motif length, inactive modules are shown first (left) followed by the active modules (right).

b. Each motif in the library, see a, consists of an alternating sequence of Δ pixels changes and pauses (active and inactive modules). A representative motif of each module length is shown, each module is coloured according to the colormap in a. Representative motifs were chosen by determining every motif’s distribution of modules and then for each observed length, selecting the motif closest to the average module distribution (see a, insert).
Context Specific Behavioural Motifs

To test the hypothesis that zebrafish use behavioural motifs context-dependently, I first tested whether behavioural motifs are differentially enriched/constrained during the day and the night in wild-type larvae. To visualise this question, I generated a matrix of enrichment/constraint scores for each motif (Figure 8a); this matrix had 496 rows (124 animals, each with 2 days and 2 nights) and 46,554 columns (one for each motif). The motifs and samples in Figure 8a are sorted to emphasise structure, specifically that there are many motifs that show either day (leftmost: high day and low night values) or night enrichment (rightmost: low day and high night values). From this matrix and similar matrices comparing other contexts (e.g. pharmacological or mutant data), I sought to identify a minimal subset of motifs that could accurately distinguish between contexts, and which may represent the most salient context-dependent motifs.

To achieve this, I applied a supervised feature selection algorithm to the data to select the best motifs for each comparison. In particular I used the minimal-redundancy-maximal-relevance criterion (mRMR) algorithm (Hanchuan Peng, Fuhui Long and Ding, 2005), which explicitly seeks features with low dependency upon each other. For this dataset this is particularly important as the motifs are highly nested and thus dependent upon each other. To illustrate this point, consider if the best motif for a comparison was a two-module sequence; in the absence of minimal-redundancy, the second ‘best’ motif may simply be a four-module pattern consisting of a repeat of the first motif, and so on.

Thus, for any given comparison I started with an enrichment/constraint scored matrix of samples by motifs (e.g. Figure 8a), to which I applied the mRMR algorithm to select the best 250 motifs (see Materials & Methods, Computing, Behavioural Data Analysis, Supervised Motif Selection for more information, page 91). To determine how many of these motifs were necessary for accurate classification, I trained linear discriminate analysis classifiers as sequential mRMR motifs, from 1 up to 250, were added. Briefly, for each subset (i) of mRMR motifs (e.g. 1-2, 1-3,... 1-250), a classifier received a matrix of enrichment/constraint scores from two classes (e.g. day and night) by i (mRMR motifs). This data was randomly divided into 10 sets (partitions). For each of these sets (validation data), linear discriminant analysis was applied to the combined data from the other sets (training data) to find the linear combination of motifs that best separated the two classes. Finally, for each of the 10 partitions, the class of each validation data point was predicted based upon the training data, and the percentage of incorrectly predicted validation points was recorded (classification error). Generally, this approach produced classification error curves with a characteristic shape, as motifs are added, classification error initially decreases to trough at a minimum, and then begins to rise again as the classifier overfits the data (Appendix Figure 7a, page 117). I determined the necessary number of motifs and resultant error from these curves by smoothing them and finding the minimum classification error. To determine how well these classifiers performed, for each comparison I also generated majority class classifiers, which stringently perform as well as the ratio (in terms of number of samples) between the two classes, i.e. if 90% of the data is of group A, the classifier will have an error of 10% (± standard
To identify motifs that varied with time I first analysed the wild-type day and night data (Figure 8a). Through this approach I found that only 15 motifs were required to separate the data with only 0.2% (0.63% Std) classification error (Figure 8a), compared to a majority class classifier with 50% error (Appendix Figure 7b, page 117, and Appendix Table 1, page 123). The day enriched motifs consisted of high amplitude movements interspersed with short pauses, while the night enriched motifs contained low amplitude movements and long pauses (Figure 8b). Next, I applied the same approach to the wild-type data between days/nights 5 and 6 of development. In both cases the classifiers achieved roughly 20% error using around 90 motifs (Appendix Figure 7b, page 117, and Appendix Table 1, page 123). This result suggests that there are changes in motif usage even over just 24 hours of development, though these changes are far less prominent than those across the day/night. To identify motifs with finer-grain temporal patterns, I shuffled the wild-type motif counts and calculated enrichment/constraint scores for each animal every hour. Initially I divided each day and night in half and compared morning/evening and early/late night motif enrichment/constraint. This approach achieved roughly 35% error for both comparisons, using 229 (morning/evening) and 26 (early/late night) motifs each (Appendix Figure 7b, page 117, and Appendix Table 1, page 123). Similar to the developmental data this result, suggests a lack of major changes in motif usage across these periods. Next, I analysed each hour separately. Aside from the hours around the lighting transitions, 09:00-10:00 and 23:00-24:00, none of these classifiers performed better than their majority class models (Appendix Figure 7b, page 117, and Appendix Table 1, page 123), which suggests that across the day or the night, motif usage is fairly consistent. At the transitions, where the classifiers did achieve good performance, this approach identified motifs with startle-like patterns (Figure 8c). Thus, revealing the types of motifs that lead to changes in compressibility at the lighting transitions (Figure 6b). Together these results demonstrate that motif usage varies between the day and the night, but aside from the lighting transitions, is relatively consistent within these periods.

Next, I sought to identify dose-dependent or dose-specific pharmacological motifs. To achieve this, I used the mRMR algorithm to select motifs unique to each condition (control or dose of compound). I started by comparing control to all melatonin dosed larvae, irrespective of dose. Remarkably, using only the single best motif for this comparison already achieves only 10% classification error (Figure 9a) compared to the 25% (±4.42%) majority class classifier (Appendix Figure 7c, page 117, and Appendix Table 2, page 124). This motif consists of large movements and short pauses, and its enrichment/constraint score decreases dose dependently (Figure 9a). Extending this single best motif approach to each dose identified dose-dependent as well as dose-specific behavioural motifs (Figure 9a). A good example of the latter is the 10µM motif, two long pauses broken by a small sequence of active bouts, which is enriched at only 3 and 10µM doses (Figure 9a). Broadening the search to more motifs eventually separated each condition with 0-2.78% classification error (Appendix Figure 7c, page 117, and Appendix Table 2, page 124). Applying this approach to the PTZ
data achieved perfect classification, 0% error, between all conditions (Figure 9b, Appendix Figure 7c, page 117, and Appendix Table 2, page 124). In this case the second-best control vs PTZ motif (Figure 9b) is particularly interesting as it is enriched in PTZ-dosed larvae but constrained in control larvae, suggesting that this motif is highly seizure specific (Figure 9b). These pharmacological results demonstrate the ability of this approach to discriminate between the behavioural effects of small molecule doses, even within the same order of magnitude, a finding that will likely enhance predictive pharmacology approaches (Lamb et al., 2006; Rihel et al., 2010; Hoffman et al., 2016) by facilitating highly accurate predictions of dose.

Finally, I wanted to test whether hcrtr mutants expressed differential behavioural motifs, where other metrics had failed to detect any strong phenotypes. For example, based upon human and rodent literature, where loss of hypocretin is associated with narcolepsy (Lin et al., 1999), it could be reasonable to expect abnormal transitions between active and inactive bouts. Applying this approach to the hypocretin mutant data achieved reasonable performance when discriminating between hcrtr+/- and hcrtr-/- during both the day and night, though weaker performance when classifying hcrtr+/- and hcrtr+/+ (Appendix Figure 7, page 117, and Appendix Table 2, page 124). This result suggests that loss of the hypocretin receptor impacts motif usage; however, no motifs with large differences in enrichment/constraint between genotypes, as may have been expected from other models, were particularly evident.

Collectively, these results reveal that behavioural motifs are used context dependently and demonstrate how a supervised learning approach can be used to parse out subtle differences in motif usage between different times of the day/night cycle, doses of drugs, and genetic mutants.
Figure 8. Supervised Learning Identifies Day/Night Motifs

a. Enrichment/constraint scores for all 46,554 motifs (x-axis) for each fish during day/night 5 and 6 of development (y-axis). To emphasise structure motifs are sorted first by their average day night difference (from day to night enriched left to right), then separately day and night by fish, finally each motif’s enrichment/constraint score, is Z-scored for ease of visualisation.

b. Left: the 15 day/night mRMR motifs module sequences are shown numbered by the order in which they were selected by the algorithm, motifs are sorted by day-night enrichment/constraint score difference (middle). The long pauses at the end of motifs 5 and 14 are cropped at 10s (arrows). Middle: each dot is 1 of 124 wild-type animals, scores above zero are coloured light blue, below dark blue. Overlaid is a mean and standard deviation. Right: a tSNE embedding of this 15-dimensional motif data into a 2-dimensional space, each circle represents a single day (light blue) or night (dark blue) sample.

c. Representative motif temporal dynamics; shown are motifs 1 (day) and 2 (night) from b, as well as a startle-like motif. Left: each motif’s module sequence. Right each motif’s mean enrichment/constraint score each hour, rescaled to 0-1.
Figure 9. Pharmacologically-relevant Behavioural Motifs

a. Left: module sequences for the single best motif for each melatonin comparison. Modules are coloured as elsewhere. Middle: for each dose’s single best motif, enrichment/constraint scores are shown for every dose, on a log x-axis. Each animal is shown as a dot, with a mean and standard deviation overlaid per dose. Right: a tSNE embedding from an expanded space of 912 unique motifs, to 2 dimensions, each animal is shown as a single dot underlaid by a shaded boundary encompassing all animals in each condition. See legend for middle and right panel condition colouring.

b. PTZ data displayed as in a, aside from that the x-axis in the middle panel is linear. Note that the control motif and corresponding enrichment/constraint score shown is mRMR motif 2, not 1, for this comparison. The tsne projection is from an expanded space of 338 unique motifs, to 2 dimensions. See legend for middle and right panel condition colouring.
Chapter Two: Behaviour

**Chd8 and bckdk Mutant Behaviour**

As demonstrated by the results above, this suite of analysis tools is capable of quantifying a broad range of behaviours in a hypothesis free manner, making it suited to studying a range of questions. The pipeline is, however, particularly well suited to the study of autism models as any differences in repetitive behaviour will be described by the compression metric, while any ‘comorbid’ phenotypes will be captured by other aspects of the pipeline. Seizures for example would be detected as obvious differences in motif usage, such as the control vs PTZ motif (Figure 9b) while sleep-related phenotypes would be captured by inactive module usage across the day/night cycle, akin to the melatonin data (Appendix Figure 5a, page 115). To provide proof of principal for this approach, I gathered behavioural data from two mutants in autism-associated genes one de novo mutation (chd8) and one rare inherited allele (bckdk) (please see Materials & Methods, Animal Husbandry, Fish Lines for more information on these mutants, page 82).

Neither chd8+/− nor chd8−/− mutants had prominent differences in any bout parameters (Appendix Figure 8a, page 118) or module probabilities (Appendix Figure 9a, page 119) compared to chd8+/+. Examining each module’s mean frequency per genotype over time, however, akin to Figure 4b, revealed that in the late evening chd8−/− had a higher frequency of short inactive modules, and correspondingly a lower frequency of longer inactive modules, compared to both chd8+/− and chd8+/- (Figure 10a). This difference was particularly evident around 22:30 p.m. (lights on: 09:00 a.m. to 23:00 p.m.) in inactive modules 2 (higher evening mean frequency in chd8−/−) and 4 (lower evening mean frequency in chd8−/−). Furthermore, these module frequencies were notably similar between genotypes during both the morning and the night (Figure 10a). Together these results reveal a lack of activity differences or gross changes in module probabilities in chd8 mutants, save for an evening specific reduction in inactive bout lengths.

bckdk−/− mutants had slight differences in active bout structure (Appendix Figure 8b, page 118): specifically, a higher mean active bout length during the day and the night, and higher mean active bout total during the day, though showed no difference in active module probabilities (Appendix Figure 9b, page 119). Both bckdk+/− and bckdk−/− mutants, however, showed day/night dependent differences in inactive module probabilities (Appendix Figure 9b, page 119). bckdk−/− mutants a had higher day probability of inactive module 1 than both bckdk+/− and bckdk−/− (Figure 10b). Additionally, both bckdk+/− and bckdk−/− mutants had higher night probabilities of inactive module 5 than bckdk+/- though did not differ from each other. As with chd8, these results reveal a lack of overt phenotypes in bckdk mutants, aside from time dependent shifts in inactive module probabilities. However, while both mutants manifest altered inactive bout lengths, each mutant’s phenotype is unique.

Though both mutants had distinct behavioural phenotypes in terms of their inactive bout structure, I hypothesised that both would share a common phenotype of increased repetitive behaviour, a characteristic feature of autism in humans. To quantify this, I compressed the modular sequences from these datasets in blocks of 500 modules. Neither chd8+/− nor chd8−/− mutants showed any difference in compressibility compared to chd8+/- (Figure 11a).
bckdk<sup>−/−</sup> mutants however were more compressive than bckdk<sup>+/−</sup> during both the day and the night (Figure 11a). bckdk<sup>−/−</sup> mutants showed an intermediate phenotype and did not differ significantly from either bckdk<sup>+/+</sup> or bckdk<sup>−/−</sup> (Figure 11a). Together these results demonstrate that bckdk<sup>−/−</sup> mutants, but not chd8 mutants, display increased repetitive behaviours during both the day and the night. Future work should aim to establish the predictive value of compression as a metric for repetitive behaviour. A good way to approach this question would be to test how a range of doses of risperidone and aripiprazole, compounds often prescribed to treat repetitive behaviours in autism despite weak evidence of efficacy (Jesner, Aref-Adib and Coren, 2007; Hirsch and Pringsheim, 2016), would alter compression. This approach could also be extended by examining the effects of these compounds upon bckdk mutant compressibility.

Finally, I hypothesised that the autism mutants would behave differently in terms of their behavioural motifs. To test this hypothesis, I first generated an expanded motif library by compressing the autism datasets full module sequences and merging the motifs from the autism dataset with those from the test dataset. This process expanded the motif library from 46,554 (test dataset) to 74,340 (combined test and autism datasets) unique motifs. Qualitatively this library had a similar structure to the test library, in terms of both motif lengths and the distribution of modules at each motif length. Finally, within this expanded motif space I trained classifiers to discriminate between pairs of genotypes (e.g. wild-type and heterozygous). Interestingly all classifiers outperformed their null models (Figure 11b and Appendix Table 3, page 125), though notably these results were not as striking as the day and night or pharmacological results (Appendix Figure 7b-c, page 117, and Appendix Tables 1 and 2, page 123), as the autism classifiers generally required more motifs and had higher classification errors. In chd8 mutant animals some comparisons achieved good performance with relatively few motifs. For example, comparing chd8<sup>+</sup> to chd8<sup>−/−</sup> during the night required only 27 motifs to achieve 18% (8.05 Std) classification error, compared to a majority class classifier with 38% (3.78 Std) performance (Figure 11b and Appendix Table 3, page 125). In bckdk mutants the classifiers tended to require more motifs (Figure 11b and Appendix Table 3, page 125). These results demonstrate that heterozygous and homozygous mutants for both genes organise their behaviour differently from typical animals. The level of classifier performance, however, suggests a lack of overt motif differences, such as those that may be expected if the mutants were displaying spontaneous seizures or other striking behavioural abnormalities at this level.
Figure 10. ASD Mutant Inactive Module Phenotypes

a. The average frequency of inactive module 2 (upper) and 4 (lower) during day (white background) and night (dark background) in the *chd8* dataset at 6dpf. Shown is a mean per genotype smoothed with a 15 minute running average. Time points relevant to Figures 13 and 14 are marked on the x-axis. n = 51, 96, 32 - *chd8*+/+, *chd8*+/− and *chd8*−/−.

b. Mean *bckdk* module probabilities for inactive modules 1 (left panel) and 5 (right panel). The module probabilities for each animal are shown as dot coloured by genotype, across days (white background) and nights (dark background) 5 and 6. Each genotypes mean value is marked by a black cross and a black line connects the mean values per time window. *bckdk*−/− had significantly higher inactive module 1 probabilities during the day compared to both *bckdk*+/+ and *bckdk*+/− (p < 0.05). *bckdk*+/+ and *bckdk*−/− both had significantly higher inactive module 5 probabilities during the night than *bckdk*+/+ (p < 0.05), though did not differ from each other. Bars show significantly different (*) comparisons. The circles on the end of the bars correspond to the genotypes compared.

Statistics are Dunn-Sidak corrected four-way ANOVA, adjusted for the following factors: development and experimental repeat. n = 73, 116, 75 - *bckdk*+/+, *bckdk*+/− and *bckdk*−/−.
Figure 11. Compression of ASD Mutant Behaviour

a. Mean day and night compressibility, per 500 modules per animal (pale lines), overlaid is a day and night population mean and standard deviation (bold). chd8 mutants (left panel), show no difference in compressibility. bckdk\(^{-/-}\) mutants (right panel), show higher compressibility than bckdk\(^{-/-}\) (p = 0.01). bckdk\(^{-/-}\) mutants do not differ significantly from either bckdk\(^{+/+}\) (p = 0.48) or bckdk\(^{-/-}\) (p = 0.14). P-values are from a Dunn-Sidak corrected four-way ANOVA adjusted for following factors: day/night, development and experimental repeat. Horizontal black bar, highlights the significant difference (*) between bckdk\(^{+/+}\) and bckdk\(^{-/-}\).

b. The classification error (%) for each ASD classifier. Real classifiers (colour) are shown as a mean and standard deviation from 10 fold cross validation. To illustrate performance majority class classifiers (grey) are shown as value and standard error of proportion. Horizontal orange and grey lines delineate chd8 and bckdk classifiers, for either combined or separate day (light blue underline) and night (dark blue underline) data. The number of motifs chosen for each classification and exact values for each classifier are detailed in Appendix Table 3 (page 125).
Discussion

Low-dimensional Representations of Behaviour

Low dimensional representations of behaviour, such as the Δ pixels metric employed here, clearly result in a loss of information, for example direction of movement or posture. Such metrics do however facilitate screening approaches and/or long-term tracking and in these contexts have provided biological insight into the molecular targets of small molecules (Rihel et al., 2010) and genetics of ageing (Churgin et al., 2017).

I built upon previous work using this assay, reviewed in (Barlow and Rihel, 2017; Oikonomou and Prober, 2017), by analysing the resultant data on a frame by frame basis, rather than in one minute bins. Conceptually this is an important step as larval behaviour is structured at sub-second timescales. Practically, this approach resulted in a substantial increase in behavioural resolution, which I quantified through direct comparisons of individual fish data as measured by the two methods (Appendix Figure 1c, page 110). This comparison revealed that the previous approach misses on average 1/3rd of the inactive periods ≥ 60s. This finding must be considered when interpreting results obtained through the older method.

This new approach to the data handling and analysis revealed several insights into both the behaviour of wild-type animals across time and the influence of small molecules upon larval behaviour. Furthermore, it revealed a novel mutant phenotype lower active bout length, standard deviation and total in hcrtr⁻/⁻ mutants (Appendix Figure 2d, page 112). This result supports behavioural and imaging data that implicates hypocretinergic activity in modulating swim bout amplitude in zebrafish (Naumann et al., 2010).

Despite the advantages of this approach over previous work, two predominant limitations must be considered. Firstly, within the set-up larvae have constrained space, thus it is likely that their swimming patterns will differ from those of larvae in larger arenas. They may for example not execute their typical slaloming swim pattern (Dunn et al., 2016) or may truncate bouts as they collide with the walls of the arena. One potential way to tackle this limitation would be to decrease the number of animals per experiment while increasing each animal’s arena size. Alternatively, it would be interesting to monitor the activity of groups of larvae in a single arena to see if they coordinate their inactive periods across the day/night cycle in some way. The second limitation that must be considered is the loss of information from the Δ pixels metric. Indeed, a recent study using centroid tracking in 96 well plates revealed that larvae show a day/night location preference within the well, and furthermore uncovered a mutant difference in this metric (Thyme et al., 2018). These results demonstrate that even within the confined space of a 96 well plate, location is an informative metric to record. It is likely that even more detailed behavioural measures, for example eye and tail angles, would yield similar insights. Such metrics could be extracted by skeletonization or even through the use of an autoencoder applied to the raw video frames from each well (Johnson et al., 2016).
Modular Descriptions of Behaviour

A key idea in ethology is that behaviour consists of stereotyped modules arranged into motifs (Lashley, 1951; Tinbergen, 1963). While early behavioural studies described behaviour in this manner through manual observations (Dawkins and Dawkins, 1976), recent advances in machine vision and learning have automated these processes (Todd, Kain and de Bivort, 2017). The behaviour of larval zebrafish is particularly well suited to these approaches as individual bouts can be easily segmented from the data and it is relatively easy to acquire many examples from even a single animal due to the high frequency of their movement (Kim et al., 2017). Leveraging these advantages, recent work using high-speed cameras and postural analysis, uncovered a locomotor repertoire of 13 swim types in larval zebrafish (Marques et al., 2018), although inactive bouts were not considered.

From the test dataset, my analysis approach identified 5 active and 5 inactive modules. As may be expected the active modules correspond to different shapes of Δ pixels changes (Figure 3a) and thus reflect swim bouts of different amplitudes and lengths. The inactive modules capture the range of pause lengths that larvae exhibit in this assay, from sub-second to minutes long (Figure 3a). Interestingly all modules are used with reasonably high and similar probability by all wild-type animals, demonstrating that these modules represent a set of common larval behaviours (Figure 4a). Furthermore, the temporal (Figure 4b) and pharmacological (Appendix Figure 5, page 115) shifts in these probabilities illustrates that module usage can be flexibly re-organised dependent upon behavioural context (Wiltschko et al., 2015). Collectively these results demonstrate the relevance and utility of these partitions in describing larval behaviour in this assay. The volume of the dataset and challenges inherent to any clustering solution, however do warrant critical appraisal.

One potential limitation is the clustering approach itself, which raises two issues. Firstly, it is possible that there are bout types not captured by these modules. For example, given the visual appearance of PTZ-induced seizures in fish (Baraban et al., 2005) it seems reasonable to assume that they should have formed a distinct module. One potential reason for this could be that the simple bout features extracted fail to capture all of the potential variance between bouts. As such an improved approach could be to extract more hand-engineered features or even to simply cluster the raw Δ pixels changes by aligning the start of each bout.

An alternative reason could be the volume of data, ≈ 30,000,000 bouts of each type, which necessitated clustering only a sample of the data. To sample data, it is necessary to decide both which and how many points to sample, factors that will affect any clustering solution. To limit the impact of these factors upon my results, I took an evidence accumulation-based approach (Fred and Jain, 2002, 2005), which allowed me to combine the results of hundreds of clustering solutions, each of which was built from an independent sample, in terms of both which and how many points were sampled from the data. An additional benefit of this approach is that while each solution identifies clusters with a Gaussian shape, combining results enables the detection of clusters of arbitrary shape, relaxing this prior assumption. While this approach likely limits the potential impact of sampling problems, it is not an ideal solution and it remains likely that rare bout types could have been missed.
through this approach. One potential solution to this problem would be to employ a two-step clustering approach, similar to that described in Marques et al., (2018). Such an approach could even allow for every bout to be clustered, if modified as follows. Firstly, data from each animal, an average of 118,318 bouts of each type per wild-type fish, could be clustered independently. Secondly, the clusters obtained from each animal could be compared and grouped to identify common bout types across animals. Together an expanded feature space and two-step clustering approach would likely enable the detection of a distinct seizure module, and perhaps other rare bout types and should be the focus of future work in this area. However, an alternative interpretation of the lack of a clear seizure module is that the visual appearance of PTZ-induced seizures derives predominantly from an alteration in active to inactive transitions, rather than from a novel bout type. This interpretation is supported by the clear seizure motif identified by the compression algorithm (Figure 9b).

A second limitation of the approach above is the assumption that all bouts are stereotyped and can consequently be fit into a module. This is a problem as it is reasonable to assume that some proportion of bouts are non-stereotyped and should be excluded from the classification, through some measure of good-ness of fit for example. In many approaches, this would be prudent (Berman, 2017). In trial based experiments, for example, simply excluding trials with poor behavioural fits may be reasonable. However, for assessing behaviour over time it would raise a problem, i.e. how should a non-stereotyped behaviour be represented in a sequence of behaviours over time? One solution could be to define a module of un-stereotyped behaviour, such that the sequence input to the compression algorithm would contain both stereotyped modules and one un-stereotyped module. On a motif level, motifs including this module may represent some aspect of behaviour, alternatively interpreting these sequences as stereotyped motifs may be unreasonable and thus they could be subsequently ignored. In terms of compression, the presence of this additional module would contribute to the compressibility of the sequence. This would arguably lead to inaccurate results, though it is hard to determine if this would be better or worse than instead assuming all bouts to be stereotyped, and thus including bouts with poor module fits. Consequently, how to best handle these bouts remains an area for further exploration.

An interesting solution to the clustering-associated challenges described above would be to instead compress each animal’s raw Δ pixels trace. This process would simultaneously identify both modules (now short sequences of Δ pixels activity, inactivity or combinations of the two) and motifs. Thus, eliminating the need for a distinct clustering step. Such an approach would, however, require the compression of sequences approximately 26 times as long, from an average of 236,636 modules to 6,308,514 frames per wild-type animal, and thus would currently be computationally intractable. Furthermore, through this approach two behavioural sequences would only be considered to be the same motif if every frame in the two were identical. Such a definition is likely overly stringent and would make comparisons between animals challenging.
Determinism and Stochasticity in Behaviour

In some contexts, it is beneficial for animals to execute coordinated patterns of behaviour. For example, larval zebrafish will efficiently search an environment by executing sequences of left and right turns (Dunn et al., 2016). In other contexts, more random behaviour will be advantageous. Predictable evasive actions, for example, could be exploited by predators (Maye et al., 2007). With these examples in mind, it should be apparent that all animal behaviour can be described on a scale of determinism and stochasticity, and that where on this scale an animal’s behaviour at any time falls will vary with behavioural context.

To map zebrafish behaviour in different contexts onto this scale I applied a compression algorithm to each animal’s modular sequence (Figure 5). Through this approach I found that wild-type behaviour was more compressive during the day than the night (Figure 5b and Figure 6a). This finding echoes recent work in Drosophila that revealed higher temporal predictability during the day than the night, as well as in females (Fulcher and Jones, 2017). A potential explanation for these findings comes from work in C. elegans (Gomez-Marin, Stephens and Brown, 2016) that demonstrated that animals who transition slowly between modules tend to be less compressive. As both zebrafish (Figure 1b) and Drosophila (Geissmann, Beckwith and Gilestro, 2018) transition more quickly between modules during the day than the night, it seems likely that this can explain the differences in compressibility/predictability. Collectively these findings reveal a potential behavioural law (Brown and de Bivort, 2017) across species, dictating that long module dwell times decrease behavioural predictability. One mechanism for this control could be that the constituent modules in a fast response are directly ‘chained’ such that once the first module is executed the remaining modules become far more likely. Conversely, longer timescale behaviours, such as rest, perhaps need to be less precise, so could be controlled on a module by module basis. Irrespective, however, of potential models, findings such as these provide a starting point to dissect these systems.

For future efforts applying compression to behavioural data, there are two avenues left to explore: what compression heuristic to use and how to compress data from multiple animals. Following the work of Gomez-Marin, Stephens and Brown (2016), I have here defined the best motif at any iteration as the most compressive, which represents a balance between the motif’s length and frequency. While this metric generally leads to the best compression (Nevill-Manning and Witten, 2000), it would be interesting to explore other measures, for example just frequency or length, and to examine what aspects of behaviour they would capture. The second avenue relates to comparisons between animals. Here, each animal is compressed individually, identifying motifs, which are later grouped into a common library. Whilst computationally tractable, this approach prevents certain comparisons across animals, for example identifying the most compressive motif across all fish. This issue could be solved by compressing a single sequence, containing all of the animal’s modular sequences joined end to end, with breakpoints to prevent inter-animal motifs. Compressing this long sequence would, however, be computationally demanding.
Contextual Behavioural Motifs

Animal behaviour consists of patterns that evolve over time. One approach to quantitatively describe these patterns is to calculate module transition probabilities. That is, given module A at time T, what is the probability of observing module A or B at time T + 1? Such approaches require a definition of the timescale at which to look, i.e. T + X, and consequently capture behaviour at only a single timescale. However, animal behaviours evolve over multiple timescales in parallel, from sub-second (Wiltschko et al., 2015) to time of day (Fulcher and Jones, 2017) and even with age (Churgin et al., 2017; Stern, Kirst and Bargmann, 2017). Quantitative proof of this idea comes from Berman, Bialek and Shaevitz (2016) who demonstrated behavioural structure at multiple timescales, even up to 20 minutes, in Drosophila behaviour. Consequently, compression offers a favourable alternative to transition models as it enables the simultaneous detection of behavioural motifs at a range of timescales.

Compressing and merging the motifs from the test dataset generated a library of 46,554 motifs (Figure 7a), each of which describes an alternating sequence of movements and pauses (Figure 7b). The motifs ranged in length from 0.1 seconds to 11.3 minutes, revealing the range of timescales at which larval behaviour is organised. Furthermore, module usage varied with motif length (Figure 7a). For example, longer motifs had lower probabilities of using the longer inactive modules (Figure 7a). This finding illustrates one way in which longer module dwell times may reduce compressibility (by shortening motif lengths). Additionally, such structure suggests clues as to the underlying control of these sequences. For example, perhaps the longer inactive modules lengths preclude mechanisms from linking them into motifs.

Comparing motif enrichment/constraint scores between contexts demonstrated both the utility of this approach and also revealed several insights into zebrafish behavioural patterns. For example, in the melatonin data it was possible to identify both dose-dependent and dose-specific motifs (Figure 9a). This result was particularly interesting, as while the former suggests an underlying mechanism that is continuously modulated, for example fraction of melatonin receptors bound, the latter indicates that different doses may have discrete effects upon neuronal circuitry, for example by binding low affinity receptors at higher doses. Given these results it would be exiting to generate and compare whole-brain activity maps for each dose to try to uncover common and unique neuronal correlates. Furthermore, comparing these neuronal activity maps to wild-type patterns of neuronal activity at night could identify normal sleep circuits that are engaged by melatonin.

The discriminatory power of this approach (Figure 9), which was capable of distinguishing between the behavioural effects of even small differences in compound doses, has great potential for future pharmacological studies. With a large library of small molecules, for example, it would be possible to use the motif space for predictive pharmacology drug discovery approaches, (Lamb et al., 2006; Rihel et al., 2010; Hoffman et al., 2016). One interesting question to address would be whether different epileptogenic compounds induce seizures with common or unique motifs. Similarly, one could compare the spontaneous
seizures observed in some zebrafish mutants to these compounds, which could suggest clues as to their underlying mechanisms.

Despite the ability of this approach to capture motifs at a range of timescales, from subseconds to minutes, in larval behaviour, it remains possible that there are longer timescales behaviours not described by this approach. This is due to the fact that for faster computation of compression, it is necessary to look for motifs of only up to a certain length. Here I looked for motifs up to 10 modules long, though through nesting obtained motifs up to 20 modules long. In future it would be interesting to try to identify longer motifs, to uncover the full range of timescales at which larval behaviour is organised.

**ASD Mutant Behaviour**

To test if zebrafish mutants in autism-associated genes showed behavioural abnormalities, I employed a high-throughput, long-timescale behavioural assay (Rihel, Prober and Schier, 2010), reasoning that such an approach would allow for the detection of a broad range of phenotypes. This concept was supported by the results from the test dataset, which demonstrated the ability of this assay and analytical approach to detect several autism-relevant phenotypes including activity/inactivity differences (Figure 2b), seizures (Figure 9b) and altered behavioural repetitiveness (Figure 5b).

*chd8*⁻/⁻ zebrafish exhibited a highly specific alteration in their behaviour: a decrease in the length of their inactive bouts during the evening (Figure 10a). Future behavioural work should aim to explore potential mechanisms for this phenotype, especially if it relates to circadian, light sensing or sleep mechanisms. To test if the phenotype is related to circadian mechanisms mutants entrained to a light/dark cycle could be placed into constant conditions (e.g. constant dark or light). If the mutants still exhibited reduced inactive bout lengths in the evening, this could suggest that the phenotype is circadian-related. In this case testing the level of wild-type chd8 gene or protein expression over the day/night cycle could be informative. To test if the phenotype is related to differences in light sensitivity a light/dark preference paradigm could be employed, even across the entire day/night cycle. Interestingly in 2 of 3 studies Chd8 mutant mice show a preference for dark over light areas (Katayama et al., 2016; Platt et al., 2017; Suetterlin et al., 2018), perhaps suggesting that altered light sensitivity could also explain the zebrafish chd8⁻/⁻ phenotype. Testing the final possibility, that the chd8⁻/⁻ phenotype is sleep related, would be informative, as in humans it remains unclear if CHD8 mutations are directly causal for sleep disruption (Bernier et al., 2014). Determining which, if any, of the behaviours described here correspond to sleep will, however, require further behavioural analysis in wild-type animals. One way to approach this question would be to deliver open-loop stimuli and to see if there is any correlation between prior inactive bout length and response, as well as any relationship between this data and the inactive modules or night-enriched motifs (van Alphen et al., 2013). Another approach would be to use a closed loop paradigm to deprive animals of specific sleep bout lengths or perhaps even specific behavioural motifs, and to see how this impacts their behaviour (Geissmann,
Together these additional behavioural tests would provide clues as to the mechanisms underlying the $chd8^{-/-}$ mutant phenotype.

I had hypothesised that all of the autism-mutants would display more repetitive behaviours, as measured by compressibility, though neither $chd8^{+/+}$ or $chd8^{-/-}$ mutants showed any difference by this metric. Interestingly, however $Chd8$ mutant mice do also not show any differences in repetitive behaviour, as assayed by grooming, marble burying or nest building (Katayama et al., 2016; Gompers et al., 2017; Platt et al., 2017; Suetterlin et al., 2018). Together, these results in mice and zebrafish suggest either a unique role relating to repetitive behaviour for the $CHD8$ gene in humans, or that the repetitive behaviours observed in $CHD8$ patients are not directly genetically determined.

In humans’ homozygous mutations in the $BCKDK$ gene have been linked to neurobehavioral deficits including autistic features (García-Cazorla et al., 2014), and in some cases autism with epilepsy (Novarino et al., 2012). This link is supported by two mouse studies, which describe repetitive hind limb clasping and seizures when $Bckdk^{-/-}$ mice are hung by the tail (Joshi et al., 2006; Novarino et al., 2012). Here several behavioural abnormalities were evident in $bckdk$ mutant zebrafish, including day/night differences in inactive module probabilities (Figure 10b), aberrant behavioural motifs (Figure 11b) and even increased repetitive behaviours (Figure 11a). Establishing whether these differences can be explained by the presence of spontaneous seizures remains an interesting future direction. One way to address this would be to compare $bckdk^{-/-}$ behaviour to that of larvae dosed with PTZ. Alternatively, as it is theoretically possible that spontaneous seizures could differ from PTZ-induced seizures, studying the $bckdk^{-/-}$ enriched motifs themselves may be informative.

In mice (Joshi et al., 2006; Novarino et al., 2012; García-Cazorla et al., 2014), it has been shown that dietary supplementation of branched chain amino acids (BCAA) ameliorates $Bckdk$ mutant phenotypes. Thus, an interesting future experiment would be to test if the addition of BCAAs would rescue some, or even all the BCKDK mutant zebrafish phenotypes. If this were the case, the zebrafish would be an ideal model in which to address several questions, for example mapping individual BCAAs to phenotypes. Conversely, seeing if reduction of larval BCAA concentrations by the administration of a targeted small molecule (Tso et al., 2013) recapitulates the mutant phenotypes, could provide additional insight into the developmental/acute nature of the phenotype.

Ultimately these mutant phenotypes provide proof of concept for the study of autism-associated genes in zebrafish, and a starting point for dissecting their aetiology. Furthermore, this work establishes a generalizable framework with which it will be possible to address fundamental questions about the genetics of autism, once expanded to a library of mutants in autism-associated genes.
Chapter Three: Circuits

“The brain-waves recorded on neatly squared paper in trembling peaks and troughs are the mirrors of the combined thought-pulses of billions of cells. Theoretically, analysis should reveal the thoughts and emotions of the subject, to the last and least”

Isaac Asimov, Second Foundation.

Mapping Behaviour to Circuits

In Marr’s framework for understanding how systems execute complex information-processing tasks (1982), three levels of understanding are necessary: Computational Theory – understanding the logic of how the task is executed; Algorithm and Representation – understanding how the systems inputs can be transformed to outputs; and Hardware Implementation – understanding how the algorithm could be physically realized. As applied to the work here, behavioural analysis provided insight into the structure of zebrafish behaviour across the day/night cycle (see Behaviour, page 25). These results can be considered as Computational Theory, as they reveal the logic by which zebrafish organise their behaviour. From this data, it is possible to gain algorithmic and representational understanding by generating hypotheses about the control of this logic. Consider for example the dose-dependent reduction in the enrichment/constraint of the single best 0.01µM melatonin motif (Figure 9a). The logic of this behaviour suggests that its control mechanism is being continuously modulated. Thus together, these first two levels provide understanding that can help to guide the search for a physical representation of the mechanism. Here I will focus upon identifying potential neuronal correlates for behavioural phenotypes (their Hardware Implementation) under the assumption that differential behaviour will be reflected by differences in neuronal activity. Extending this approach to the melatonin example, the focus would be upon identifying a brain area with dose-dependently modulated activity.

While Marr’s framework can help to formulate hypotheses about the type of activity patterns we may expect for a phenotype, it is not obvious where in the brain to look for these signals. Consequently, a hypothesis free approach (in terms of brain area) is preferable. Zebrafish are ideally suited to such approaches for two reasons. Firstly, at 6 days post-fertilisation the zebrafish brain contains only approximately 100,000 neurons and is only roughly 600 x 300 x 300µm (length x width x depth) in size, making it easy to acquire whole-brain imaging data. Secondly, at larval stages zebrafish are optically translucent; thus, fluorescence can be imaged in fixed tissue without the need for dissection or even non-invasively in live animals.

Leveraging these advantages, a range of tools are available for mapping behaviour to circuits in zebrafish at a whole brain level. These methods can essentially be divided into static (‘snapshots’ of neuronal activity) and dynamic (real-time monitoring of neuronal
activity) approaches. There are three predominant examples of static methods. Firstly, in-situ hybridisation staining for c-fos, a gene whose expression is up-regulated in response to neuronal activity (Hunt, Pini and Evan, 1987). Secondly, immunostaining for the proteins phosphorylated and total extracellular signal-regulated kinase (pERK and tERK), the ratio of which (pERK/tERK) (Randlett et al., 2015) can be used as a readout of neuronal activity as ERK is phosphorylated in response to membrane depolarization (Rosen et al., 1994). The final static approach that has been employed is photo-conversion of a genetically encoded calcium sensing protein (CaMPARI), which irreversibly changes fluorescence when elevated intracellular Ca²⁺ (due to neuronal activity) and experimenter delivered UV light are coincident (Fosque et al., 2015). This protein can be expressed throughout the zebrafish brain by the use of a pan-neuronal promoter and thus whole-brain ‘snapshots’ of neuronal activity can be acquired. Notably these static methods allow for animals to be freely-behaving up until the ‘snapshot’ is acquired. In the case of c-fos and pERK/tERK this involves sacrificing the animals, which can be done rapidly by transferring them to fixative. For CaMPARI acquiring the snapshot involves exposing the larvae to UV light (Fosque et al., 2015). The mainstay dynamic approach is GCaMP based imaging, in which a genetically encoded calcium indicator, whose fluorescence increases with intra-cellular Ca²⁺, is pan-neuronally expressed. Primarily, this approach has been applied to animals tethered under a microscope (Ahrens et al., 2012), though more recently it has been extended to free swimming zebrafish by compensating for the fishes motion in real-time as it swims under the microscope (Kim et al., 2017).

Having acquired neuronal activity data, it is often desirable to compare this data between samples or to pair it with information from other sources, such as anatomical labels. While these steps can be carried out manually, recent work in zebrafish has automated these processes. Specifically either colorimetric (Allalou et al., 2017) or fluorescently labelled brains (Ronneberger et al., 2012; Marquart et al., 2015; Randlett et al., 2015) can be aligned, in three dimensions, to a common reference brain via a process termed registration. To achieve these alignments, most registration algorithms make use of both rigid, for example rotation and translation, and non-rigid transformations, including local sample deformation. Registration facilitates two broad categories of comparisons. Firstly, it enables statistical comparisons between samples, including in-situ hybridisation data (Allalou et al., 2017), immunostaining patterns (Randlett et al., 2015) or even brain morphology (Gupta et al., 2018). Secondly, it allows novel data to be compared to libraries of existent data (atlases), including anatomical labels, staining patterns, transgenic lines (Ronneberger et al., 2012; Marquart et al., 2015; Randlett et al., 2015) and even data from serial electron-microscopy (Hildebrand et al., 2017). Together static and dynamic methods for mapping neuronal activity, and registration techniques form a powerful toolkit for mapping behaviour to circuits in zebrafish.

For dissecting behavioural phenotypes from this assay, I reasoned that a good approach would be to employ a static method to generate whole-brain activity maps from freely behaving animals. The static component was important as it facilities testing multiple conditions (for example doses of drugs or different genotypes) and time-points. The freely behaving aspect was important given a lack of evidence that tethered zebrafish behaviour over long timescales
will recapitulate free swimming behaviour. With these requirements in mind I decided upon pERK/tERK based activity mapping (Randlett et al., 2015) as it presented several advantages over the other static methods. Compared to c-fos, pERK/tERK mapping has faster temporal dynamics generating signals within 5 minutes of activation (Randlett et al., 2015). Compared to CaMPARI (Fosque et al., 2015), pERK/tERK mapping doesn’t require photo-conversion by UV light (Fosque et al., 2015), which zebrafish find highly noxious and will thus alter their behaviour and likely neuronal activity (Guggiana-Nilo and Engert, 2016).

pERK/tERK mapping consists of three steps (Appendix Figure 11, page 121). Firstly samples are fixed, dissected and stained for phosphorylated and total extracellular signal-regulated kinase (pERK and tERK) (Materials & Methods, page 94). Secondly samples are imaged (Materials & Methods, page 95) and registered to a reference brain using both rigid and non-rigid transformations (Materials & Methods, page 87). This step is carried out using the tERK channel, as this protein is broadly expressed with some regional variability throughout the larval brain, then these transformations are applied to the pERK channel. Finally, voxel-wise statistical comparisons of pERK/tERK ratio are made between groups (for example genotypes) to identify areas with differential levels of neuronal activity. In zebrafish it is necessary to normalise pERK by tERK level, as tERK is non-homogeneously expressed throughout the larval brain (Randlett et al., 2015). The validity of this ratio as a measure of neuronal activity in zebrafish was established by comparisons of pERK/tERK ratio to compounds with established effects upon neuronal activity (Randlett et al., 2015); for example, PTZ induced seizures increased pERK/tERK ratio pan-neuronally, while anesthetic induced sedation reduced pERK/tERK ratio in all regions aside from the olfactory epithelium. Furthermore, comparisons of dynamic neuronal activity (measured by GCaMP fluorescence) and pERK/tERK ratio in individual samples (fixed immediately after GCaMP imaging), revealed a strong correlation between these two measures (Randlett et al., 2015). It should, however, be noted that as ERK has many functions (Pearson et al., 2001), some pERK/tERK signal may not be due to neuronal activity. Consequently, here this approach is predominantly employed as a first-pass measure to identify neuronal correlates, whose definitive causal link to behaviour will need to be determined by further experiments.

Here I first describe work to establish these tools in the lab, including registering fluorescent in-situ hybridisation data and comparing c-fos to pERK/tERK patterns. Finally, I present my findings from comparing wild-type and chd8 mutant whole-brain activity maps at three different times of the day/night cycle.
3D Sample Alignment

As discussed later quantifying registration accuracy is challenging. Thus, initially I simply sought to set-up and qualitatively test the registration pipeline (Appendix Figure 11, page 121). For this purpose, I acquired whole-brain stacks from samples stained fluorescently for galanin mRNA and tERK for the registration channel (Materials & Methods, page 93). I choose galanin as it has a well-defined and relatively sparse expression pattern consisting of two-distinct bilaterally symmetrical populations, from which it would be easy to spot obvious registration artefacts. I compared galanin expression between seven individual larvae taken from two independent experiments and imaging sessions (Figure 12a). Qualitatively one can see that the registration produces a great deal of overlap between the samples, as seen from both a top-down as well as side on projection. It is notable however that not all the stains are perfectly overlapping, though in these cases the patterns are only misplaced by at most a few cell bodies. Furthermore, it is difficult to determine if such disparities are due to the registration process itself or individual variation in galanin expression. To illustrate the insights that can be gained from having data registered to an atlas I used the Z-Brain Browser (Randlett et al., 2015) to explore galanin’s expression pattern. This approach revealed that one population is located in the dorsal, rostral preoptic area and the other in the ventral part of the intermediate hypothalamus. Overall this data suggested that the registration pipeline (Appendix Figure 11, page 121) could qualitatively achieve reasonable sample alignment.

Whole Brain Activity Mapping

To test the whole-brain activity mapping method described in (Randlett et al., 2015), I collected pERK/tERK data from pitpnc1a (a lipid transported expressed exclusively in the brain) mutant larvae, which are known to show upregulated c-fos expression in several brain areas (Ashlin et al., 2018). Comparing the results from the two methods qualitatively demonstrated general agreement with extensive overlap in areas including the dorsal and ventral telencephalon and hypothalamus (Figure 12b and Figure 12c). pERK/tERK activity mapping, however, revealed a far more widespread increase in pitpnc1a\(^{-/-}\) neuronal activity than detected by c-fos, uncovering additional upregulated areas including the hindbrain (Figure 12b and Figure 12c). Mapping this data using the Z-Brain atlas (Randlett et al., 2015) also revealed that several of the areas with the most strongly upregulated pERK/tERK signal in pitpnc1a\(^{-/-}\) mutants, are implicated in the control of arousal, including: hypocretinergic and dopaminergic neurons in the hypothalamus (Figure 12b and Figure 12c). These findings are broadly consistent with the increased behavioural activity of these mutants (Ashlin et al., 2018). Though sparse a few voxels in the telencephalon also showed a lower pERK/tERK ratio in pitpnc1a\(^{-/-}\) mutants, a difference not detected from comparisons of colorimetric c-fos staining (Figure 12b and Figure 12c). Ultimately, this result demonstrates the accuracy of the pERK/tERK mapping method in detecting differences in neuronal activity, and furthermore highlights this methods increased sensitivity and power over manual comparisons of colorimetric c-fos staining.
Figure 12. Establishing Registration Tools

a. Fluorescent galanin mRNA expression in 7 individual 6dpf brains. Each sample is individually coloured and is shown both pre (left) and post-registration (right), using anti-tERK as the reference channel for registration. Samples are shown as maximum intensity projections, from a dorsal-ventral (upper panels) and lateral view (lower panels).

b. Representative colorimetric c-fos expression in pitpnc1a<sup>-/-</sup> (upper panels) and pitpnc1a<sup>+/+</sup> 6dpf larvae (lower panels) n = 5 of each genotype. Brains are shown from a ventral-dorsal (left panels) and lateral view (right panels). R - right, L - left, V - ventral, D - dorsal, Tel - telencephalon, Hb - hindbrain, Hyp - hypothalamus. Black arrows show areas of agreement with pERK/tERK mapping (see c). Adapted from Ashlin et al., 2018.

c. pERK/tERK comparison of dark-reared, 6dpf pitpnc1a<sup>-/-</sup> (n = 15) and wild-type (n = 12) brains. Green indicates increased signal, while magenta indicates decreased signal, in mutant brains relative to wild-type. Data is shown as a thresholded maximum intensity projection overlain on a maximum intensity projection of the Z-Brain tERK reference (grey). R - right, L - left, V - ventral, D - dorsal. White arrows show areas with notable agreement with c-fos results (see b). Magenta arrows highlight down-regulated voxels.
Comparing Whole Brain Activity in ASD Mutants

chd8 Algorithm and Representation

I sought to determine if pERK/tERK mapping could be used as a tool to uncover differences in autism-mutant neuronal activity. Such differences would provide a starting point for uncovering the mechanisms underlying their aberrant behaviour, and would enable comparisons of whole-brain activity across mutants to identify commonly affected circuits.

Due to its specificity I decided to focus upon mapping the chd8-/- behavioural phenotype (reduced evening inactive bout lengths) (Figure 10a) to circuits. Considering potential Algorithms and Representations for this phenotype allowed me to generate three hypotheses about the neuronal data based upon the behavioural data. Firstly I hypothesised that chd8-/- neuronal activity would differ from chd8+/- during the evening and would be similar at other timepoints. To test this hypothesis I reasoned that samples should be taken during the morning, the evening and the night. By visually examining the behavioural data (Figure 10a) and determining when chd8-/- behaviour was most different to chd8+/- and chd8+/- I decided upon an evening timepoint of 22:30 p.m. (lights on: 09:00 a.m. to 23:00 p.m.). By looking for time-points with similar behaviour between genotypes, I decided upon 09:30 a.m. and 06:30 a.m. (Figure 10a). Note that in terms of development the timepoints were spaced 09:30 a.m., 22:30 p.m. and 06:30 a.m. Secondly I hypothesised that chd8+/- neuronal activity would be similar to chd8+/- Therefore, to test this hypothesis I decided to sample brains from a heterozygous in-cross and compare mutants from all three genotypes. Finally, based upon the observation that short inactive bout lengths reduce their frequency across the day (Figure 4b), I hypothesised that areas with altered chd8-/- signal, would modulate their activity over time in chd8+/- and chd8+/-.

chd8 Physical Realization

Approach

To test these hypotheses, I executed the following protocol in triplicate for each of the three timepoints. Briefly, free swimming 6-day post fertilisation larvae from a heterozygous in-cross were fixed at each timepoint under the appropriate lighting conditions (i.e. light at 09:30 a.m. and 22:30 p.m. and dark at 06:30 a.m.), each samples brain was dissected out and samples were genotyped (see Materials & Methods, page 92), sorted by genotype, stained using antibodies against pERK and tERK (see Materials & Methods, page 94) then imaged using a two-photon microscope (see Materials & Methods, page 95). After imaging samples were genotyped again and samples of incorrect or unclear genotype were discarded. Finally, all brains were registered to the Z-Brain tERK reference brain and activity maps were generated and compared across genotypes and timepoints (Randlett et al., 2015). Ultimately, this approach resulted in a final dataset consisting of 265 whole-brain 2-channel stacks, with an average of 29 brains per genotype and timepoint.
Results

Heterozygous mutants (\textit{chd8}^{+/-}) showed no overt differences in neuronal activity compared to \textit{chd8}^{+/+} at any of the three time points (Figure 13). This finding is consistent with the similarity of their behaviour to \textit{chd8}^{+/+}, and supports pERK/tERK ratio as a metric for detecting behaviourally relevant differences in neuronal activity. Homozygous mutants (\textit{chd8}^-/-), however, showed altered neuronal activity in the late evening, and little difference at either the morning or night timepoints (Figure 13). Specifically, at 22:30 p.m. \textit{chd8}^-/- pERK/tERK signal was increased compared to \textit{chd8}^{+/+} in a small area in the right hemisphere (Figure 13). Using the Z-Brain atlas (Randlett \textit{et al.}, 2015) I determined this area to be the most ventral and lateral portion of the tectum stratum periventriculare (Figure 13). Interestingly, more dorsal and medial areas of the tectum stratum periventriculare generally contain the cell bodies of tectal neurons involved in visual processing (Del Bene \textit{et al.}, 2010). Thus, the difference in activity in this area may indicate a difference in visual processing in \textit{chd8}^-/- mutants. This finding raised two interesting questions. Firstly, how does activity in these voxels vary over time in the different genotypes and secondly, why is this difference asymmetric?

To address these questions, I started with the voxels (each 4.7μm$^3$) with increased \textit{chd8}^-/- pERK/tERK signal at 22:30 p.m. as defined by the Z-Brain analysis (Randlett \textit{et al.}, 2015). These voxels (only 0.08% of the entire brain) were predominantly located in the ventral and lateral portion of the tectum stratum periventriculare though do not form a single continuous area (Figure 13). As measured by pERK/tERK, each voxel’s signal theoretically ranges from 0, in which case no ERK is phosphorylated, to 1, in which case 100% of ERK is phosphorylated. From each of the 265 brains in the dataset I extracted these voxels and calculated their mean pERK/tERK value per brain (Figure 14a). To gain insight into the asymmetry of the \textit{chd8}^-/- pERK/tERK signal (Figure 13), I identified the corresponding voxels on the left side of the brain by horizontally reflecting the 3-dimensional increased \textit{chd8}^-/- pERK/tERK signal and additionally extracted these voxels from all 265 brains (Figure 14b). Finally, I compared these mean pERK/tERK values by using a 3-way ANOVA (factors: genotype, time and hemisphere) and a Dunn-Sidak post-hoc test to identify significant differences. This approach revealed several insights into the neuronal activity of these voxels across genotypes, time and hemisphere.

The first insight was into the pERK/tERK level of these voxels. Interestingly, at all-time points irrespective of genotype, the activity of these voxels was relatively high, falling around the 70% percentile of all pERK/tERK signals from this dataset. Comparing the values for each timepoint and hemisphere revealed that the only significant difference between genotypes was at the 22:30 p.m. timepoint at which \textit{chd8}^-/- mutants had higher mean signal than \textit{chd8}^{+/+} in the right hemisphere (Figure 14a). This finding both supports the results of the Z-Brain analysis (Figure 13) as well as the validity of representing the data as a simple mean value per animal. Interestingly, \textit{chd8}^-/- signal at 22:30 p.m. in the right hemisphere showed an intermediate phenotype and did not differ significantly from either \textit{chd8}^{+/+} or \textit{chd8}^-/- (Figure 14a). Additionally, \textit{chd8}^-/- trended towards higher activity in these voxels at all timepoints in both hemispheres, though none of these differences were significant as determined by either this approach or the Z-Brain analysis.
Next, I focussed upon the temporal pERK/tERK signal of these voxels. In $chd8^{+/+}$, signal in these voxels varied over time in both hemispheres, being highest in the morning and then decreasing significantly by 22:30 p.m. At 06:30 a.m. activity resembled 22:30 p.m. in both hemispheres though did not differ significantly from either of the other timepoints (Figure 14). The temporal pattern in $chd8^{-/-}$ was qualitatively similar to $chd8^{+/+}$ though was only significantly decreased in the left hemisphere between the morning and the other two timepoints (Figure 14b). In contrast, $chd8^{+/-}$ signal remained constant over time in both hemispheres (Figure 14).

Finally, I focussed upon the asymmetry of the altered $chd8^{-/-}$ signal by using separate 2-way ANOVAs for each genotype (factors: time and hemisphere). In $chd8^{+/+}$ there was a significant effect of time, no gross difference between hemispheres and no interaction between time and hemisphere, demonstrating that in $chd8^{+/+}$ there is no asymmetry in the morning to evening reduction of pERK/tERK signal in these voxels (Figure 14). In $chd8^{+/-}$ pERK/tERK signal was significantly modulated across time and significantly higher in the right hemisphere, though there was no interaction between time and hemisphere (Figure 14). These findings reveal that, unlike $chd8^{+/+}$, en masse pERK/tERK activity is asymmetric in these voxels in $chd8^{+/-}$, though like $chd8^{-/-}$ the morning to evening reduction in signal is not asymmetric. Finally, in $chd8^{-/-}$ there was no significant effect of time, and although signal was significantly higher in the right hemisphere there was no iteration between hemisphere and time (Figure 14). These results demonstrate that $chd8^{+/-}$ pERK/tERK activity in these voxels does not vary with time, though like $chd8^{+/-}$ pERK/tERK signal is asymmetric, being higher in the right than the left hemisphere.

Collectively, these results provide insight into the two questions posed above about the pERK/tERK signal of these voxels. Regarding their activity over time, these results reveal that in both $chd8^{+/+}$ and $chd8^{+/-}$ activity in these voxels decreases from morning to evening (Figure 14). In contrast, in $chd8^{-/-}$ mutants pERK/tERK signal is constant over time. Regarding the asymmetry, they reveal that in both heterozygous and homozygous animals signal is elevated on the right compared to the left hemisphere (Figure 14). Together, these two differences, the morning/ evening decrease in $chd8^{+/+}$ and increased $chd8$ mutant signal in the right hemisphere give rise to the $chd8^{-/-}$ neuronal phenotype.

One caveat that should be kept in mind with this approach however, is that the transformation of voxels from the right to the left side of the brain can only be considered an approximation to the corresponding area. Furthermore, it is possible that the differences identified at any of these levels (genotype, time or hemisphere) could diverge at the level of individual voxels, as compared to this en masse approach. For example, hypothetically some of these voxels could increase not decrease their activity morning to evening in wild-type animals. Consequently, exploring this data on a voxel-wise basis remains an interesting direction for future work. Furthermore, extending this approach to study every voxel in the brain could reveal neuroanatomical areas with time-dependent differences in neuronal activity, even across hemispheres, thus providing a starting point for identifying sleep and circadian related neuronal circuits on a whole-brain level.
Figure 13. Whole-Brain Activity Mapping in chd8 Mutants
Voxels with significantly higher (green) or lower (magenta) pERK/tERK signal are shown overlain on the Z-Brain reference tERK stack (grey). Comparisons are shown between heterozygous and wild-type animals (upper row), and homozygous and wild-type animals (middle row), from brains sampled at 09:30 a.m. (left), 22:30 p.m. (middle) and 06:30 a.m. (right). Green and magenta colouring indicates higher or lower signal in mutant, compared to wild-type brains. Data is shown as maximum intensity projections from a dorsal-ventral view, as well as from a lateral view. L - left, R - right, V - ventral and D - dorsal. n denotes the number of mutant samples used for each comparison, the number of wild-type samples for each comparison is as follows: 09:30 a.m. - 32, 22:30 p.m. - 29, and 06:30 a.m. - 30.
Figure 14. Mapping Neuronal Activity Across Genotypes, Time and Hemisphere

a. At each time-point each brains mean pERK/tERK value for the voxels shown in the right panel (green - extracted voxels, grey - maximum intensity projection of Z-Brain tERK reference) is shown as a circle, coloured by genotype. Black crosses mark each genotypes mean, coloured lines connect each genotypes mean across time. All significant differences from a Dunn-Sidak corrected 3-way ANOVA (factors: genotype, time, hemisphere) are marked with a bar and a *, the circles on the ends of the bar shown the genotypes compared. L - left, R - right. The number of brains of each genotype at each time-point is the same as in Figure 13.

b. As in a but voxels from the left-side of the brain were extracted from a horizontally reflected mask (see right panel). Data in a and b was compared for each genotype independently by an additional 2-way ANOVA comparison (factors: time and hemisphere, with time and hemisphere interaction). Significant differences (p < 0.05) were: chd8+/+ - time. chd8+/+ - time and hemisphere. chd8+/- - hemisphere.
Discussion

Registration

The use of registration algorithms in zebrafish research is an exciting development as it enables both statistical comparisons between samples as well as the mapping of novel data to existent datasets. Indeed, here these techniques provided information on typical galanin expression as well as time of day and mutant differences in whole-brain neuronal activity patterns.

One unresolved question for these techniques, however, is how to quantify registration accuracy, i.e. how well can a given sample be aligned to a reference brain? One approach to this problem has been to measure distances between landmarks in registered samples, for example the retinal outer plexiform layer (Ronneberger et al., 2012) or Mauthner cell (Randlett et al., 2015). This approach however fails to capture the quality of the registration across the whole brain and furthermore it seems likely that such easily identifiable landmarks will achieve better alignment than other, less well-defined areas of the brain. An alternative approach has been the use of either whole-brain (Marquart et al., 2015) or landmark based correlation metrics (Marquart et al., 2017) between samples. Correlation in this context can, however, be a misleading measurement, as increased correlation can result from excessive sample distortion (Rohlfing, 2012), and consequently alignments with high correlation must be manually inspected for excessive deformations, introducing a subjective element to this process.

The challenge of quantifying registration accuracy generates several issues including how to choose a reference sample, how to tune registration parameters and finally how to compare quality between registration algorithms. Here I simply followed the Z-Brain framework and used the computational morphometry toolkit to align my samples to its tERK reference brain using similar registration parameters (Randlett et al., 2015). As is evident from the aligned galanin expression patterns and c-fos to pERK/tERK comparison (Figure 12), this approach produced reasonable results. Two methods would however likely improve the quality of the registration approach used here. Firstly, it may be beneficial to use one of my own wild-type samples as a reference brain. This is due to several factors including the potential genetic and development differences between wild-type zebrafish in independent laboratories, as well as potential differences in staining and imaging procedures. An alternative solution to this problem, would be to generate an average wild-type brain to use as a reference, for example through iterative shape averaging (Rohlfing et al., 2001; Allalou et al., 2017). Secondly, an alternative registration algorithm could be used (Marquart et al., 2017). Conversely, quantifying the benefit gained from these approaches would be challenging, and ultimately the use of follow up experiments to test the findings from the current approach would likely be more informative.
Activity Mapping in chd8 Mutants

To test the hypothesis that aberrant chd8 behaviour is caused by changes in neuronal activity, I applied the pERK/tERK method to chd8 mutants at a set of behaviourally relevant time points. This approach revealed elevated neuronal activity in part of the right ventral tectum stratum periventriculare in chd8⁻/⁻ mutants (Figure 13). This difference was particularly evident during the late evening as in chd8⁺/+ larvae, this area’s activity decreased from a morning peak, to a low evening level, while in chd8⁻/⁻ animals activity remained consistently high across time, particularly in the right hemisphere (Figure 14). These findings corroborate the behavioural evidence (Figure 10a), which predicted that any potential differences in chd8⁻/⁻ neuronal activity would be most evident in the late evening and absent in chd8⁺/+, and thus support a link between the neuronal and behavioural data.

Together, these datasets reveal a correlation between the pERK/tERK signal of this area and short inactive bouts. This link supported by two lines of evidence. Firstly, in chd8⁺/+, pERK/tERK signal is highest in the morning (Figure 14) when inactive bouts are short (Figure 10a) and decreases in the evening when inactive bouts are longer (Figure 14). Interestingly, signal in this area does not decrease further during the night as inactive bouts lengthen, perhaps suggesting that it has a minimal level of pERK/tERK signal or alternatively that it is only involved in controlling inactive bout length during the day. Secondly, in chd8⁻/⁻ increased pERK/tERK signal in this area in the evening correlates with the shorter inactive bouts that they exhibit at this time (Figure 10a). Together these lines of evidence suggest that this area is involved in the control of inactive bout lengths in typical animals, and that the mechanism regulating the temporal pattern of this area’s neuronal activity is disrupted in chd8⁻/⁻ mutants.

Alternative explanations for these findings should, however, also be considered. In particular, it is feasible that such a difference could result from altered chd8⁻/⁻ brain morphology, that when registered to fit the wild-type reference brain leads to a higher pERK/tERK ratio in this area. This potential caveat is of concern due to the area’s lateral location (Figure 13), its relatively consistent signal in chd8⁻/⁻ across time (Figure 14), and the fact that in humans (Bernier et al., 2014) and mice (Katayama et al., 2016; Gompers et al., 2017; Platt et al., 2017; Suetterlin et al., 2018) CHD8 mutations lead to macrocephaly. Conversely, however, measurements of interocular distance (5dpf) and acylated-tubulin area (4dpf) in this chd8 mutant have detected no gross differences in head size or brain area (Hoffman Laboratory, Yale – personal communication). Regardless, it is at least hypothetically possible that a small local difference in brain morphology could account this phenotype. As such, quantitatively comparing brain morphology in the mutants, as in (Gupta et al., 2018), remains an important future direction, and a step that should always be carried out when making comparisons between mutant and wild-type brains registered to a common reference. An alternative solution to this problem, and an area for future exploration, would be comparing pERK/tERK signal in unregistered brains. A reasonable approach to this analysis would be to first register the samples to a common reference, then to transform neuroanatomical areas from the reference brain back to the original samples and finally compare area-wise pERK/tERK signals between groups.
With the correlative nature of these findings and this potential caveat in mind, future experiments should aim to decisively determine this area’s role in regulating inactive bout lengths in wild-type animals and if this area’s altered activity underpins the \( chd8^{-/-} \) behavioural phenotype. Initial focus should be upon testing the necessity and sufficiency of this area in modulating inactive bout lengths. For this purpose, a suitable approach would be to record the behaviour of larvae both prior to and following laser ablation of this area. Another approach would be the focal electroporation of chemo- (Chen et al., 2016) or optogenetic channels (Singh, Oikonomou and Prober, 2015) which would allow for this area’s activity to be modulated in real time during recordings of free-swimming behaviour. An alternative strategy to electroporation would be to identify a transgenic line with expression specific to this area. A specific transgenic line would aid in consistently targeting the correct area for ablation or the generation of a stable line, negating the need for electroporation, and would thus facilitate these experiments. The best approach to identifying potential driver lines would be to make use of the zebrafish atlases to explore existent lines, as well as to identify any neuronal subtypes in this area, which could form the basis for novel driver lines (Ronneberger et al., 2012; Marquart et al., 2015, 2017; Randlett et al., 2015; Gupta et al., 2018). Regardless of the methodology, testing the effects of these manipulations in both wild-type and \( chd8 \) mutant animals, even uni- or bi-laterally, would likely provide mechanistic insight. For example, if stimulating neuronal activity on the right, but not the left side of the wild-type brain shortened inactive bout lengths, this could suggest that even in wild-type animals this area has an asymmetric function, despite the similar temporal patterns of pERK/tERK signal in the two hemispheres. An alternative approach to determining the causality of these links would be to behaviourally manipulate inactive bout lengths and to then use pERK/tERK mapping to study if pERK/tERK signal in this area was modulated as predicted. One hypothesis, for example, would be that melatonin (which induces long active bout lengths) would reduce activity in this area.

Assuming that neuronal activity in this area is involved in regulating inactive bout lengths, and that its aberrant activity generates the \( chd8^{-/-} \) phenotype, the next vector of study would be uncovering the mechanisms linking this area to behaviour. \( CHD8 \) itself is a chromatin remodelling factor (Thompson et al., 2008) thought to regulate the expression of hundreds of genes. Consequently, a plethora of mechanisms could be imagined linking \( chd8 \) to altered neuronal activity in this area. Thus, focussing experimental efforts on three aspects of the phenotype would be informative. Firstly, experiments could explore if the \( chd8^{-/-} \) mutant phenotype is of a developmental or acute nature. One approach to answering this question would be to generate a conditional \( chd8 \) mutant, such that \( chd8 \) expression could be retained during most of development but then removed at a later timepoint. An alternative approach would be to focus upon studying the neuroanatomy of the area; for example, mapping the area’s inputs and outputs or cell types in wild-type and mutant animals could identify a difference in neuroanatomy or connectivity, reflective of altered development. Secondly, potential mechanisms for the temporal nature of the phenotype could be explored, for example, time of day-dependent \( chd8 \) gene expression. To illustrate this possibility, consider hypothetically if \( chd8 \) expression were inversely correlated with neuronal activity in this area.
In this case the wild-type behavioural and neuronal data could be explained by low chd8 expression in the morning (driving increased neuronal activity in this area) then elevated chd8 expression in the evening, leading to decreased neuronal activity in this area. In this scenario, chd8⁻/⁻ mutant data could be explained by the persistent lack of chd8 signalling, which would lead to constant high activity in this area. To test this hypothesis, one could use qPCR to measure the level of chd8 from wild-type brains at same time-points used for the pERK/tERK mapping. Finally, the functional asymmetry of the chd8⁻/⁻ phenotype could be explored. One possibility, for example, is that in wild-type brains chd8 expression is particularly strong in this area or differs across hemispheres. To address this question fluorescent in-situ hybridisation data for chd8 could be registered to the Z-Brain and aligned with the pERK/tERK activity map. Furthermore, correlating the pERK/tERK signal against the level of chd8 expression on a voxel-wise basis could reveal a more general role for chd8 in regulating neuronal activity, perhaps even in specific areas. Alternatively, given the function of CHD8, an informative approach could be comparing RNA-sequencing data across mutant and wild-type brains, this approach could be extended to compare hemispheres, to single-cell RNA sequencing in the area identified by the pERK/tERK mapping and perhaps even to the multiple timepoints tested.

Overall these findings are the first steps towards uncovering the mechanisms linking the autism-associated gene chd8 to circuits and behaviour. Ultimately, however, they demonstrate the potential of this whole-brain activity mapping approach for uncovering common or divergent differences in neuronal activity across zebrafish mutants in autism-associated genes.
Chapter Four: General Discussion

“To hear of such things is to image a God who had long since abdicated all obligations toward his creation making a triumphant return in the guise of a software, an alpha and omega of zeros and ones”

Mark O’Connell, To Be a Machine.

The work described herein predominantly focused on developing tools for quantifying behaviour at scale and on mapping mutant behavioural abnormalities to circuits in zebrafish. At a behavioural level the tools developed are a substantial improvement over previous methods and accordingly provided insight into how zebrafish behaviour varies across the day/night cycle and with drug exposure. Furthermore, this analysis approach revealed behavioural phenotypes in mutants in two autism-associated genes, *chd8* and *bckdk*. At a circuit level, pERK/tERK mapping identified increased neuronal activity in the right, ventral tectum stratum periventriculare in *chd8/-* mutants, providing a starting point for dissecting how this autism-associated mutation impacts the brain and behaviour. Collectively, the results presented here provide proof of concept for these tools and start to demonstrate their use in mapping autism-associated genes to circuits and behaviour. Future work should focus on applying this framework to a library of mutants in autism-associated genes. In the longer term, work should aim to extend this approach to other screening datasets, for example a library of pharmacological compounds. Together, such datasets could be used to identify phenotypic suppressors in these autism mutants that could eventually form the basis for novel targeted therapeutics.

**Autism-associated Genes**

**Genetic Screening**

Initially, future work should focus on extending this approach to build an atlas of behaviour, neuroanatomy and neural activity for a library of mutants in autism-associated genes. While testing mutants in all 831 autism-associated genes would likely be challenging, a logical way to prioritise genes for study would be based upon their strength of association with autism. For example, [gene.sfari.org](http://gene.sfari.org) currently lists only 86 autism-associated genes as being either high confidence or strong candidates, as defined generally as genes identified from case control comparisons in multiple studies. With this atlas it would be possible to attempt to identify phenotypic subtypes of genes whose mutants share particular behavioural, neuroanatomical or neuronal activity features. The identification of phenotypic subtypes would be useful as it would be likely that genes within a subtype would share underlying phenotype-causing mechanisms. Thus, future work could focus upon uncovering the mechanisms underlying a
representative gene or small number of genes of each subtype, and furthermore it is possible that subtype specific treatments could be developed. Thus, eventually based upon these findings autism cases could be genetically sub-typed and treated. Given the multi-faceted nature of this prospective atlas (which can be thought of as a matrix of mutants by behavioural and neuronal features), and the aim of discovering structure within it, the best approach to identifying potential subtypes would be to apply unsupervised learning to the atlas (Figure 15). From this approach two results would be possible: one possibility would be that no phenotypic subtypes would exist, the second would be that clear phenotypic subtypes would exist.

The first possibility would be that no phenotypic subtypes exist, which could itself be for two reasons. One reason could be that all mutants would share a common phenotype, for example increased brain activity in an area, though no other phenotypes (Figure 15, solution 1). This result would suggest that all autism-associated genes are linked through a commonality, in this case altered neuronal activity in a brain area, though it would remain possible that the underlying mechanisms for this phenotype could be diverse across genes. Consequently, in this scenario, a productive approach would be to subtype the genes by function, i.e. synaptic genes, chromatin modellers etc, and to then study how a representative gene of each class impacts the phenotype of interest. The second reason would be that all mutants could show distinct phenotypes and thus no subtypes would be identifiable (Figure 15, solution 2). This result could suggest that each mutation defines its own subtype of autism, necessitating the future study of each genes disease-causing mechanisms.

The second possible result would be that all genes could be classified into subtypes (Figure 15, solution 3), for example perhaps mutants with seizures would share a common neuroanatomical abnormality, while mutants with increased repetitive behaviours would have altered neuronal activity in a specific area. This scenario would suggest that distinct phenotypic subtypes of mutations exist, perhaps each underwritten by an independent mechanism. In this case an appropriate approach would be to select and prioritise a representative gene from each subtype for further study.

These scenarios where either no or all autism-associated genes can be classified into phenotypic subtypes delineate the extreme ends of the possible solution space for this problem. Instead an intermediate result, whereby some genes have unique phenotypes and some fall into subtypes (Figure 15, solution 4), is the most likely. This is due to the phenotypic heterogeneity within autism, which is suggestive of distinct underlying mechanisms, as well as available evidence from network approaches, which suggests that en masse autism risk genes do converge into subtypes based upon protein-protein interactions (Krumm et al., 2014), developmental timepoints (Willsey et al., 2013) and neuroanatomical structures (Ellegood et al., 2014).

One caveat that should be kept in mind with this circuit focussed approach is the possibility that commonalities could exist at other levels, for example, peripheral nervous system deficits, which have been shown be involved in driving behavioural phenotypes including
anxiety-like behaviour and social interaction deficits in at least one genetic mouse model of autism (Orefice et al., 2016). As evidenced by this example, however, deficits at another level will likely translate to behavioural or even circuit level phenotypes. Thus, mutants sharing a commonality at an unobserved level may still form a distinct subtype, from which this mechanism could later be uncovered.

Given the generalisable nature of this framework, two future genetic screening approaches would be informative. Firstly, the focus of this work has been upon mapping the impact of rare inherited and single de novo mutations upon behaviour and neuronal circuits. Such work is a crucial first step towards developing suitable frameworks for addressing these questions and for uncovering the mechanisms linking single genes to autism. These classes of mutations, however, are estimated to account for only a small fraction of autism-risk, particularly compared to the addictive effects of common variants (Gaugler et al., 2014). Consequently, once future work has gained insight into these mutational classes, the study of combinations of mutants may be informative. The second approach would be to compare the autism-mutant atlas to data from other genetic screens. One direction would be to compare data between mutants in genes from different psychiatric disorders, for example schizophrenia or bipolar disorder. This approach could reveal common or divergent mechanisms across disorders, as have already been suggested from comparisons of transcriptomic data (Gandal et al., 2018). One of the diagnostic features for schizophrenia, for example, is disorganized behaviour (American Psychiatric Association, 2013). Thus, it would be interesting to test the compressibility of schizophrenia mutants. One hypothesis from this approach would be that while autism-associated mutants would tend to be more compressive (more repetitive behaviours), schizophrenia mutants would tend to be less compressive (less repetitive/organised behaviour). Another hypothesis would be that genes common to the two disorders may both be expressed in areas involved in the control of repetitive behaviours, though their loss may modulate neuronal activity in these areas in opposite directions. Ultimately, such comparisons across disorders would provide insight into the common and unique mechanisms by which genetic aberrations translate to behavioural phenotypes in such disorders.

Importantly, while this atlas of behavioural and neuronal data for mutants in autism-, and perhaps other disease-associated genes would provide clues as to the potential mechanisms involved and could help to prioritise genes for further study, additional steps would be necessary to uncover the actual mechanisms by which these mutations impact behaviour.
Figure 15. Phenotypic Sub-typing in a Prospective ASD Atlas

Work-flow depicting how extending the behavioural and neuronal data acquisition and analysis frameworks presented here, to study a library of mutants in autism-associated genes would generate an autism-atlas (genes x behaviour and neuronal features). Applying unsupervised learning to this atlas would generate one of four possible solutions, shown in a hypothetical solution space. In each panel each of 90 mutants (dots) is shown in a reduced 2-dimensional behaviour and neuronal feature space. Black - mutants with no subtype, red and blue - mutants with a phenotypic subtype. Solution 1 - no phenotypic subtypes exist as all mutants share a single feature (here a neuronal feature), but all differ in other measures (here behaviour). Solution 2 - no phenotypic subtypes exist as all mutants differ by all metrics. Solution 3 - all mutants can be classified into phenotypic subtypes (here two). Solution 4 - some mutants can be classified into phenotypic subtypes but others are distinct. Behaviour and neuronal data are taken from other figures, autism atlas data is fictive. Solution space data was generated by sampling random data and/or data from Gaussian distributions.
Uncovering Mechanisms

To uncover the precise mechanisms by which autism-associated genes affect behaviour, a selection of general and case specific experiments will be necessary. In terms of general experiments, an informative starting point of study for many mutants would be the use of additional behavioural assays to provide clues as to the underlying mechanisms. For example, in the case of the zebrafish cntnap2 mutant phenotype, night-time hyperactivity (Hoffman et al., 2016), behavioural experiments with altered light cycles could be used to determine if the phenotype is circadian or light dependent. In the former case, it may be useful to see if expression of the cntnap2 gene itself is circadian in wild type animals. In the latter case, comparing the expression of photoreceptors, as in (Davies et al., 2015), between mutants and wild type animals may be informative. Another question of general interest across mutants will be determining the extent to which phenotypes are developmental or acute in nature. One approach to addressing this question would be to use a conditional knock-out system to selectively remove genes later in development. Of the resultant phenotypes, those concordant with the original knock-out animals would be putatively acute, while discordant phenotypes would likely be developmental. In Chd8 mutant mice, for example, it was shown that down-regulating Chd8 expression specifically in the nucleus accumbens recapitulated their germline Chd8 mutant’s improved motor learning, demonstrating an acute role for Chd8 signalling in this phenotype (Platt et al., 2017). A final question of general interest across mutants is the extent to which phenotypes are of neuronal origin. One way to address this question would be to test the behavioural and neuronal effects of brain specific conditional knock-outs. An alternative approach would be the use of peripheral nervous system knock-outs; as in Orefice et al., (2016), who demonstrated that resorting Mecp2 expression, a gene associated with Rett syndrome, exclusively in somatosensory neurons of Mecp2 mutant mice, restored some but not all behavioural phenotypes.

In terms of case (either gene or phenotypic subtype) specific experiments, one starting point would be to consider the annotated functions of these genes. In the case of chd8 for example, a chromatin remodelling gene, many rodent studies have focussed upon using ChIP- and RNA-sequencing to compare mutant and wild-type animals (Katayama et al., 2016; Gompers et al., 2017; Platt et al., 2017; Suetterlin et al., 2018). Generally, these studies have detected only small magnitude changes in gene expression between genotypes. On the whole, however, these genes do converge on neurodevelopmental functions, suggesting that a focus upon studying neurodevelopment or brain morphology in Chd8 mutants may be informative. In zebrafish chd8 mutants, whole-brain RNA-sequencing at 4dpf has revealed similar results (Hoffman Laboratory – personal communication), though given the chd8 data presented here, comparisons of either intra-hemispheric, or even single cell (in the area with altered pERK/tERK signal in chd8/-/- mutants), RNA-sequencing may be illuminating.

Collectively, such experiments would begin to uncover the mechanisms by which autism-associated genes impact neuronal activity and behaviour. This knowledge would improve our understanding of behavioural genetics and could be leveraged to develop targeted therapeutics through hypothesis driven approaches. Examples, such as the failure of treating...
Chapter Four: General Discussion

Fragile X with mGluR5 antagonists (Erickson et al., 2017) however, demonstrate the challenges inherent to this approach. Consequently, it is worth considering the fact that many clinically approved drugs were discovered by serendipity and that their mechanisms were worked out later, or in some cases still remain unknown. Thus, hypothesis-free pharmacological screening approaches to identify potential therapeutics, whose mechanisms could be determined later, may be more fruitful.

Predictive Pharmacology

One approach to identifying pharmacological compounds of interest in autism-models would simply be to screen a library of small molecules in the mutants and to identify drugs which suppressed relevant phenotypes. In rodent models, this approach would likely be unfeasible. Zebrafish are, however, well suited high-throughput pharmacological screening (Ghosh and Rihel, 2015). To date, zebrafish behavioural pharmacology screens have tested the effects of as many as 6,000 compounds in free swimming larvae (Rihel et al., 2010) and 14,000 in younger animals (Kokel et al., 2010). Despite this, however, screening this many compounds in libraries of mutants would be challenging. Consequently, a predictive pharmacology approach in which compounds of interest could be pre-selected for testing from existent libraries would be preferable.

An example of this predictive pharmacology approach comes from Hoffman et al., (2016) in which a library of 550 compounds (filtered from an overall set of 6,000 tested compounds) with quantified effects on zebrafish sleep/wake behaviour (Rihel et al., 2010) was leveraged to predictively identify molecules to ameliorate the night-time hyperactivity of cntnap2 mutants. This involved a three-step process. Firstly, several behavioural features were measured from cntnap2 mutants, such that their phenotype (night-time hyperactivity) could be represented as a ‘behavioural fingerprint’ describing their behaviour across multiple parameters. Secondly, fingerprints were generated to describe the behavioural effects of the 550 pharmacological compounds from the existent library (Rihel et al., 2010). Finally, anti-correlation of the cntnap2 fingerprint to the pharmacological finger-prints identified compounds with the opposite phenotype, i.e. that specifically reduced night-time activity. Interestingly these compounds were enriched for estrogenic agonists. Applying these compounds to the mutants demonstrated not only that they could suppress the mutant phenotype at doses that do not affect wild-type behaviour, but also that they do so better than risperidone (Hoffman et al., 2016) and have no effects on other behaviours, such as day-time activity. Overall, this result demonstrates three main advantages of predictive pharmacology, as compared to traditional screening approaches. Firstly, a predictive approach is far quicker and less labour intensive, meaning that it could feasibly be applied to a whole library of mutants. Secondly, as the approach considers multiple behavioural parameters, compounds with a very specific behavioural rescue are identified (i.e. other behaviours are unaffected). This could be considered as akin to identifying compounds with minimal side-effects. Finally, while one compound may perform best when tested in the mutant, the approach simultaneously identifies many compounds of interest. In cntnap2 mutants, for example,
estrogenic agonists were generally enriched amongst the anti-correlating compounds. Thus, advantageously if one of these compounds failed at later stages of study or was unsafe in humans, other compounds (for example the second or third best) could be focused upon.

With these advantages in mind, future work should aim to apply this methodology to a library of mutants in autism-associated genes. This screen would produce two broad results of interest. Firstly, by correlating the mutant phenotypes to pharmacological compounds, i.e. identifying drugs that copy mutant phenotypes, potentially disrupted molecular pathways could be identified. In the case of the cntnap2 mutants it was shown that NMDA antagonists correlated with the mutant phenotype (i.e. induced night-time hyperactivity) and mutants were behaviourally more sensitive to these compounds. This finding implicated altered NMDA signalling in cntnap2 mutants, despite a lack of obvious glutamatergic deficits from neuroanatomical data (Hoffman et al., 2016). Comparing this data across mutants could identify commonly implicated signalling pathways, which would be difficult to identify through other means and would provide additional features for classifying the mutant library into distinct subtypes (Figure 15).

The second result of interest would be the compounds anti-correlating with each mutant phenotype. In cntnap2 mutants these compounds were enriched for estrogenic agonists, which were subsequently shown to rescue the mutant phenotype at doses with no effect on control animals (Hoffman et al., 2016). This result not only demonstrates the ability of this approach to rapidly identify phenotypic suppressors but also suggests that estrogenic signalling may be impaired in this mutant. Extending this anti-correlation approach to the entire autism-mutant library could uncover additional signalling pathways of interest. Furthermore it is possible that it could identify suppressors shared by several mutants or genes within a phenotypic subtype, providing a starting point for studies searching for common autism therapeutics.

A quick avenue to identifying these compounds would be to assess the behaviour of the autism mutants using the older analysis method, such that they could be directly compared to the available pharmacological data (Rihel et al., 2010). An alternative route however would be to screen a new drug library and assess the data using the improved behavioural analysis approach presented herein; the benefits of which, in terms of pharmacology, are evident from the melatonin and PTZ data (Figure 9). This approach would be preferable not only due to its substantially improved behavioural resolution, but also as it would include autism-relevant metrics such as compressibility. In this latter case, two factors should be considered when designing the new drug library.

Firstly, the aim of the library should be to cover the full space of possible phenotypes in this assay, such that correlating and anti-correlating compounds can be identified for any mutant phenotype. Clearly predicting the behavioural effects of untested compounds would be challenging, though the likelihood of covering the phenotypic space could be increased by selecting compounds with a diverse range of mechanisms and by simply screening as many compounds as possible. Secondly, while screening non-clinically approved compounds could
identify novel molecules of interest, given the high costs and low likelihood of these being translated to clinic (Corsello et al., 2017), a focus on drugs that are already known to be safe in humans may be beneficial. With these two features in mind a good selection would be the drug repurposing hub (Corsello et al., 2017), a hand curated panel of 4,707 compounds, all of which are already marketed or in clinical development.

An alternative approach to identifying compounds of interest would be to carry out a genetic screen, using for example heat shock promoters to overexpress secreted peptides (Chiu et al., 2016), to generate an additional database to which mutant data could be compared (Figure 15). If for example, the behaviour of a mutant and overexpressed peptide was highly correlated, this could suggest that an inhibitor of this peptide’s receptor would normalise the mutant’s behaviour. Thus, such an approach would provide additional data which could guide pharmacological target selection.

Importantly, while suppression of zebrafish behavioural phenotypes, for example increased night-time hyperactivity in cntnap2 mutants (Hoffman et al., 2016), may seem far removed from treating autism in humans, it is already known that treating autism’s co-morbidities, such as sleep disorders, can improve core symptoms (Herrmann, 2015; Singh and Zimmerman, 2015). Consequently, identifying compounds in zebrafish with potential relevance to any of autism’s comorbidities could be beneficial. Furthermore, beneficial core-comorbidity relationships could easily be explored in zebrafish. One direction in which this could work would be in testing suppressors identified through this assay on other behavioural phenotypes. If cntnap2 mutants for example demonstrated social deficits as assayed through either social preference (Dreosti et al., 2015) or shoaling (Tang et al., 2018), one could test if estrogenic compounds could also acutely rescue this phenotype, or even if normalising the mutant’s night-time behaviour for several nights or even the animal’s entire lifetime prior to the social assay would have an impact. Conversely, however, it is possible that the control of human social behaviour is divergent from other animals; consider for example the lack of social deficits, or repetitive behaviours, in Chd8 mutant mice (Katayama et al., 2016; Gompers et al., 2017; Platt et al., 2017; Suetterlin et al., 2018). With the beneficial impacts of treating autism’s comorbidities and the potential complexities of translating behaviours between animals, I would argue that the effect, or lack thereof, of compounds upon social behaviour should not be taken as the final word in deciding upon promising compounds for autism.

An alternative direction for such work would be initially using higher throughput screening methods to identify phenotypic suppressors and then testing if these compounds would hold true at older ages. One example of a high-throughput screening assay is the motor response that embryos show to a high intensity light stimulus at 28 hours post fertilisation, upon which the effects of 14,000 compounds have already been established (Kokel et al., 2010). A good proof of concept trial for this approach would be testing if the elfn1 schizophrenia mutant, which shows an increased startle response to light at 5/6 days post fertilisation (Thyme et al., 2018) also shows a phenotype at 28 hours post fertilisation, if so an anti-correlation approach within this assay could be used to identify suppressors that could be tested at both ages.
Together these future directions delineate how phenotypic suppressors could be identified for mutants in autism-, and other disease-associated genes in zebrafish. Were this possible, what would the next steps be?

From Zebrafish to Humans

Let us assume that the approach outlined above succeeded and that it enabled the identification of phenotypic suppressors, which may or may not be common across mutations, in autism-associated genes, what would be the path to translating these findings from zebrafish to humans?

One direction to explore would be to learn more about the suppressor’s mechanism of action, for example which receptors were necessary for its effect. This question could be tackled in zebrafish by genetically removing or chemically blocking putative target receptors or by testing if other compounds of the same class had the same effect. If for example, two estrogenic agonists had divergent effects, this could suggest that the effect of interest is not mediated via the oestrogen receptor. Determining the mechanism of the compound would be important as it could allow for the selection of a more targeted compound, perhaps with fewer side effects. For example, continuing the illustration above, once the mechanism has been determined, it could be worth identifying a compound with no effect on estrogenic signalling, to avoid estrogen-related side effects.

Another vector to explore would be testing these compounds in rodent or other genetic animal models. In cases where phenotypes agreed across models it would be prudent to test if the suppressors identified in zebrafish could also normalise the same phenotypes in other animals. Notably, however, even in cases where phenotypes were non-concordant across models, it could still be informative to test these compounds and then to observe their effects on a variety of behavioural paradigms. For example, it is feasible that a compound that decreased abnormal repetitive behaviour in a zebrafish model could ameliorate other phenotypes in a mouse model even if it lacked increased repetitive behaviours.

Having been shown to be efficacious in zebrafish and potentially other animal models as well as safe in humans, the final step would be a randomised double-blind placebo-controlled trial in which autistic cases would be given either the drug or a placebo and monitored for improvement, with both participants and investigators unaware of each participant’s group. Such a study would however, face both practical and conceptual challenges.

Practically it should be kept in mind that very few cases harbour each mutation. For example, despite having one of the strongest associations with autism of any single gene, CHD8 currently accounts for only 0.21% of cases. Consequently, pharmaceutical trials may only be feasible if common mechanisms, or subtypes consisting of large numbers of patients, can be identified. Furthermore, these small numbers could make enrolling enough participants challenging, even in the scenario where a common drug could be identified across multiple genetic lesions. An additional practical challenge would be determining the age at which
to test the drug, a problem compounded by the fact that novel drugs need to be tested in adults first, particularly in vulnerable populations (Erickson et al., 2017). This challenge underscores the potential benefits, as discussed above, of testing the compound in multiple behavioural assays at different ages, though it is difficult to determine if this would actually be an effective solution to this problem.

Conceptually the main challenge would lie in selecting an appropriate outcome metric to monitor for improvement. Clearly if a compound had been selected in zebrafish based upon its ability to decrease repetitive behaviour and results were concordant in another animal model, then this would be an important outcome to measure in a trial. In this case, a direct comparison could even be made between zebrafish and human compressibility, as the compression algorithm applied to zebrafish behaviour here has already been applied to human video data (Lee et al., 2013). If, however, the phenotype was less obviously relatable, for example altered pauses of a certain length, deciding upon an appropriate outcome to measure could be challenging. In this case, a good solution could simply be to measure a range of autism-relevant metrics. One approach to this would be to use a standardised scale that measures autism relevant behaviours, such as the aberrant behaviour checklist (Aman et al., 1985), which was used in the mGluR5 Fragile X trials (Erickson et al., 2017).

**Final Thoughts**

Uncovering the underlying mechanisms by which genetic lesions lead to behavioural disorders, such as autism spectrum disorder, remains challenging. The work presented here aims to provide a generalisable framework for uncovering these mechanisms, with a focus on quantitative, high-throughput methods with which it will be possible to quickly screen genetic and pharmacological libraries. Together these datasets will provide insight into how these genes affect neuronal circuits and behaviour and will hopefully one day enable the identification of translatable pharmacological therapeutics for autism and other genetic disorders.
Materials & Methods

Animal Husbandry
Adult zebrafish were reared by the UCL Fish Facility on a 14:10 light:dark cycle (lights on: 09:00 a.m. to 23:00 p.m.). To obtain embryos, pairs of adult males and females were isolated overnight with a divider that was removed at 09:00 a.m. the following morning. After a few hours, fertile embryos were collected and sorted, under a bright-field microscope, into groups of 50 embryos per 10 cm Petri dish, filled with fresh fish water (0.3g/L Instant Ocean). Plates were kept in an incubator at 28.5°C on a 14:10 light:dark cycle. Each day, using a Pasteur pipette under a bright-field microscope, debris was removed from the plates and the fish water replaced. All work was in accordance with the Animal Experimental Procedure Act (1986) under Home Office Project Licence 70/7612.

Fish Lines
Throughout the term wild-type animals refers to AB x TUP LF animals (UCL Line 573). This line was used for the wild-type experiments, as well as the Melatonin and PTZ dose response curves. Mutant experiments were always carried out on embryos collected from a heterozygous in-cross, with larvae genotyped later (see Materials & Methods, page 92). Four mutant lines were used herein and can be genotyped using the following PCR and/or KASP primers. All KASP primers are available for purchase upon request from LCG Genomics. Note that for the KASP flanking sequences alternative alleles are shown in square brackets and in bold, with a forward slash indicating a deletion in the alternative allele. Note that in all three KASP assays the genotypes will be assigned the following colours by the software (see Materials & Methods, page 92) wild-type – orange, heterozygous – green, homozygous – blue.

hcrtr
ZFIN ID: hu2098 (Yokogawa et al., 2007). Identified from an ethylnitrosourea-mutagenized screen. UCL Line 2114.

PCR Forward Primer (Rihel database, 6) : CCACCCGCTAAAATTCAAAAGCACTGCTAAC
PCR Reverse Primer (Rihel database, 7): CATCACAGACGGTGAACAGG
PCR Information: Digest PCR products with Ddel at 37°C to produce a 170bp band in the wild type animals and in mutants 140 and 30bp bands

KASP Assay ID: 554-0090.1
**KASP Flanking Sequence:**

ACCGCTGGTATGCGATCTGCCACCCGCTAAAATTCAAAAGCACTGCTAAA[A/T]GAGCCCGCAAGAGCATCGTGCTGATCTGGCTGGTGTCCTGCATCATGATG

**chd8**

chd8 ex8 5del. Generated by Ellen Hoffman (Yale) using zinc finger nucleases. UCL Line 2250.

PCR Forward Primer (Rihel database, 321): TATGATTCATTCTCCCAGTTTG

PCR Reverse Primer (Rihel database, 322): TACCTACAATGGCTCCTGTGAGT

PCR Information: Wild type band – 250bp

KASP Assay ID: 1310.24

**KASP Flanking Sequence:**

TATGATTCATTCTCCCAGTTTGAGAGAGTTTGACTCACTGATATATGAGTTTGTTTCTTACTCTT-GATCTCTTCTCCACAGCACCATCACTCTAGGGGTAAGAAAGAAAGAAACCGCTTTCTGAT-CACCTAGACGTAGTCTCAGCCCCACTTTGGCG[CCCGC/]ACACTGGGAGGAAGAAATGTCACAG-GTGTAACTTGTGCTGCACATTTGCTTATTAGTGAACACTACAGGGAGCCATTTGAGGTACTGTCGGCAG-TGTGGCAGAATCTTGTAGGGCAATTGGAAGAAAAACACAAACCTCATTATTTAAACTCTGCTTTTGAT- TTTGTCCTTCCACACAG

**bckdk**

bckdk ex3 10ins. Generated by Ellen Hoffman (Yale) using zinc finger nucleases. UCL Line 2247.

PCR Forward Primer: AATCGCGTTTTTGCTTTTAAT

PCR Reverse Primer: ACCCAATATCCCCTCTGAGAAT

PCR Information: Wild type band – 277bp

KASP Assay ID: 1310.24

**KASP Flanking Sequence:**

AATCGCGTTTTTGCTTTTAATTCAATACAGCAGCAGTGCCAGGTATCTACACAAAGAGCTGCGCTC-GTATTGGCCCATCG[GAATTAACGG]CATTAAGGGGCTTCCGACTCCACTGTTATTATCTTCTTCTCCTATTATCATTATTTCTTGTTGATGTTGGTATGTTGCTTTATCTACAGGTACCGTCTCTCCTACAAGCTCCACTGTTAGTTGGAAGGAGCTCCACTTCTTCGATAT-TCTCAGAGGGATATGGGT
**pitpnc**


PCR Forward Primer: TCTGTCCGTCGCTCTCTTC

PCR Reverse Primer: AGGCTTTTCGTCACGTAG

**Behavioural Setup**

For all behavioural experiments a Pasteur pipette was used to transfer single zebrafish larvae (aged 4-5 days post fertilisation) into the individual wells of a clear 96-square well plate (7701-1651; Whatman, Clifton, NJ), then each well was filled with 650µl of fish water. For experiments longer than 24 hours larvae were plated at 4 days post fertilisation (dpf) and tracking was started the same day. For the duration of these experiments, evaporated fish water was replaced each morning between 09:00-09:30 a.m. For the wild-type experiments, each plate was covered with a plastic lid (4311971, Applied Biosystems) to prevent evaporation and negate the need to replenish the fish water. For the 24-hour small molecule experiments (melatonin and PTZ), larvae were plated at 5dpf and the plates were left overnight in a 28.5°C 14:10 light:dark incubator. The following morning each plate was transferred to a behaviour setup where larvae were dosed, immediately after which, between 09:00 and 10:00 a.m., behavioural recordings were started and run for 24 hours.

To record each animal’s behaviour, each plate was placed into a Zebrabox (ViewPoint, LifeSciences) running quantization mode with the following manually determined settings: detection sensitivity -- 15, burst -- 50 and freezing -- 4. All experiments were conducted on a 14:10 light:dark cycle (lights on at 09:00 a.m. to 23:00 p.m.) with constant infrared illumination and an ambient temperature of 28.5°C. All experiments were recorded at 25Hz, aside from the chd8 experiments which were recorded at 15Hz. Larvae were tracked for 24-75 hours, after which all larvae unresponsive to touch with a 10µl pipette tip were excluded from subsequent analysis. Following this, larvae were euthanised with an overdose of 2-Phenoxyethanol (Acros Organics).

**Computing Hardware**

A desktop computer with 16GB of RAM was used for most data analysis, figure production and writing. Two-time intensive functions, Batch_Compress.m and Batch_Grammar_Freq.m, were run in parallel, with a worker for every animal, on the UCL Legion Cluster (Research Computing Services, UCL).
Software Sharing

Unless otherwise stated all software used for data handling, analysis and the production of figures is available at https://github.com/ghoshm.

Processing Behavioural Data

See Appendix Figure 10 (page 120) for a flow diagram describing behavioural data acquisition and analysis. All custom behavioural analysis software was written and run in MATLAB 2016b-2018a (MathWorks), all code is available at https://github.com/ghoshm/Structure_Paper. Note that the suffixes .m and .mat denote MATLAB code and MATLAB data files respectively.

Behavioural data was recorded by subtracting subsequent pairs of frames from each other and determining the number of pixels that changed intensity within each well between each pair of frames termed Δ pixels. To acquire behaviour data, each Zebrabox was setup using ViewPoint’s ZEBRALAB software (version 3.22), which outputs an .xls and a .raw file (ViewPoint specific format) per experiment. Each behaviour .xls file was reorganised into a .txt file using the function perl_batch_192.m (Jason Rihel). For each experiment a .txt metadata file assigning each animal to an experimental group, for example genotype, was manually produced. To replicate the old analysis methodology, as in Appendix Figure 1c (page 110), behaviour and metadata .txt files were input to the function sleep_analysis2.m (Jason Rihel).

To assess data on a frame by frame basis each experiment’s .raw file, output from ViewPoint, was exported within the ZEBRALAB software to thousands of .xls files. Each .xls file contained 50,000 rows and 21 columns, with data from any given well listed approximately every 192 rows, as the setup always assumes recordings are from two 96-well plates. This formatting is, however, only approximate as infrequently the well order is erroneously non-sequential, these rows were termed ordering errors. Each .xls file is formatted with 21 columns, of which 3 contain useful data: type – notes when ViewPoint defined data acquisition errors occurred, location -- denotes which well the data came from, and data1 – records the Δ pixel value from that well for that time point.

The function Vp_Extract.m was used to reformat the .xls files from each experiment to single frame by fish matrices, from which each animal’s behaviour was quantified, see Behavioural Data Analysis, Δ Pixels for details (page 88). Vp_Extract.m requires three inputs to be selected: a folder containing the .xls files, a .txt behaviour file output from perl_batch_192.m and a .txt metadata file. To ensure that each animal has the same number of frames, frames with ViewPoint defined errors or ordering errors (which are automatically detected by Vp_Extract.m) are discarded. A maximum Δ pixels value can be set and active bouts containing even a single frame with a higher Δ pixel value than this are set to zero for the entire duration of the bout. Here a maximum Δ pixels threshold of 200 was set, this value was determined from manual inspection of the test dataset as well as by comparisons of this data to data recorded from plates with no animals in. Time periods during which water is being replenished are automatically detected and set to a Δ pixel value of zero. These time periods are noted
and excluded from later analysis. The function outputs .mat files for subsequent analysis. Either single or multiple .mat files output from Vp_Extract.m were input to Vp_Analyse.m and Bout_Clustering.m.

Vp_Analyse.m was used to compare general activity levels and bout features across time and between groups. The function has two options. The first allows for specific days and nights of interest to be cropped from the data. The second determines how experimental repeats are handled, treating the data as either a single merged dataset or as separate datasets. In the latter case each experimental repeat is plotted with the same colour scheme as the first experiment, with progressive shading for each repeat. Additionally, the N-way ANOVA comparisons include a repeat factor, which can be used to determine if results are consistent across experimental repeats, see Behavioural Data Analysis, Δ pixels, for details (page 88). Vp_Analyse.m outputs two statistics results structures: twa -- N-way ANOVA comparison results, and kw – Two-sample Kolmogorov-Smirnov test results. Vp_Analyse.m outputs figures showing each groups activity (e.g. Figure 2) and bout features (e.g. Appendix Figure 2, page 111) over time.

The script Bout_Clustering.m was used to cluster all active and inactive bouts into behavioural modules, as well as to compare the resultant modules. To cluster the data an evidence accumulation approach is used (Fred and Jain, 2002, 2005) implemented by the custom MATLAB function gmm_sample_ea.m. Bout_Clustering.m produces figures (e.g. Appendix Figures 3 and 4, page 113) and statistically compares the modules, see Behavioural Data Analysis, Clustering (page 89). The MATLAB workspace output from Bout_Clustering.m can be input to either Bout_Transitions.m or Bout_Transitions_Hours.m.

The function gmm_sample_ea.m clusters data using an evidence accumulation approach (Fred and Jain, 2002, 2005) through which the results of multiple Gaussian Mixture Models are combined to generate an aggregate solution. This process is executed through the following six steps. Firstly, a sample of ‘probe points’ are randomly sampled from the data. The number of probe points to sample is user defined. Secondly, values of K and sample sizes are uniformly sampled from user set ranges. The values of K are used to set the number of mixture components for each mixture model. The sample sizes determine the number of points, randomly sampled from the data, that each mixture model is fit to. Thirdly, iteratively, a Gaussian Mixture Model is fit to the sampled data with K components. Each probe point is assigned to the component with the highest corresponding posterior probability and evidence is accumulated on the probe points, evidence is defined as pairwise co-occurrences in the same component. Fourthly, the evidence accumulation matrix is hierarchically clustered, and the final number of clusters is determined by using the maximum differentiated linkage distance to cut the resultant dendrogram. The linkage metric used is a user-defined option. Fifthly, the clusters are normalised for size by randomly sampling the number of points in the smallest cluster, from each cluster. Finally, all data points are assigned to these final size normalized clusters using the mode cluster assignment of the k-nearest neighbours, with k being user defined.
The script `Bout_Transitions.m` takes the MATLAB workspace output from `Bout_Clustering.m` as an input and compresses each animal’s full module sequence to generate a library of behavioural motifs, the number of occurrences of each motif are counted and normalised by comparison to paired shuffled data and finally, a supervised learning algorithm is applied to identify context specific behavioural motifs. See Behavioural Data Analysis, Hierarchical Compression and Supervised Motif Selection for analytical details (page 90). For two-time intensive steps: hierarchical compression of full module sequences (`Batch_Compress.m`) and normalising the behavioural motif counts (`Batch_Grammar_Freq.m`) data is manually copied (via MobaXterm, Personal Edition v10.5) to UCL Legion Cluster (Research Computing Services, UCL) and processed in parallel with a worker for every fish. MATLAB code for hierarchical compression is described in Gomez-Marin, Stephens and Brown (2016). MATLAB code for submitting these jobs to Legion, analysing data and retrieving results is available at [https://github.com/ghoshm/Legion_Code](https://github.com/ghoshm/Legion_Code). Ultimately, `Bout_Transitions.m` outputs a library of behavioural motifs and motif related figures (e.g. Figures 7-9).

The script `Bout_Transitions_Hours.m` compresses blocks of 500 modules for statistical comparison, uses the motif library from `Bout_Transitions.m` to count the occurrence of each motif every hour, normalises these counts to paired shuffled data and finally uses supervised learning to identify hour specific behavioural motifs. As with `Bout_Transitions.m` behavioural motifs are normalised, via `Batch_Grammar_Freq.m`, using UCL Legion Cluster. `Bout_Transitions_Hours.m` outputs figures (e.g. Figure 5) and statistics. See Behavioural Data Analysis, Hierarchical Compression and Supervised Motif Selection for analytical details (page 90).

**Processing Imaging Data**

See Appendix Figure 11 (page 121) for a flow diagram describing image processing and analysis. All custom imaging analysis software was written and implemented in MATLAB 2016b-2018a (MathWorks), all code is available at [https://github.com/ghoshm/2p_Imaging_Analysis](https://github.com/ghoshm/2p_Imaging_Analysis).

Prairie View software (version 5.3) was used to control a custom two-photon microscope (Bruker), with which whole-brain data was acquired in the form of individual .tif slices. The script `2p_Process.m` was used to batch process this data to make it compatible with Computational Morphometry Toolkit (CMTK, [https://www.nitrc.org/projects/cmtk](https://www.nitrc.org/projects/cmtk)) registration to the Z-Brain atlas (Randlett et al., 2015). Specifically, `2p_Process.m` compiles each samples slices into stacks, if necessary (user specified option) flips the stack to be arranged ventro-dorsal, filters each stack for noise, and saves each brain as both a .tif and .nrrd stack. The .nrrd stacks are automatically assigned pixel size metadata that is provided in a separate .xlsx sheet. The .nrrd related steps require MIJ (a Java package for controlling ImageJ from MATLAB, [http://bigwww.epfl.ch/sage/soft/mij/](http://bigwww.epfl.ch/sage/soft/mij/)). Following this, each stack was manually rotated in Fiji ([https://fiji.sc/](https://fiji.sc/)) to lie rostro-caudal from the top to the bottom of the image.
After each stack has been manually rotated, the script `2p_Process.m` registers each rotated stack to the Z-Brain reference (Randlett et al., 2015) using CMTK through Cygwin (a Linux emulator, https://www.cygwin.com/). For CMTK the following command string was used: `-a -w -r 0102 -l af -X 52 -G 80 -C 8 -R 3 -A "--accuracy 0.4 --auto-multi-levels 4" -W "--accuracy 1.6" -T 4`. Both prior to and following registration all stacks were visually inspected for quality and samples of poor quality were excluded from further analysis. Next, `2p_Process.m` emulates the Z-Brain script `PrepareStacksForMAPMapping.ijm` by using MIJ to down sample and blur the stacks. This step was modified such that each channel’s maximum pixel intensity was set to the top 0.1% of its pixel intensities prior to 8-bit conversion, which improves the dynamic range of the resulting stacks. Finally `pERK/tERK` activity maps were generated using the Z-Brain scripts (Randlett et al., 2015).

**Behavioural Data Analysis**

**Δ Pixels**

At the acquisition stage, Δ pixels data is filtered by the software (ViewPoint) such that each frame for a given well is scored as either zero or higher. In the absence of movement within a well, and hence no pixels changing intensity, Δ pixels values of zero are recorded. These periods were termed inactive bouts and were defined as any single or consecutive frames with Δ pixels values equal to zero. The length of each inactive bout was used as a descriptive feature. When there is movement within a well, pixels change intensity and Δ pixels values greater than zero are recorded. These periods were termed active bouts and were defined as any single or consecutive frames with Δ pixels values greater than zero. Six features were used to describe each active bout: length, mean, standard deviation, total, minimum and maximum. These features, as well as the number of active bouts, percentage of time spent active and total Δ pixels activity, were compared between conditions, e.g. day and night and dose of drug, in two ways using the function `Vp_Analyse.m`.

To compare the distribution of values for each feature between conditions, a probability density function (pdf) was fit to each animal’s data and the mean shape of each condition’s pdf was compared using a Two-sample Kolmogorov-Smirnov test (e.g. Appendix Figure 2a, page 111). To compare each features average values between conditions, mean values were taken from each animal, and N-way analysis of variance was computed. The following factors, when relevant, were included and full interaction terms were calculated: condition - e.g. mutant and wild-type, time - e.g. day and night, development - defined as a consecutive day and night, and experimental repeat – i.e. which experimental repeat a datapoint came from. For experiments with multiple repeats, the lack of an interaction effect between the comparison of interest and experimental repeat factor was considered as evidence of a consistent result.
Clustering

To cluster the initial test dataset, the script Bout_Clustering.m was used. First matrices of bouts by features were constructed. Active – 30,900,018 x 6. Inactive – 30,900,418 x 1. To prepare the active data for clustering each animal’s data was individually normalised by calculating z-scores using equation 1, which illustrates how every bout (i) from each animal (f) is normalised by first subtracting the mean of this animal’s bout features $\bar{x}_i$ from the bout $x_i$ and then dividing by the standard deviation of each bout feature for this animal $\sigma_f$.

Equation 1:

$$ Z_i = \frac{x_i - \bar{x}_f}{\sigma_f} $$

Active bout features across all animals were then centred by subtracting each feature’s mean value from every bout, and principal component analysis (PCA) was used to reduce the data to 3 dimensions, the knee point of the scree plot, which together explain 97.5% of the variance (Appendix Figure 3a-b, page 113).

Next, the active and inactive bouts were separately clustered using an evidence accumulation based approach (Fred and Jain, 2002, 2005), implemented by the function gmm_sample_ea.m (page 85). Firstly, 40,000 probe points were randomly sampled from the data. Next, for 200 iterations, another group of points were randomly sampled and fit with a Gaussian mixture model with a random number of clusters. For each iteration, these two parameters varied uniformly in the following ranges: the number of points sampled -- 40,000 to 100,000; the number of clusters fit -- 2 to 20. Each mixture model was fit using MATLAB’s fitgmdist function (MATLAB, Statistics and Machine Learning Toolbox) with full, regularized, independent covariance matrices and initialised using the k-means++ algorithm (Arthur and Vassilvitskii, 2007). Each mixture model was fit 5 times and the one with the largest log-likelihood was retained. Once each model had been fit, each probe point was assigned to the component with the largest posterior probability, and evidence in the form of pairwise occurrence in the same cluster, was accumulated on the probe points. Once the 200 mixture models had been fit, average link clustering was applied to the evidence accumulation matrix and the final number of clusters determined based on maximum cluster lifetime. Finally, the size of the resultant clusters was normalised by randomly selecting the number of points in the smallest cluster from each cluster (5,983 active and 614 inactive bouts) and all points were assigned to these size normalised clusters using the mode cluster assignment of the 50 nearest neighbours for every point.

ASD mutant data was fit into the resultant clusters from the test dataset as follows. The new active bouts (n = 39,486,104), were parameterised, centred by subtracting the test datasets mean feature values, and then projected into the test PCA space. For the inactive bouts (n = 39,486,484) both the test and new bouts were converted from frames to seconds. Finally, as with the test data, new bout cluster assignments were determined from the mode cluster
assignment of the 50 nearest neighbours for every new point to the original size normalised clusters from the test dataset.

**Hierarchical Compression**

Clustering reduced each animal’s behaviour to a non-repetitive sequence of active and inactive bouts, termed modules. On average this reduced each wild-type sequence length by 96%; from 6,308,514 frames to 236,636 modules, easing the computational demands of compressing these sequences.

To compress modular sequences an offline compressive heuristic (Nevill-Manning and Witten, 2000) was used. At each iteration (i) of the algorithm the most compressive motif was defined as the motif which made the most savings, defined as a balance between the length of the motif (W) and the number of times it occurred in the sequence (N), which also considered the combined cost of adding a new motif to the dictionary (W + 1) and of introducing a new symbol into the sequence (+N) at every occurrence of this motif in the sequence (equation 2).

\[ \text{Savings}_i = WN - (W + 1 + N) \]

The overall compressibility of a given input sequence was calculated by summing these savings across all iterations and dividing this total by the length of the original input sequence (in modules). This process resulted in a compressibility metric that ranged from 0-1 (low-high compressibility). To reduce computational time, motifs of a maximum of 10 modules long were sought, although the hierarchical nature of the algorithm enabled the identification of longer motifs through nesting. To generate the common motif libraries, the motifs obtained from compression of every animal’s full module sequence (Batch_Compress.m) were merged, and then all unique motifs were kept (Bout_Transitions.m). To generate sets of paired control sequences for every animal, each animal’s module sequence was divided into sequential day and night or hourly segments and the modules within each of these windows was shuffled 10 times, maintaining the active/inactive transition structure (Bout_Transitions.m). As compressibility varies non-linearly with uncompressed sequence length (Appendix Figure 6b, page 116), to enable comparisons between samples with different numbers of modules, non-overlapping blocks 500 modules long were compressed (Bout_Transitions_Hours.m).
Supervised Motif Selection

To identify both which and how many motifs were required to distinguish between behavioural contexts (e.g. day and night) the following approach was employed, executed by the function Batch_Grammar_Freq.m. Firstly, the number of occurrences of every motif, from the common motif library, was counted in every real and shuffled modular sequence. Next, to calculate enrichment/constraint scores for every motif the deviation of the real from shuffled counts, as well as the deviation of each shuffle from the other shuffles, was calculated. For a given animal and time window, i.e. day or night, the mean number of times motif (i) was counted in the shuffled data $s_i$ was subtracted from the real number of counts $x_i$ and divided by the standard deviation of the shuffled counts $\sigma_{si}$.

Equation 3:
$$ Z_i = \frac{x_i - \bar{s}_i}{\sigma_{si}} $$

When comparing the shuffled data to itself, each shuffle (now $x$) was excluded from $\bar{s}_i$ and $\sigma_{si}$. Infinite values occurred when there was no standard deviation in the $\sigma_{si}$ counts and thus $\sigma_{si}$ equalled zero. For subsequent working, infinite values were replaced with a constant value of ± 3.32. This value was chosen as equation 3 will always output this value when there is no standard deviation in the shuffled counts and $x_i$ is included in the calculation of $\sigma_{si}$. Note that in the real test data, infinite values only constituted 2.2% of all enrichment/constraint scores.

For any given comparison, motif library enrichment/constraint scores for the relevant animals were formatted into a matrix of samples by motifs (e.g. Figure 8a). Scores for each motif (column) were normalised by subtracting each column’s mean score and dividing by each columns standard deviation. A supervised feature selection algorithm (Hanchuan Peng, Fuhui Long and Ding, 2005) was applied to these matrices, to select the top 250 maximally relevant and minimally redundant (mRMR) motifs. To determine how many of these motifs were necessary for accurate classification linear discriminant analysis classifiers were trained on this data, using 10-fold cross validation as sequential mRMR motifs were added, and classification error mean and standard deviation were calculated. The MATLAB function fitcdiscr (Statistics and Machine Learning Toolbox) was used to implement these steps. Finally, to determine how many motifs were necessary for a given comparison, classification error curves were smoothed with a running average 3 motifs wide and the number of motifs at which the minimum classification error was obtained was identified. To evaluate classifier performance, the results of each classifier were compared to a majority class classifier, whose performance depended upon the ratio of samples of each class. For example, in a dataset with two labels at a ratio of 0.1 : 0.9, the majority class classifier would consistently assign the latter label and
achieve a classification error of 10% (± standard error of proportion).

Genotyping

DNA Extraction
DNA extraction was performed using HotSHOT DNA preparation (Truett et al., 2000). Larval samples were transferred to the individual wells of a 96-well PCR plate. Excess liquid was pipetted from each well before applying 50µl of 1x base solution (1.25M KOH, 10mM EDTA in water). Plates were heat sealed and incubated at 95°C for 30 minutes then cooled to room temperature before the addition of 50µl of 1x neutralisation solution (2M Tris-HCL in water).

PCR
PCR reactions were carried out as described below. Primer sequences for each mutant are detailed in Fish Lines (page 82). The following reaction mixture was prepared per sample, on ice in a 96-well PCR plate: 18.3µl PCR mix (2mM MgCl₂, 14mM pH 8.4 Tris-HCl, 68mM KCl, 0.14% Gelatin in water, autoclave for 20 minutes, cool to room temperature, chill on ice, then 1.8% 100mg/ml BSA and 0.14% 100mM d [A, C, G, T ] TP), 0.5µl of forward and reverse primers (20 µM), 5.5µl water, 0.2µl of Taq polymerase and 3.0µl of DNA were added. Next, each plate was heat sealed and placed into a thermocycler, set with the following program: 95°C --5 minutes, 44 cycles: 95°C -- 30 seconds, 57°C -- 30 seconds and 72°C -- 45 seconds, then 72°C -- 10 minutes and 10°C until collection. Finally, samples were mixed with 6x loading buffer (Colourless buffer: Ficoll-400 - 12.5g, Tris-HCl (1M, pH 7.4) – 5ml, EDTA (0.5M) – 10mL, to 50ml in pure water. Heat to 65°C to dissolve. Per 10ml of colourless buffer 25mg of both xylene cyanol and orange G were added, then diluted to 6x) and ran on agarose gels (1-2%) with 4% GelRed (Biotium).

KASP
To facilitate high-throughput experiments and the detection of small genetic insertions or deletions, KASP genotyping primers (LGC Genomics) were used. KASP primer results were validated by comparison to PCR based genotyping or Sanger sequencing (Source BioScience) of samples from each KASP classified genotype. KASP genotyping was carried out in white, low profile PCR plates on ice with six wells allocated 50:50 for positive and negative controls. The following reaction mixture was prepared per sample: 3.89µl of 2x KASP reaction mix, 0.11µl KASP primers, 1.0µl water and 3.0µl DNA. Plates were then heat sealed and placed into a thermocycler with the following thermal cycling program: 94°C -- 15 minutes, 10 cycles: 94°C -- 20 seconds, 61-53°C (dropping 0.8°C per cycle) -- 60 seconds. 26 cycles: 94°C --20 seconds, 53°C -- 60 seconds, then 10°C until collection.

In brief, KASP works as follows. Each allele-specific forward primer’s tail sequence corresponds to one of two FRET cassettes, during thermal cycling the tail sequence’s complement is generated, the fluorescently labelled oligos bind this sequence and now, no longer quenched
emit fluorescence. Thus, homozygous alleles generate one of two possible fluorescent signals while heterozygous samples produce a mixed signal. Following thermal cycling I used a fluorescence reader (Bio-Rad CFX96 Real-Time System) and Bio-Rad CFX Manager software (version 3.1) to automatically determine each sample's genotype from a 2d scatter plot of fluorescence in each channel. A heterozygous in-cross will for example produce a cluster of points for each resulting genotype from which outlying samples, of unclear genotype, can be excluded either automatically (by the software) or by manual curation.

**Iconography**

Analytical figures were generated in MATLAB 2016b-2018a. Imaging figures were prepared in Fiji. Each figure legend details how data is presented. All figures were formatted in InDesign CS6 (Adobe).

**Immunohistochemistry**

**Fluorescent In-situ Hybridisation**

For fluorescent in-situ hybridisation larvae were euthanised with an overdose of 4% tricaine. Then, the following protocol was carried out by Aisling O'Sullivan under my supervision. Samples were placed into fixative (4% paraformaldehyde (PFA) in phosphate buffered saline (PBS) with 4% sucrose), and left overnight on a rotating shaker at 4°C. Samples were dissected the next day and then dehydrated, through quarterly methanol/PBTween (PBS with 0.1% Tween) solutions into 100% methanol, in which they were stored overnight at -20°C. Next samples were rehydrated into PBTween at room temperature, permeabilised with 0.1% proteinase K (in PBTween) for 5 minutes and then left for 30 minutes in fixative. Samples were pre-hybridised in hybridisation buffer (50% formamide, 5 x SSC, 0.1% Tween, 5mg/mL torula yeast RNA, 50µg/mL heparin and 5% Dextran) at 67°C for an hour, then hybridised with galanin probe (0.0025% in hybridisation buffer) overnight at 67°C. The next day the probe was removed, and samples were subject to a series of hot washes at 65°C, in 0.1% Tween in pure water with the following modification per wash: 2 x 30 minutes in 50% formamide (2 x SSC), 1 x 15 minutes (2 x SSC) and 1 x 30 minutes (0.2% SSC).

Next samples were blocked with monoclonal antibody solution for 1 hour at room temperature, then incubated in 0.1% anti-DIG POD (Boehringer) at 4°C overnight. The following day samples were incubated in TSA plus Cy3 (2% in amplification buffer) and left to develop. Once the samples had developed the brains were washed in PBSTween and the steps below were followed, from the blocking step onwards, using Tween instead of Triton, for tERK immunohistochemistry.
**pERK/tERK Immunohistochemistry**

For pERK/tERK immunohistochemistry the Z-Brain protocol (Randlett et al., 2015) was followed, with minor modifications. Briefly larvae were fixed overnight on a rotating shaker at 4°C in 4% PFA in PBS with 4% sucrose, permeabilised in 0.05% Trypsin-EDTA on ice for 45 minutes, blocked for an average of 5 hours at room temperature in PBT (PBS plus 0.25% Triton) with 2% normal goat serum, 1% bovine serum albumin (BSA) and 1% dimethyl sulfoxide (DMSO), then incubated over sequential nights at 4°C in primary (Cell Signalling 4370 and 4696, 1:500) and secondary antibodies conjugated with Alexa fluorophores (Life Technologies, 1:200) in PBT plus 1% BSA and 1% DMSO.

*chd8* pERK/tERK experiments were conducted in triplicate for each timepoint. To minimise the effects of genetic or development variation, larvae for each *chd8* pERK/tERK experiment were always collected from breeding a single *chd8*+/− pair. Note that the pair differed between experimental repeats to ensure that results were independent of the particular parental pair. For the *chd8* pERK/tERK experiments a water permeable filter (Cell Culture) was placed at the bottom of each well of a 6-round well plate filled with fish water. Then a Pasteur pipette was used to transfer approximately 15 larva to each well of the plate, which was left in a 28.5°C light:dark incubator for several hours until 15 minutes prior to fixation, at which time each plate was moved to a lab bench to minimise the effects of handling upon the staining. At the fixation timepoint each filter, and thus the larvae, was manually transferred directly to a second 6-well plate filled with fixative. For the 06:30 a.m. timepoint these steps were carried out in the dark using a red head-torch, taking care to expose the larvae to as little light as possible.

During dissection each larva’s skull and eyes were placed into a 96-well PCR plate with 25µl of base solution (see Methods, Genotyping, DNA Extraction, page 92) per well. Each brain with most of the larva still attached, was placed into the corresponding well of a 96-well behaviour plate with 100µl PBS per well, then stored at 4°C. To determine each samples genotype DNA was extracted from the skull’s and eyes and KASP genotyping was carried out (see Methods, Genotyping, KASP, page 92). Based upon the KASP results, brains were sorted into genotype matched tubes and the immunostaining steps above were carried out. Immediately prior to mounting for imaging (see Methods, Microscopy, page 95) each samples brain was cut from its body using a scalpel. Finally, following imaging a scalpel was used to cut each sample from the slide into a 96-well PCR plate, prepared with 25µl of base solution in each well, and samples were again genotyped by KASP (see Method, Genotyping, KASP, page 92). Following this, imaged samples of incorrect or unclear genotype were excluded from further analysis.
Microscopy
For imaging samples were mounted in 1.5% low melting temperature agarose and imaged with a custom two-photon microscope (Bruker) with a 20x water immersion objective (Olympus XLMPLFLN, 1.00 NA). For imaging an 800nm laser at 40mW power was used to image a 1024 x 1024 plane with 0.584 x 0.584µm sized pixels and a pixel dwell time of 1.2µs. Each plane was scanned 8 times and averaged by the acquisition software (Prairie View, version 5.3). Sequential planes were acquired with a 2µm z-step size.

For the chd8 pERK/tERK experiments, genotype matched pairs of samples were mounted on slides, and across each imaging session genotype slides were imaged in rotation, e.g. 2 x chd8+/+ brains, 2 x chd8+/− brains, 2 x chd8−/− brains, 2 x chd8−/+ etc.

Pharmacology
0.15M melatonin and 1M pentylenetetrazole (Sigma - M5250 and P6500) stock solutions were made in DMSO and sterile water respectively. Behavioural testing concentrations for each compound were selected based upon (Rihel et al., 2010). For behaviour experiments each animal was dosed, in a well with 650µl of fish water, with 1.3µl of either vehicle control or compound at 500x concentration, resulting in a 1 in 500 dilution and thus the desired testing concentration.
“You’ll find, my friend, that what you love
will take you places you never dreamed you’d go.”

Tony Kushner, Angles in America


Bibliography, p -102


Lashley, K. S. (1951) ‘The Problem of Serial Order in Behavior’. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior; the Hixon Symposium (pp. 112-146).


Appendix of Figures
Appendix Figure 1. Behavioural Set-up & Analysis

a. Schematic of the Viewpoint behavioural set-up. Note that aside from the computer, the set-up is fully enclosed. Not shown to scale. IR - infra-red, LED - light emitting diode.

b. A fictive illustration of zebrafish behaviour (blue line) in a binary sense of active or inactive at each time point. Two minutes of data are shown divided by a black dashed vertical line. The old analysis approach would score both minutes as 20 seconds of activity, thus failing to capture that fact that there are twice as many bouts in the second minute and furthermore would completely miss the 60 second period of inactivity in between. This latter loss leads to a discrepancy in the number of periods ≥ 60s between the old and new methods (see c).

c. The number of inactive periods ≥ 60s for each of 124 wild type animals is shown, as measured by the old and new analytical approaches. Data is from each animals entire recording period (4-7dpf). Data for each animal is shown as a pale blue line, overlaid with a population mean and standard deviation. Insert shows the percentage of the new counts, detected by the old method per animal. Each animal’s data is shown by a circle and an orange cross marks the population mean.
Appendix Figure 2. Bout Features

a. Bout feature distributions during the day (light blue) and the night (dark blue). For the probability curves each animal’s data was fit with a probability density function (pdf), shown is a mean pdf (bold line) and standard deviation (shaded surround), with a log scale on the x-axis. For the scatter plots each fish’s average value across day/night 5/6 of development is shown as a light blue (day) or dark blue circle (night), an orange cross marks each population’s mean. Only the mean day and night active bout total and inactive bout length pdfs were consistently significantly different across three independent experiments (p < 0.01; Two-sample Kolmogorov-Smirnov test). n = 124 wild-type larvae.

b. Melatonin bout feature means. A mean was taken per animal, per feature and day/night at 6dpf. Shown is a population mean and standard error of the mean, during the day (white background) and the night (grey background). Control - DMSO. n = 24 controls then n = 12 per dose.
Appendix of Figures, p -112

Appendix Figure 2. Bout Features

c. PTZ bout feature means. A mean was taken per animal, per feature and day/night at 6dpf. Shown is a population mean and standard error of the mean, during the day (white background) and the night (grey background). Control - H2O. n = 24 controls then n = 10 (2.5mM), n = 9 (5mM) and n = 9 (7.5mM).

d. hcrtr bout feature means as in c, from days (white background) and nights (grey background) 5 to 6 of development. hcrtr+/− mutants had significantly lower mean values compared to both hcrtr+/+ and hcrtr−/+ for the following active bout features: length, standard deviation and total (p < 0.05 for all comparisons, Dunn-Sidak corrected four-way ANOVA, adjusted for the following factors: day/night, development and experimental repeat). No features differed significantly between hcrtr+/+ and hcrtr−/+. n = 39, 102 and 39, for hcrtr+/+, hcrtr−/+ and hcrtr−/− respectively.
**Appendix Figure 3. Module Details**

**a.** Scree plot showing the percentage of variance explained by each principal component from the active bout data. The first 3 principal components, the knee point of the curve, were kept for subsequent analysis and the colours of these points refer to **b**.

**b.** Each of the 3 retained components coefficients for the different active bout parameters are shown. Coloured as in **a**.

**c.** The active bouts within each module were fit by a Gaussian. Each bout is shown in a 3d space of PC1, PC2 and probability. Each bout is coloured by its module assignment (see legend in **d**).

**d.** Probability density functions for each bout feature. The legend panel indicates each modules colour. All features are shown with a log scale on the x-axis.
Appendix Figure 4. Evidence Accumulation Based Clustering

a. Evidence accumulation matrix for the 40,000 active probe points (matrix dimensions are thus 40,000 by 40,000). A higher E.A. index indicates a higher frequency of pairwise occurrences in the same cluster across 200 Gaussian mixture models. This matrix was clustered hierarchically, and a maximum lifetime cut was made to determine the final number of modules. The dendrogram above shows all 40,000 leaves and is coloured by mean module length from shortest (lightest) to longest (darkest) as in other figures.

b. Evidence accumulation matrix as in a, for the inactive bouts.
Appendix Figure 5. Melatonin, PTZ and hcrtr Module Probabilities

a. Melatonin module probabilities during day (upper panels) and night (lower panels) 6dpf for both the active (left) and inactive (right) modules. Shown is a mean and standard error of the mean for each group, coloured according to the legend. Active modules are sorted by average wild type day probability, from highest to lowest, based upon wild-type data in Figure 4a. Inactive modules are sorted by increasing mean length. Control - DMSO. n = 24 controls then n = 12 per dose.

b. PTZ data as in a, with H2O (control). n = 24 controls then n = 10 (2.5mM), n = 9 (5mM) and n = 9 (7.5mM).

c. hcrtr data as in a, with mean values across 5 and 6dpf. No module probabilities differed significantly between genotypes (full four-way ANOVA, with the following factors: genotype, day/night, development and experimental repeat). n = 39, 102 and 39, for hcrtr+/+, hcrtr+- and hcrtr-- respectively.
Appendix Figure 6. Hierarchical Compression Metrics

a. The compressibility (y-axis) of the real wild-type data is higher than the paired shuffled data (p < 10^{-15}; two-way ANOVA, real vs shuffled data, no significant interaction with experimental repeat factor). Each animal’s data is shown as a pale blue line. Overlaid is a mean and standard deviation. Insert: the mean difference in compressibility between each fish’s real and shuffled data, each fish is shown by a circle and the orange cross marks the mean.

b. The compressibility (y-axis) of the real wild type data varies non-linearly with uncompressed sequence length, each fish is shown as a circle.

c. Probability density functions (pdfs) showing the probability of observing motifs at different enrichment/constraint scores rounded to whole numbers and summed at values above or below ± 4. Each wild type animal is depicted by a single pale blue (real data) and 10 black (shuffled data) lines, overlaid in bold are mean pdfs. The insert shows the higher kurtosis of the real data compared to the shuffled data (pale lines)(p < 10^{-271}; two-way ANOVA, real vs shuffled data, no significant interaction with experimental repeat factor). Overlaid is a population mean and standard deviation.
Appendix Figure 7. Motif Classifier Performance

a. Classification error (%) from linear classifiers separating wild-type day and night behaviour using motif enrichment/constraint scores, as sequential mRMR motifs, from 1-250, are added (x-axis). The average error is shown in light blue, overlaid in darker blue is a running average 3 motifs wide. The broken black lines show the minimum of the smoothed data to be at 15 motifs where the classification error is 0.2%.

b. Wild-type temporal classifier performance. Real classifiers (colour) are shown as a mean and standard deviation from 10 fold cross validation. Majority class classifiers (grey) are shown as value and standard error of proportion. Each classifiers data is listed on the x-axis. D - day, N - night, M/E - morning/evening, E/LN - early/late night. The number of motifs chosen for each classification and exact values for each classifier are detailed in Appendix Table 1 (page 123).

c. *hcrt* Melatonin and PTZ classifier performance, as in b. For *hcrt* comparisons grouped classifiers are shown, as well as separate day (light blue underline) and night (dark blue underline) classifiers. For melatonin and PTZ only day data was compared. Classifier details can be found in Appendix Table 2 (page 124).
Appendix Figure 8. ASD Mutant Bout Features

a. chd8 bout feature means from days and nights 5 to 6 of development. A mean was taken per animal, per feature and day/night. Shown are genotype means and standard error of the mean, during the day (white background) and the night (grey background). No metrics differed significantly between genotypes (full four-way ANOVA, with the following factors: genotype, day/night, development and experimental repeat). n = 51, 96, 32 - WT, Het, Hom.

b. bckdk bout feature means as in a. Compared to bckdk+/+, bckdk−/− mutants had significantly higher mean active bout length during the day and the night, and higher active bout total during the day (p < 0.05, Dunn-Sidak corrected four-way ANOVA, adjusted for the following factors: development and experimental repeat). bckdk+/− did not differ significantly from either bckdk+/+ or bckdk−/− by any of these metrics. n = 73, 116, 75 - WT, Het, Hom.
Appendix of Figures, p -119

Appendix Figure 9. ASD Mutant Module Probabilities

a. Mean chd8 module probabilities during day (upper panels) and night (lower panels) 5 and 6dpf for both the active (left) and inactive (right) modules. Shown is a mean and standard error of the mean for each genotype, coloured according to the legend. The active modules are sorted by average wild type day probability, from highest to lowest, using wild-type data from Figure 4a. The inactive modules are sorted by increasing mean length. No module probabilities differed significantly between genotypes (full four-way ANOVA, with the following factors: genotype, day/night, development and experimental repeat). n = 51, 96, 32 - WT, Het, Hom.

b. Mean bckdk module probabilities as in a. There were no significant differences between genotypes for the active module probabilities. bckdk\textsuperscript{-/-} had significantly higher inactive module 1 probability during the day compared to both bckdk\textsuperscript{+/+} and bckdk\textsuperscript{-/-} (p < 0.05). bckdk\textsuperscript{+/+} and bckdk\textsuperscript{-/-} both had significantly higher inactive module 5 probabilities during the night than bckdk\textsuperscript{+/+} (p < 0.05), though did not differ from each other (see Figure 10b). Statistics are Dunn-Sidak corrected four-way ANOVA, adjusted for the following factors: development and experimental repeat. n = 73, 116, 75 - WT, Het, Hom.
Appendix Figure 10. Behaviour Analysis Pipeline

Flow diagram depicting the steps of the behaviour analysis pipeline. Data is output from the behavioural set-up (ViewPoint) in the form of a .xls file. perl_batch_192.m organises this data to a .txt format. Experiment metadata (e.g. animal genotypes) is supplied in the form of a .txt file. The old analysis method uses sleep_analysis2.m to produce figures and statistics from these two .txt files. The new method exports .raw data from ViewPoint to produce .xls files. Vp_Extract.m reorganises these, using .txt data, to a .mat file which can be input to either Vp_Analyse.m or Bout_Clustering.m. Vp_Analyse.m produces figures and statistics. Bout_Clustering.m uses the clustering function gmm_sample_ea.m to assign data to modules, produce figures and calculate statistics, it’s output can be input to Bout_Transitions.m which compresses full modular sequences by calling Batch_Compress.m and Batch_Grammar_Freq.m. The motifs identified from this approach can be input to Batch_Transitions_Hours which compresses 500 module chunks and uses Batch_Grammar_Freq to count motif occurrences per hour. With the exception of the old analysis method (sleep_analysis2.m), two example figures are shown for each figure producing step. All code can be run locally, though for speed several steps (indicated in green) are best run on a cluster computer.
Appendix Figure 11. Imaging Analysis Pipeline
Flow diagram depicting the steps of the imaging analysis pipeline. Whole brain imaging data is output from a custom 2-photon microscope, via Prairie View software, in the form of individual .tif slices. The script 2p_Process.m organises these slices into stacks, filters the stacks to remove noise and assigns pixel width metadata, provided in the form of an .xlsx file. This process outputs both .tif and .nrrd stacks for each channel of each sample. These stacks must be manually rotated in Fiji to lie in the same orientation as the Z-Brain reference. Now the script 2p_Process.m can register each stack to the Z-Brain reference, a two step process in which both affine and warp transformations are first calculated for the tERK channel (grey) then applied to the pERK channel (magenta). For whole-brain activity mapping 2p_Process.m separately down-samples each channel, blurs each with a 2d Gaussian filter, and then divides each samples pERK by tERK channel. Finally these pERK/tERK stacks are compared by Z-Brain code, which generates differential activity maps. Additional analysis of these maps is also possible. Each processing step is illustrated by following a single wild-type brain through the pipeline, the same single slice is shown at each step (middle slice in top triptych). The Z-Brain reference is shown as a maximum intensity projection.
Appendix of Tables
## Appendix Table 1. Wild Type Motif Classifier Performance

A table showing the performance of each classifier. Each classifier sought to separate the data shown in the comparison column, e.g. Day/Night. For the hours comparisons each hour was compared to data from all other hours grouped together. For each comparison 250 motifs were chosen by mRMR then a smaller number were retained (see Motifs column) based on classification error curves (see appendix Figure 7a, page 117). Cv – 10 fold cross validated. Std – standard deviation across the 10 folds. Mc – majority class classifier.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Motifs (Number)</th>
<th>Cv Error (%)</th>
<th>Cv Error Std (%)</th>
<th>Mc Error (%)</th>
<th>Mc Error of Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day/Night</td>
<td>15</td>
<td>0.20</td>
<td>0.63</td>
<td>50.0</td>
<td>2.25</td>
</tr>
<tr>
<td>Night 5/Night 6</td>
<td>85</td>
<td>19.76</td>
<td>8.09</td>
<td>50.0</td>
<td>3.18</td>
</tr>
<tr>
<td><strong>Day Hours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09-10</td>
<td>102</td>
<td>6.39</td>
<td>1.23</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>10-11</td>
<td>1</td>
<td>7.37</td>
<td>0.31</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>11-12</td>
<td>5</td>
<td>7.20</td>
<td>0.23</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>12-13</td>
<td>9</td>
<td>7.06</td>
<td>0.34</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>13-14</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>14-15</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>15-16</td>
<td>1</td>
<td>7.11</td>
<td>0.14</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>16-17</td>
<td>1</td>
<td>7.09</td>
<td>0.15</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>17-18</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>18-19</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>19-20</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>20-21</td>
<td>3</td>
<td>7.11</td>
<td>0.27</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>21-22</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td>22-23</td>
<td>1</td>
<td>7.14</td>
<td>0.12</td>
<td>7.14</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Night Hours</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23-24</td>
<td>23</td>
<td>0.69</td>
<td>0.47</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>24-01</td>
<td>177</td>
<td>9.84</td>
<td>1.83</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>01-02</td>
<td>5</td>
<td>9.92</td>
<td>0.51</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>02-03</td>
<td>88</td>
<td>9.72</td>
<td>1.18</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>03-04</td>
<td>1</td>
<td>10.00</td>
<td>0.17</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>04-05</td>
<td>22</td>
<td>9.92</td>
<td>0.47</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>05-06</td>
<td>1</td>
<td>10.00</td>
<td>0.17</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>06-07</td>
<td>1</td>
<td>10.00</td>
<td>0.17</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>07-08</td>
<td>3</td>
<td>9.84</td>
<td>0.34</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td>08-09</td>
<td>1</td>
<td>10.00</td>
<td>0.17</td>
<td>10.0</td>
<td>0.60</td>
</tr>
<tr>
<td><strong>Morning/Eveining</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>33.21</td>
<td>2.32</td>
<td>50.0</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td><strong>Early/Late Night</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>36.37</td>
<td>2.18</td>
<td>50.0</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Appendix Table 2. *hcrtr* and Pharmacological Classifier Performance

A table showing the performance of each classifier. Each classifier sought to separate the data shown in the comparison column, e.g. *hcrtr*+/+ (WT) and *hcrtr*+/− (Het). For the pharmacological comparisons each condition was compared to the rest of the conditions grouped together, aside from the control comparison, control data was excluded from these comparisons. For each comparison 250 motifs were chosen by mRMR, then a smaller number were retained (see Motifs column) based on classification error curves (see Supplemental figure 7a, page 117). Cv – 10 fold cross validated. Std – standard deviation across the 10 folds. Mc – majority class classifier. WT - *hcrtr*+/+, Het – *hcrtr*+/−, Hom – *hcrtr*−/−.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Motifs (Number)</th>
<th>Cv Error (%)</th>
<th>Cv Error Std</th>
<th>Mc Error (%)</th>
<th>Mc Error of Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>hlrtr</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day and Night</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• WT/Het</td>
<td>173</td>
<td>25.53</td>
<td>6.77</td>
<td>27.66</td>
<td>1.88</td>
</tr>
<tr>
<td>• WT/Hom</td>
<td>83</td>
<td>24.68</td>
<td>6.07</td>
<td>50.00</td>
<td>2.83</td>
</tr>
<tr>
<td>• Het/Hom</td>
<td>235</td>
<td>24.65</td>
<td>3.76</td>
<td>27.66</td>
<td>1.88</td>
</tr>
<tr>
<td>Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• WT/Het</td>
<td>80</td>
<td>19.50</td>
<td>9.60</td>
<td>27.66</td>
<td>2.66</td>
</tr>
<tr>
<td>• WT/Hom</td>
<td>195</td>
<td>16.67</td>
<td>7.50</td>
<td>50.00</td>
<td>4.00</td>
</tr>
<tr>
<td>• Het/Hom</td>
<td>55</td>
<td>22.70</td>
<td>7.02</td>
<td>27.66</td>
<td>2.66</td>
</tr>
<tr>
<td>Night</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• WT/Het</td>
<td>79</td>
<td>16.31</td>
<td>6.38</td>
<td>27.66</td>
<td>2.66</td>
</tr>
<tr>
<td>• WT/Hom</td>
<td>53</td>
<td>12.82</td>
<td>9.58</td>
<td>50.00</td>
<td>4.00</td>
</tr>
<tr>
<td>• Het/Hom</td>
<td>76</td>
<td>15.96</td>
<td>7.27</td>
<td>27.66</td>
<td>2.66</td>
</tr>
<tr>
<td>Melatonin (Day)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Control</td>
<td>40</td>
<td>0.0</td>
<td>0.0</td>
<td>25.00</td>
<td>4.42</td>
</tr>
<tr>
<td>• 0.01μM</td>
<td>89</td>
<td>1.39</td>
<td>4.52</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>• 0.1μM</td>
<td>192</td>
<td>1.39</td>
<td>4.52</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>• 1μM</td>
<td>132</td>
<td>2.78</td>
<td>6.02</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>• 3μM</td>
<td>97</td>
<td>0.0</td>
<td>0.0</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>• 10μM</td>
<td>250</td>
<td>2.78</td>
<td>6.02</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>• 30μM</td>
<td>133</td>
<td>0.0</td>
<td>0.0</td>
<td>16.67</td>
<td>4.39</td>
</tr>
<tr>
<td>PTZ (Day)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Control</td>
<td>26</td>
<td>0.0</td>
<td>0.0</td>
<td>46.15</td>
<td>6.91</td>
</tr>
<tr>
<td>• 2.5mM</td>
<td>55</td>
<td>0.0</td>
<td>0.0</td>
<td>35.71</td>
<td>9.06</td>
</tr>
<tr>
<td>• 5mM</td>
<td>162</td>
<td>0.0</td>
<td>0.0</td>
<td>32.14</td>
<td>8.83</td>
</tr>
<tr>
<td>• 7.5mM</td>
<td>104</td>
<td>0.0</td>
<td>0.0</td>
<td>32.14</td>
<td>8.83</td>
</tr>
</tbody>
</table>
Appendix Table 3. ASD Classifier Performance
A table showing the performance of each classifier. Each classifier sought to separate the data shown in the comparison column, e.g. *chd8*+/− (WT) and *chd8*−/− (Het). For each comparison 250 motifs were chosen by mRMR, of which a smaller set were retained (see Motifs column) based on classification error curves (see Appendix Figure 7a, page 117). Cv – 10 fold cross validated. Std – standard deviation across the 10 folds. Mc – majority class classifier.