A study of the Chd family of ATP-dependent chromatin remodellers in *Dictyostelium discoideum*

James L. Platt

Cardiff School of Biosciences

Cardiff University

PhD Thesis

Submitted: September 2013

Acknowledgements

First and foremost, I would like to thank my supervisors, Adrian Harwood at Cardiff University and Alan Kimmel at the National Institutes of Health, for their support and guidance throughout the four years. I've been lucky to be able to do great science on two continents for which I would like to thank the Wellcome Trust–NIH programme for funding the project.

I'd like to thank everyone in the Harwood and Kimmel labs as well as members of the Dale, Watson, Kent and Oliver labs for their help and advice throughout the project.

Thanks to Nick Kent, Ryan Dale and numerous other people I pestered for bioinformatics advice. Harold Smith at the NIDDK sequencing core has been an invaluable source of knowledge on library preparation.

Finally, I would like to thank my parents for supporting me and Nicola for following me around the world and staying up late talking whilst we were on opposite sides of the Atlantic.

Abstract

The CHD family of ATP-dependent chromatin remodellers, which are present in all eukaryotes, utilise energy from the hydrolysis of ATP to alter nucleosome structure. The multiple CHD proteins have previously been implicated in transcription and developmental regulation, although the relationship between different family members is not well understood. The CHD family in *H.sapiens* contains nine proteins divided into three subfamilies. Whereas *S.cerevisiae* possess a single CHD, the *Dictyostelium discoideum* genome contains three CHD genes: one subfamily I member, related to human CHD1/2, and two subfamily III members, related to human CHD7/8. *Dictyostelium* fills a gap in the complexity scale, providing a smaller complement of CHD family members than mammalian models whilst still possessing a multicellular development stage and retaining a compact manipulatable genome.

I have studied the expression patterns and created knockout mutants for all three Chd family members. The *chd* mutants have different developmental phenotypes that correlate with their expression profiles. RNA-seq analysis of the mutants showed that each Chd protein is responsible for regulating the expression of distinct classes of genes. Mapping nucleosomes genome wide in wild-type cells revealed that nucleosomes are structured similarly to higher eukaryotes. Mapping of nucleosomes in *chdC*-null suggested that ChdC regulates nucleosome repeat length in a subset of genes in a developmentally regulated manner.

This thesis presents a comprehensive study of the Chd family in *Dictyostelium* and, to my knowledge, the first in vivo study of the nucleosome remodelling activity of a Chd subfamily III member.

Figure 0.1: Word cloud abstract

Word cloud representation of 'A study of the Chd family of ATP-dependent chromatin remodellers in *Dictyostelium discoideum*'. Created using tagxedo.com.

Contents

Chapter 1:

Introduction

1.1 Chromatin and the CHD family

Chromatin remodellers alter chromatin structure, which regulates DNA-dependent processes such as transcription, DNA replication and repair. The first ATP-dependent chromatin remodeller was discovered in 1984 in two separate screens for regulators of mating-type switching and of sucrose fermentation in yeast (Stern *et al*., 1984, Neigeborn and Carlson, 1984). Subsequently four main families of ATP-dependent chromatin remodellers were discovered: SWI/SNF, ISWI, CHD and INO80. The CHD (Chromo Helicase DNA-binding) family was discovered in a mouse cDNA screen for DNA-binding proteins (Delmas *et al*., 1993). CHD proteins appear to be present in the vast majority of eukaryotes. Whilst *S.cerevisiae* has only one CHD protein (Chd1), mammalian systems have nine, split into three subclasses. The action of CHD1 has been extensively studied and has been well established as being involved in transcriptional elongation (Stokes *et al*., 1996, McDaniel *et al*., 2008). It is thought to replace nucleosomes after the transcriptional machinery has passed through the gene (Radman-Livaja *et al*., 2012, Smolle *et al*., 2012).

Study of the CHD proteins has been limited in general and study of the CHD family in higher multicellular eukaryotes has thus far been even more limited because many of the mutants are embryonically lethal (Khattak *et al*., 2002, Hurd *et al*., 2007, Bosman *et al*., 2005). Lethality of mutants and the disorders that accompany mutations in CHD proteins (e.g. CHARGE syndrome (Vissers *et al*., 2004), neuroblastoma (Thompson *et al*., 2003)) highlight their importance in growth and development. Comprehensive studies of the CHD family in

development and the specificity of individual CHD proteins have not taken place. *Dictyostelium* was previously identified as having three distinct CHD proteins (Rogers, 2010, Platt *et al*., 2013). *Dictyostelium* fills the complexity gap between unicellular yeast and more complex multicellular eukaryotes, providing strong established genetic tools, a small complement of CHD proteins and a multicellular developmental cycle. The aim of this study is to use the simple but animal-like organism *Dictyostelium discoideum* to investigate roles of CHD proteins in development, regulation of gene expression and nucleosome positioning.

1.2 DNA, histones and chromatin structure

The double helix structure of DNA was first described by James Watson and Francis Crick in 1953 (Watson and Crick, 1953) but it was not until 1974 that the accepted structure of the nucleosome was established (Kornberg, 1974, Olins and Olins, 1974). The nucleosome consists of ~147bp of DNA wrapped 1.7 times around the protein core, commonly called the histone octomer (Figure 1.1), which is made up of two copies of each histone (H2A, H2B, H3 and H4). The nucleosome is the basic repeating unit of chromatin in eukaryotic cells. The nucleosome provides protection to the DNA and has the ability to form complex higher-order structures (Figure 1.2) to compact large amounts of genomic DNA into the nucleus whilst providing a system for accessibility (Woodcock and Dimitrov, 2001). As well as packaging and protecting the DNA, the structure of chromatin, and individual nucleosome positions, provides a regulatory function, controlling the access of binding proteins to the DNA sequence and therefore

impacting on transcription, DNA repair, recombination and replication (Ehrenhofer-Murray, 2004).

Individual nucleosomes are separated by linker DNA, forming what is known as the 10nm fibre, which explains the beads-on-a-string structure as observed in vitro by electron microscopy (Olins and Olins, 1974). The linker DNA length varies widely between organisms and even in different tissues of the same organism from 10 to 80bp (Felsenfeld and Groudine, 2003, Valouev *et al.*, 2011). The core histones (H2A, H2B, H3 and H4) are highly conserved throughout eukaryotes (Baxevanis and Landsman, 1996). All contain the highly conserved C-terminal histone fold, a 'helix turn helix turn helix' structure that allows for dimerisation (Arents and Moudrianakis, 1995, Baxevanis *et al.*, 1995). The core histones have long, unstructured N-terminal tails that protrude from the nucleosome core; these tails can be covalently modified and interact with other nucleosomes and other nuclear proteins (Wolffe and Hayes, 1999). There are also H1 and H5 histones that are known as linker histones. H1 has been shown to bind the entry and exit sites of the DNA on the nucleosome, allowing the formation of higher-order structures that eventually form the classic mitotic chromosome (Allan *et al.*, 1981, Carruthers *et al.*, 1998). There are other nucleosomes known as non-canonical histones that are variations on the core nucleosomes and have a specific function and/or chromosomal location. The H3 variant CENP-A replaces one or both of the H3 histones only at centromeres. It contains an N-terminal domain which allows kinetochore binding (Sullivan *et al.*, 1994). H2 variant H2AX is used as a marker of double-stranded DNA breaks

(Rogakou *et al*., 1998). H2AZ is found in nucleosomes at the promoter of genes (Mavrich *et al*., 2008).

Chromatin has been described as being divided into two broad types, heterochromatin and euchromatin. Euchromatin is a very open, accessible form of chromatin, containing genes that are being expressed as well as genes that are not, and is thought to be mainly in the 10nm fibre structure (Grewal and Jia, 2007). Whereas heterochromatin is a tightly packed, inaccessible structure, containing repressed genes, and is made up of higher-order chromatin structures (Grewal and Jia, 2007). Both euchromatin and heterochromatin are actively maintained in these states. As well as higher-order packing of nucleosomes, the position of individual nucleosomes can also affect DNA accessibility and gene expression. The DNA is so tightly wrapped on the nucleosome that it is inaccessible to some factors, thus nucleosome positions can affect activities such as transcription factor binding (Martinez-Campa *et al*., 2004, Boeger *et al*., 2004).

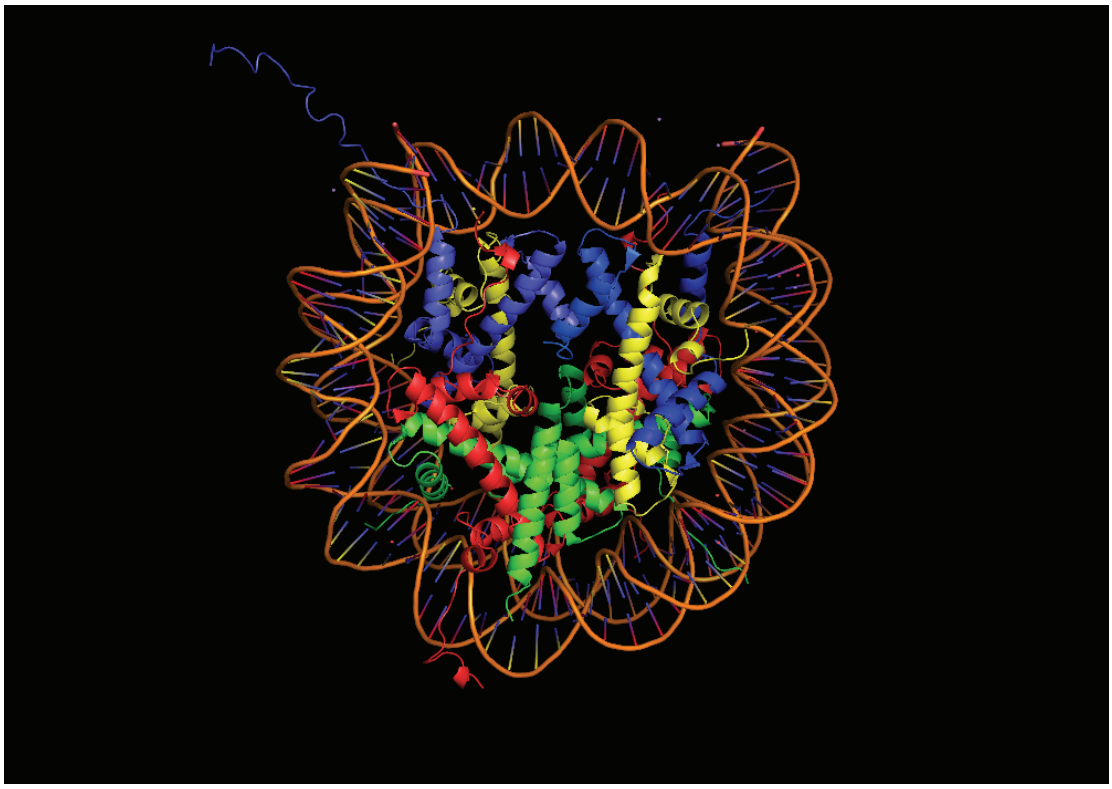


Figure 1.1: The nucleosome

Eight histones make up the core histone octomer: two copies of H2A, H2B, H3 and H4. The nucleosome consists of ~147bp of DNA wrapped around the histone octomer 1.7 times. The unstructured N-terminal histone tails protruding from the core are able to interact with other nucleosomes and proteins. Protein Data Bank ID: 1AOI, (Luger *et al*., 1997).

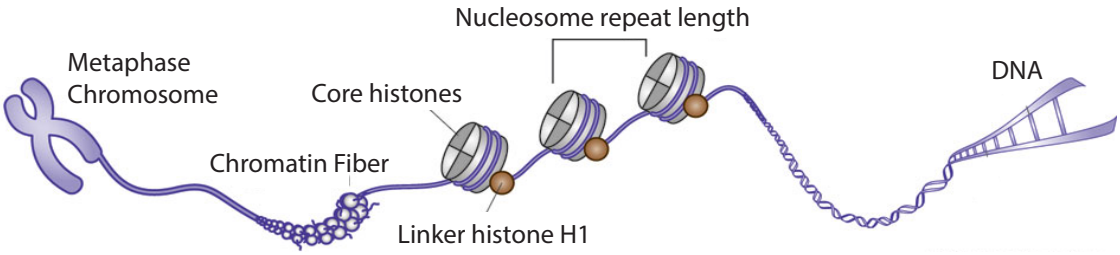


Figure 1.2: Chromatin structure from DNA to metaphase chromosomes

DNA is wrapped 1.7 times around the nucleosome. This makes up the fundamental 10nm beads-on-a-string fibre. This fibre is thought to be folded further to make the 30nm fibre. Further folding compresses the DNA into its fully compacted state during metaphase in mitosis. Adapted from Jakovcevski and Akbarian, 2012.

1.3 Chromatin remodelling

Chromatin can be physically changed or modified, altering the access of DNA-binding factors to the DNA sequence, converting between euchromatin and heterochromatin, moving nucleosomes or recruiting factors to a region of chromatin. This is important for regulation of DNA replication and repair, production and maintenance of gene expression patterns during differentiation, as well as dynamic regulation of other genes. There are two broad ways in which chromatin can be remodelled: 1) Covalent histone modification, whereby the N-terminal tails of histones are subject to post-translational modifications (acetylation, methylation, phosphorylation, ubiquitination, ADP ribosylation, SUMOylation) by specific enzymes (Strahl and Allis, 2000), 2) ATP-dependent chromatin remodelling, where complexes slide or evict nucleosomes, loop DNA off the nucleosome or exchange core histones for non-canonical histones (Clapier and Cairns, 2009). See Figure 1.3.

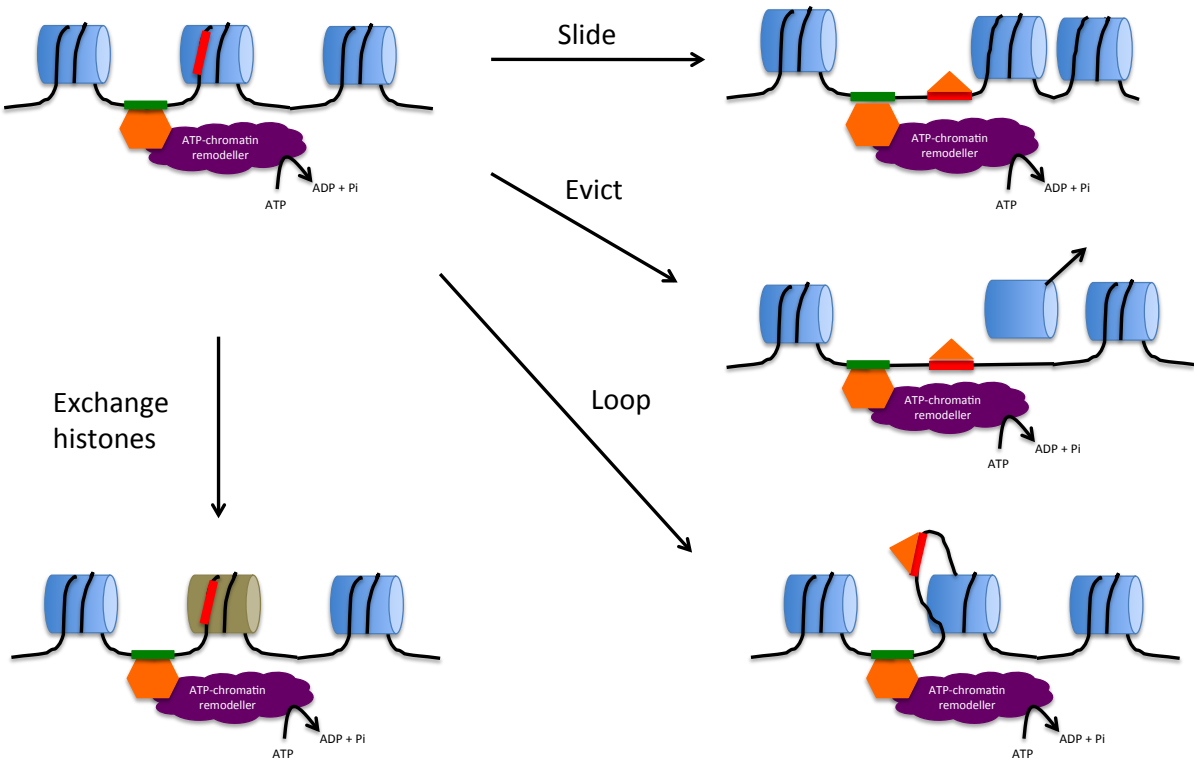


Figure 1.3: Mechanisms of ATP-dependent chromatin remodelling

ATP-dependent remodellers are recruited to chromatin and use the energy from the hydrolysis of ATP to physically alter chromatin structure. This can be achieved by sliding nucleosomes along the DNA, evicting the nucleosome in its entirety from the DNA, looping the DNA that is on the nucleosome off or by exchanging canonical histones for non-canonical histones. This impacts on DNA accessibility or, in the case of histone exchange, the interaction with other proteins.

1.4 Covalent histone modification

Specific enzymes target and modify specific amino acid residues in the N-terminal histone tails. Currently known modifications include acetylation, methylation, phosphorylation, ubiquitination, ADP ribosylation and SUMOylation. These marks are added or removed from histone tails by different enzymes. These modifications can either change intra- or inter-nucleosomal interactions. Acetylation of lysine residues neutralises the histone's positive charge, loosening interactions with the negatively charged phosphate backbone and making the DNA wrapped around the nucleosome more accessible (Hong *et al*., 1993). Hyperacetylation of histone tails is associated with actively transcribing genes (Allfrey *et al*., 1964, Dion *et al*., 2005). Some marks are generally thought to be repressive marks (H3K9me3) and some activating (H3K4me3). Although contrary to previous ideas of activating and repressive marks, genome-wide analyses have started to show that this may not be completely the case. In fact, the state of the chromatin or the activation of a gene may be much more complex dependent on the combination of histone marks present at a particular region, which is termed the 'histone code' (Strahl and Allis, 2000, Wang *et al*., 2008, Bernstein *et al*., 2012).

Histone modifications can also be used as molecular marks or 'flags' to recruit other proteins to the chromatin. These secondary proteins can then have a further function either to repress or open up the region of chromatin they are associated with (Bhaumik *et al*., 2007). Histone marks are bound by specific proteins that contain appropriate binding domains. Bromodomains can bind

acetylated histone marks; chromodomains can bind methylated histone marks. Differences in the binding pockets of domains allow them to bind different residues. The chromodomain in heterochromatin associated protein 1 (HP1) binds H3K9me2/3 (Lachner *et al*., 2001, Bannister *et al*., 2001), whereas the chromodomains of the human CHD1 protein bind H3K4me2/3 (Flanagan *et al*., 2005).

1.5 ATP-dependent chromatin remodelling

ATP-dependent chromatin remodellers use the energy from the hydrolysis of ATP to physically move the nucleosome position or change the structure of the nucleosome. These remodelling complexes can alter or remodel nucleosomes by sliding or evicting nucleosomes, looping DNA off the nucleosome or exchanging core histones for non-canonical histones (Figure 1.3). These chromatin remodellers are most commonly the core of a much larger complex containing the ATPase domain. These complexes can be in excess of 1MDa in size, containing other binding proteins and catalytic proteins. The SWI2/SNF2 complex is ~2MDa in size (Carlson and Laurent, 1994). The CHD3/4 proteins are part of the NuRD (Nucleosome Remodelling and histone Deacetylase) complex, which also contains histone deacetylases that remove acetyl groups from histone tails (Tong *et al*., 1998, Wade *et al*., 1998).

The first ATP-dependent chromatin remodeller to be discovered, SWI2/SNF2 was identified in *Saccharomyces cerevisiae* in two separate mutation screens for aberrant mating-type switching (SWI) and sucrose fermentation (Sucrose

NonFermentable) (Stern *et al*., 1984, Neigeborn and Carlson, 1984). SWI2 was responsible for regulating expression of HO, an endonuclease required for mating-type switching, and SNF2 was responsible for regulating expression of SUC2, an invertase involved in sucrose metabolism. Once sequenced, SWI2 and SNF2 were found to be the same gene (Peterson and Herskowitz, 1992), thus it is referred to as SWI2/SNF2. Experiments revealed that SWI2/SNF2 has DNA-dependent ATPase activity (Laurent *et al*., 1993) that has the ability to alter histone–DNA interactions (Hirschhorn *et al*., 1992). Further experiments determined SWI2/SNF2 to be the core of a 2MDa protein complex with other SWI and SNF proteins (Cairns *et al*., 1994, Peterson *et al*., 1994). The complex was later found in humans (Khavari *et al*., 1993, Wang *et al*., 1996) and is conserved in eukaryotes.

Since the discovery of SWI/SNF, other families of ATP-dependent chromatin remodellers have been discovered. The ISWI family was first discovered in *Drosophila* embryo extracts using an in vitro transcription factor accessibility assay (Tsukiyama and Wu, 1995). It was named imitation switch (ISWI) because of the similarity of the core ATPase domain to that of SWI2/SNF2 (Tsukiyama and Wu, 1995). The CHD (Chromodomain-helicase DNA-binding protein) family was first identified in a mouse cDNA screen for DNA-binding proteins; the CHD1 protein was identified as having features of SWI2/SNF2 and Polycomb/HP1 proteins (Delmas *et al*., 1993). The fourth and most recent family to be discovered, INO80 was identified some time after the first three families in *Saccharomyces cerevisiae* as it was found to be required for activating

transcription in response to low inositol levels (Ebbert *et al*., 1999). It was also identified independently through similarities to ISWI (Shen *et al*., 2000).

The four families share the common SNF2 ATPase domain and display high sequence similarity in this region (Gorbalenya and Koonin, 1993, Clapier and Cairns, 2009). What defines the different families is the presence of different key characteristic domains (Figure 1.4) outside of the SNF2 ATPase region (Clapier and Cairns, 2009). SWI/SNF contains a DNA-binding HSA (helicase-SANT) domain and a C-terminal bromodomain, which binds acetylated histone tails and is in a complex of 8 to 14 subunits (Hassan *et al*., 2002). ISWI contains C-terminal SANT and SLIDE (SANT-like ISWI domain). Jointly, the two domains together bind an unmodified histone tail and DNA respectively (Grüne *et al*., 2003). ISWI is found in the smallest complexes containing only 2 to 4 subunits (Corona and Tamkun, 2004). The CHD family contains characteristic N-terminal double chromodomains, which in human CHD1 bind methylated H3K4 (Flanagan *et al*., 2005) and often a DNA-binding domain (Marfella *et al*., 2006). Depending on the family member and species, CHD proteins act alone or in a group of up to 10 subunit protein complexes (Clapier and Cairns, 2009). The INO80 family, which includes SWR1, also contains a HSA domain, but its defining feature is the 'split' ATPase domain, the ATPase domain consists of two parts, DExx and HELICc domains, which in the INO80 family are split by a much larger distance than in the other families (Ebbert *et al*., 1999, Mizuguchi *et al*., 2004). This 'insertion' contains the binding site for other proteins, whilst maintaining ATPase activity, and unlike the other families shows strong ATPase activity in the absence of DNA (Shen *et al*., 2000). INO80 can be found in large complexes (>10 proteins) with various actin-related proteins (Shen *et al*., 2000).

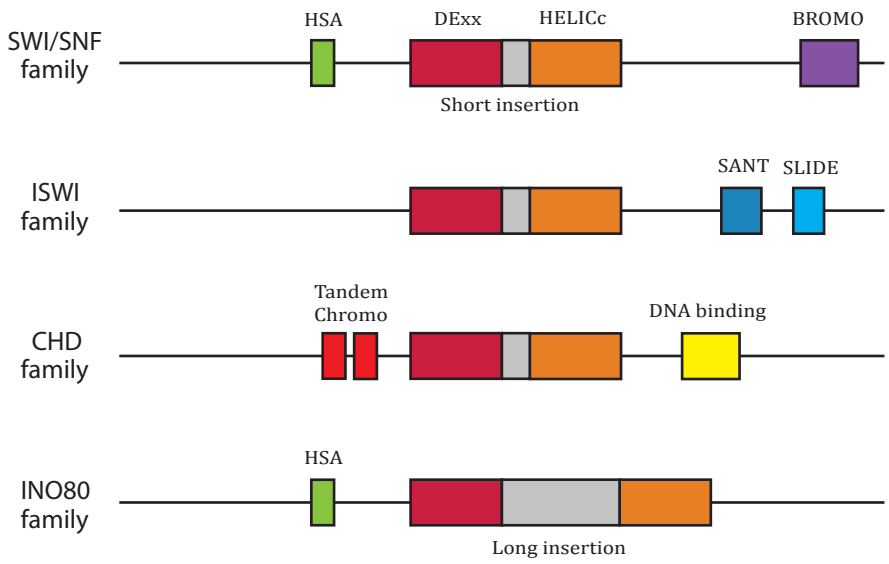


Figure 1.4: ATP-dependent chromatin remodeller families

The four main families are SWI/SNF, ISWI, CHD and INO80. All have a central ATPase domain made up of a DExx and a HELICc domain. INO80 family is defined by having a longer insertion between the two. Each family has unique domains that are labelled HSA (helicase-SANT), Bromo (bromodomain) and Tandem-chromo (pair of chromodomains).

1.6 Studying nucleosome positions

Shortly after the isolation of histones their modification was shown to be important in regulating gene expression (Allfrey *et al*., 1964). Then after the discovery of the 'beads on a string' (Olins and Olins, 1974, Oudet *et al*., 1975, Woodcock *et al*., 1976) and the repeating structure of the nucleosome with DNA wrapped around (Hewish and Burgoyne, 1973, Kornberg, 1974), interest gathered on the positions of nucleosomes and their impact upon higher-order packaging of chromosomal DNA, transcription, replication and repair of DNA. Nucleosomes were found to inhibit transcriptional initiation by RNA polymerase in vitro (Knezetic and Luse, 1986, Lorch *et al*., 1987). Initial studies utilised nuclease digestion of DNA and radiolabelling on extrachromosomal fragments in vitro. Micrococcal nuclease (MNase) is an endonuclease that can only cut linker DNA between nucleosomes. The nucleosome provides protection from MNase, thus producing a 'footprint' of where the nucleosome was (Axel, 1975). MNase digestion and indirect end-labelling became a standard way to map nucleosomes in vivo on individual loci (Kent and Mellor, 1995, Wu and Winston, 1997). With the advent of high-throughput approaches, microarrays were used to map nucleosomes at 10–20bp resolutions (Yuan *et al*., 2005, Lee *et al*., 2007, Whitehouse *et al*., 2007, Ozsolak *et al*., 2007). Subsequently high-throughput sequencing approaches were then used to map nucleosomes at up to 1bp resolution and genome wide on a variety of organisms: *S.cerevisiae* (Albert *et al*., 2007), *C.albicans* (Field *et al*., 2009), *S.pombe* (Lantermann *et al*., 2010), *D.melanogaster* (Mavrich *et al*., 2008), *H.sapiens* (Schones *et al*., 2008), *C.elegans* (Valouev *et al*., 2008), *D.discoideum* (Chang *et al*., 2012, Platt *et al.,* in

preparation), *A.thaliana* (Chodavarapu *et al*., 2010), *M.musculus* (Li *et al*., 2011) and *P.falciparum* (Westenberger *et al*., 2009).

Nucleosomes are globally phased relative to the 5' start of genes, in the majority of organisms. The +1 nucleosome is the most positioned with each subsequent nucleosome less so. Nucleosome-free regions exist at the 5' and 3' end of most genes and generally intronic and intergenic regions have lower nucleosome occupancy levels than exonic regions (Andersson *et al*., 2009, Schwartz *et al*., 2009, Tilgner *et al*., 2009). Nucleosomes show preference for certain sequences over others; in particular nucleosomes are unable to form on top of stiff stretches of homopolymeric poly(dA:dT) (Nelson *et al*., 1987, Segal and Widom, 2009). It has been suggested that DNA sequence plays a significant role in nucleosome positioning (Segal *et al*., 2006, Kaplan *et al*., 2009). Although DNA sequence alone is not enough to reconstitute nucleosome patterns in vitro, only nucleosome-free regions were reconstituted via salt dialysis (Gkikopoulos *et al*., 2011). Whole cell extract and ATP are required to recover nucleosome patterns in vitro (Zhang *et al*., 2011). These experiments demonstrate that ATP-dependent chromatin remodellers are essential for establishing, maintaining and enabling dynamic movement of the nucleosome positions. Zhang *et al*. (2011) were able to reconstitute nucleosome positioning in vitro suggesting transcription and DNA replication are not necessary for nucleosome phasing on genes.

Genome-wide nucleosome mapping of chromatin remodeller mutants has begun to dissect the roles of individual and combinations of chromatin remodellers (Gkikopoulos *et al*., 2011) in nucleosome positioning: ISW2 is important for +1

nucleosome placement (Whitehouse *et al*., 2007), ISWI is involved in the positioning of nucleosomes in the coding region (Tirosh *et al*., 2010) and Chd1 is involved in the replacement and positioning of nucleosomes over the gene body (see section 1.10.1). As yet, no nucleosome map of any of the other CHD family members has been produced.

1.7 CHD family

CHD1 was first identified in a mouse cDNA screen for DNA-binding proteins (Delmas *et al*., 1993). It was identified as having features of SWI2/SNF2 (the ATPase domain) and features of Polycomb/HP1 proteins (the chromodomains), as well as a DNA-binding domain. Uniquely it contained double chromodomains, a feature only found in CHD ATP-dependent chromatin remodellers. Subsequently CHD proteins were found in a variety of eukaryotes (Hall and Georgel, 2007). There are nine CHD proteins in mammals. They are divided into three subfamilies: subfamily I contains CHD1 and CHD2 (Woodage *et al*., 1997); subfamily II contains CHD3, CHD4 (Woodage *et al*., 1997) and CHD5 (Thompson *et al*., 2003); subfamily III contains CHD6 (Schuster and Stöger, 2002), CHD7 (Vissers *et al*., 2004), CHD8 (Ishihara *et al*., 2006) and CHD9 (Shur and Benayahu, 2005). The subfamilies are divided based on their accessory domains outside the core CHD domains (Hall and Georgel, 2007). Subfamily I is the archetypal CHD protein, containing the three original domains – tandem chromodomains, ATPase/helicase domain and a DNA-binding domain – as well as a DUF4208 (Domain of Unknown Function) domain (Marfella and Imbalzano, 2007). Subfamily II is characterised by having N-terminal PHD (Plant HomeoDomain)

domains, some of which have been shown to bind histone tails (Musselman *et al*., 2011, Oliver *et al*., 2012, Paul *et al*., 2013). Subfamily III is a more heterogeneous class; they are generally larger than the proteins in the other groups. Most contain a SANT/SLIDE domain that is thought to interact with histone tails and DNA. CHD7, 8 and 9 all have a BRK domain (Marfella and Imbalzano, 2007), another domain of unknown function only found in these CHD proteins and BRG1, the human ortholog of SWI2/SNF2. BRK is found exclusively in metazoans (Allen *et al*., 2007).

1.8 CHD Core domains

1.8.1 Double chromodomains

The chromodomain (chromatin organisation modifier) was originally characterised in HP1 and polycomb (Paro and Hogness, 1991). The chromodomain is a 50a.a. sequence that is evolutionarily conserved and found in proteins involved in processes connected with chromatin. The double chromodomain is characteristic of CHD proteins (Woodage *et al*., 1997).

Chromodomains can recognise methylated lysine residues in histone tails (Yap and Zhou, 2011). Human and *D.melanogaster* CHD1 bind to H3K4me2/3 but *S.cerevisiae* Chd1 does not (Sims *et al*., 2005, Flanagan *et al*., 2005, Flanagan *et al*., 2007). Although the chromodomains in *D.melanogaster* CHD1 are not important for localisation, they are required for enzymatic activity (Morettini *et al*., 2011). Crystal structures of *S.cerevisiae* Chd1 suggest that the chromodomains inhibit

ATPase activity by only enabling activation in the presence of nucleosomes but not naked DNA (Hauk *et al*., 2010). The chromodomains of Mi-2 (CHD3/4) in *D.melanogaster* do not bind H3K4 and are implicated in binding DNA (Bouazoune *et al*., 2002). It has been suggested that the CHD proteins can be grouped based on sequence differences in their chromodomains, which impact on their function (Flanagan *et al*., 2007).

1.8.2 ATPase domain

The ATPase motor, the SNF2 domain named after the first chromatin remodeller studied, is well conserved though evolution (Winston and Carlson, 1992). The energy provided by the ATPase activity is used to slide nucleosomes or displace histones (Becker and Hörz, 2002, Durr *et al*., 2006). ATPase activity in CHD proteins is stimulated by either nucleosomes or DNA or in some cases both, and activity has been shown in CHD1 (Tran *et al*., 2000), dMi-2 (Bouazoune *et al*., 2002), CHD6 (Lutz *et al*., 2006), CHD8 (Thompson *et al*., 2008) and CHD9 (Shur and Benayahu, 2005). Recent work has shown that the other domains in the CHD protein act to regulate the ATPase and nucleosome sliding ability of the SNF2 domain (Hauk *et al*., 2010, Ryan *et al*., 2011, Patel *et al*., 2011).

1.8.3 DNA-binding domain

The DNA-binding domain, once thought to be unique to subfamily I, has been shown to be made up of a SANT and SLIDE domain, two related DNA-binding domains (Ryan *et al*., 2011). Combinations of SANT and SLIDE domains have

been found in the other CHD subfamilies (Ryan *et al*., 2011). In fact the SLIDE domain of the CHD1 homologue in *Dictyostelium* is more closely related to metazoan CHD1 than fungi is (Supplemental Figure 6, Ryan *et al*., 2011). The DNA-binding domain of CHD1 has been shown to have little sequence specificity but displays a preference for AT-rich DNA (Delmas *et al*., 1993, Stokes and Perry, 1995). CHD9 also binds AT-rich DNA (Shur and Benayahu, 2005). It was suggested CHD1 binds to the minor groove of DNA (Stokes and Perry, 1995, Ryan *et al*., 2011), although crystal structure of the DNA-binding domain with DNA suggests it may bind either major or minor grooves (Sharma *et al*., 2011). Extranucleosomal DNA is required for binding in CHD1 (Ryan et al., 2011, Sharma *et al*., 2011). The DNA-binding domain of Chd1 increases the efficiency of nucleosome sliding (Ryan *et al*., 2011). Work from others also showed the DNA-binding domain was required to increase efficiency and directionality of nucleosome sliding (McKnight *et al*., 2011).

1.9 CHD expression and developmental role

1.9.1 CHD subfamily I

CHD subfamily I is the most studied subgroup. For the most part research has focused on CHD1 and most progress has been made in *S.cerevisiae*, which has only one CHD protein. In *S.cerevisiae* a knockout of *chd1* was viable and produced no growth phenotype, although it did have a minor resistance to 6-azauracil suggesting its involvement in transcription (Woodage *et al*., 1997, Tsukiyama *et al*., 1999). Deletion of CHD1 in *Drosophila* was also non-lethal but did cause a

wing margin phenotype and infertility in both males and females (McDaniel *et al*., 2008). Human CHD1 and CHD2 mRNA was expressed in all major tissues tested (Woodage *et al*., 1997). Mouse CHD2 was also highly expressed in heart tissue (Marfella *et al*., 2006).

CHD1 is upregulated upon reprogramming of fibroblasts into pluripotent stem cells (iPSC) and shRNA knockdown of CHD1 in these cells decreases the efficiency of reprogramming (Grskovic *et al*., 2007, Gaspar-Maia *et al*., 2009). Knockdown of CHD1 resulted in loss of pluripotency in embryonic stem cells (ESC) and increases in heterochromatic regions (Gaspar-Maia *et al*., 2009).

Deletion of the DNA-binding domain of CHD2 in mice results in perinatal death in homozygotes. Heterozygotes had decreased neonatal viability and an increase in non-neoplastic lesions to major organs (Marfella *et al*., 2006), demonstrating CHD2 is vital for proper organ development. No mouse CHD1 knockout has been reported.

1.9.2 CHD subfamily II

This subfamily was originally discovered when the Mi-2 protein was identified as an autoantigen of the human disease dermatomyositis (Seelig *et al*., 1995). Human CHD3 mRNA was expressed in most tissues studied (Woodage *et al*., 1997). In situ hybridisation in mouse neonates found CHD4 mRNA expressed in the thymus, kidney, hemopoietic foci in the liver, specific areas of the brain, mucosal epithelium and in hair follicles (Kim *et al*., 1999). CHD3 was found in the

same tissues at a lower level (Kim *et al*., 1999). *D.melanogaster Mi-2a* (CHD3) *Mi-2*β (CHD4) mRNA were both present in the embryos, *Mi-2a* in the early to mid stages of development (0-9H) and *Mi-2*β in the mid to late stages (3-24H) and both are present in the ovary. Mutants were embryonic lethal (Khattak *et al*., 2002). Analysis of *C.elegans* CHD3 and CHD4 demonstrated both are essential for development and are partially redundant (von Zelewsky *et al*., 2000).

Expression of CHD5 mRNA was limited to neural-derived tissues (fetal brain, total brain, cerebellum) and the adrenal gland. Expression was absent in all other tissues (Thompson *et al*., 2003). CHD5 was originally identified as the tumour suppressor in neuroblastoma (see section 1.11), a tumour of the peripheral sympathetic nervous system frequently affecting infants and children (Thompson *et al*., 2003). Human CHD5 is sometimes placed in subfamily III despite containing characteristic PHD domains.

1.9.3 CHD subfamily III

The third subfamily is the most diverse group. CHD6 mRNA was found expressed in cancer cell lines originating from a variety of different tissues (Schuster and Stöger, 2002) and enriched in various parts of the mouse brain (Lathrop *et al*., 2010). Homozygous deletion of one exon encoding part of the ATPase domain of CHD6 in mice produced viable fertile offspring with only mild behavioral defects: impaired co-ordination and balance (Lathrop *et al*., 2010).

Mutations of CHD7 are found in CHARGE syndrome (see section 1.11), a severe developmental disorder (Vissers *et al*., 2004). CHD7$^{-/-}$ mice are embryonic lethal (Hurd *et al*., 2007, Bosman *et al*., 2005). CHD7 is ubiquitously expressed in the developing human fetus by 22d, further into development becoming more specific to the CNS and reducing to tissues affected in CHARGE-syndrome (Sanlaville *et al*., 2006). Using an in vitro model for human neural crest cell migration and differentiation, knockdown of CHD7 was shown to be important for neural crest cell formation and migration (Bajpai *et al*., 2010). CHD7 is critical for olfactory neural stem cell proliferation and regeneration of olfactory sensory neurons (Layman *et al*., 2009). Inactivation of CHD7 in adult neural stem cells in mice reduces neurogenesis (Feng *et al*., 2013). CHD7 is also initially expressed ubiquitously in the zebrafish embryo, with expression later in development persisting only in the brain and retina (Patten *et al*., 2012).

CHD8 is expressed in a variety of tissues in the adult mouse (Ishihara *et al*., 2006). It is involved in the function of the *H19* DMR insulator required for the imprinted expression of *IGF2* (Ishihara *et al*., 2006). CHD8 is also involved in modulating the β-catenin signalling response (Thompson *et al*., 2008). A fragment of the N-terminal end of CHD8 called Duplin, which contains the double chromodomain, was shown to be highly expressed during early mouse development but low in newborns and Duplin$^{-/-}$ mice died in early embryogenesis (Nishiyama *et al*., 2004).

CHD9 (or CReMM), originally cloned from mesenchymal stromal cells (Shur and Benayahu, 2005), interacts with skeletal tissue-specific promoters – CBAF1,

biglycan, collagen-II, oesteocalcin and myosin – in a differential manner (Shur *et al*., 2006). A study in bone marrow-derived osteogenic cell line showed that CHD9 was robustly expressed in proliferating cells (Marom *et al*., 2006).

Table 1.1: Summary of the expression and developmental roles of CHD proteins in different organisms

	Organism	Expression/Phenotype
Chd1	*S.cerevisiae*	6-azauracil resistance (Woodage *et al*., 1997).
CHD1	*D.melanogaster*	Wing margin phenotype and infertility in males and females (McDaniel *et al*., 2008).
CHD1	*H.sapiens*	mRNA widely expressed in all tissues tested (Woodage *et al*., 1997). Required for reprogramming of fibroblasts into iPSC and maintaining pluripotency in ESC (Grskovic *et al*., 2007, Gaspar-Maia *et al*., 2009).
CHD2	*H.sapiens*	mRNA expressed in all tissues tested (Woodage *et al*., 1997).
CHD2	*M.musculus*	Homozygous deletion of DNA-binding domain results in perinatal death (Marfella *et al*., 2006).
CHD3	*H.sapiens*	mRNA expressed in all tissues tested (Woodage *et al*., 1997).
Mi-2*a*	*D.melanogaster*	mRNA present in the embryo and early to mid stages of development. Mutant was embryonic lethal (Khattak *et al*., 2002).
CHD3	*C.elegans*	Essential for development and partially redundant with CHD4 (von Zelewsky *et al*., 2000).
CHD3	*M.musculus*	mRNA expressed in the thymus, kidney, hemopoietic foci in the liver, specific areas of the brain, mucosal epithelium and in hair follicles (Kim *et al*., 1999).
Mi-2β	*D.melanogaster*	mRNA present in the embryo and mid to late stages of development. Mutant was embryonic lethal (Khattak *et al*., 2002).
CHD4	*C.elegans*	Essential for development and partially redundant with CHD3 (von Zelewsky *et al*., 2000).

CHD4	*M.musculus*	mRNA expressed in the thymus, kidney, hemopoietic foci in the liver, specific areas of the brain, mucosal epithelium and in hair follicles (Kim *et al*., 1999).
CHD5	*H.sapiens*	mRNA limited to neural-derived tissues (Thompson *et al*., 2003).
CHD6	*H.sapiens*	mRNA expressed in a variety of different cancer cell lines (Schuster and Stöger, 2002).
CHD7	*H.sapiens*	Ubiquitously expressed in 22d fetus, becoming more specific to the CNS later (Sanlaville *et al*., 2006). Important for neural crest cell formation and migration (Bajpai *et al*., 2010).
CHD7	*M.musculus*	Homozygous knockout is embryonic lethal whilst heterozygote recapitulates CHARGE symptoms (Hurd *et al*., 2007, Bosman *et al*., 2005). Inactivation of CHD7 in adult neural stem cells in mice reduces neurogenesis (Feng *et al*., 2013).
CHD7	*D.rerio*	Ubiquitously expressed in the embryo, with expression later in development persisting only in the brain and retina. Knockdown recapitulates CHARGE-like symptoms (Patten *et al*., 2012).
CHD8	*H.sapiens*	Required for the imprinted expression of *IGF2* (Ishihara *et al*., 2006). Modulates the β-catenin signalling response (Thompson *et al*., 2008).
CHD8	*M.musculus*	Expressed in a variety of tissues in the adult mouse (Ishihara *et al*., 2006). Duplin the N-terminal end of CHD8 is highly expressed during early mouse development and Duplin$^{-/-}$ mice die in early embryogenesis (Nishiyama *et al*., 2004).
CHD9	*H.sapiens*	Interacts with skeletal tissue-specific promoters (Shur *et al*., 2006). Osteogenic differentiation (Marom *et al*., 2006).

1.10 CHD function

1.10.1 CHD subfamily I

Early work showed despite having chromodomains like HP1, CHD1 did not bind heterochromatic regions (Stokes and Perry, 1995) and the *D.melanogaster* homologue was shown to localise to interbands and chromatin puffs (decondensed regions and sites of highly active transcription) in polytene chromosomes (Stokes *et al*., 1996). CHD1 also co-localises with elongating (phosphorylated at ser2) RNA polymerase II (McDaniel *et al*., 2008). *S.cerevisiae* Chd1 binds a variety of transcriptional elongation factors (Simic *et al*., 2003), fitting with the observed resistance to 6-azauracil that initially suggested an involvement in transcription (Woodage *et al*., 1997, Tsukiyama *et al*., 1999). A role for Chd1 in transcriptional termination has also been demonstrated (Alen *et al*., 2002).

Chd1 has been shown to be able to slide nucleosomes (Stockdale *et al*., 2006), assemble nucleosomes and form regularly spaced nucleosome arrays in vitro (Lusser *et al*., 2005, Pointner *et al*., 2012). Sliding requires linker DNA and slides nucleosomes towards the middle and not to the end of a piece of DNA (Stockdale *et al*., 2006, Hauk *et al*., 2010). The DNA-binding domain is thought not to be essential for sliding but determines the direction Chd1 slides a nucleosome (McKnight *et al*., 2011).

Genome-wide nucleosome mapping in *S.cerevisiae chd1*-nulls shows loss of nucleosome phasing within the gene body but the +1 nucleosome remains unaffected (Gkikopoulos *et al*., 2011). Similar results were found in *S.pombe* (Pointner *et al*., 2012). Limited changes in gene expression accompany nucleosome changes, although an increase in antisense transcription is observed in mutants for *chd1*, more so in double *isw1* and *chd1* mutants (Cheung *et al*., 2008, Smolle *et al*., 2012) and in the *S.pombe* homologues (Pointner *et al*., 2012, Shim *et al*., 2012). Chd1 is thought to be responsible for replacement and stabilisation of nucleosomes in the gene body after transcription, preventing cryptic transcription (Radman-Livaja *et al*., 2012, Smolle *et al*., 2012).

Early Chromatin ImmunoPrecipitation (ChIP)-chip studies in *S.cerevisiae* and *S.pombe* showed Chd1 binding at promoters and over the open reading frame of transcribed genes (Simic *et al*., 2003, Walfridsson *et al*., 2007). A more recent ChIP-chip experiment in mouse embryonic stem cells found binding at active promoters, which overlapped with polII binding (Gaspar-Maia *et al*., 2009). A ChIP-seq experiment in *S.cerevisiae* found Chd1 binding to NFR and gene bodies. They found no relationship with steady-state gene expression levels but did find a positive correlation of binding with histone turnover and active transcription rate suggesting the high histone turnover is a consequence of transcriptional elongation (Zentner *et al*., 2013). Although, another recent ChIP-exo (ChIP-seq with an exonuclease digestion step) study in *S.cerevisiae* was unable to determine binding sites (Yen *et al*., 2012).

The double chromodomain in human CHD1 binds the H3K4me mark suggesting a mechanism for recruitment to active genes. However chromodomains in *S.cerevisiae* Chd1 do not bind H3K4me (Sims *et al*., 2005) and *D.melanogaster* CHD1 chromodomains are not important for localisation to sites of active transcription but are required for enzymatic activity (Morettini *et al*., 2011). In *S.cerevisiae* it has been suggested that Chd1 is recruited to transcribed genes by its interaction with the PAF complex, which interacts with polII and regulates initiation and elongation (Simic *et al*., 2003). The crystal structure of *S.cerevisiae* Chd1 suggests that the chromodomains inhibit ATPase activity only enabling activation when a nucleosome is present rather than naked DNA (Hauk *et al*., 2010) with the former being the preferred substrate (Tran *et al*., 2000) having ~10 times higher ATPase activity (Hauk *et al*., 2010).

A wealth of information points to CHD1 being involved in transcriptional elongation. Other studies have shown it may have other roles. Human CHD1 has been shown to bind transcriptional repressors NCoR as well as histone deacetylases (HDAC) (Tai *et al*., 2003). Chd1 was also found to interact with components of the mRNA splicing complex, and overexpression of Chd1 affects alternate splicing (Tai *et al*., 2003). In humans CHD1 is also important for deposition of CENP-A at centromeres (Okada *et al*., 2009). *S.pombe* Chd1 (Hrp1) was shown to be involved in histone CENP-A placement at centromeres during early S-phase (Walfridsson *et al*., 2005), although ChIP-seq of Chd1 showed that it doesn't bind at the *S.cerevisiae* centromere (Zentner *et al*., 2013). CHD2, which is absent in *S.cerevisiae*, has remained understudied.

1.10.2 CHD subfamily II

One or both of CHD3 and CHD4 (*D.melanogaster* dMi2 α/β) are found in the NuRD (Nucleosome Remodelling and histone Deacetylase), a repressive complex that also has histone deacetylase (HDAC) activity (Tong *et al*., 1998, Wade *et al*., 1998, Zhang *et al*., 1998, Xue *et al*., 1998). The NuRD complex contains proteins that are associated with repressive functions: HDAC1, HDAC2, MBD3 (Methyl-CpG Binding Domain), RbAp46, RbAp48 (Retinoblastoma Associated proteins), MTA1, MTA2, MTA3 (Metastasis Associated proteins). It is thought that different NuRD complexes occur with different subunits present (Bowen *et al*., 2004). The ATPase activity of CHD3/4 in *D.melanogaster/X.laevis/H.sapiens* is stimulated by chromatin, not by DNA or histones alone (Wade *et al*., 1998, Brehm *et al*., 2000, Wang and Zhang, 2001). dMi-2 has been shown to bind nucleosomes and to be able to slide nucleosomes in vitro towards the middle of a piece of DNA (Brehm *et al*., 2000, Guschin *et al*., 2000). Subfamily II CHD proteins are characterised by their dual PHD domains. The PHD domains of CHD4 also bind histone H3 tails bivalently. Each PHD domain binds a separate H3 tail, and this binding is required for repression by CHD4 (Musselman *et al*., 2011). The PHD domains and the chromodomains both regulate ATPase activity of human CHD4 (Watson *et al*., 2012).

CHD5 has been observed in a NuRD-like complex (Potts *et al*., 2011). ChIP-seq revealed 60% of CHD5 binding sites occur within 2Kb of the transcription start site (TSS) (Paul *et al*., 2013). Mutations in the PHD domain prevented binding and repression of selected cancer progression genes (Paul *et al*., 2013). The PHD

domains of CHD5 have also been shown to bind unmodified histone H3 tails (Oliver *et al*., 2012, Paul *et al*., 2013).

1.10.3 CHD subfamily III

CHD6 was shown to be a DNA-dependent ATPase that co-localises with hypo- and hyperphosphorylated forms of RNA polymerase II (Schuster and Stöger, 2002, Lutz *et al*., 2006). Consistent with roles in transcriptional control CHD6 has been shown to act as a coactivator with Nrf2, a basic leucine zip transcription factor binding to antioxidant response elements (AREs) in the promoters of its target genes to activate transcription (Nioi *et al*., 2005). CHD6 also acts as a corepressor with HPV31 E8--E2C to repress the HPV E6/E7 promoter (Fertey *et al*., 2010).

CHD7 was shown to have ATP-dependent nucleosome sliding activity in vitro and requires DNA outside of the nucleosome to function (Bouazoune and Kingston, 2012). All of the three core domains are required for remodelling activity. Some remodelling activity was present without the BRK domain. CHD7 could still remodel nucleosomes without N-terminal histone tails, although at a lower rate than with tails. Mutations found in patients were recapitulated in vitro. Two missense mutations found in CHARGE patients in the chromodomain resulted in a modest decrease in remodelling activity. A third missense mutation found in CHARGE and idiopathic hypogonadotropic hypogonadism completely ablated remodelling (Bouazoune and Kingston, 2012).

CHD7-binding sites map preferentially to enhancers that regulate ESC-specific genes (Schnetz *et al*., 2010). CHD7 was suggested to have a modulating repressive function on a subset of ESC specific genes. Although, CHD7 is not completely repressive, as other gene sets are repressed in the absence of CHD7 (Schnetz *et al*., 2010). CHD7 binds the Sox2 transcription factor and has overlapping genomic binding sites and regulates the expression of common gene sets, including genes mutated in Alagille, Pallister-Hall, and Feingold syndromes (Engelen *et al*., 2012). CHD7 has also been shown to interact with the human PBAF complex (hSWI/SNF-β) and share a subset of genomic locations (Bajpai *et al*., 2010). 89% of co-occupied regions were within 1Kb from a TSS. Their binding is important for expression of *Sox9*, *Twist* and *Slug* transcription factors (Bajpai *et al*., 2010). Interestingly CHD7 has also been shown to bind CHD8 and missense mutations in the binding region found in CHARGE syndrome interrupt this interaction (Batsukh *et al*., 2010, Batsukh *et al*., 2012).

CHD8 has been shown to interact with the transcriptional repressor/insulator protein CTCF by yeast two-hybrid (Y2H), and pull-down experiments showed CTCF selectively bound the BRK domain of CHD8 (Ishihara *et al*., 2006). CHD8 binding was demonstrated at *BRAC1*, *c-myc* and *H19* insulator loci. A luciferase assay utilising the *H19* insulator sequence showed CHD8 was required for enhancer-blocking activity. Knockdown of CHD8 caused reactivation of the endogenous *IGF2* gene on the maternal chromosome, suggesting that CHD8 is involved in the function of the *H19* DMR insulator for the imprinted expression of *IGF2* (Ishihara *et al*., 2006).

CHD8 also binds β-catenin and β-catenin responsive genes (*axin2*, *c-myc*, *DKK1*) (Thompson *et al*., 2008), after stimulation of the Wnt pathway (by lithium or Wnt3A), recruiting histone H1, which downregulates the response (Nishiyama *et al*., 2012). CHD8 only binds upon stimulation of the pathway, suggesting it modulates sensitivity to the to Wnt/β-catenin signalling (Nishiyama *et al*., 2012).

CHD9 has been shown to have DNA-dependent ATPase activity and to be able to bind AT-rich DNA (Shur and Benayahu, 2005). CHD9 has been show to bind to tissue-specific promoters (Shur *et al*., 2006) and interact with GR (Glucocorticoid receptor) (Marom *et al*., 2006) and PPARalpha (peroxisome-proliferator activated receptor) (Surapureddi *et al*., 2006). Studies to date have been limited to certain tissues.

1.11 CHD family in disease states

The maintenance and dynamic control of gene expression is important for development, differentiation and responding to external cues. As such perturbations in the regulation of gene expression result in unwanted consequences. Germ line mutations in transcriptional regulators can result in developmental disorders and diseases, whereas somatic mutations resulting in gene expression changes can cause unwanted growths and malignancies.

Cancer research initially focused on single mutations such as point mutations, amplifications or deletions in genes in pathways essential for tumour progression (insensitivity to external cues, cell cycle progression, apoptosis)

(Wang *et al*., 2007b). It is becoming clear that alterations in gene expression through histone modification and chromatin remodelling play important roles in tumour progression (Wang *et al*., 2007a, Wang *et al*., 2007b). Large cancer genome sequencing projects (The Cancer Genome, Wellcome Trust; The Cancer Genome Atlas, NIH) have highlighted the frequency of somatic mutations in epigenetic regulators in cancer (Dawson and Kouzarides, 2012). Small molecule inhibitors of epigenetic modifiers are an active area of investigation for cancer treatment, with some already on the market (DNMT, Azacytidine; HDAC inhibitors, Vorinostat), although the current approach provides little specificity and the mechanism is somewhat contentious (Dawson and Kouzarides, 2012, Minucci and Pelicci, 2006).

Roles for CHD proteins are beginning to emerge in cancer. CHD1 has been identified as the 5q21 tumour suppressor, a region deleted in 13–26% of prostate cancer (Liu *et al*., 2011, Huang *et al*., 2011, Burkhardt *et al*., 2013), and mutations of various CHDs (CHD1, CHD2, CHD3, CHD4, CHD7, CHD8) have been observed in gastric and colon cancers (Kim *et al*., 2011). The CHD5 gene, which lies in the short arm of human chromosome 1, is frequently deleted in neuroblastoma (Brodeur *et al*., 1977, Brodeur *et al*., 1981, Okawa *et al*., 2008) and has been found to be mutated (Gorringe *et al*., 2008, Lang *et al*., 2011, Robbins *et al*., 2011) or silenced (Wang *et al*., 2009) in a number of different cancers. Deletions of CHD5 in mice confirmed its role as a tumour suppressor, as heterozygous mice displayed spontaneous tumourogenesis (Bagchi *et al*., 2007). The PHD domains' interaction with unmodified H3 was shown to be essential for

its tumour-suppressor function (Paul *et al*., 2013). CHD7 mutations were also found in a form of small-cell lung cancer (Pleasance *et al*., 2010).

CHD5, unlike other CHDs, was observed expressed only in the brain rather than a wide range of tissues (Potts *et al*., 2011). A link to Alzheimer's disease was suggested when knockdown of CHD5 in rats resulted in misregulation of genes associated with aging and Alzheimer's (Potts *et al*., 2011). Mutations in CHD8 have been identified as a risk factor for autism spectrum disorders (O'Roak *et al*., 2012, Neale *et al*., 2012). 1.2% of epileptic encephalopathies, a group of epilepsies with poor prognosis, are caused by mutations in *CHD2* (Carvill *et al*., 2013).

Perhaps the most well-known disease associated with CHD proteins is CHARGE syndrome, a severe developmental disorder associated with mutations in CHD7 (Vissers *et al*., 2004). CHARGE is an acronym of the symptoms that define the syndrome: **C**oloboma of the eye, **H**eart defects, **A**tresia of the choanae, **R**etardation of growth, **G**enital and **E**ar abnormalities including hearing loss (Pagon *et al*., 1981). However, not all patients display all symptoms. CHD7 is mutated in 60–80% of CHARGE patients (Lalani *et al*., 2006, Janssen *et al*., 2012). $CHD7^{-/-}$ mice are embryonic lethal (E10.5) and $CHD7^{+/-}$ recapitulates some of the CHARGE symptoms (Hurd *et al*., 2007, Bosman *et al*., 2005). This is in line with the human disorder being the result of a haploinsufficiency (Bartels *et al*., 2010). Mutations are distributed throughout the gene, the majority being nonsense or frameshift mutations; ~8% are missense (Janssen *et al*., 2012). Bouazoune and Kingston (2012) showed that missense mutations found in CHARGE patients

resulted in loss of nucleosome-remodelling activity to varying degrees in vitro, although the authors only studied three mutations in the conserved chromodomain. Mutations are scattered throughout the gene (http://www.chd7.org). Mutations elsewhere in the protein could reduce CHD7 action through protein stability or inhibiting binding of accessory proteins. Mutations in CHD7 have also been observed in Kallmann syndrome (KS) (Jongmans *et al*., 2009) and idiopathic hypogonadotropic hypogonadism (IHH). It has been suggested IHH/KS represents a milder allelic variant of CHARGE syndrome (Kim *et al*., 2008). *CHD7* polymorphisms in humans have also been linked to scoliosis (Gao *et al*., 2007).

Table 1.2: Summary of human diseases associated with CHD proteins

	Disease	**Reference**
CHD1	Prostate cancer	(Liu *et al*., 2011, Huang *et al*., 2011, Burkhardt *et al*., 2013)
CHD2	Epileptic encephalopathy	(Carvill *et al*., 2013)
CHD3/4	Dermatomyositis	(Seelig *et al*., 1995)
CHD5	Neuroblastoma	(Brodeur *et al*., 1977, Brodeur *et al*., 1981, Okawa *et al*., 2008)
CHD7	CHARGE syndrome	(Pagon *et al*., 1981, Vissers *et al*., 2004)
	Kallmann syndrome and idiopathic hypogonadotropic hypogonadism	(Jongmans *et al*., 2009, Kim *et al*., 2008)
	Small-cell lung cancer	(Pleasance *et al*., 2010)

1.12 *Dictyostelium discoideum* as a model organism

Dictyostelium discoideum, the social amoeba, is a non-metazoan eukaryote, discovered by Kenneth Raper in 1933 (Raper, 1935) in a forest in North Carolina. *Dictyostelium* are in the Amoebazoa supergroup, division Mycetozoa (Baldauf *et al*., 2000) that is thought to have split from metazoans after the split from plants (Figure 1.5). *Dictyostelium* are soil-dwelling amoeba that feed on bacteria in the wild. When nutrients are readily available, *Dictyostelium* grows as a single-celled organism but upon nutrient depletion cells aggregate together to form a multicellular organism (Gaudet *et al*., 2008). Initially upon starvation cells signal each other using extracellular cyclic AMP (cAMP) to aggregate together to form a mound of cells. These cells then begin to undergo a differentiation programme making different specialised cell types. The mass of cells undergo several defined morphogenic stages, which ultimately forms the fruiting body structure (Figure 1.6). A fruiting body can contain up to ~200,000 cells and is composed of a spore head containing desiccated spores, which is supported by the stalk and a basal disc that attaches the structure to the substratum (Figure 1.6). Strains that can also grow in axenic culture (nutrient-rich media without bacteria) have been produced (Sussman and Sussman, 1967, Watts and Ashworth, 1970) and are now widely used over non-axenic strains, and the fast (24 hours) developmental cycle can be reproduced in the laboratory.

Dictyostelium discoideum has been actively studied for almost 60 years and is one of the National Institutes of Health (Bethesda, Maryland, USA) model organisms for biomedical research. Its simple developmental cycle and haploid genome

allow for easy manipulation of its genetics and dissection of biological problems (Gaudet *et al*., 2008). Uncoupled growth and developmental stages allow the study of both independently, and a mutation affecting the latter doesn't affect the efficiency of isolating and propagating the mutant. *Dictyostelium* has a close phylogenetic relationship with metazoa, and possesses much of the complexity seen in metazoan cell signalling and cell biology (Williams and Harwood, 2003).

Dictyostelium is genetically tractable and provides a wealth of molecular biology tools. Homologous recombination occurs at high frequency (De Lozanne and Spudich, 1987, Williams, 2010). This combined with its haploid genome and LoxP-cre system (Faix *et al*., 2004) makes multiple knockouts and knock-ins readily available (Faix *et al*., 2013). Studies on *Dictyostelium* have enabled breakthroughs in various fields; it enabled early studies on actin mechanics (Clarke *et al*., 1975), presented the first descriptions of a eukaryotic cell chemoattractant (Konijn *et al*., 1967) and a cell–cell adhesion protein (Muller and Gerisch, 1978), and was the first eukaryote in which individual gene transcription was visualised in living cells (Chubb *et al*., 2006b).

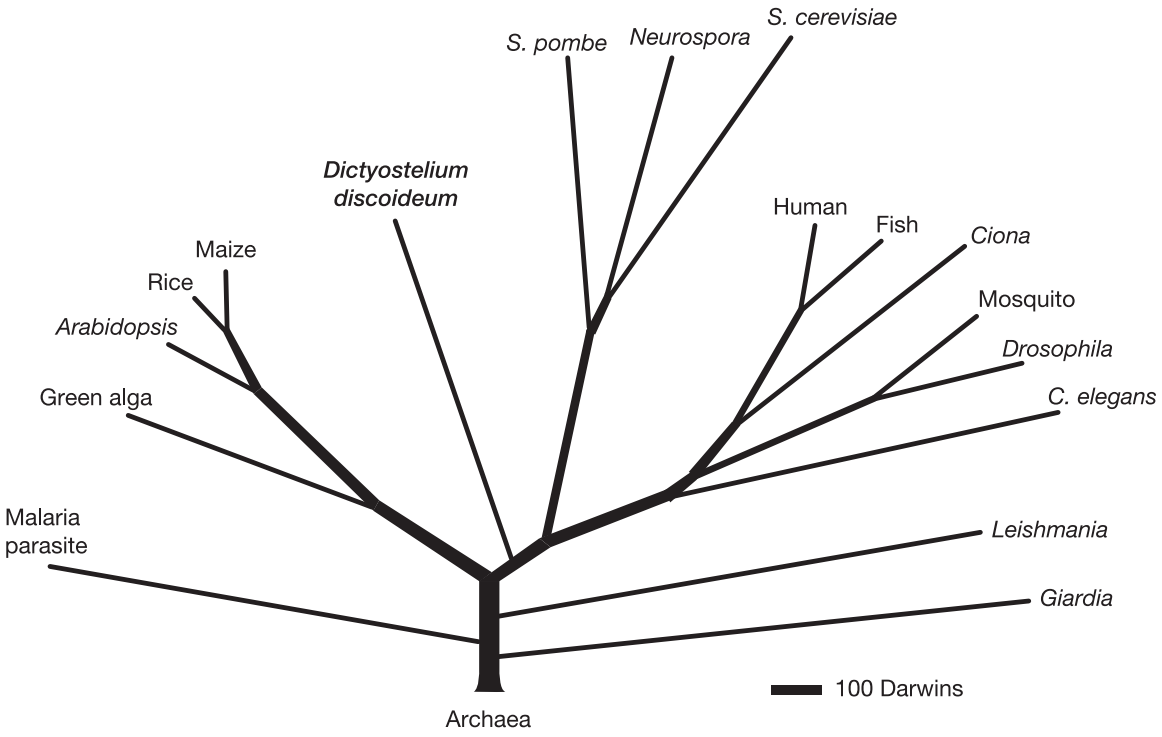
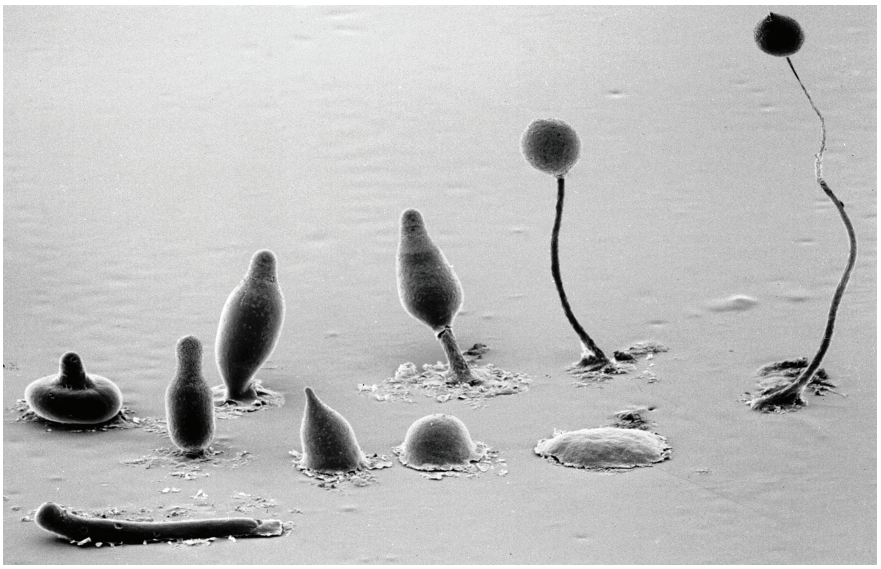


Figure 1.5: Eukaryotic phylogenetic tree

A phylogenetic tree based on the proteome showing the estimated divergence of *Dictyostelium discoideum*. This demonstrates *Dictyostelium* has a closer relationship to animals than plants (Eichinger et al., 2005).

A



B

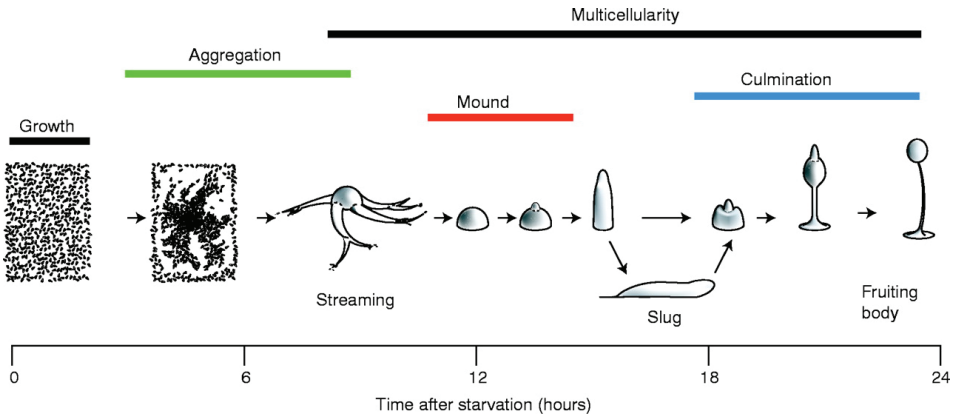


Figure 1.6: The development life cycle of *Dictyostelium discoideum*

A. Electron microscopy pictures show the *Dictyostelium* development cycle. Upon starvation cells aggregate together to form a mound that then forms what is known as a finger, which falls over to form the slug, which will migrate for a period of time until it forms a second mound. This then forms the fruiting body, which is made up of a spore containing head, a stalk that supports the spore head and a basal disc that contacts the substratum. Copyright M.J. Grimson & R.L. Blanton Biological Sciences Electron Microscopy Laboratory, Texas Tech University. **B.** A schematic diagram showing the approximate times of the developmental cycle from the point of starvation in wild-type cells (Coates and Harwood, 2001).

1.13 *Dictyostelium discoideum* genomics

Dictyostelium discoideum has a fully sequenced 34Mbp genome (Eichinger *et al*., 2005) currently thought to encode ~12,750 protein-coding genes (Gaudet *et al*., 2011), spread across six chromosomes and contains an extrachromosomal ribosomal RNA palindrome. *D.discoideum* was the first of the dictyostelids to have its genome sequenced; subsequently *Dictyostelium purpureum* (Sucgang *et al*., 2011), *Dictyostelium fasciculatum* and *Polysphondylium pallidum* (Heidel *et al*., 2011) have been sequenced, which allows cross-species comparisons (Basu *et al*., 2013).

D.discoideum has a haploid genome, although due to short interphase and G_1 stages of the cell cycle, most of the time is spent in G_2 with a duplicated genome (Muramoto and Chubb, 2008). Its haploid genome and ability for high-frequency homologous recombination was exploited early (De Lozanne and Spudich, 1987, De Lozanne, 1987). The *Dictyostelium* genome is gene rich, containing a large number of genes with similarity to those of higher metazoans. In fact it contains a greater number of related human disease genes than yeast (Urushihara, 2009). Its simple life cycle and genetics coupled with characteristics of higher metazoans make it ideal to study conserved cellular, biochemical and genetic processes present in higher eukaryotes.

The genome of *D.discoideum* is AT rich. Overall it has 22.4% GC content; in coding regions this is higher at 27% but drops to 15% in intergenic regions and 12% in introns. Genes have a mean length of 1.7Kb and a mean intergenic length

of ~750bp. 73% of genes are spliced; of these genes there is an average of 1.9 introns per gene and a mean length of 146bp per intron (Eichinger *et al*., 2005).

Studies of *Dictyostelium* chromatin have been limited in the past. More recently the completion of the genome sequence and new technologies have created interest in these studies in *Dictyostelium*. The completion of the genome has shown the *Dictyostelium* genome to contain well-conserved homologues of many chromatin proteins (Eichinger *et al*., 2005). *Dictyostelium* contains genes for the canonical histones (H2A, H2B, H3 and H4) and the linker histone H1 and various versions of histones including H2AZ. *Dictyostelium* has many histone modifiers, including histone methyltransferases (HMTase), histone demethylases, histone acetyltransferases (HAT) and histone deacetylases (HDAC). Modification of the N-terminal tails of histones was reported comprehensively by Stevense *et al*. (2011). The histone tails in *Dictyostelium* contain many of the same modifications found in other eukaryotes (Figure 1.7), including methylation of H3K4, H3K9, H3K36 and H3K79 acetylation of H3K9, H3K14, H3K18, H3K23 and H3K27. One notable absence is H3K27me found in higher eukaryotes (Kouzarides, 2007), although *Dictyostelium* also lacks the polycomb complex, which binds H3K27me (Stevense *et al*., 2011).

Studies of these histone-modifying enzymes have shown them to be important regulators of developmental gene expression. Levels of methylation of H3K4 have been shown to change over development both globally and in developmentally regulated genes (Chubb *et al*., 2006a). Levels of the H3K4me3 mark, associated with active genes in other organisms (Santos-Rosa *et al*., 2002),

correlate with the expression of *acaA* (adenylate cyclase) and *rasG* genes (Chubb *et al*., 2006a). Set1, the HMTase required for methylation of H3K4, was shown to be important in regulating development in *Dictyostelium*. Null mutants of Set1 displayed precocious aggregation, entering development much earlier than wild type, although only a limited number of genes were misexpressed (Chubb *et al*., 2006a). H3K4 methylation is required for the inheritance of transcriptional state (Muramoto *et al*., 2010). Centrosomes have been shown to be enriched for the heterochromatin mark H3K9me3 (Dubin *et al*., 2010). H3K79me2 was shown to be present throughout development. Loss of dot1, the HMT responsible for H3K79 methylation, resulted in delayed development and increased susceptibility to UV damage, implicating its role in DNA damage (Muller-Taubenberger *et al*., 2010) is conserved in *Dictyostelium* (Nguyen and Zhang, 2011). Although the *Dictyostelium* genome encodes eight putative histone acetyltransferases (*gcn5*, *DDB_G0275017*, *DDB_G0275159*, *DDB_G0279699*, *elp3*, *DDB_G0274269*, *taf1* and *rbbD*), currently none have been investigated. Histone deacetylase (HDAC) function has been investigated. Treatment with the HDAC inhibitor Tricostatin A (TSA) results in a hyperacetylation of histones, a slight developmental delay and a reduction in sporulation efficiency (Sawarkar *et al*., 2009).

DNA methylation, which was once thought to be absent in *Dictyostelium* (Smith and Ratner, 1991), has been shown to be added by the dnmt2 DNA methyltransferase dnmA (Kuhlmann *et al*., 2005, Katoh *et al*., 2006). Although methylation levels are very low, in the range of 0.1–0.2% of cytosines being methylated, comparable to *Drosophila melanogaster* (Gowher *et al*., 2000) but

low in comparison to animals (2–10%) (Razin and Riggs, 1980, Ehrlich and Wang, 1981). Methylation was shown to be developmentally regulated and involved in repressing retrotransposons, but developmental phenotypes were mild (Katoh *et al*., 2006) and transcriptional changes were limited (Kuhlmann *et al*., 2005).

Until now ATP-dependent chromatin remodelling has remained relatively under-studied in *Dictyostelium*. Previously three Chd family remodellers were identified (Rogers, 2010, Platt *et al.,* 2013), named *chdA* (*DDB_G0284171*), *chdB* (*DDB_G0280705*), and *chdC* (*DDB_G0293012*). Sequence analyses also identified other related *Dictyostelium* genes encoding ATP-dependent chromatin remodellers, members of the SWI2/SNF2 (*snf2a*/*DDB_G0285205* and *snf2b*/*DDB_G0271052*), ISWI (*isw*/*DDB_G0292948*), and INO80 (*ino80*/*DDB_G0292358* and *swr1* like protein *DDB_G0267638*) chromatin remodelling families. The *Dictyostelium* homologues cluster appropriately with their counterparts in yeast and humans (Figure 1.8).

Early attempts to map nucleosomes in *Dictyostelium* resulted in difficulties. The genome was not as easy to use as other organisms for nucleosome mapping, possibly due to the AT-rich nature of its genome. Despite these difficulties some studies were carried out. *Dictyostelium* was shown to have nucleosomes like other eukaryotes and a repeat length of ~170bp (Parish *et al*., 1977). Other initial studies focused on the chromatin structure of the ribosomal DNA palindrome (Ness *et al*., 1983, Edwards and Firtel, 1984), an 88Kb region present in multiple copies in each cell, containing the 17S, 25S and 5S rDNA (Maizels,

1976, Cockburn *et al*., 1978). Studies using MNase and radiolabelling found that the actively transcribed regions were less structured than non-transcribed regions (Ness *et al*., 1983). Using indirect-end labelling an accurate map of an ~10Kb region near the telomere of the rDNA was created, showing evidence of irregularly spaced nucleosomes (Edwards and Firtel, 1984). Other work on single-copy genes interpreted results based on smearing of gels. Nucleosomes were not mapped as the experiments did not utilise end labeling (Blumberg *et al*., 1991, Pavlovic *et al*., 1989). With improved protocols (Kent and Mellor, 1995) and new techniques (Platt *et al*., 2013, Kent *et al*., 2010) nucleosome mapping in *Dictyostelium* has faced renewed interest. Chang *et al*. (2012) mapped nucleosomes globally using high-throughput sequencing and showed that *Dictyostelium* has a 162–169bp repeat length and similar chromatin structure around protein coding genes as in other eukaryotes.

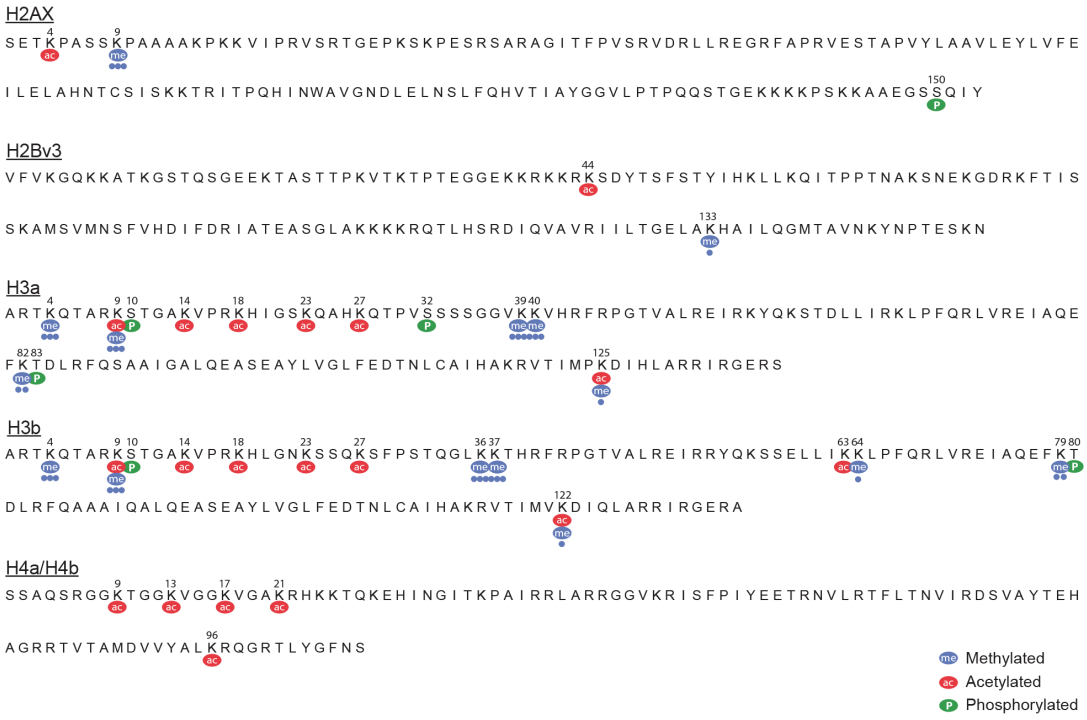


Figure 1.7: Modification of N-terminal histone tails in *Dictyostelium*

Post-translational histone modifications identified in *Dictyostelium* using mass spectrometry. Numbers of dots under the methylated lysine residues represent mono-, di- or tri-methylation (Stevense *et al*., 2011).

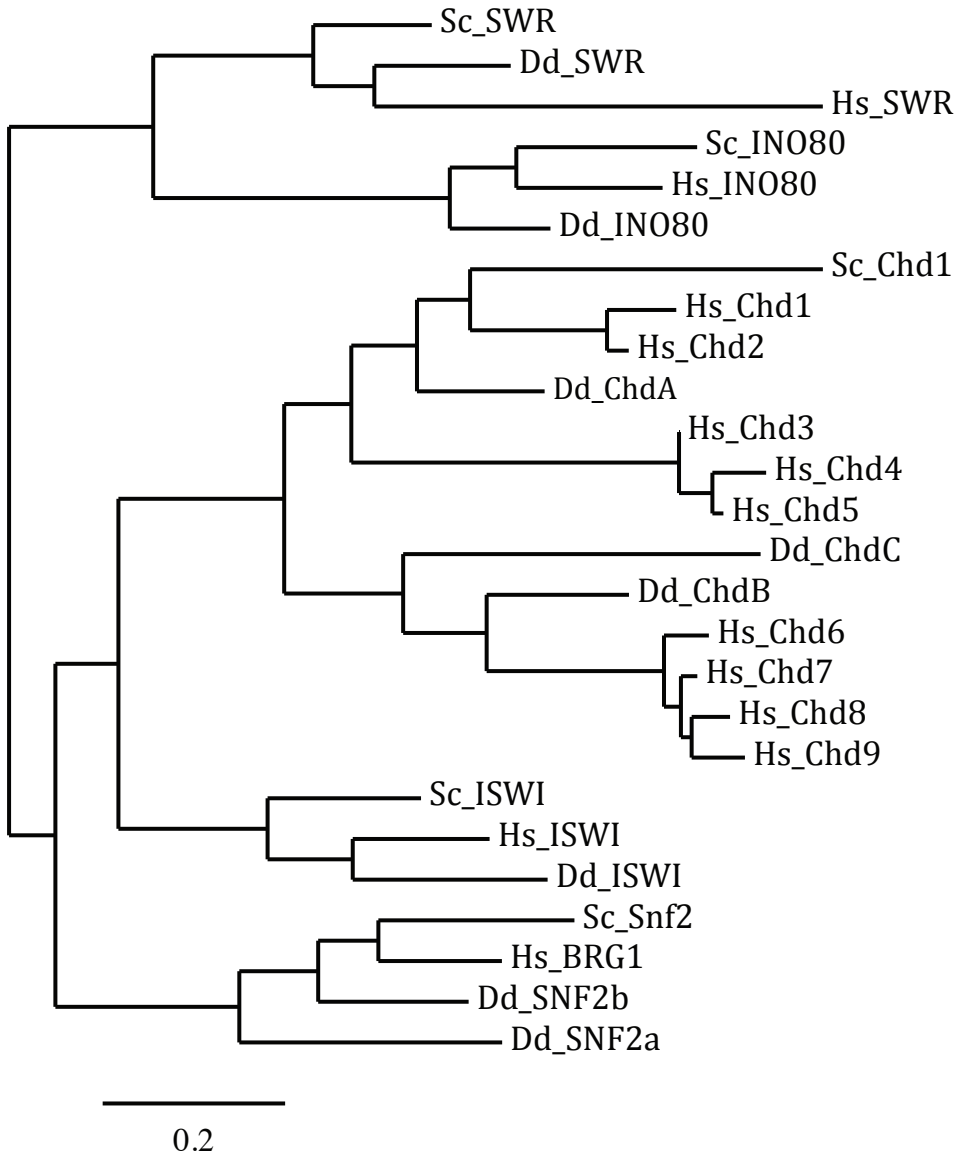


Figure 1.8: Phylogenetic tree of *Dictyostelium* ATP-dependent chromatin remodellers SWI/SNF, ISWI, CHD and INO80

A phylogenetic tree showing the clustering of *Dictyostelium* homologues of SWI/SNF, ISWI, CHD, INO80 and SWR1 proteins. Abbreviations Dd – *Dictyostelium discoideum*, Sc – *Saccharomyces cerevisiae*, Hs – *Homo sapiens*. Created using Phylogeny.fr (Dereeper *et al.*, 2008).

1.14 Chd family in *Dictyostelium*

There are three *chd* genes in *Dictyostelium*, which have been named *chdA, chdB* and *chdC*. The first member (*chdC*) of the *Dictyostelium* Chd family was discovered independently in separate insertional mutagenic screens for lithium resistance and multicellular development. Bioinformatics searches identified a further two *chd* genes encoded in the *Dictyostelium* genome: *chdA* and *chdB*.

Dictyostelium ChdA shows the strongest similarity to human CHD subfamily I, clustering with CHD1/2 through the entire protein sequence, including the C-terminal DUF4208 domain (Figure 1.9). *Dictyostelium* ChdB possesses weak C-terminal CHDCT2 domains, which are sometimes present in CHD subfamily II. However, ChdB lacks PHD finger domains characteristic of subfamily II members and clusters better with human CHD6-9 (Figure 1.9). ChdC is the largest *Dictyostelium* Chd protein, comprising >3,000 amino acids. ChdC has strongest sequence homology to the human CHD6-9 group. Although ChdC has a C-terminal SANT and SLIDE DBD motifs as in other subfamily III members, it appears to lack a conserved BRK site. The BRK domain is of unknown function, but is a presumed marker of complex metazoan development (Allen *et al*., 2007). ChdC may uniquely possess a SUMO-like amino acid sequence of unknown function. Collectively, the individual Chds of *Dictyostelium* would seem to embody three different Chd proteins.

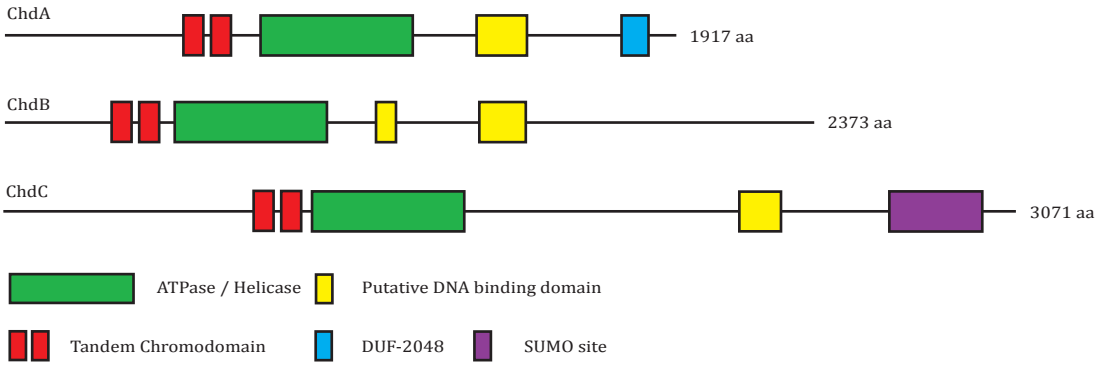


Figure 1.9: *Dictyostelium* Chd proteins

Schematic diagram showing the domain layout of the three Chd proteins in *Dictyostelium*.

1.15 Aims of this study

Chromatin and its regulation are the conserved processes essential for life, hence ATP-dependent chromatin remodellers are well conserved from the simplest to the most complex eukaryotes (Varga-Weisz, 2001). Previous studies have shown members of the CHD family to be regulators of nucleosome positioning, gene expression and development. However, the interplay between different CHD family members is not yet understood. There have been some reports of interactions between remodellers (Batsukh *et al*., 2010) or redundancy (von Zelewsky *et al*., 2000). Much focus has been on CHD1 due to its ubiquity in eukaryotes; as such other family members have taken a back seat. Many of the epigenetic modulators present in higher eukaryotes are conserved in *Dictyostelium* (Stevense *et al*., 2011). The aim of this study is to generate null mutants for each of the *chd* genes in the simple eukaryote *Dictyostelium discoideum*, determining the effect and therefore role in multicellular development. Utilising high-throughput sequencing this study will also look at the Chd proteins' role in regulating gene expression and nucleosome positioning. *Dictyostelium* fills a gap in the complexity scale, with a greater level of complexity and larger family of CHD proteins than *S.cerevisiae* whilst possessing a compact manipulatable genome. This study should advance our understanding of CHD proteins and their specificity.

Chapter 2:

Materials and Methods

2.1 General *Dictyostelium* maintenance

Dictyostelium were grown axenically in HL-5 medium (14g/L peptone, 7g/L yeast extract, 13.5g/L glucose, 0.5g/L KH_2PO_4, 0.5g/L Na_2HPO_4, Formedium, Norfolk, UK) supplemented with 100μg/ml Streptomycin-sulfate at 22°C for normal culture and with the addition of other selection antibiotics where appropriate. Cells were also grown on a lawn of *Klebsiella aerogenes* on SM agar plates (10g/L peptone, 1g/L yeast extract, 10g/L glucose, 1.9g/L KH_2PO_4, 1.3g/L $K_2HPO_4.3H_2O$, 0.49g/L $MgSO_4$, 17g/L Agar, Formedium, Norfolk, UK).

The Ax2 strain was used as the wild-type strain and is the parent of all mutants. All strains were replaced every 4 weeks from frozen stock. Cells were frozen in HL-5 medium with 8% DMSO.

Cells were washed and starved in KK2 buffer (2.25g/L KH_2PO_4, 0.51g/L K_2HPO_4) or on KK2 agar (KK2 buffer, 18g/L Agar). Cells were pelleted by centrifugation at 440 x g for 2 minutes.

2.2 Genomic DNA extraction

For use as PCR template, $2x10^7$ cells were pelleted and washed twice in KK2 before being resuspended in 750μl DNAzol (Life Technologies, California, USA), then 375μl of 100% ethanol was added to precipitate the DNA on ice for 10 minutes. After precipitation, centrifugation at 16,000 x g, the DNA pellet was

washed once in 70% ethanol, air dried and resuspended in 50μl 8mM NaOH and 7.5μl 0.1M HEPES buffer (pH 7.5).

For quick extraction of DNA for screening of transformed clones, cells grown in a 96-well plate were resuspended in the media in the well. 10μl was transferred to a PCR tube and 10μl lysis buffer (50 mM KCl, 10 mM TRIS pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP40, 0.45% tween 20 and 0.8μg/μl proteinase K) was added. After 2 minutes at room temperature, proteinase K was inactivated by heating at 95°C for 1 minute.

2.3 General cloning

All plasmids constructed were screened by restriction digest using approximately 1μg DNA and restriction enzymes from either Fermentas/Thermo Scientific (Massachusetts, USA) or New England Biolabs (Massachusetts, USA). Digested plasmids were analysed by agarose gel electrophoresis. Fragments were cloned and sequenced to ensure fidelity when necessary. Ligations were carried out at room temperature for 1 hour using T4 DNA ligase (New England Biolabs, Massachusetts, USA) with a 3:1 insert–vector molar ratio.

2.4 Generation of *chd* knockout constructs

Two regions for each *chd* gene were amplified from genomic DNA using Acuzyme proofreading polymerase (Bioline, London, UK), the first 5' of and close to the ATG start codon, the second ~1Kb downstream from the first. The design, as a consequence of a double crossover during homologous recombination, would result in the replacement of the ATG start codon with a 1Kb cassette containing a blasticidin resistance gene, which will abolish any translation of the mRNA into protein. The PCR primers (table 2.1) were designed to incorporate SalI and BglII sites into the 3' end of the 5' fragment and 5' end of the 3' fragment. Equal amounts of the two PCR products were mixed and a second PCR was performed using the outermost primers to join the two fragments together. The resulting ~2Kb fragment was then ligated into the TOPO blunt II vector (Life Technologies, California, USA). The vector was then cut at the central SalI and BglII sites and the LoxP flanked blasticidin cassette, which was cut with XhoI and BamHI from pLPBLP (Faix *et al*., 2004, Faix *et al*., 2013), was ligated in. The three plasmids produced were pCR-Blunt II-TOPO-chdA+Bsr, pCR-Blunt II-TOPO-chdB+Bsr and pCR-Blunt II-TOPO-chdC+Bsr. Constructs for *chdA* and *chdB* were excised from the vector using EcoRI, *chdC* with NotI and HindIII, then sodium acetate (3M, pH 5.2)/ethanol precipitated and resuspended in water before transformation into *Dictyostelium*.

Table 2.1: Primers used to create knockout constructs

In bold, restriction sites. Underlined, overlap region.

Gene	Primer	Sequence 5' to 3'
chdA	5' Fragment 5' primer	AAGAAAAATTAAGAAAACATATCCATGAA
	5' Fragment 3' primer	**AGATCT**TAC**GTCGAC**AACAATATCAAACTAATGAACATATCAAAG
	3' fragment 5' primer	**GTCGAC**GTA**AGATCT**GAATCAGAATCAGATTCAGATTATGAAC
	3' Fragment 3' primer	TAAACTCTCATAAGCATCCCATGTATT
chdB	5' Fragment 5' primer	GTTCATCATTTCTTTTTTTGGATG
	5' Fragment 3' primer	**AGATCT**TAC**GTCGAC**ATTCGATTTGTGTGAGAAAAGAAA
	3' fragment 5' primer	**GTCGAC**GTA**AGATCT**GATCCAATTGAAAGAATTAAAAAAGAA
	3' Fragment 3' primer	TTAGTGATCTTTCCACTAGCACCA
chdC	5' Fragment 5' primer	TTCCACAAATATCTTAAAAATGTAATCA
	5' Fragment 3' primer	**AGATCT**TAC**GTCGAC**TTCTTGTAATTAAATTTTCTTCTTCCC
	3' fragment 5' primer	**GTCGAC**GTA**AGATCT**GCTGATGAAAAGAAACCCAATAAT
	3' Fragment 3' primer	AAAGAAGGTGGTACAAATTGAGGT

2.5 Transformation of *Dictyostelium*

Constructs were transformed by electroporation as previously described (Gaudet *et al*., 2007). Briefly, $1x10^7$ cells were washed twice in cold electroporation buffer (KK2, 50mM sucrose) before finally being resuspended in 800μl volume and transferred to a 4mm electroporation cuvette (Cell Projects, Kent, UK). For integration plasmids 15μg of DNA was added to the cells and incubated on ice for 10 minutes. Then cells were electroporated at 1kV followed by another incubation on ice for 10 minutes. 8μl of salt solution (0.1M $MgCl_2$, 0.1M $CaCl_2$) was then added to the cells and left at 22°C for 15 minutes in a 10cm petri dish, before adding HL5 media containing only 100μg/ml streptomycin-sulfate.

After 24 hours, transformations of integrating plasmids were cloned into 5 x 96-well plates in the presence of 10μg/ml blasticidin S. After 10–14 days clones were observed in wells, clones were screened (see section 2.6) and positive clones transferred to a larger plate.

2.6 Screening knockout clones

Blasticidin-resistant clones were screened by performing PCR on genomic DNA extracted using a proteinase K method (see section 2.2). PCR was performed with GoTaq polymerase (Promega, Wisconsin, USA), using primers that targeted sequences inside the knockout region (table 2.2). PCR products were then run on

a 1% agarose gel. The absence of a band indicated a positive clone. Positive knockout clones were confirmed using western blotting.

Table 2.2: Primers used in screening

Gene	Primer	Sequence 5' to 3'
chdA	5' primer	GTATACACCACCTCCGTTACAACAA
	3' primer	TGAATATGTTATTCTCTTTGCCTGTC
chdB	5' primer	GGGTAGGTAAAATTATTCACTATCG
	3' primer	TCATCACTTTCATCGATTGTATTAT
chdC	5' primer	CAAGAATGACTGTTTCAAAGGTATT
	3' primer	AATGATTGATCATCGTCACTCTTTA

2.7 Western blotting

Knockout strains were also confirmed by western blotting. Cells were harvested and washed twice in KK2 before lysis in NuPAGE LDS-sample buffer (Life Technologies, California, USA) supplemented with 5% β-mercaptoethanol and protease inhibitors (Roche, Basel, Switzerland) before boiling for 10 minutes to denature the proteins. Proteins from $5x10^5$ cells were loaded per well on a NuPage 3–8% Tris-Acetate gel (Life Technologies, California, USA). Western blotting of proteins onto a nitrocellulose membrane was carried out using standard methods. The CHD proteins are large proteins. To ensure transfer of

proteins from the SDS-PAGE gel to the nitrocellulose membrane blotting was carried out at a low voltage (15mV) and low temperature (4°C) overnight. The membrane was blocked in 5% milk in TBS/T (tris-buffered saline, 0.1% tween 20) incubated with primary antibodies in 5% milk solution for 1 hour at room temperature, washed and incubated with the appropriate HRP-linked secondary antibodies in TBS/T for a further hour. The membrane was incubated with Pierce ECL western blotting substrate (Thermo Scientific, Massachusetts , USA) for 5 minutes. Membranes were then exposed in the GeneGnome imager (Syngene, Cambridge, UK). When required, intensities of bands were calculated using the GeneTools software (Syngene, Cambridge, UK).

Antibodies used were polyclonal anti ChdA and ChdB antibodies that were raised in rabbit (70-day protocol) by immunising with peptides from the C-terminal end of both proteins and affinity purified with the original peptides (Openbiosystems/Thermo Scientific, Massachusetts , USA). Polyclonal anti ChdC antibodies were a gift from Dr Benjamin J. Rogers. Specificity of the antibodies were ensured by testing against cell lines containing the protein of interest, cells containing a TAP (Tandem Affinity Purification)-tagged version of the protein and knockout cells (see appendix A).

2.8 Analysis of development by time lapse

$5x10^6$ cells were washed twice in KK2 and placed on KK2 agar in a 6-well plate, allowed to settle for 15 minutes and the residual KK2 buffer was aspirated off. Time-lapse movies were created using an Olympus 1X71 (Tokyo, Japan) inverted

microscope with an automated stage taking images every 1–2 minutes for 20–24 hours and stitching individual images together with ImageJ (Schneider *et al*., 2012).

2.9 Development on nitrocellulose filters

Cells were washed twice in KK2 and placed on a 0.45μm hydrophilic nitrocellulose filter (Merk-Millipore, Massachusetts, USA) in 10μl spots and images were taken at key points in development using the Zeiss Stemi SV11 (Oberkochen, Germany) stereo microscope.

2.10 Analysis of slug migration

Cells were washed twice in KK2 and placed in a line at one end of a 0.45μm hydrophilic nitrocellulose filter and placed in a slug can, a device that ensures *Dictyostelium* cells are only exposed to a small amount of unidirectional light. Upon development during the migratory slug stage, the slug migrates towards the light leaving a trace of where it has been. The filters were left in the slug can for 48 hours, after which images of the whole filter were taken.

2.11 Analysis of chemotaxis towards cAMP

Twice-washed cells were placed in KK2 and shaken for 5 hours whilst being subjected to 100nM pulses of cAMP every 6 minutes to induce a chemotactic state. After 5 hours cells were added to the Zigmond chamber (Zigmond, 1988)

(Neuro Probe, Maryland, USA) where a cAMP gradient is established using 1μM cAMP (in KK2) in one dam and KK2 in the other dam. The Zigmond chamber was left for 15 minutes for the gradient to establish before analysis. Cells were imaged on an Olympus 1X71 inverted microscope using DIC (Differential Interference Contrast) in 6 fields of view simultaneously. An image was captured every 10 seconds for 15 minutes. DIAS 3.4.1 software (Soll Technologies Inc, Iowa, USA) was used to analyse the data (Heid *et al*., 2005, Wessels *et al*., 2006). Each cell was outlined in each frame and the path the cell took was determined. Average speed, directionality and chemotactic index (Kay *et al*., 2008) were calculated for each cell over the 15 minutes. A minimum of 100 cells for each cell type were analysed over three separate experiments. Statistical significance was determined using the Kruskal-Wallis test (Kruskal and Wallis, 1952) and Mann-Whitney test (Mann and Whitney, 1947).

2.12 DAPI staining

For analysis of nuclei number per cell, cells were seeded onto poly-L-lysine-coated coverslips and allowed to settle and adhere for 15 minutes. Then fixed in 4% formaldehyde in phosphate-buffered saline (PBS). Coverslips were then washed once in PBS and mounted on slides using Vectorshield mounting medium (Vector Labs, California, USA) mixed with 1μg/ml 4',6-Diamidino-2-phenylindole dihydrochloride (DAPI). Images were then taken using a Zeiss Axiovert 200M (Oberkochen, Germany) confocal microscope. Nuclei number were then counted in >200 cells for each cell type.

2.13 Bioinformatic analysis of ATP-dependent remodellers

Identification of ATP-dependent chromatin remodellers in *Dictyostelium* was achieved by searching dictybase (Basu *et al*., 2013, http://dictybase.org/) and using both blastp and tblastn (Camacho *et al*., 2009) searches of the *Dictyostelium* genome. Phylogenetic trees were created using phylogeny.fr software aligning full-length protein sequences (Dereeper *et al*., 2008).

2.14 Gene expression analysis

2.14.1 RNA extraction and sequencing

RNA was extracted from $5x10^7$ cells at different developmental time points, using TRIzol reagent (Life Technologies, California, USA) following the manufacturer's instructions. For each cell line and time point RNA was extracted from two cell populations as a technical repeat. RNA quality was checked on the Bioanalyzer 2100 (Agilent Technologies, California, USA) prior to library preparation. Libraries for the Illumina (California, USA) next-generation sequencing were then produced using the RNA extracted, following the manufacturer's instructions. Briefly, 1μg of total RNA was polyA+ enriched using poly(dT) magnetic beads. The mRNA was fragmented using alkaline hydrolysis and reverse transcribed to create cDNA. Adaptors were then ligated to the ends of the dsDNA. PCR with primers reciprocal for the adaptor sequences was used to enrich for dsDNA with adaptors ligated to both ends. Libraries were then sequenced on the Illumina Genome Analyzer IIx or the Hiseq2000 sequencing

platform, at the NIDDK genomics core (NIH, Bethesda, Maryland, USA) generating 36–50bp sequences from one (single-end) end of the sequence. The sequences were passed through the CASAVA v1.7 pipeline to filter out erroneous reads.

2.14.2 Bioinformatic analysis of gene expression

Reads passing quality control were aligned to the genome using gene models downloaded from dictybase (Basu *et al*., 2013) with Tophat (Trapnell *et al*., 2009, http://tophat.cbcb.umd.edu/). Reads that align to genes were counted using HTSeq (Anders and Huber, 2010, http://www-huber.embl.de/users/anders/HTSeq/) and differential expression between samples was determined using the R (R Development Core Team, 2012) package DESeq (Anders and Huber, 2010, http://www-huber.embl.de/users/anders/DESeq/). GO term analysis of the differentially expressed genes was carried out using Orange (Curk *et al*., 2005, http://orange.biolab.si/) or Cytoscape (Smoot *et al*., 2011) with the BiNGO plugin (Maere *et al*., 2005). When referring to expression patterns of individual genes through development, this is either taken from cited literature or the dataset present on dictyexpress (Rot *et al*., 2009, http://www.ailab.si/dictyexpress,). Comparison of gene expression boxplots were drawn in R. Gene expression heatmaps were clustered and rendered in R using heatmap.2 function in the gplots package (Warnes, 2012).

2.15 Chromatin structure analysis

2.15.1 Chromatin digestion – preparation of nucleosomal DNA

Chromatin digestion was carried out using a modification of the basic method (Kent and Mellor, 1995) described in detail in Platt *et al*., 2013. $1x10^8$ cells at different time points in development were washed on ice once in KK2 and once in 100mM sorbitol and resuspended in 400μl of digestion buffer (100mM sorbitol, 50mM NaCl, 10mM Tris-HCl (pH 7.5), 5mM $MgCl_2$, 1mM $CaCl_2$, 1mM β-mercaptoethanol, 0.5mM spermidine, 0.1% Nonidet P-40), a permeabilisation buffer allowing micrococcal nuclease (MNase, Affymetrix/USB, California, USA) to penetrate the cell and nuclear membranes. Chromatin was digested with MNase at varying concentrations (0–300 units) for 2 minutes at 37°C. 40μl stop solution was added (5% SDS, 250mM EDTA) to lyse the cells and terminate MNase activity. DNA was then extracted using a phenol/chloroform (sample:phenol:chloroform 2:1:1 ratio) extraction. The supernatant was transferred to a new tube containing 15μl RNAse A (10mg/ml) and incubated for 30 minutes at 37°C. DNA was then phenol/chloroform extracted and sodium acetate/ethanol precipitated and resuspended in TE buffer (10mM TRIS pH 8.0, 1mM EDTA). 10% of the final sample was run on a 1.5% agarose gel to check the quality of the digest. Two technical replicate samples were combined and run on a 1.2% agarose gel and DNA in the range of 50bp to 1,000bp was excised from the gel. The agarose was placed in 0.45μm cellulose acetate spin column (Sigma-Aldrich, Missouri, USA), and freeze-thawed (10 minutes, -20°C to room temperature) and then spun at 16,000 x g for 10 minutes to isolate the DNA. The

DNA solution was phenol/chloroform extracted and treated with 100 units T4 polynucleotide kinase (New England Biolabs, Massachusetts, USA) for 30 minutes at 37°C to remove 3' phosphoryl groups left by MNase. DNA was extracted once more with phenol/chloroform, precipitated with sodium acetate and ethanol, washed with 80% ethanol, dried and re-suspended in TE (pH 7.5). 5µg of DNA was processed following the manufacturer's instructions to produce libraries to sequence. Briefly, double-stranded DNA (dsDNA) was end-repaired and adaptors were then ligated to the ends of the dsDNA. PCR with primers reciprocal for the adaptor sequences were used to enrich for dsDNA with adaptors ligated to both ends. Libraries were then sequenced on the Illumina Hiseq2000 sequencing platform, at the NIDDK genomics core (NIH, Bethesda MD, USA) generating 76bp sequences from both ends (paired-end) of the sequence. The sequences were passed through the CASAVA v1.7 pipeline to filter out erroneous reads.

Two control digests were carried out and sequenced identically to the chromatin digestions. Previously isolated genomic DNA (naked DNA) was re-suspended in digestion buffer and incubated with 0–5U MNase at 25°C for 30 seconds. Naked DNA was also sonicated to a range of fragments found in the MNase digest. DNA from both controls was cleaned up, isolated in the 50–1,000bp region and sequenced as above.

2.15.2 Bioinformatic analysis of nucleosome positions

Informatics involved in the elucidation of nucleosome positions was described in Kent *et al*., 2010. Scripts (SAMparser.pl, SAMhistogram.pl, SiteWriter_Full_10bin_1.pl and ClusterWriter_Full.pl) used to map nucleosomes were modified from there to suit the *Dictyostelium* genome.

Reads were trimmed to 36bp and aligned to the genome using Bowtie v1.7.3 (Langmead *et al*., 2009), http://bowtie-bio.sourceforge.net/). Reads were aligned using a maximum insert size of 5,000 (-maxins 5000), allowing zero mismatches in the seed sequence (-n 0), reporting only one alignment per read (-k 1), reporting the hits of the highest quality/stratum (-best), bowtie outputted aligned reads in SAM format (--sam).

Perl script SAMparser_dicty.pl was used to parse reads from the SAM file output by bowtie selecting reads that are spaced 150bp ±30bp apart, representing the size of a nucleosome. The mid-point genomically between the two reads was determined and was taken to be where the nucleosome dyad was. SAMhistogram-dicty.pl then tallies all the nucleosome dyads across the genome to produce a histogram along the genome showing where the dyads of nucleosomes were in the population of cells studied. SAMhistogram-dicty.pl produces an .sgr file which can be loaded into Integrated Genome Browser (Nicol *et al*., 2009, http://bioviz.org/igb/) allowing nucleosome positions to be visualised.

SiteWriter_Full_10bin_1.pl was used to produce CFD graphs (Cumulative Frequency Distribution) it requires the .sgr file and a second file that lists the details (chromosome, position, strand) of features of interest – Start codon, Stop codon, etc. SiteWriter_Full_10bin_1.pl parses the .sgr file for regions surrounding (e.g. +/- 1Kb) the features of interest combining dyad frequencies in each bin, which was normalised to total dyad frequencies across the entire region (+/- 1Kb of feature), the resulting data was rendered in R. ClusterWriter_Full.pl works similarly to SiteWriter_Full_10bin_1.pl except features/genes are not combined instead are outputted onto separate lines of a file that can be clustered by Cluster 3 (de Hoon *et al*., 2004), heatmaps were then produced using Java Treeview (Saldanha, 2004).

The normalised CFD graph was produced by dividing nucleosome dyads by 'dyads' derived from MNase digested naked DNA and then passing the data through SiteWriter_Full_10bin_1.pl. ATGCN-content.pl calculates A,T,G,C,N content surrounding features listed, using a 15bp window.

Differentially positioned nucleosomes between two cell lines were calculated using PeakMarkCompare.pl (Nicolas Kent), which determines nucleosome positions and heights and comparing them to those of another nucleosome map. A nucleosome position difference was called when non-matching nucleosome summits of >2 fold or >10bp mismatch in peak summit position. Genes were defined as having different nucleosome structure if three or more nucleosomes were differentially positioned in the first 1000bp of the gene in both biological

repeats. The significance of the overlap between two gene lists was determined using the hypergeometric distribution test (Fury *et al*., 2006).

Chapter 3:

Exploring the roles of the Chd family in *Dictyostelium discoideum* development

3.1 Introduction

Two separate genetic screens in *Dictyostelium* isolated a mutant for a CHD chromatin remodeller, one for lithium resistance, and another for late development mutants. Previous sequence searches of the *Dictyostelium* genome discovered the presence of a further two *chd* genes. These three genes were named *chdA* (*DDB_G0284171*), *chdB* (*DDB_G0280705*) and *chdC* (*DDB_G0293012*). *chdC* was the original gene isolated in the two independent screens. The mutant isolated in the Harwood lab (Cardiff University, Wales, UK) via a REMI (Restriction-Enzyme Mediated Integration) screen was termed REMI-LisG and found to be lithium resistant, chemotaxing towards cAMP normally in the presence of lithium (Keim-Reder, 2006). Subsequent verification of the mutant by using homologous recombination to produce a null mutant, ΔLisG/*chdC*-null, surprisingly displayed a different phenotype (Rogers, 2010). The *chdC*-null chemotaxed poorly in a gradient of cAMP, however when treated with lithium chemotaxis was fully restored. Development was arrested at mound stage not producing fruiting bodies. A separate *chdC*-null mutant was isolated in the Kimmel Lab (National Institutes of Health, Maryland, USA) that displayed the identical phenotype with mound arrest, defects in cell patterning and differentiation (Platt *et al*., 2013).

Dictyostelium provides a unique system to study CHD proteins as it contains a small number (three) of CHD proteins, has a simple development cycle and has a small (34Mb) genetically amenable genome. To study the CHD family I produced

knockout mutants of each of the *chd* genes in *Dictyostelium*. Their roles in development and gene expression were subsequently investigated.

3.2 Chd family of ATP-dependent chromatin remodellers in *Dictyostelium*

Dictyostelium has three *chd* genes that encode proteins that were named ChdA, ChdB and ChdC. ChdA shows the strongest similarity to human CHD subfamily I, clustering with CHD1/2 through the entire protein sequence, including the C-terminal DUF4208 (Figure 1.8 and 1.9). *Dictyostelium* ChdB possesses weak C-terminal CHDCT2 domains, which are generally characteristic of CHD subfamily II. However, ChdB lacks the N-terminal PHD finger present in most subfamily II members and globally aligns better with human CHD6-9 proteins (Figure 1.8). ChdC is the largest *Dictyostelium* CHD, comprising 3,071 amino acids. ChdC has the strongest sequence homology to the CHD6-9 group (Figure 1.8). Although ChdC has a C-terminal SANT and SLIDE DBD motifs as in other subfamily III members, it lacks a conserved BRK domain (Figure 1.9). The BRK domains, of unknown function, are presumed markers of complex metazoan development (Allen *et al*., 2007). The BRK domain is also found in human BRG1, a homologue of *S.cerevisiae* SWI/SNF that lacks a BRK domain but is considered functionally equivalent (Ryan and Owen-Hughes, 2011). ChdC also possesses a SUMO-like amino acid sequence of unknown function. The individual Chds of *Dictyostelium* would seem to embody different CHD class types.

3.3 Developmental expression of the Chd family in *Dictyostelium*

To investigate if the expression of *chd* genes is developmentally regulated, publicly available RNA-seq data was used (dictyexpress.org, Rot *et al*., 2009). The three *chd* genes have separate expression patterns throughout development in *Dictyostelium discoideum* (Figure 3.1). Performing BLAST searches on the other recently sequenced Dictyostelid genomes, *Dictyostelium purpureum* (Sucgang *et al*., 2011), *Dictyostelium fasciculatum* and *Polysphondylium pallidum* (Heidel *et al*., 2011), identified homologues of all three *D.discoideum chd* genes, demonstrating their common ancestor possessed three *chd* genes 0.6 billion years ago. RNA-seq data for *D.purpureum* (Parikh *et al*., 2010) shows as well as having three *chd* genes, these genes display strikingly similar expression patterns to those in *D.discoideum* (Figure 3.1). Notably the patterns are delayed in *D.purpureum,* fitting with observations that aggregation is ~4 hours delayed in *D.purpureum* but produces fruiting bodies within the 24-hour period (Parikh *et al*., 2010).

To study the expression of Chd proteins through development antibodies were raised towards the three Chd proteins in *D.discoideum*. *Dictyostelium* were developed on nitrocellulose and whole cell proteins were extracted every two hours for a 24-hour period. Western blotting was then used to monitor the protein levels through development (Figure 3.2). Relative levels of expression were calculated from the intensities of the bands from western blotting (Figure 3.2). As found with the RNA-seq data, Chd protein expression throughout development varies, demonstrating Chd proteins are developmentally expressed.

All genes are expressed in growing cells: ChdB and ChdC peaks in expression earlier in development with expression declining during late development; ChdA is expressed throughout development, peaking late in development. The protein levels through development show similarities with mRNA levels previously observed. Differences between mRNA levels and protein levels might be explained by differences in translation rates, mRNA stability, protein stability and other post-transcriptional processes (Schwanhausser *et al*., 2011). RNA-seq data was also taken from Ax4 cells grown on bacteria two variables which may affect gene expression (Rot *et al*., 2009).

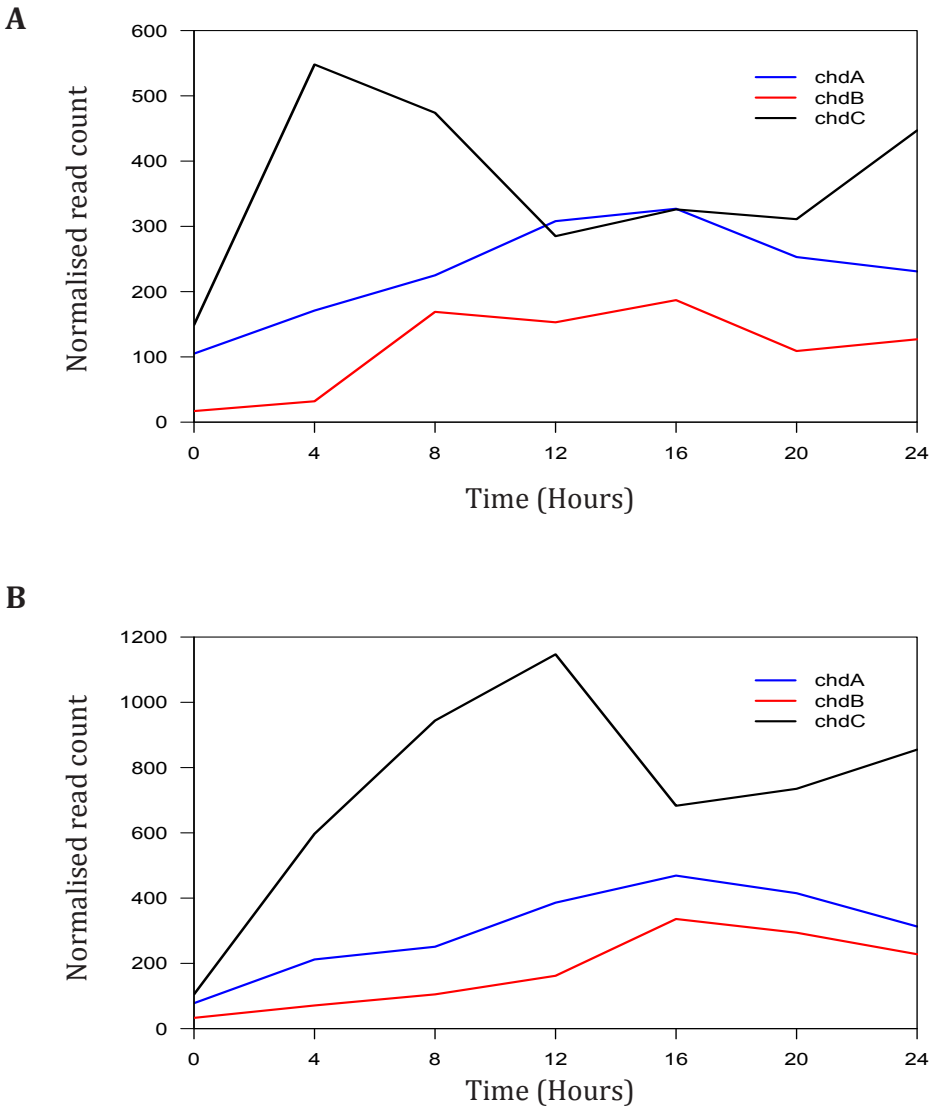


Figure 3.1: Expression of *chd* RNA throughout development in *Dictyostelium discoideum* and *Dictyostelium purpureum*.

A. RNA levels from dictyexpress.org (Rot *et al.*, 2009) for *chdA*, *chdB* and *chdC* in *Dictyostelium discoideum* throughout 24 hours of development. **B.** RNA levels from dictyexpress.org for *DPU_G0057338/chdA*, *DPU_G0071826/chdB* and *DPU_G0062114/chdC* in *Dictyostelium purpureum* throughout 24 hours of development.

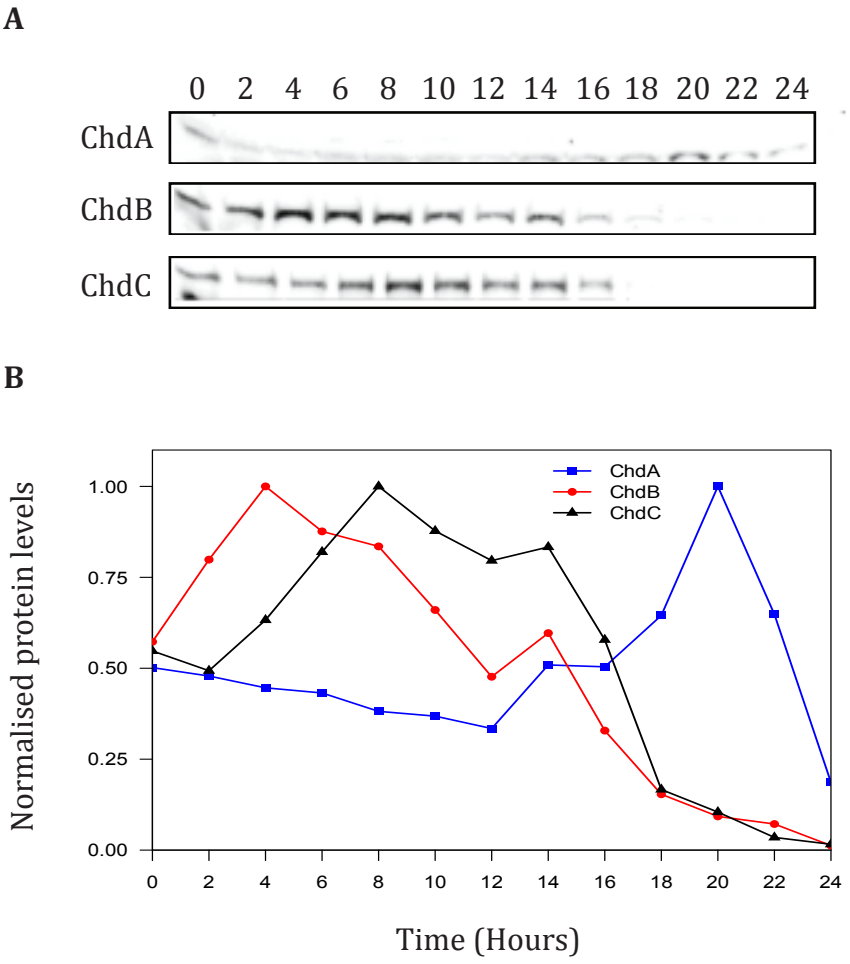


Figure 3.2: Chd protein levels throughout development in *Dictyostelium discoideum*.

A. Whole cell extract was prepared from an equal number of developing cells at two-hour intervals for 24 hours. Extracts were run on an SDS-PAGE gel and probed with antibodies towards each individual Chd protein. **B.** Quantified intensities of bands were calculated and normalised to the highest expression for each Chd protein.

3.4 Generating *chd*-null mutants

In order to study the normal roles of each Chd protein in *Dictyostelium*, three separate gene knockout cell lines were produced. Null mutants were produced by directed homologous recombination, leading to the insertion of a blasticidin selectable marker cassette that is flanked by LoxP sites (Faix *et al*., 2013). The knockout constructs were designed to remove a ~1Kb region of the start of the gene, which includes the ATG translational start site (Figure 3.3). Clones were originally screened by PCR and later confirmed by western blotting using the antibodies from above. *chdA*-null and *chdB*-null were new null mutants. Two previous *chdC*-null mutants were isolated in the Ax3 background (Kimmel lab) and Ax2 background (Harwood lab) using separate strategies. Here a third *chdC*-null mutant in Ax2 background was created using a third knockout strategy described above. Three separate nulls were created for *chdA* and six for *chdB* to ensure consistency of initial phenotype and to rule out secondary mutations.

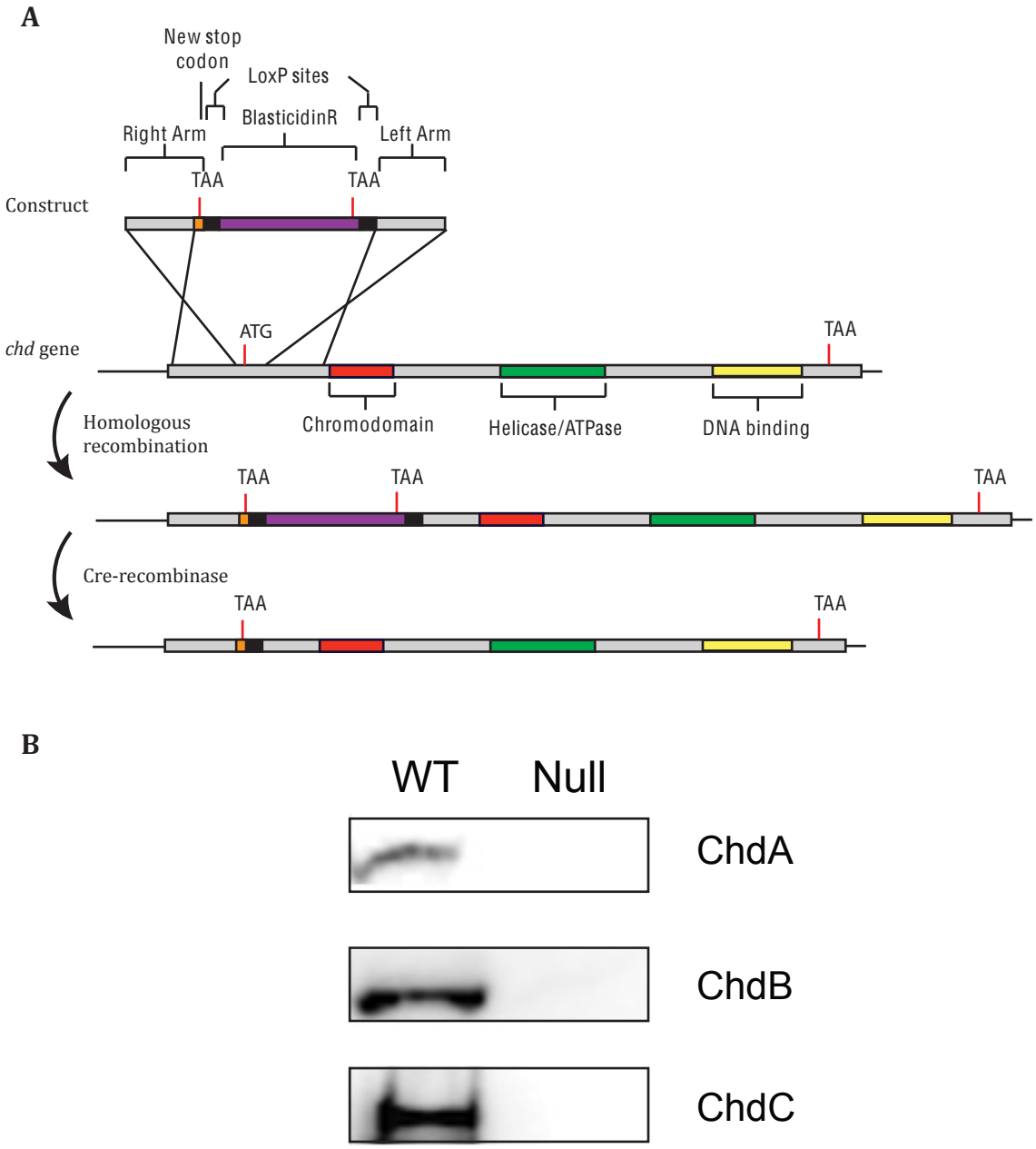


Figure 3.3: Generation of *chd*-null strains.

A. Schematic representation of the knockout construct and strategy used to produce null strains with use of homologous recombination, inserting a blasticidin resistance cassette. The blasticidin cassette can be removed using cre-recombinase enabling the creation of multiple knockouts. **B.** Western blotting on whole cell extracts from growing cells confirming the knockout of *chdA*, *chdB* and *chdC*.

3.5 The roles of Chd proteins in *Dictyostelium* growth

To study how the loss of Chd proteins impacted on normal cell growth in *Dictyostelium*, cells were counted over a time course to calculate doubling time (Figure 3.4). Ax2 cells in axenic media (HL5 + glucose, Formedium) has a doubling time of ~10 hours, *chdA* and *chdB*-null cells had wild type doubling times of approximately 11 hours each, whereas *chdC*-null cells had a severe growth delay with a doubling time of 22 hours. To determine whether the growth defect was due to impaired cytokinesis, *chdC*-null cells were stained with DAPI to determine whether they were multi-nucleate. Counting revealed no unusual number of multi-nucleate cells compared to Ax2 cells (Figure 3.4).

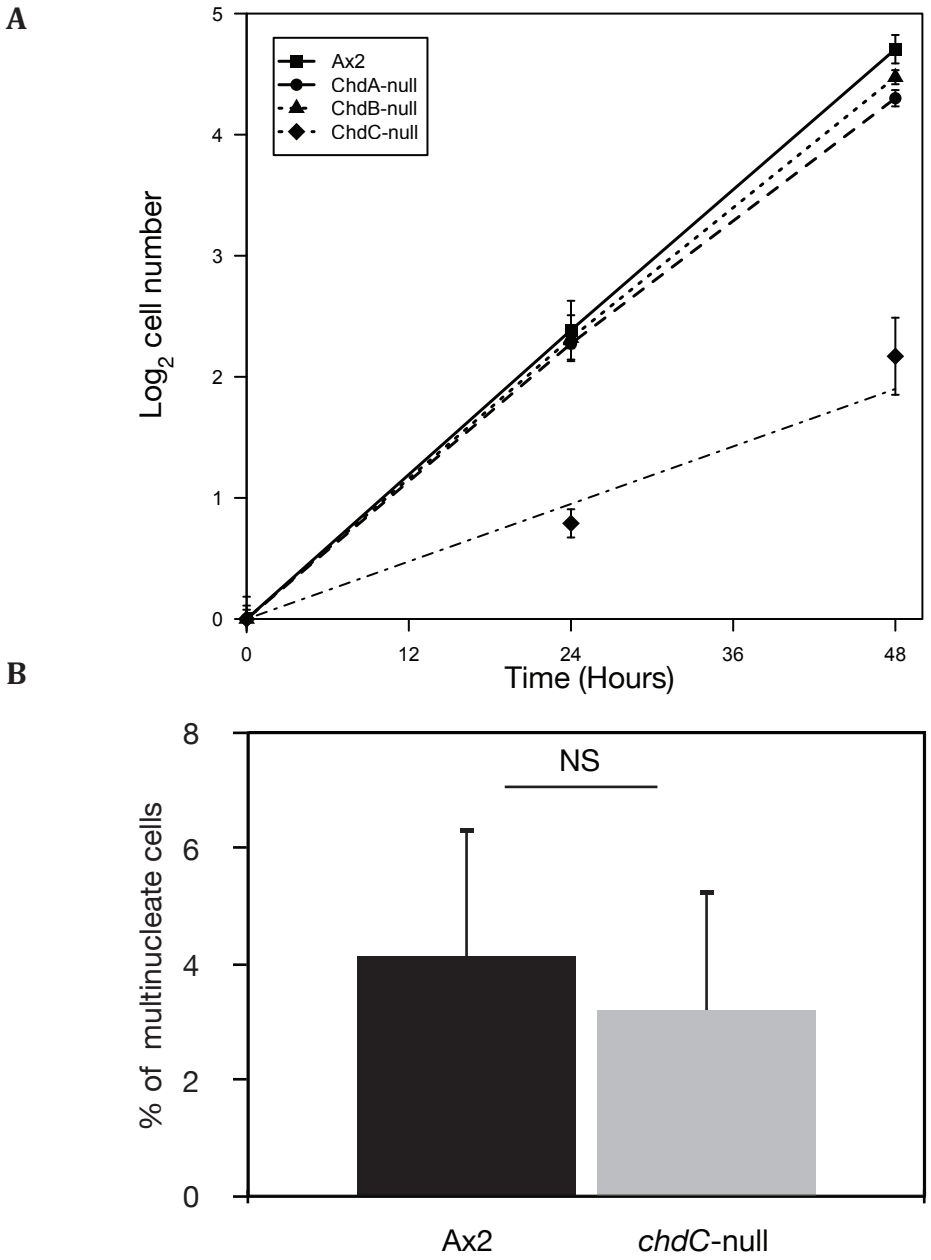


Figure 3.4: Growth rates of *chd*-nulls

A. Ax2, *chdA*, *chdB* and *chdC*-null cells were grown in log phase growth and counted at various times to evaluate growth rate. Increase in cell number was normalised to the initial cell density and plotted on a $\log_2$ scale ± standard deviation from three seperate experimental repeats. **B.** *chdC*-null cells were fixed and stained with DAPI to count the number of nuclei per cell. Bar chart shows percentage of cells with two or more nuclei per cell ± standard deviation from three seperate experimental repeats. Mann-Whitney test for significance was used.

3.6 The roles of Chd proteins in *Dictyostelium* development

Earlier it was shown that *chd* mRNA and protein levels are developmentally regulated, suggesting they are important for development. To address the question of whether Chd proteins regulate development, cells were developed on nutrient deficient KK2 agar plates (Figure 3.5).

chdA-null cells displayed a slight delay on entering development of ~1-2 hours and a similar delay on completing development. They also appeared to form longer slugs and stay in the migratory slug stage for long periods (Figure 3.5, 14 Hrs). To further address the *chdA*-null slug phenotype, cells were developed on nitrocellulose filters whilst subjected to a small amount of unidirectional light in a slug can. Under these conditions Ax2 cells migratory slugs are inhibited and culminate, whereas *chdA*-nulls form slugs at high frequency, which migrate towards the light (Figure 3.5). This increased propensity to form migratory slugs suggests *chdA*-nulls have a culmination defect.

chdB-nulls display a heterochrony phenotype with a delay upon entering development of ~3-4 hours (Figure 5.3, 18–22 Hrs). The rest of development appeared to be normal.

chdC-nulls display a short delay upon entering development, taking ~2 hours longer than wild-type to aggregate. During aggregation *chdC*-nulls also fail to form clear streams (Figure 3.5, 8 Hrs) that also fail to break up, resulting in large

mounds, which never progress past the mound stage producing no fruiting bodies (Figure 3.5, 12–48 Hrs).

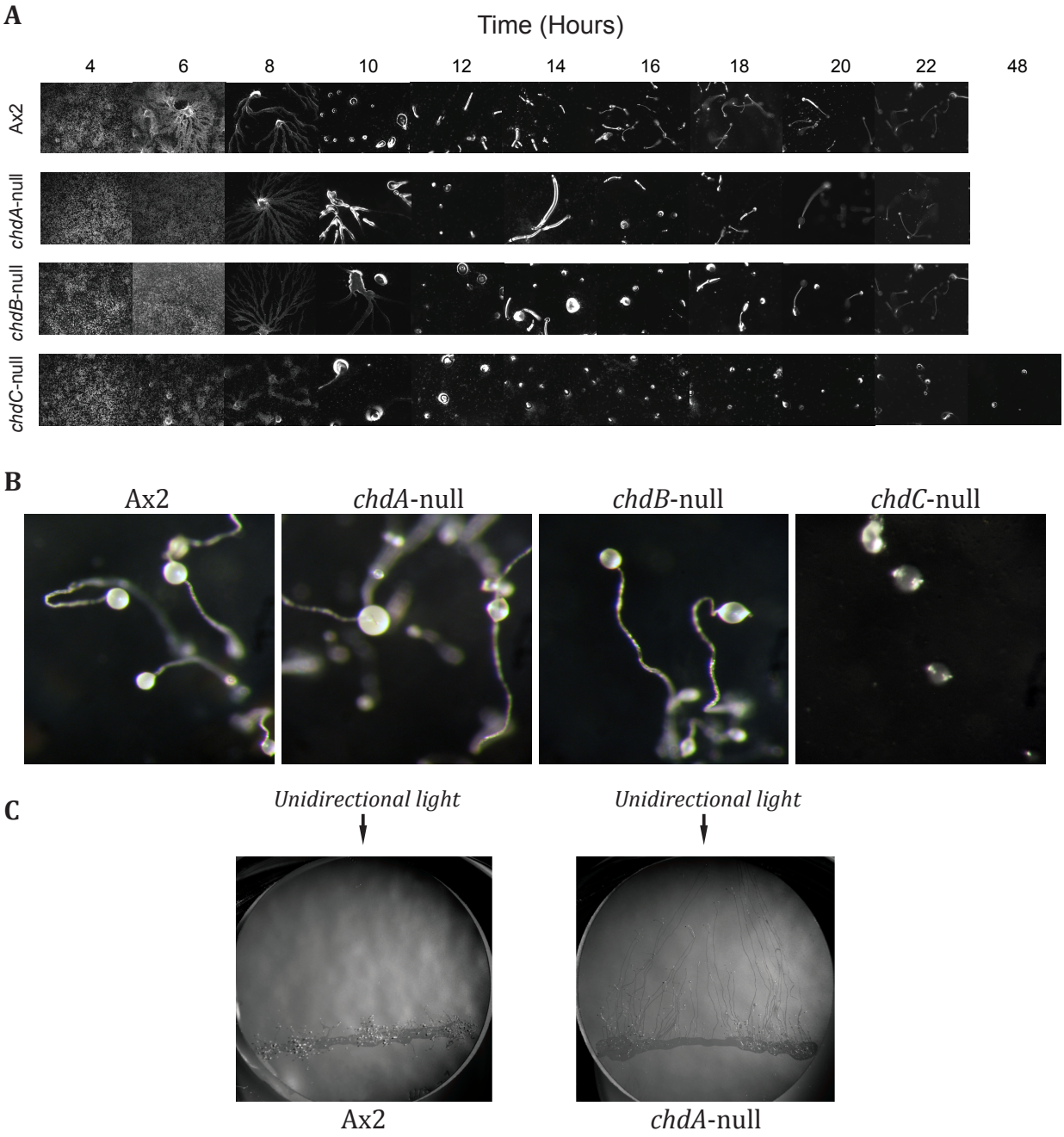


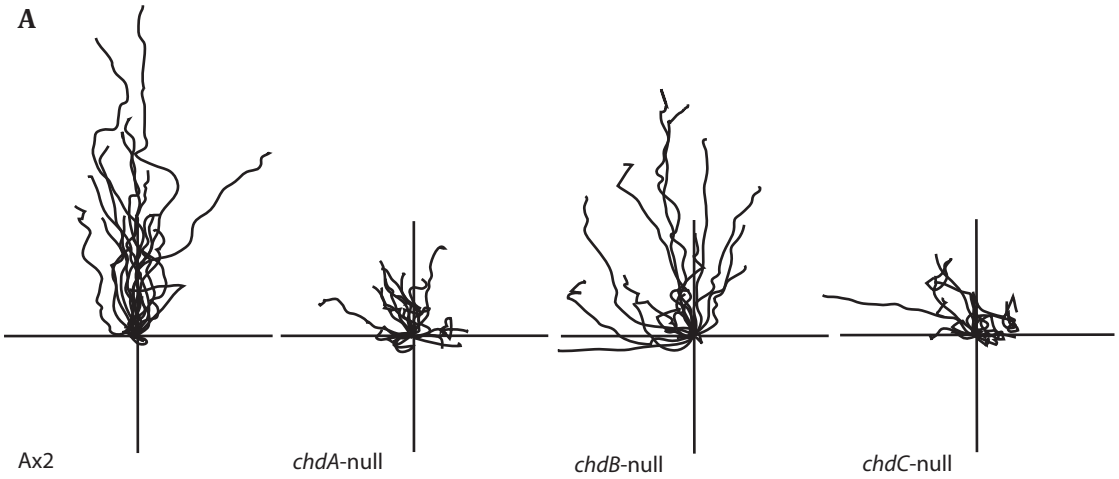
Figure 3.5: *Chd*-nulls have developmental phenotypes

A. Axenically grown Ax2, *chdA*, *chdB* and *chdC*-null cells were washed in KK2 and plated on KK2 agar. Images were taken at intervals during development as shown. **B.** 30-hour developed cells from A. **C.** Axenically grown Ax2 and *chdA*-null cells were washed in KK2 and placed in a line on nitrocellulose filters and subjected to a small amount of unidirectional light in a slug can. Images were taken after 48 hours. Tracks show the route migratory slugs took towards the light.

3.7 Chd roles in chemotaxis

During development *Dictyostelium* cells use extracellular cyclic adenosine monophosphate (cAMP) to signal other cells to move towards each other and to aggregate to produce the mound stage. During aggregation cells move towards the gradient of cAMP (chemotax) and release cAMP at their posterior side, propagating the cAMP signal. Using an in vitro system it is possible to replicate cells chemotaxing towards a gradient of cAMP. By placing separate cells in an artificial cAMP gradient one can study the response of individual cells to cAMP and eliminate factors such as their ability to establish and propagate the gradient. Previously *chdC*-null mutants had been identified as chemotaxing poorly towards cAMP. It was important to assess whether the other *chd* mutants also had chemotaxis defects. *chdB*-nulls chemotax normally towards cAMP, with a chemotactic index (CI) akin to Ax2 cells (Figure 3.6). In contrast, both *chdA* and *chdC*-null cells had significant chemotaxis defects: reduced chemotactic index, reduced cell speed and increased cell turning (Figure 3.6). However, in neither *chdA* nor *chdC*-null cells was the ability to chemotax completely lost.

A



B

	Directionality	SEM Directionality	Speed	SEM Speed µm/min	CI	SEM CI
Ax2	0.65	0.05	13.21	0.94	0.8	0.06
chdA-null	0.37	0.03	7.39	0.47	0.45	0.09
chdB-null	0.58	0.07	13.17	1.45	0.69	0.10
chdC-null	0.33	0.04	7.02	0.65	0.44	0.08

C

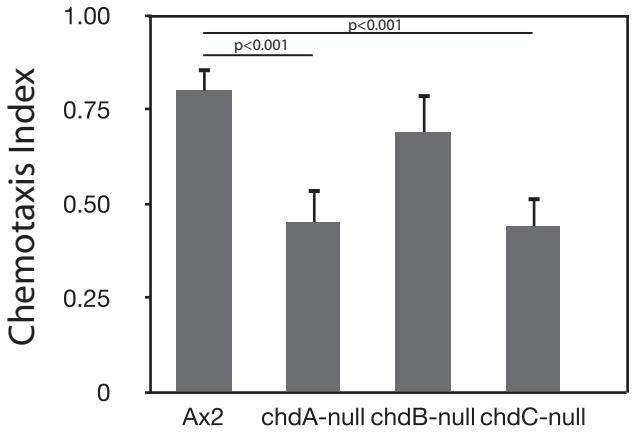


Figure 3.6: *chd*-nulls chemotaxing towards cAMP

Axenicly grown Ax2, *chdA*, *chdB* and *chdC*-null cells were washed and resuspended in KK2 and pulsed for 5 hours with cAMP. Cells were imaged for 15 minutes in a gradient of cAMP. **A.** Traces represent the trajectory random individual cells took. **B.** Calculated values for chemotactic index, speed and directionality of the population of cells. **C.** Bar plot of chemotactic index; the values plotted are mean ± standard error of the mean. Calculations were made from a minimum of 100 cells from three separate experiments. The Mann-Whitney test for statistical significance was used.

3.8 Using RNA-seq to globally measure gene expression

ATP-dependent chromatin remodellers play an important role in nucleosome positioning and the regulation of gene expression. As a way of assessing the Chd family's role in regulation of gene expression, RNA-seq was used to look genome wide at the expression levels of mRNA for all protein-coding genes. RNA was isolated from both log phase growing cells and cells starved and pulsed with cAMP for 5 hours to achieve a chemotactically competent state and then poly(dA) enriched before library preparation and sequencing on the Illumina platform.

To demonstrate the reproducibility of RNA-seq data FPKM (Fragment Per Kilobase of exon per Million reads) values were plotted from two experimental repeats of Ax2 cells from growing cells and 5-hour pulsed cells (Figure 3.7). FPKM values were highly reproducible with high coefficients of determination (R^2) values. Data from biological repeats in growing cells had a higher R^2 value than equivalent 5-hour pulsed data. Although pulsing with cAMP synchronises development, cells may still be at slightly different stages resulting in more variable gene expression data than growing cells. Relative expression levels also correlate with previously published data.

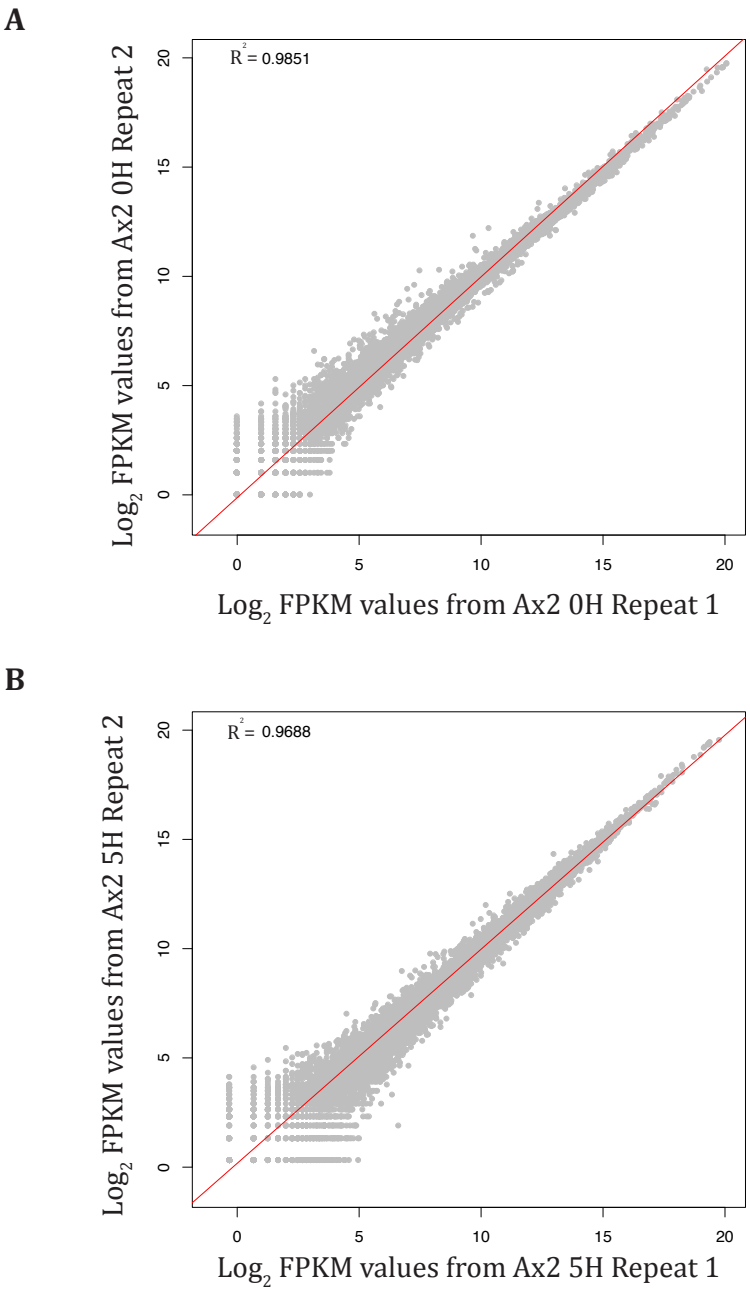


Figure 3.7: Reproducibility of RNA-seq data

Scatter plot of log2 FPKM expression values calculated for biological repeat 1 and repeat 2 of **A.** growing Ax2 cells and **B.** 5-hour cAMP pulsed Ax2 cells. Red line, x=y. Coefficients of determination (R^2) values are indicated.

3.9 Distinct gene sets are mis-regulated in separate *chd*-nulls

Initially comparing gene expression (FPKM) in each of the *chd*-nulls against Ax2 in growing cells shows widespread alterations with 10–15% (See gene numbers in Figure 3.8) of the genome misexpressed ≥2-fold (p≤0.05). Then by comparing gene expression in the *chd*-nulls against each other it was apparent that the genes that are misregulated in one *chd*-null are not identical to those in the other *chd*-nulls (Figure 3.8). Genes that were misexpressed in any of the *chd*-nulls when compared to Ax2 were selected, clustered and their $\log_2$ fold change in expression displayed in a heatmap format (Figure 3.8). The conclusions that can be drawn from analysis of the heatmap are that there is very little overlap between the genes misregulated in different *chd*-nulls and there is limited antagonistic regulation of genes by different Chd remodellers. Combined, the data shows in the absence of Chd proteins the expression of distinct gene sets are disrupted.

Comparing growing cells to chemotactically competent, five-hour pulsed cells reveals that, as *Dictyostelium* cells initiate development, large (>50%, ≥2-fold at p≤0.05) global changes in gene expression manifest in Ax2 cells (Figure 3.9). Expected gene expression terms such as ribosome biogenesis are enriched for downregulated genes (Figure 3.9). Upregulated genes are enriched for multicellular organism development GO terms (Figure 3.9).

Changes in gene expression in 5-hour pulsed cells are further augmented in the *chd*-nulls, where individually 15–30% of all genes are misexpressed ≥2-fold

($p \leq 0.05$) compared to Ax2 (Figure 3.10). As with the RNA-seq data from growing cells, there was limited overlap between genes misexpressed in the *chd*-nulls and little antagonistic regulation (Figure 3.10), again showing in the absence of Chd proteins the expression of distinct gene sets are disrupted.

In both growing and 5-hour pulsed cells there were no compensatory profile changes for *chdA*, *chdB or chdC* in the various mutant strains, demonstrating that the separate remodellers are unlikely to regulate each other.

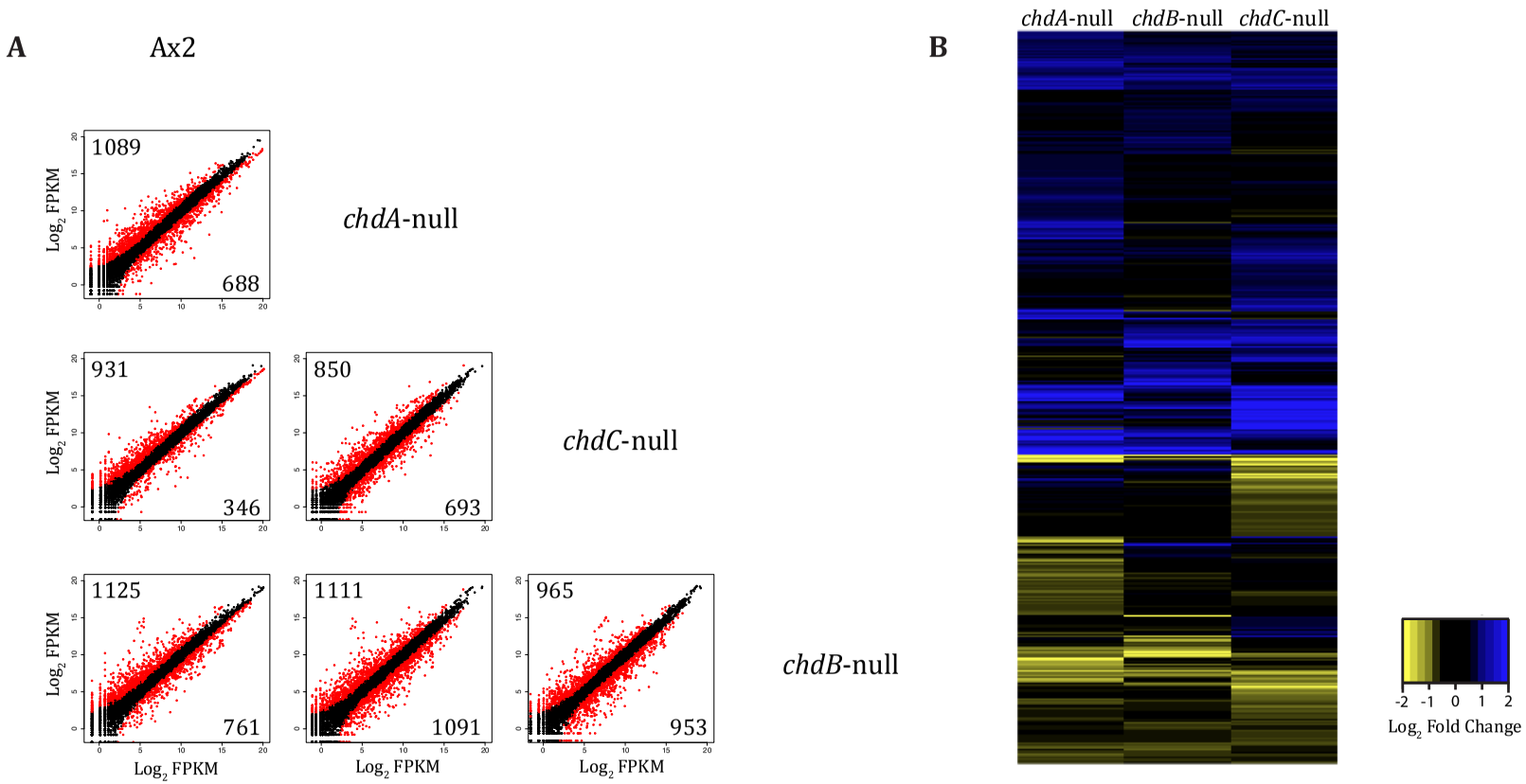


Figure 3.8: Chd proteins regulate distinct genes sets in growing cells

A. Scatter plots of $\log_2$ FPKM expression values calculated for Ax2, *chdA*, *chdB* and *chdC*-nulls in growing cells plotted against each other. Black points are genes with identical expression values between the two cell lines, <2-fold change or p>0.05. Red points are genes that are differentially expressed between the two cell lines, ≥2-fold change and p≤0.05. **B.** Clustered heatmap representation of the $\log_2$ fold change in expression level of 3,313 genes that were identified as misexpressed in *chdA*, *chdB* or *chdC*-nulls against Ax2 cells in growing cells.

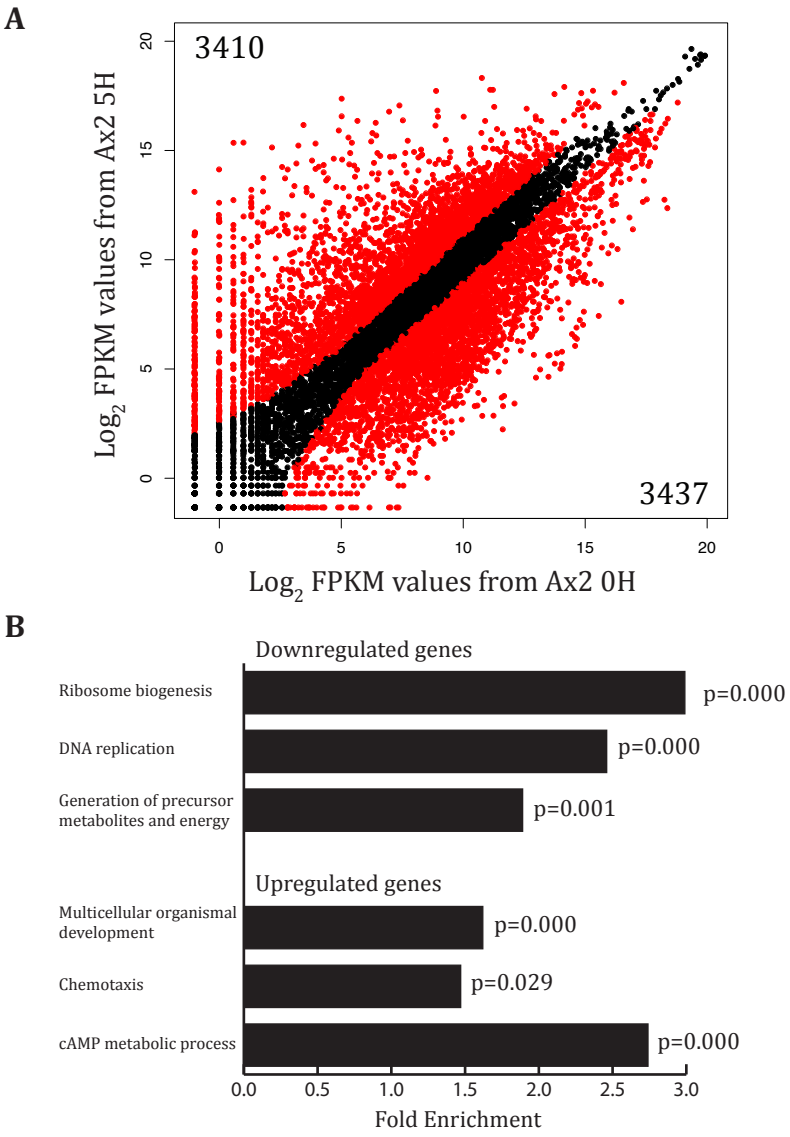


Figure 3.9: Gene expression changes in wild type Ax2 cells from growth to a partially developed state

A. Scatter plot of $\log_2$ FPKM expression values calculated for growing Ax2 cells against 5-hour cAMP pulsed Ax2 cells. Black points are genes with identical expression values between the two states, <2-fold change or p>0.05. Red points are genes which are differentially expressed between the two states, ≥2-fold change and p≤0.05. **B.** Gene ontology analysis of gene expression changes in 5-hour pulsed cells relative to growing cells. Selected enriched GO terms in downregulated and upregulated genes are shown. Hypergeometric test for significance used.

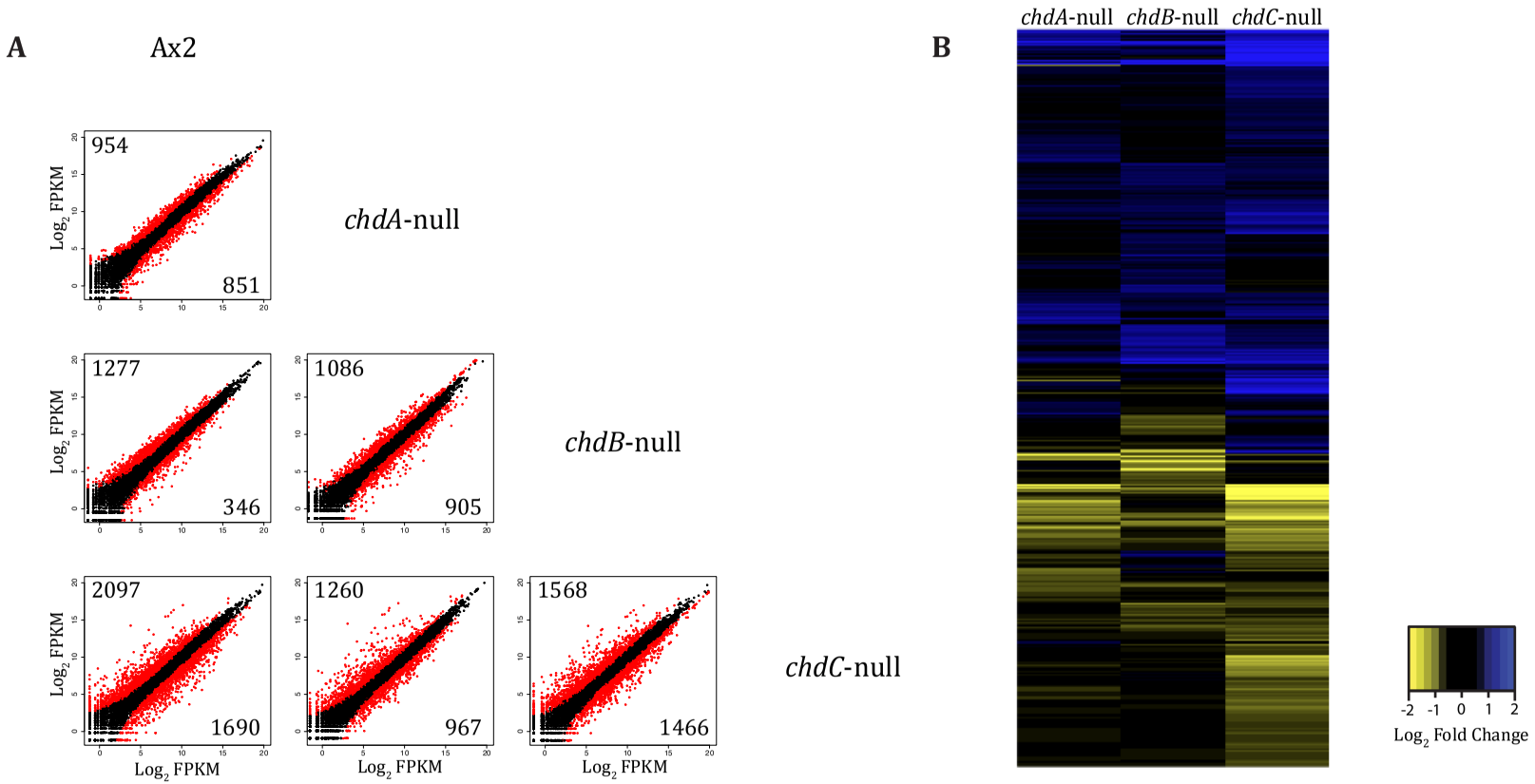


Figure 3.10: Chd proteins regulate distinct genes sets in chemotacticaly competent cells

A. Scatter plots of $\log_2$ FPKM expression values calculated for Ax2, *chdA*, *chdB* and *chdC*-nulls in 5-hour pulsed cells plotted against each other. Black points are genes with identical expression values between the two cell lines, <2-fold change or p>0.05. Red points are genes that are differentially expressed between the two cell lines, ≥2-fold change and p≤0.05. **B.** Clustered heatmap representation of the $\log_2$ fold change in expression level of 5,104 genes that were identified as misexpressed in *chdA*, *chdB* or *chdC*-nulls against Ax2 cells in 5-hour pulsed cells.

3.10 Gene expression changes in growing cells

Gene Ontology (GO) term enrichment within the misexpressed genes in the *chd*-nulls was used with the aim of trying to dissect the phenotypes previously observed. For complete GO term enrichment diagrams see Appendix A.

Despite a large number of upregulated genes in *chdA*-nulls no GO term enrichment was observed, although notably the transcription factors *dimB* and *hbx3* were derepressed. Within the GO terms enriched for downregulated genes were terms for cytoskeleton organisation and vesicle trafficking/lysozyme function. *dscA-1*, *dscC-1*, *dscD-1*, *lmcA* and *lmcB* genes, which are involved in the transition from growth to early development (Berger and Clark, 1983, McPherson and Singleton, 1992), are all downregulated in *chdA*-nulls (Figure 3.11), which may explain the delayed entry into development.

Derepressed genes in *chdB*-nulls also had no GO term enrichment. GO terms enriched in downregulated genes in *chdB*-null were rRNA processing, cytoskeleton organisation and vesicle trafficking. The jumonji transcription factor *jcdB* is upregulated and *agnB*, which encodes argonaute, the catalytic component of the RNA-induced silencing complex (RISC), is downregulated. In contrast to *chdA*-nulls *lmcA* and *lmcB* were upregulated. *dscA-1*, *dscC-1* and *dscD-1* are unaffected (Figure 3.11).

GO terms enriched in the downregulated genes in *chdC*-nulls included genes expressed in response to salt and bacteria, and peptidoglycan synthesis. *ctnA*,

cf45-1, and *cf60* represent three out of the four gene members of the secreted countin complex that potentiate inter-cell signalling and chemotactic movement (Brock and Gomer, 1999, Brock *et al*., 2003, Brock *et al*., 2006) and are downregulated in *chdC*-null (Figure 3.11). Interestingly, like *chdA*-nulls, *dscC-1*, *dscD-1*, *lmcA* and *lmcB* but not *dscA-1*, are downregulated (Figure 3.11). These downregulated genes would fit with cells not producing a correct starvation response and as such a delayed entry into development. Downregulation of the countin complex also results in larger mounds as observed in the *chdC*-null.

GO analysis provided no indication as to the cause of the slow growth phenotype in *chdC*-nulls. Instead genes that might explain the phenotype, such as cell cycle and energy regulation genes, were selected. There was no effect on expression of cell cycle genes (Figure 3.12). Interestingly many nuclear genes that encode mitochondrial components had decreased expression in *chdC*-nulls (Figure 3.12) including components of the mitochondrial import complex (timm/tom, *timm10*) and the electron transport chain (cytochrome C, *cytC*). Expression of Low mitochondrial function therefore may be the cause of slow growth in *chdC*-null.

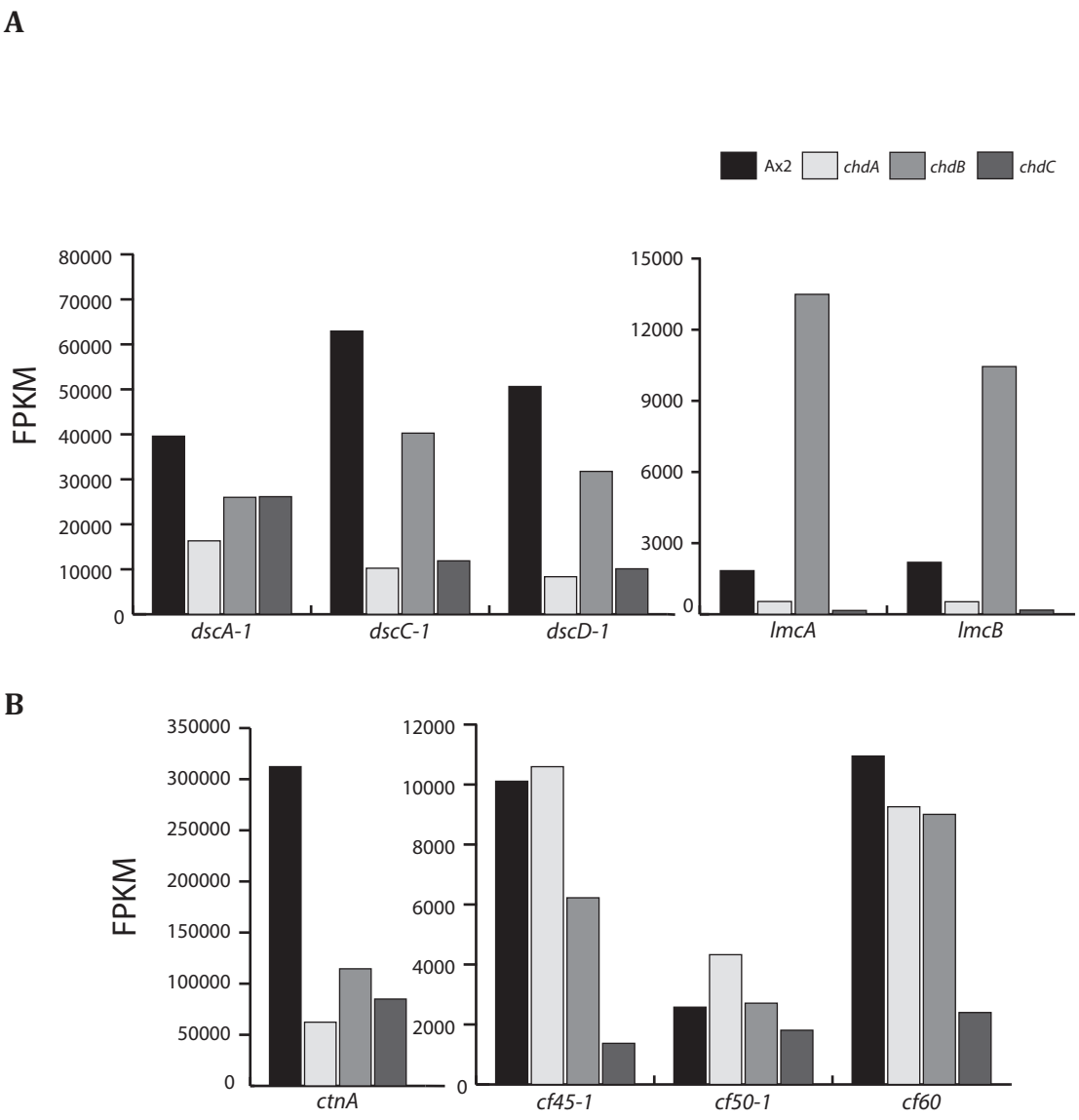


Figure 3.11: Mis-regulated early development genes in *chd*-nulls.

FPKM expression values from RNA-seq data from Ax2, *chdA*, *chdB* and *chdC*-null growing cells for A. Early development genes *dscA-1*, *dscC-1*, *dscD-1*, *lmcA* and *lmcB* and **B.** Members of the countin complex, *ctnA*, *cf45-1*, *cf50-1* and *cf60*.

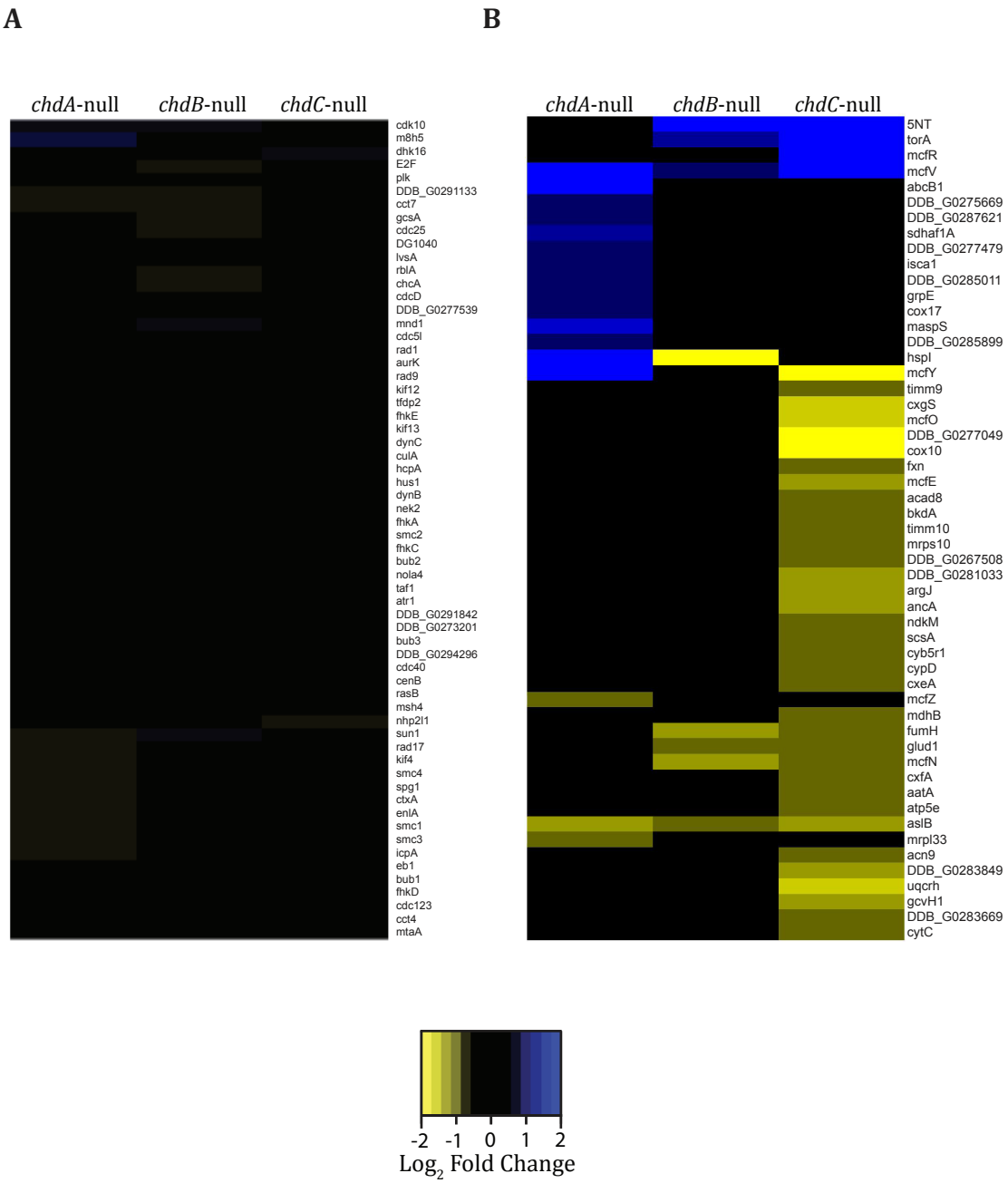


Figure 3.12: Misregulated genes in *chdC* may cause the slow growth phenotype

Clustered heatmap representation of the $\log_2$ fold change in expression level of *chdA*, *chdB* or *chdC*-nulls against Ax2 cells in growing cells **A.** 66 genes with GO terms related to cell cycle. **B.** 53 genes annotated with mitochondrial functions by GO analyses that are misexpressed in *chdA*, *chdB* or *chdC*-nulls are shown.

3.11 Gene expression changes in 5-hour pulsed cells

Gene Ontology (GO) term enrichment within the misexpressed genes in the *chd*-nulls was used with the gene expression data from 5-hour cAMP pulsed cells. The downregulated genes in *chdA*-nulls were enriched for GO terms for multicellular developmental processes. This would fit with poor chemotaxis and later developmental problems. Perhaps unsurprisingly, *chdB*-null had no enrichment of GO terms as it displays no phenotype at this stage. Upregulated genes in *chdC*-null are enriched for GO terms for rRNA processing. Downregulated genes are enriched for terms involved in glucose metabolism, ras signalling and muliticellular development. The latter would explain the developmental arrest shortly after this point in development.

To address the chemotaxis phenotypes, genes that have previously been identified as regulators of chemotaxis were selected. There are several signalling pathways that work downstream of the cAMP receptor CarA/Gα2βγ (receptor-G protein complex). Downstream of CarA there are several pathways that signal the cAMP response, including the cGMP-dependent and -independent actions of guanylyl cyclases sGC and GCA, the PIP3-dependent and -independent functions mediated via mTORC2, inter- and intra-cellular cAMP accumulation and signalling, and the interactions of Ca^{2+} and Phospholipase A_2 pathways (Kortholt and van Haastert, 2008, McMains *et al.*, 2008). 50% of the genes are misexpressed in *chdC*-nulls (Figure 3.13), explaining the defects in chemotactic response in *chdC*-nulls. A smaller yet significant number of these genes are also misexpressed in *chdA*-nulls.

cAMP signalling also directs developmentally regulated gene expression changes. Of the expression of thirteen genes that were previously classified as exclusively cAMP pulse-dependent (Iranfar *et al*., 2006), eleven were under-expressed in *chdC*-nulls and ten were under-expressed in *chdA*-nulls, whereas only one gene is affected in *chdB*-nulls (Figure 3.13). Aggregation and cell-type specific gene expression is also mediated, in part, by regulatory genes such as TgrC1/LagC, GbfA, and SrfA (Sukumaran *et al*., 1998, Escalante *et al*., 2003). Only the *chdC*-nulls show developmental arrest during aggregation and only *chdC*-nulls show impaired expression of all these essential regulatory factors.

3.12 ChdC regulation of mound progression

chdC-null presented the most severe phenotype: developing *chdC*-nulls arrested at the mound stage never completing development. Previous experiments showed under conditions that promote spore and stalk differentiation, *chdC*-nulls fail to differentiate correctly (Platt *et al*., 2013). RNA-seq was used in *chdC*-nulls and Ax2 cells at the loose mound stage. Here the mutant was morphologically matched to the wild type. RNA was taken from the Ax2 cells at the 10-hour stage of development and from the *chdC*-null at the 12-hour stage (Figure 3.14).

53% of genes annotated as prespore-specific (including *pspA* and *cotB*) and 56% of genes annotated as prestalk-specific have >2-fold reduced expression in *chdC*-nulls at the loose mound stage (Figure 3.14). This is consistent with the defects

described for prespore/spore differentiation. Downregulated genes in *chdC*-null were enriched for GO terms for regulators of cell division and metabolism. The former group (Figure 3.14) includes cell cycle genes (*cdk1*, *cdc45*, *cycB*), DNA replication genes (*polA1*, *polA3*, *polD2*, *pcna*, *rcf3*, *rcf4*, *rcf5*), and mitotic genes (*nek2*, *kif2*, *kif12*), although controversy surrounds the requirement of cell cycle progression during *Dictyostelium* development (Chen *et al*., 2004, Muramoto and Chubb, 2008). These gene sets are significantly upregulated during wild type development (Strasser *et al*., 2012).

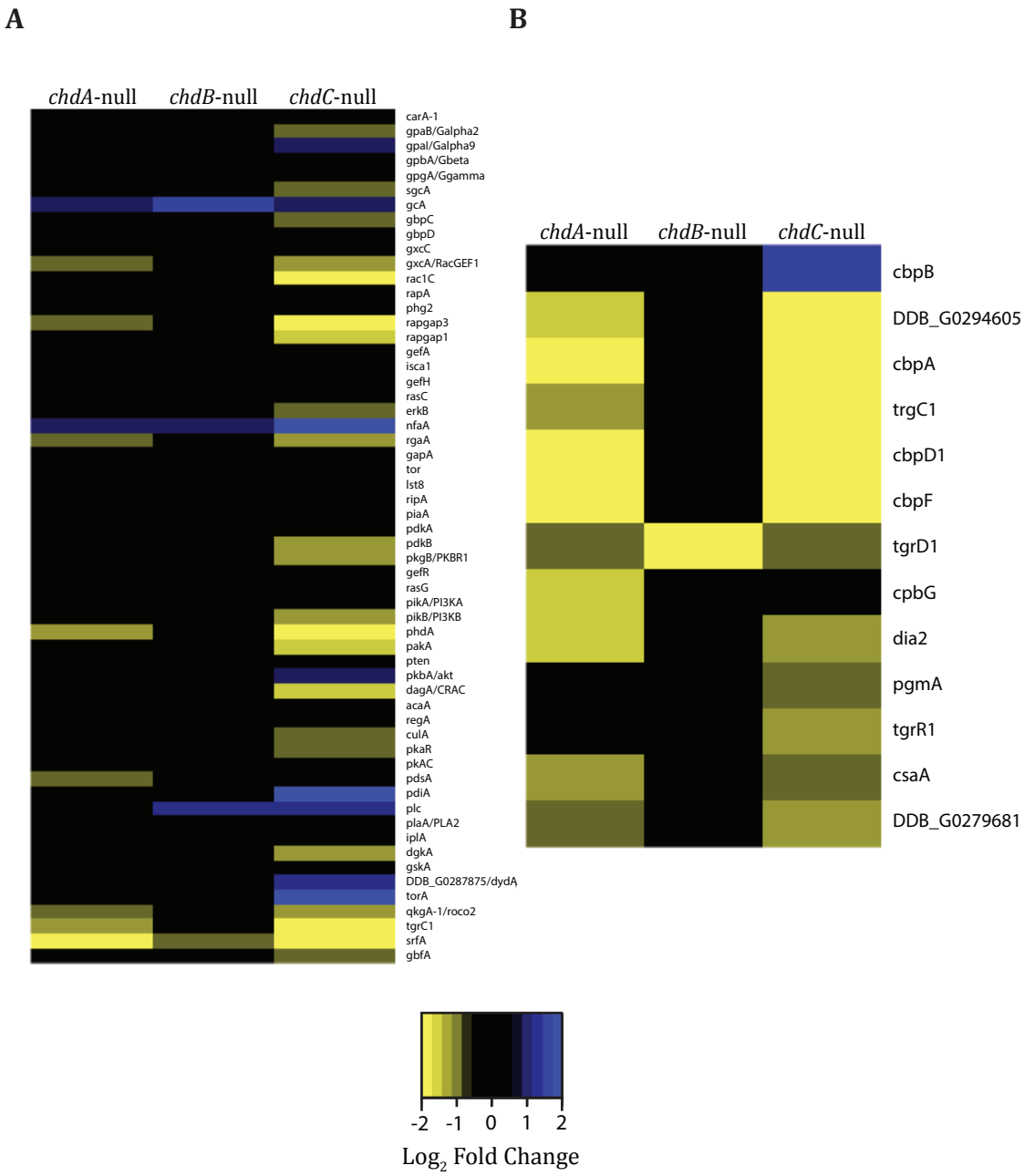


Figure 3.13: Mis-expression of chemotaxis regulators produce defects in chemotaxis

Heatmap representations of the $\log_2$ fold change in expression level of *chdA*, *chdB* or *chdC*-nulls against Ax2 cells in 5-hour cAMP pulsed cells **A.** 50 genes previously identified as part of signalling pathways in the cAMP chemotaxis response. **B.** 13 genes identified as having their developmental expression regulated by cAMP pulses (Iranfar *et al.*, 2006).

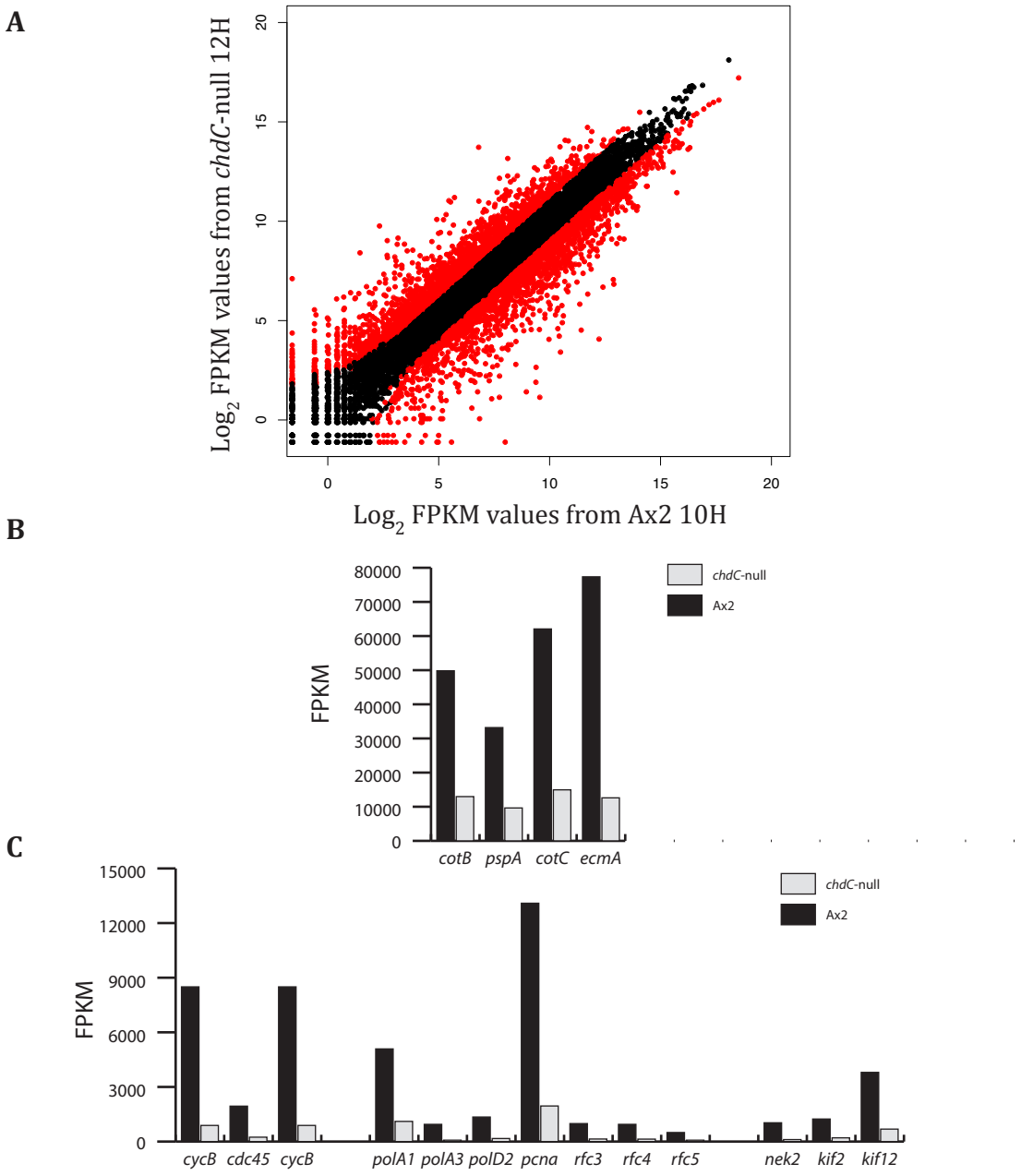


Figure 3.14: Gene expression changes in *chdC*-null at the loose mound stage

A. Scatter plot of $\log_2$ FPKM expression values calculated for loose-mound stage Ax2 cells morphologically matched to *chdC*-nulls. Black points are genes with identical expression values between the two cell lines, <2-fold change or p>0.05. Red points are genes which are differentially expressed between the two cell lines, ≥2-fold change and p≤0.05. **B.** FPKM values from selected prespore and prestalk genes at the loose mound stage in Ax2 and *chdC*-null cells. **C.** FPKM values from selected cell cycle, DNA replication and mitosis genes at the loose mound stage in Ax2 and *chdC*-null cells.

3.13 Discussion: Chd proteins are essential for proper regulation of gene expression and development in *Dictyostelium*

3.13.1 Chd proteins are developmental regulators

Dictyostelium discoideum contains three different Chd ATP-dependent chromatin remodelling proteins, which are conserved across other dictyostelids. All the members of the Chd family are expressed in growing cells and each has a different expression pattern through development. When knocked out each *chd*-null possessed a unique developmental phenotype that broadly correlated with their peak of protein expression. ChdA is expressed throughout development and has a peak of expression during late development (18–22 hours) and has a culmination defect which causes it to stay in the migratory 'slug' stage for longer than normal, and a chemotaxis defect. ChdB peak expression is during early development (2–8 hours) and expression reduces to undetectable levels by 18 hours. ChdB expression is consistent with the delay on entering development in *chdB*-nulls, although subsequent developmental timing was normal, giving it a heterochrony phenotype. ChdC has a broad peak of expression around aggregation and the mound stage (6–14 hours) and like ChdB drops to undetectable levels after 18 hours. The phenotype of *chdC*-null fits its expression pattern with both a chemotaxis defect and mound arrest. *chdC*-null cells also displayed a severe growth phenotype, although it was shown not to be due to a cytokinesis defect.

3.13.2 Separate gene sets are mis-regulated in *chd*-nulls

CHD chromatin remodellers have previously been shown to be regulators of gene expression as activators and/or repressors (Murawska and Brehm, 2011). Here high-throughput mRNA sequencing was employed to study the transcriptome of wild type and *chd*-nulls in growing and developed cells. Loss of an individual Chd member resulted in aberrant expression of up to 15% of the genome in growing cells and 30% in 5-hour cAMP pulsed cells. Large numbers of genes were both repressed and derepressed upon loss of Chd proteins, suggesting they have complex roles in both activation and repression of gene expression.

The genes misexpressed in each of the *chd*-nulls were distinctly separate. Few genes were regulated by more than one Chd remodeller, indicating that Chd proteins display specificity for the genes they regulate. Compensation by other CHDs has been observed in other systems (Williams *et al*., 2004). Interestingly upon loss of one *chd* in *Dictyostelium* no compensation, at the RNA level, by the other *chd* genes was observed, demonstrating that Chd proteins are not responsible for the expression of each other. This re-enforces their distinct roles in regulating different genes, as there appears to be no regulatory system to compensate for each other.

chdC-null has decreased expression of nuclear genes that encode mitochondrial proteins, which may cause the slow growth observed due to low mitochondrial function and therefore energy production. Mitochondrial function might be

assessed using a mitochondrial marker such as Mitotracker®. In 5-hour pulsed cells, which are chemotactically competent, misexpression of genes that encode components of cAMP signalling pathways was observed. In *chdC*-nulls, genes are misexpressed in all of the signalling pathways that signal downstream of the cAMP receptor. These signalling pathways are altered to a lesser but significant extent in *chdA*-nulls. This demonstrates that the chemotaxis defects of *chdA* and *chdC*-nulls are due to the inability to sense cAMP and transduce the signal. Further fitting with their inability to chemotax towards cAMP, almost all of the genes that were previously identified as cAMP pulse regulated are downregulated in *chdA* and *chdC*-nulls (Iranfar *et al*., 2006). Despite showing few phenotypic differences to wild type large numbers of genes were still misexpressed in *chdB*-nulls, although GO enrichment did not reveal any possible phenotypes that have not yet been explored.

RNA-seq at the loose mound stage of development showed *chdC*-nulls repressed the expression of >50% of prestalk and prespore markers, fitting with the mound arrest phenotype and previous work that showed *chdC*-nulls are unable to differentiate into stalk and spore cells. Cell cycle and DNA replication genes were also repressed in *chdC*-nulls, although it's not clear whether the cell cycle is required in development (Chen *et al*., 2004, Muramoto and Chubb, 2008). If the cell cycle is required in development, it is not clear whether repression of these genes in *chdC*-null causes mound arrest or is an effect of it.

3.13.3 Summary

In this chapter I demonstrated that Chd ATP-dependent chromatin remodellers have separate expression patterns and through knockout cell lines that I produced showed they are in turn important regulators of development. Utilising high-throughput transcriptome sequencing (RNA-seq) I found that Chd proteins regulate the expression of large and distinct gene sets, probably through both activation and repression. Genes that are misexpressed in the mutants are essential for development to proceed correctly, demonstrating Chd proteins regulate development through gene expression. This study is the first to examine comprehensively the complete Chd family in *Dictyostelium*.

Chapter 4:

Genome-wide mapping of nucleosome positions in *Dictyostelium discoideum*

4.1 Introduction

Micrococcal nuclease (MNase) is an endonuclease that cuts DNA almost independent of sequence, although it is only able to cut DNA not in contact with protein. Thus it is unable to cut DNA wrapped around a nucleosome, cutting only the linker DNA. Chromatin digested with a limited concentration of MNase can be run out on a gel producing the classic nucleosome DNA ladder with bands representing the amount of DNA protected by one nucleosome and two nucleosomes and multiples thereafter (Axel, 1975). Traditional nucleosome position analyses combined micrococcal nuclease digestion and radiolabelling to map nucleosome positions on a small scale (Kent and Mellor, 1995, Wu and Winston, 1997). Recent advances in microarrays and later high-throughput sequencing have enabled genome-wide unbiased studies of nucleosome positions (Yuan *et al*., 2005, Albert *et al*., 2007, Lantermann *et al*., 2010). Genome-wide studies have allowed general nucleosome rules and roles of chromatin remodellers to be interrogated.

4.2 Mapping nucleosomes genome wide in *Dictyostelium*

Historical approaches to map nucleosomes in *Dictyostelium* achieved limited success, the AT content of the genome possibly causing problems. With the advent of the sequencing of the *Dictyostelium* genome and more sensitive approaches, mapping nucleosomes genome wide in *Dictyostelium* has become more viable. The mapping method employed for this study used a combination of MNase digestion and Illumina paired-end sequencing. There was however one

main difference to most methods; rather than isolating just mono-nucleosomal DNA, DNA was isolated from a broad range of fragments from 50 to 1,000bp (Kent *et al*., 2010, Platt *et al*., 2013), enabling mapping of mono-nucleosomes, di-nucleosomes and subnucleosomal fragments.

Paired-end sequencing the DNA fragments results in two reads that originate from either end of a single fragment; these reads can be mapped back to the genome sequence and the fragment the pair of reads originates from can be calculated from the distance apart the two reads map genomically. To map nucleosomes, reads that mapped 150bp ± 30bp apart are selected, and the mid-point genomically between the two mapped reads is taken as the dyad of the nucleosome. With the dyad calculated for each pair of reads the average positions of all nucleosome dyads can be visualised (Figure 4.1 & 4.2). Well-positioned nucleosomes are represented by sharp peaks; the nucleosome is in a low number of different places so resulting pairs of reads produce the same calculated dyad more often. A broader peak is a less consistently positioned nucleosome; the nucleosome is variable between cells therefore MNase digests produce different fragments and the dyad calculated is in a variety of different positions.

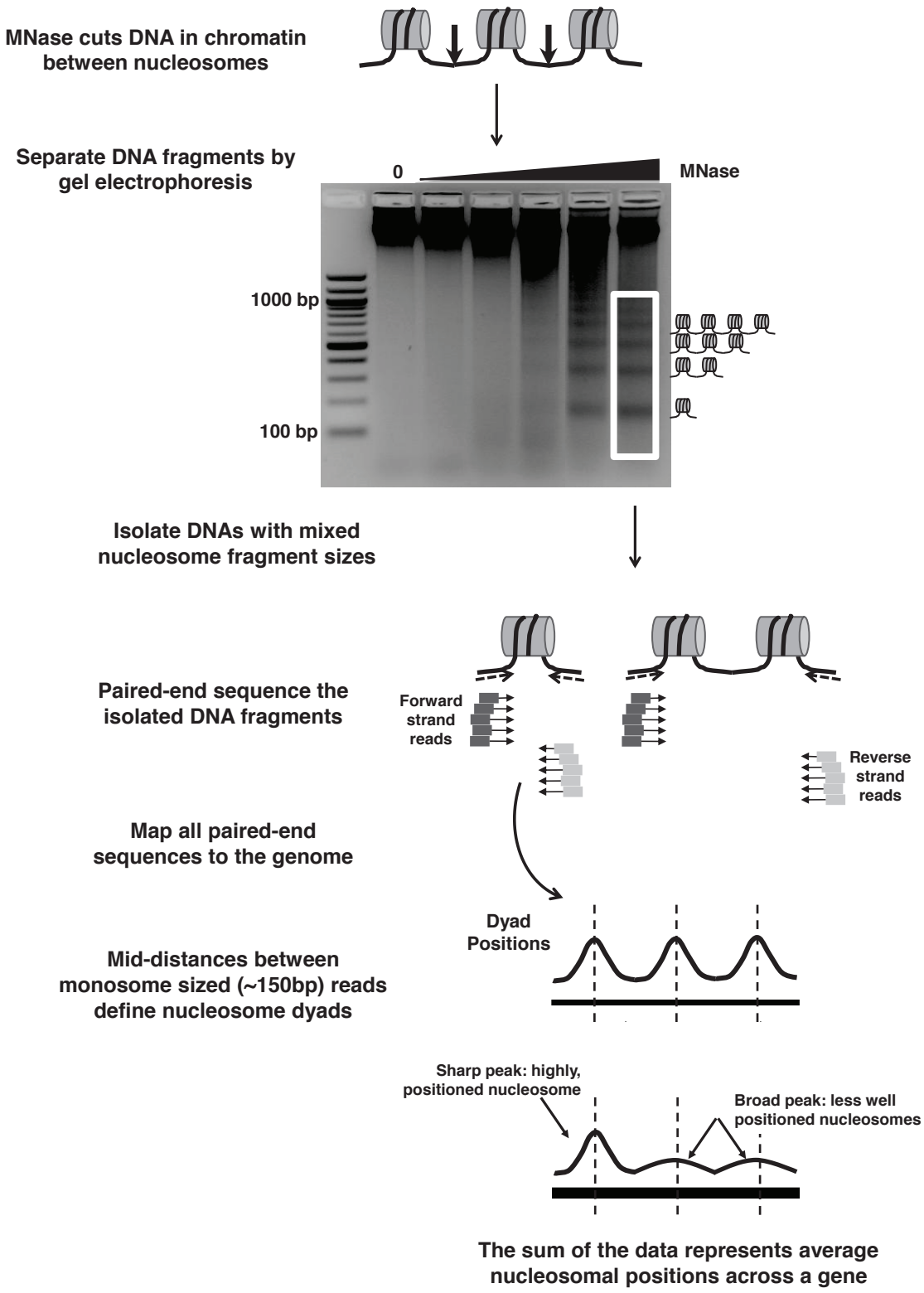


Figure 4.1: Nucleosome mapping method

See legend on next page.

Figure 4.1: Nucleosome mapping method

Chromatin is digested in whole live cells without fixation with varying amounts of micrococcal nuclease (MNase). The DNA is extracted and separated on an agarose gel. Digestion by MNase produces a characteristic ladder, with ~160bp nucleosome repeat length. The DNA (~50–1,000bp) is size selected by excision from the gel, purified and paired-end sequenced. The resulting sequences are mapped to the genome. The mid-point between two 150 ± 30bp paired reads defines the dyad of a nucleosome; the sum of all dyads assigns the average position of every nucleosome. Nucleosomes with highly similar positioning in most cells will exhibit a very sharp coverage peak, whereas more flexibly positioned nucleosomes result in broader peaks (Adapted from Platt *et al*., 2013).

4.3 Nucleosomes are positioned in *Dictyostelium*

Studying individual genes (Figure 4.2) reveals that certain nucleosomes appear to be well positioned and others are more flexibly positioned, occupying a variety of positions. In *eftud1* gene nucleosomes at the beginning of the gene are highly positioned and evenly spaced. Further into the gene nucleosome positions are more heterogeneous and less evenly spaced. Comparison of tworepeats of the micrococcal nuclease digestion found the nucleosome positions to be highly repeatable at both the individual gene and global level when all genes are aligned to their ATG (Figure 4.3).

Early work on chromatin in *Dictyostelium* focused on the rDNA palindrome as it is present in multiple copies in each cell. Ness *et al*. (1983) mapped restriction sites along the rDNA palindrome and used fragments of the palindrome generated by EcoRI digestion as a probe to look at nucleosome structure following MNase digestion. The authors found that coding regions in the rDNA palindrome had less structure than the non-coding regions. Using their digestion map of the rDNA palindrome (Ness *et al*., 1983) to identify the regions previously studied, the same observations can be made with the genome-wide nucleosome mapping technique employed for this study (Figure 4.4). Clear nucleosome peaks are seen in intergenic regions (probe III, Figure 4.4) and much noisier peaks within the coding regions (probe VII, Figure 4.4).

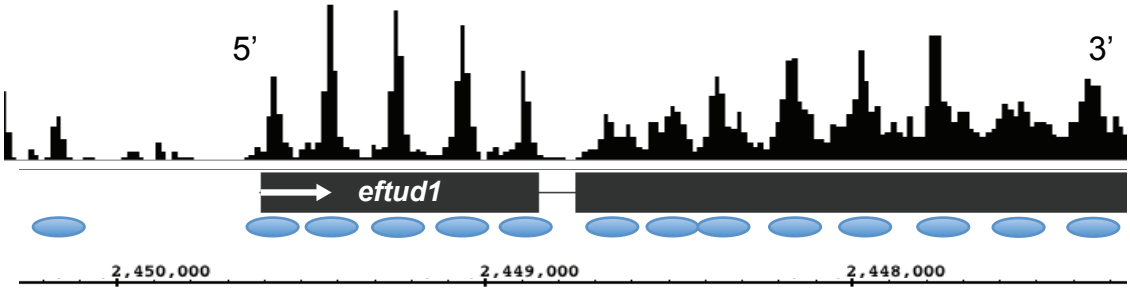


Figure 4.2: Nucleosome positions along an average *Dictyostelium* gene.

The nucleosome map of the *eftud1* gene in growing Ax2 cells. The gene model is represented by black rectangles, which represent exons. Peaks represent average nucleosome positions in the population. Ovals also represent nucleosomes.

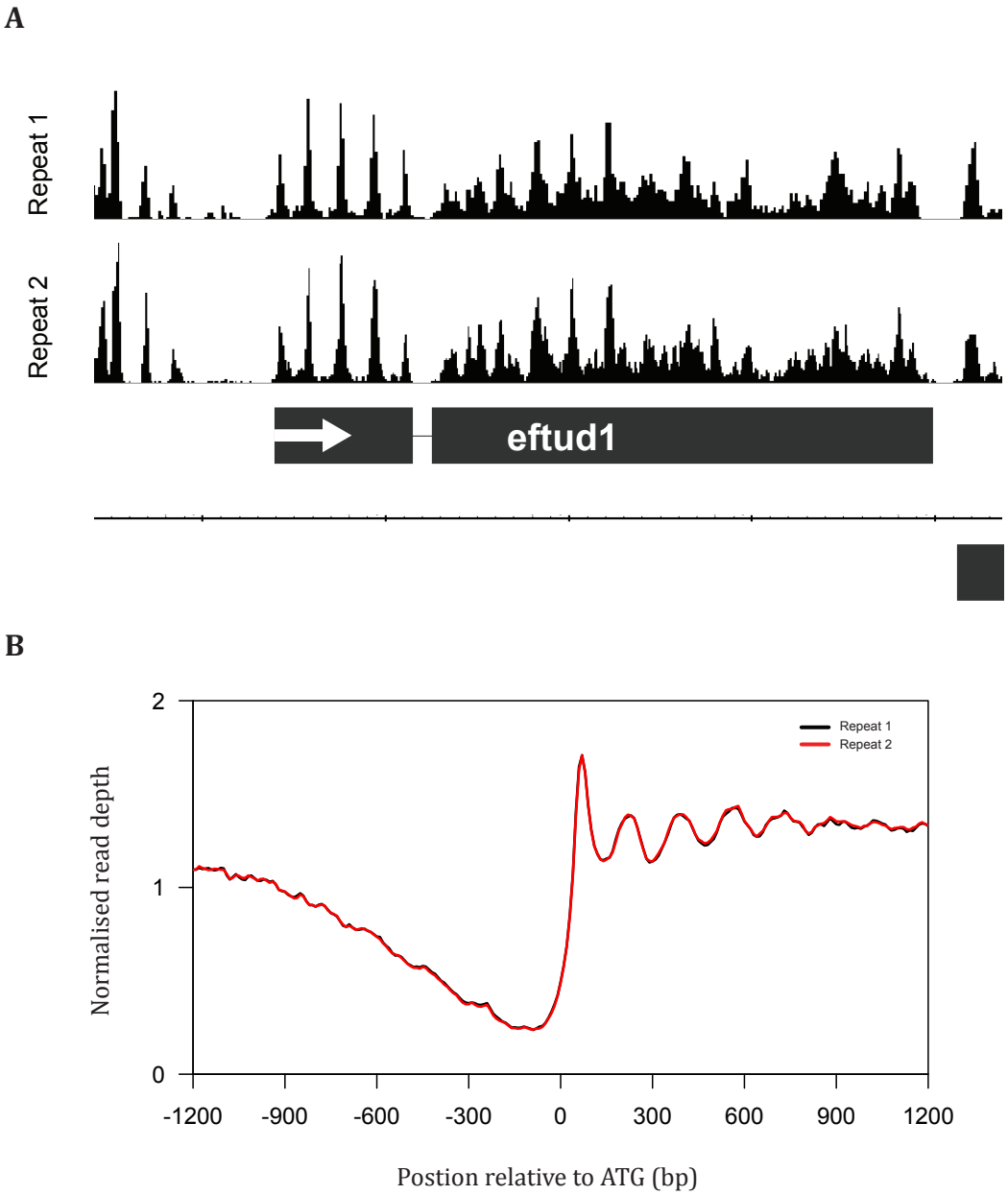


Figure 4.3: Nucleosome positioning data is highly repeatable

A. Nucleosome positions along the *eftud1* gene and **B.** Cumulative frequency distribution graph comparing nucleosome positions around ATG start sites of 12,750 genes in two independent repeats of nucleosome positions in growing Ax2 cells.

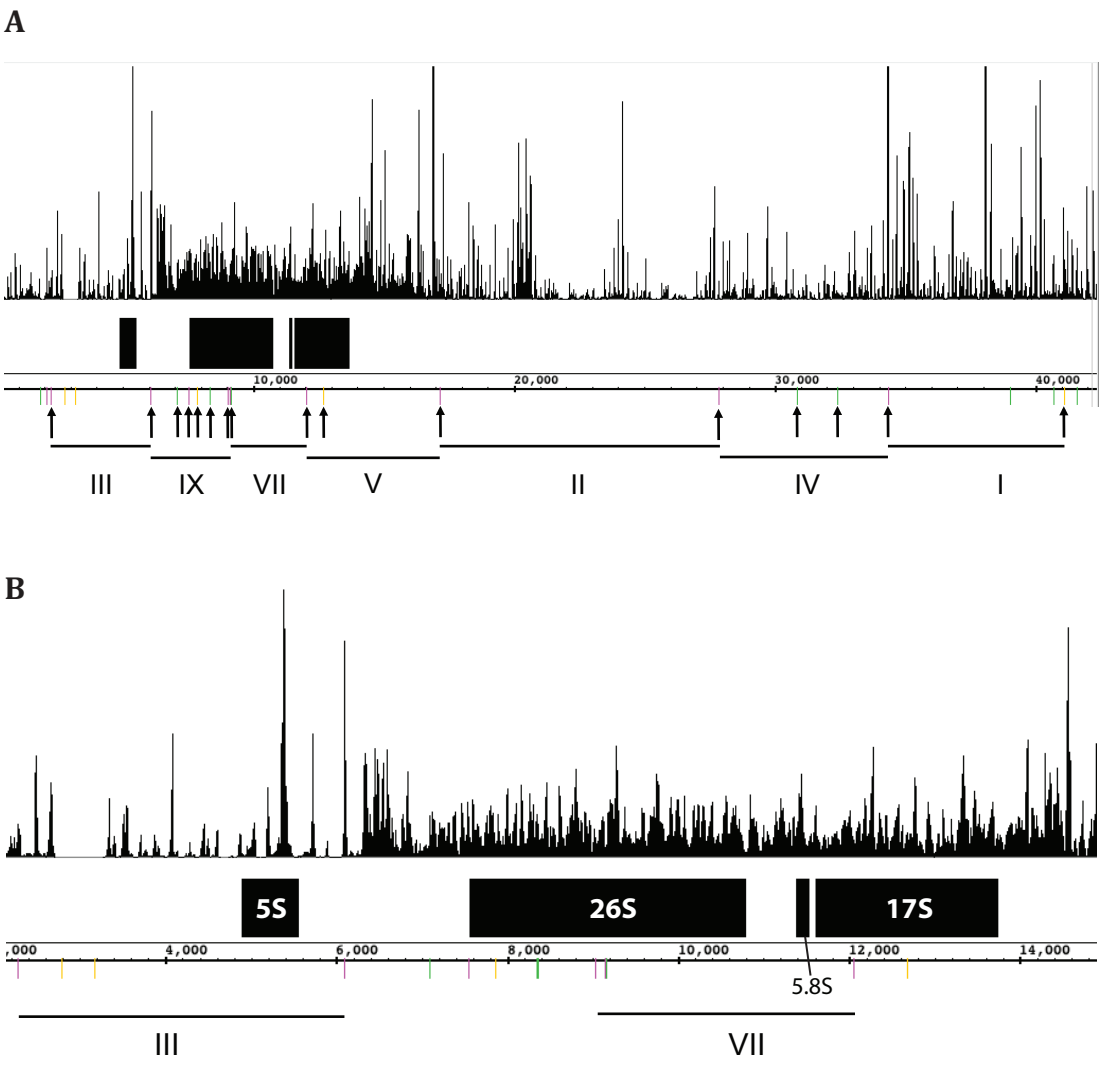


Figure 4.4: MNase-seq nucleosomes recapitulate previous findings on the rDNA palindrome in *Dictyostelium*.

A. Nucleosome positions along half of the ribosomal palindrome. Restriction sites based on sequence are displayed: Pink – EcoRI, Green – HindIII, Yellow – SalI. Arrows indicate restriction sites previously mapped by Ness *et al*. (1983). Bars represent probes previously used. **B.** Closer view of the ribosomal palindrome highlighting one intergenic region (III) and one coding region (VII).

4.4 Nucleosomes are positioned around the start of genes

To get a genome-wide view of particular features the dyad frequencies surrounding a feature can be summed and normalised to get a cumulative frequency distribution (CFD) graph, a description of the chromatin structure around the feature of interest. Transcriptional start sites (TSS) are not universally annotated in *Dictyostelium* and the majority of *Dictyostelium* 5' UTRs (untranslated regions) are very short (<50bp), hence the ATG serves as a common, close indicator for the TSS. Others have also used this approach when 5' UTR information is not available (Westenberger *et al*., 2009, Valouev *et al*., 2008). When genes are aligned to the ATG start site of all 12,750 protein-coding genes (Figure 4.5), it is clear that a large proportion of *Dictyostelium* genes have organised nucleosomes relative to the 5' end of genes. A prominent +1 nucleosome is observed with a following 3–4 well-phased nucleosome peaks with an average repeat length of 165bp. This observation is in line with data generated by an independent approach (Chang *et al*., 2012) showing a repeat length of 162bp for vegetative *Dictyostelium* cells. The phasing of nucleosomes around the start site is evident in a variety of species (Albert *et al*., 2007, Lantermann *et al*., 2010, Schones *et al*., 2008, Mavrich *et al*., 2008). With an average repeat length between dyad centres of 165bp, and the nucleosome core containing ~147bp, this assumes an average linker DNA length of ~18bp. Though Chang *et al*. (2012) suggested nucleosomes are arranged in di-nucleosomes, this was not observed in the present study. Notably, a nucleosome-free region and limited nucleosome phasing relative to the stop codon

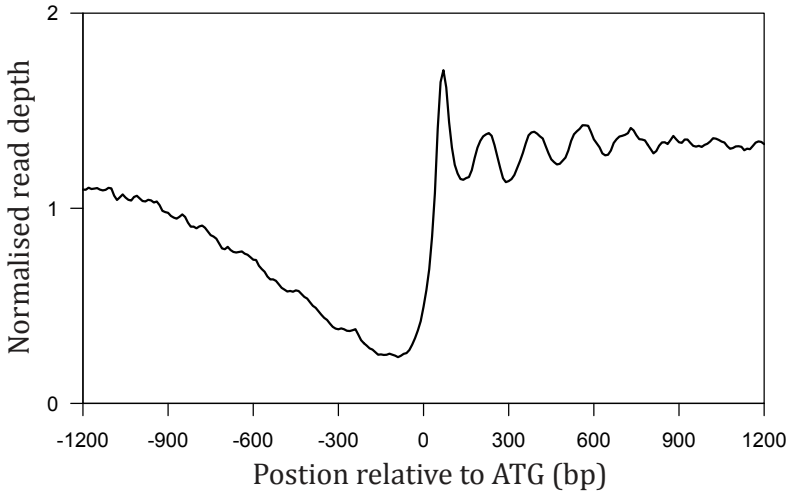
(TAA/TAG/TGA) was observed (Figure 4.5) as in other organisms (Shivaswamy *et al*., 2008, Mavrich *et al*., 2008).

Utilising a clustering and heatmap approach to display nucleosome-positioning data, each gene is represented and genes are clustered based on their overall pattern. By k-means clustering, five distinct nucleosome patterns were observed (Figure 4.6), which when combined produce the one pattern observed in the CFD graph (Figure 4.5). Each cluster had the same repeat length of ~165bp. Whether or not the different clusters encoded functional differences was explored. Gene expression within the clusters was found not to be different between clusters (Figure 4.6). Clusters 2, 3 and 5 were found to have more introns close to the 5' end than other classes (Figure 4.6). Of the genes with introns in cluster 2, 79% had their first intron within the first 300bp of the gene, well above the average of 65% genome wide. The presence of these early introns may disrupt the nucleosome pattern as introns are very short (mean 146bp) and AT rich (88%) in *Dictyostelium*. Being unable to position nucleosomes within an intron may contribute to the alternative nucleosome pattern observed in cluster 2. Introns have been shown to be nucleosome depleted in other organisms, and nucleosome positioning is thought to be involved in regulation of splicing (Andersson *et al*., 2009, Schwartz *et al*., 2009, Tilgner *et al*., 2009).

As in other organisms from the CFD graph of all genes (Figure 4.5) the +1 nucleosome is well positioned and with each subsequent nucleosome phasing tends to diminish (Shivaswamy *et al*., 2008, Mavrich *et al*., 2008, Lantermann *et al*., 2010). There are two reasons why the nucleosome phasing tends to diminish.

The first is that the +1 nucleosome is thought to be the most important and rigidly positioned and the other nucleosomes are thought to be phased relative to the +1 nucleosome (Hughes *et al.*, 2012). The second reason is that genes are variable in length, so the shorter genes have a shorter stretch of phased nucleosomes past their ATG than the longer genes (Figure 4.7) because nucleosomes in intergenic regions are less well positioned (Figure 4.5).

A



B

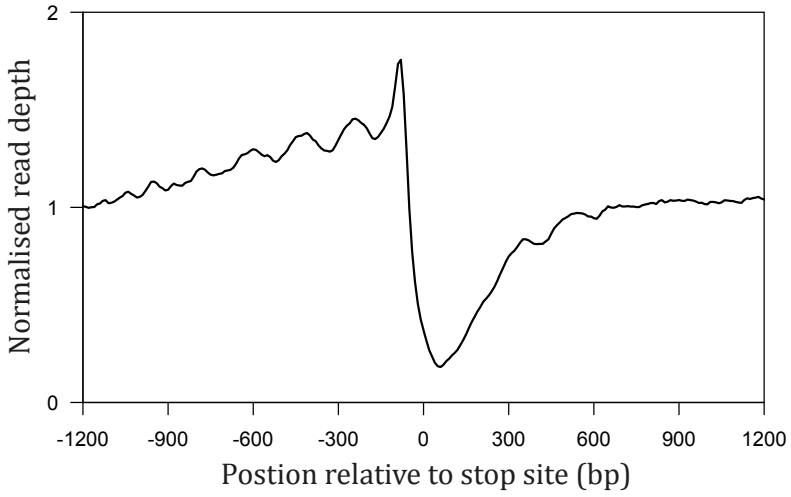


Figure 4.5: Nucleosomes are globally organised in *Dictyostelium*

Cumulative frequency distribution graph for nucleosome dyads around the **A.** ATG start site and **B.** Stop site of 12,750 protein-coding genes in Ax2 growing cells.

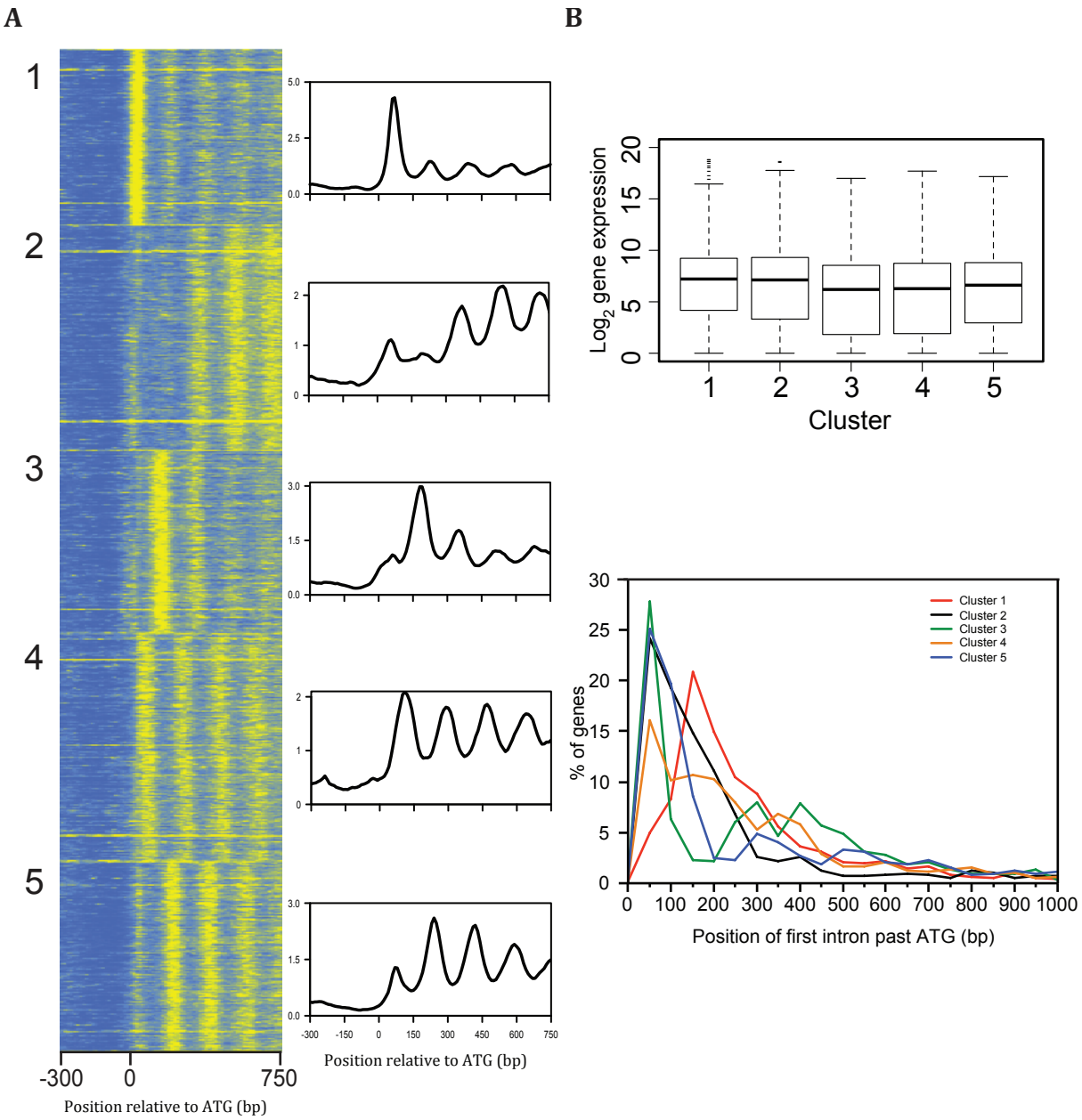


Figure 4.6: Genes cluster into five separate nucleosome patterns.

A. Heatmap representation of dyad frequencies surrounding the ATG start site of 12,750 genes clustered into five groups using a k-means method. A CFD graph for each cluster is shown. **B.** Boxplot representation of gene expression distribution in the five clusters in A. **C.** Density plot displaying the position of the first intron relative to the ATG start site in all five clusters in A.

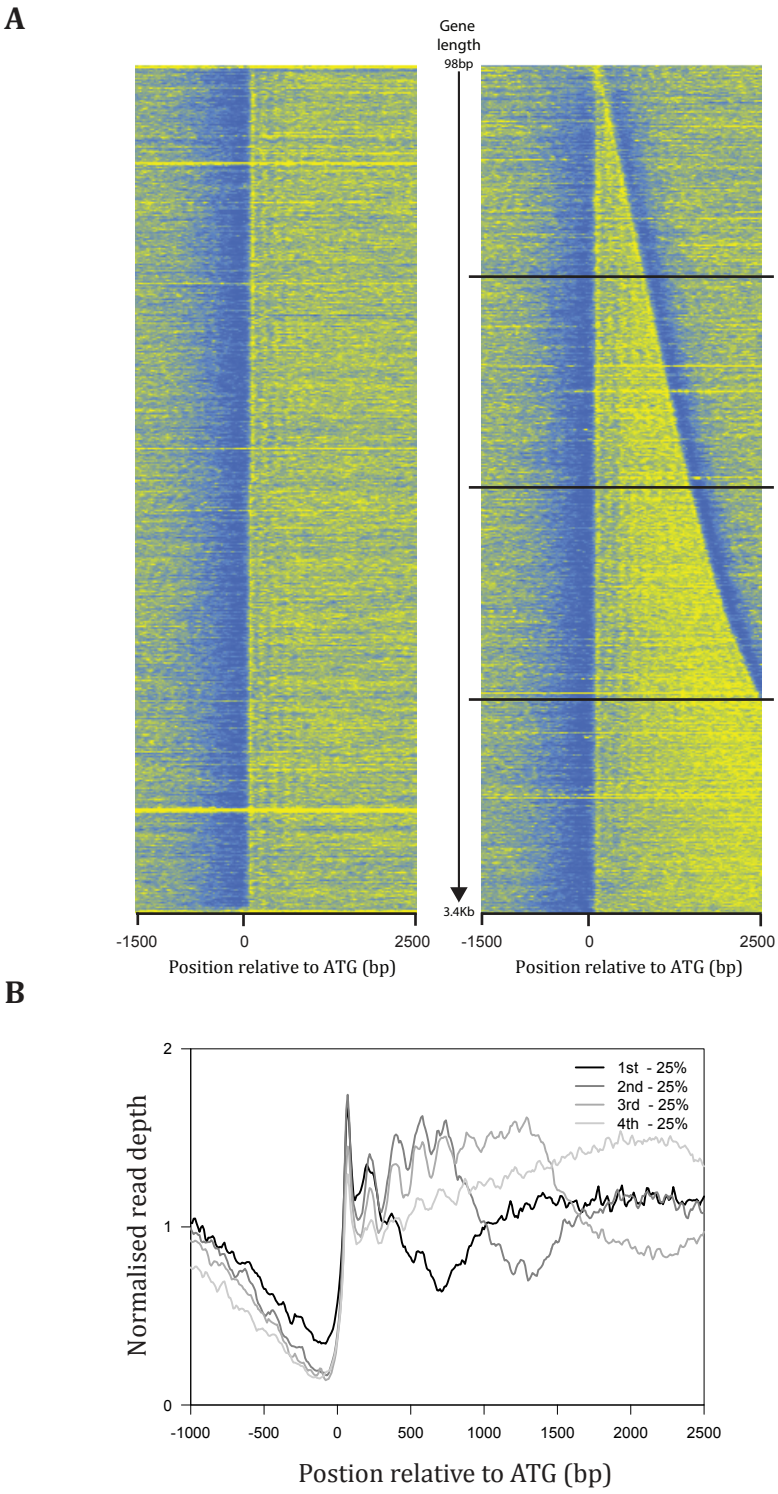


Figure 4.7: Nucleosomes remain phased in longer genes

A. Heatmap representation of dyad frequencies surrounding the ATG start site of 12,750 genes unsorted and sorted by gene length. **B.** Genes were sorted by gene length and split into fourths and displayed are CFD graphs for the four quarters.

4.5 Promoter regions of *Dictyostelium* genes are nucleosome depleted

Phasing relative to the +1 nucleosome extends to around five nucleosomes 3' into the gene. However prominent nucleosome phasing is generally limited to coding regions (Figure 4.5). Outside the coding region, the intergenic regions showed limited nucleosome phasing. Looking at nucleotide base content surrounding start sites and stop sites of genes, GC content is constant in the gene body. Sharp drops (~15%) in GC content are observed moving into intergenic regions (Figure 4.8). Overall intergenic regions are highly AT rich (85%). Large stretches of AT can reduce the efficiency of both sequencing these regions and the mappability due to their low complexity (Dohm *et al*., 2008). To assess the impact of sequencing efficiency and mappability in *Dictyostelium*, naked DNA (purified, histone-free DNA) was sonicated and sequenced identically to previous MNase digested DNA. The CFD graph from the sonicated DNA (Figure 4.9) also displays low read numbers upstream of the ATG. If all regions of the genome were equally mappable and sequenced with equal efficiency an even line would be present. Its absence suggests that the reason for the lack of obvious nucleosomes upstream of ATG may indeed be due to problems with AT content.

Micrococcal nuclease also has a preference for cutting AT di-nucleotides (Chung *et al*., 2010). With the *Dictyostelium* genome being highly AT rich, it was important to demonstrate the nucleosome peaks observed were due to the presence of nucleosomes protecting DNA rather than an artifact of MNase sequence preference or sequencing approach. Naked DNA was digested with MNase and sequenced identically to the MNase digestion of chromatin. The

nucleosome repeat pattern observed in Ax2 MNase digest (Figure 4.5) was absent in the MNase control (Figure 4.9) demonstrating nucleosome patterns observed are not artifacts of the method or the underlying DNA sequence. Again in the MNase naked DNA control there was little read/'dyad' frequency upstream of the ATG in the presumed promoter region. However, normalising dyad frequencies from nucleosomal sequence reads against the MNase-digested naked DNA control (Figure 4.9) reveals limited nucleosome phasing in the 5' promoter regions of genes, with -1 and -2 nucleosome peaks (Figure 4.9). This demonstrates that when nucleosomes are present in promoters they are phased. The -1 and +1 nucleosome dyads are separated by ~300bp, an apparent nucleosomal free/depleted region (NFR). Similar nucleosome-depleted promoter regions have been observed prominently in *S.cerevisiae* (Lee *et al*., 2007) and to some extent in *D.melanogaster* (Mavrich *et al*., 2008).

Dictyostelium has a median intergenic distance of 737bp, which is relatively large when compared to 442bp in *S.pombe* and 336bp in *S.cerevisiae* (Lantermann *et al*., 2010). In an alignment of subsets of *Dictyostelium* genes with medians of 442bp and 336bp more nucleosome structure upstream of the ATG start site is observed (Figure 4.10). This suggests that *S.cerevisiae* has such well-structured regions 5' of the NFR at least partly due to the proximity of the next upstream gene.

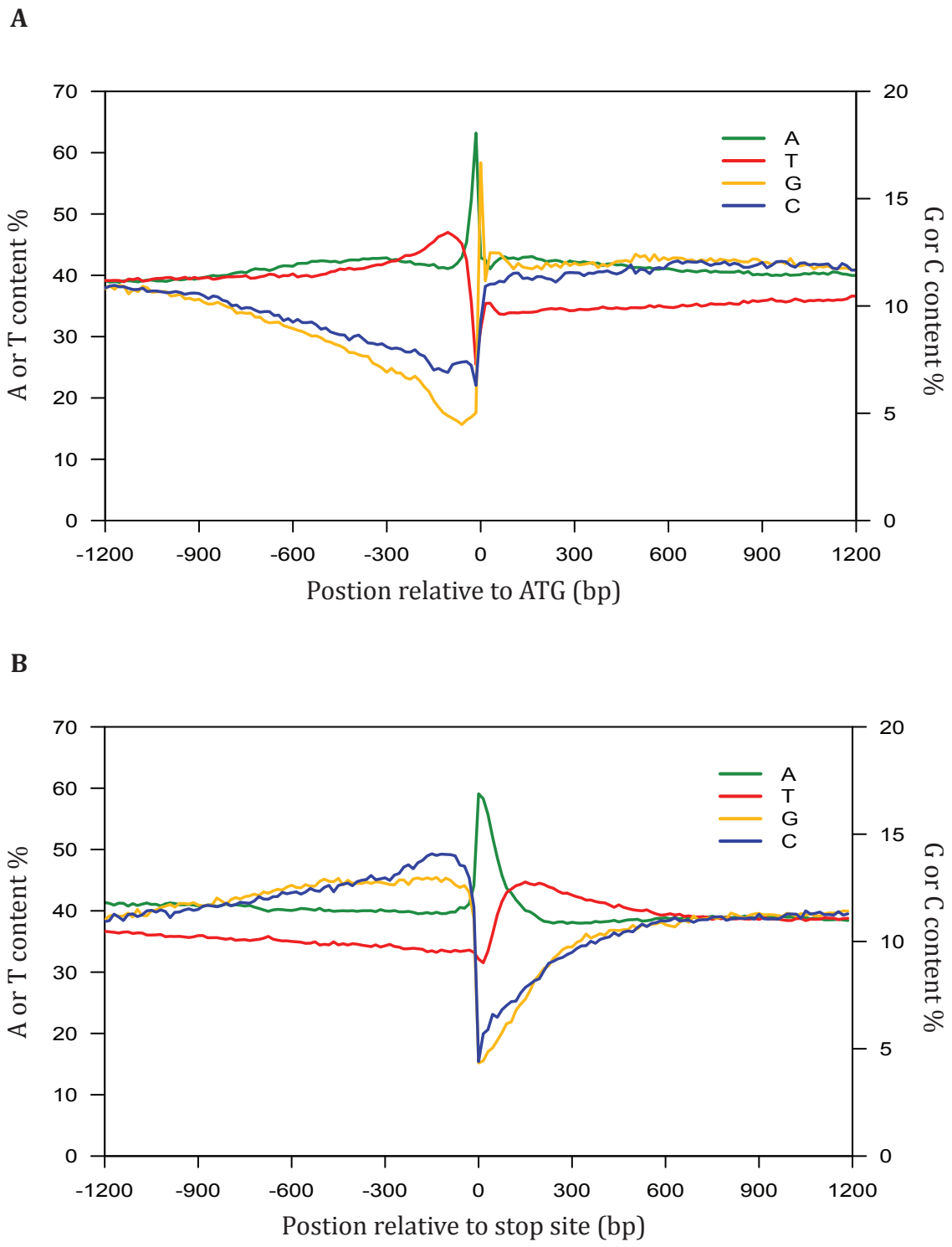


Figure 4.8: ATGC content in *Dictyostelium*

CFD graph showing ATGC content relative to **A.** ATG start site and **B.** stop site. A, T, G and C content was calculated for a 15bin window, from -1200 to +1200 relative to the feature.

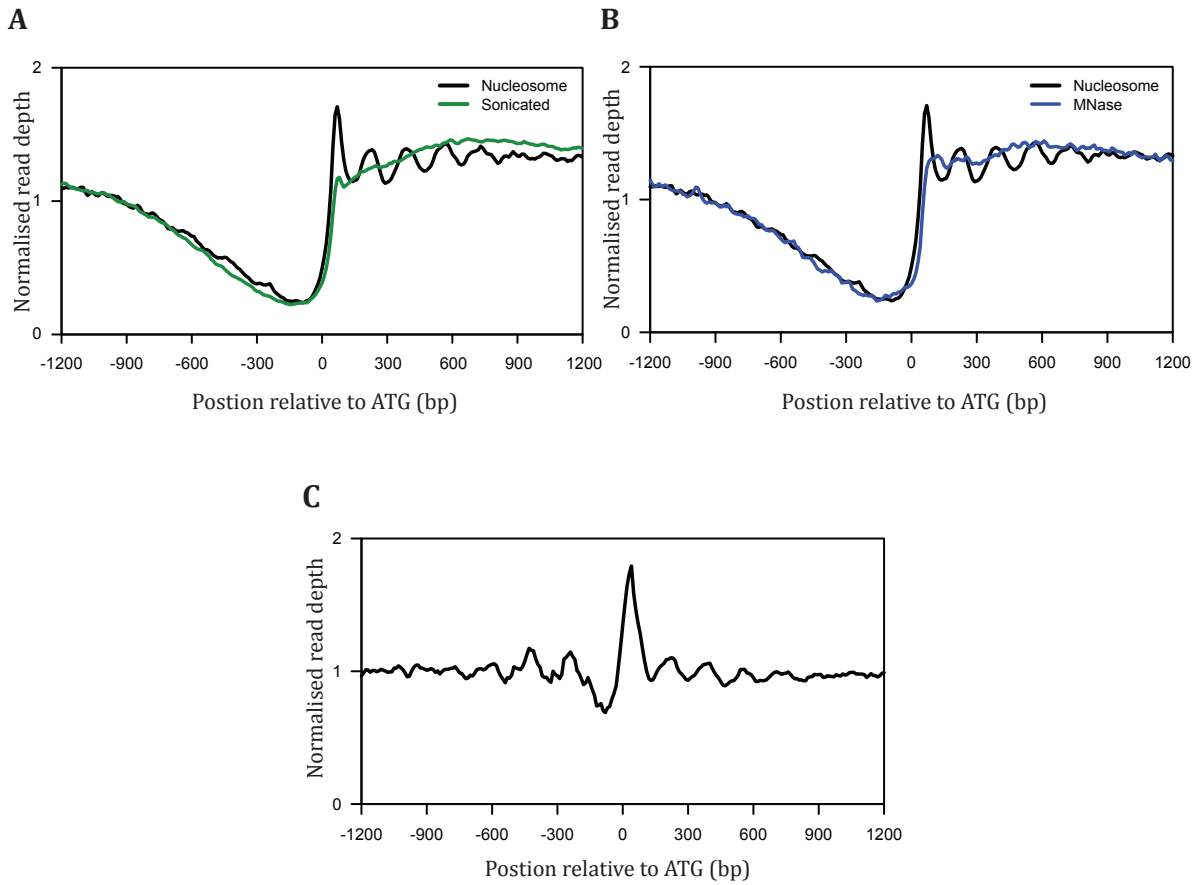


Figure 4.9: Controls for nucleosome mapping

A. CFD graph of Ax2 nucleosomes (black) around the ATG start site and 'nucleosome' dyads derived from sonicated naked DNA (green) sequenced identically to MNase digested chromatin. **B.** CFD graph of Ax2 nucleosomes (black) around the ATG start site and from 'nucleosome' dyads derived from naked DNA digested with MNase (blue) sequenced identically to MNase digested chromatin. **C.** CFD graph from nucleosome dyads from Ax2 growing cells normalised to dyads from MNase digested naked DNA from B.

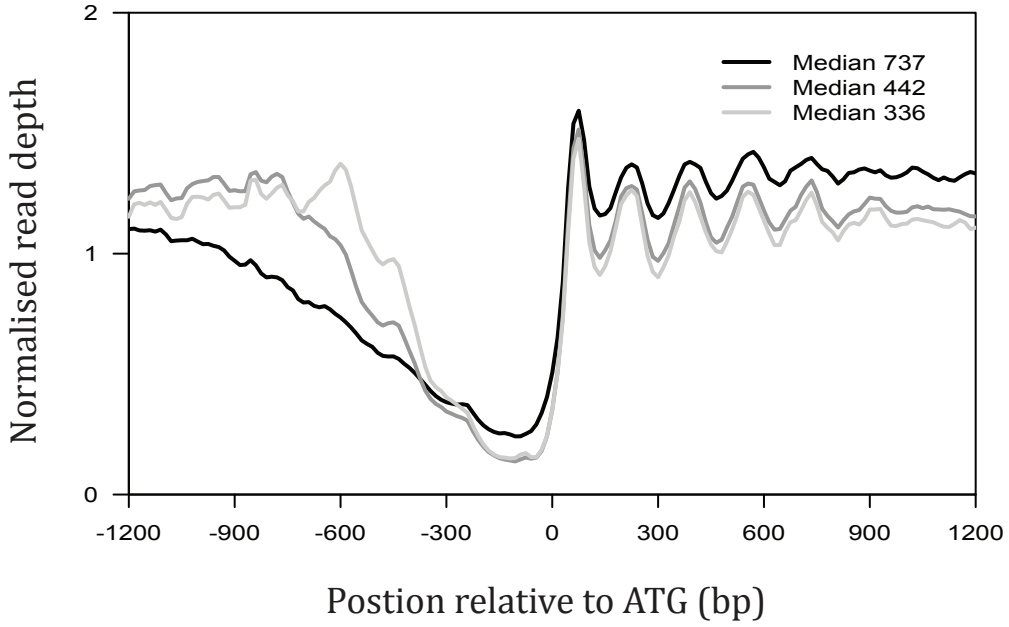


Figure 4.10: Intergenic length in *Dictyostelium*

CFD graph for nucleosome dyads around the ATG start site with subsets of the Dictyostelium genome with median intergenic distances of 737bp, 442bp and 336bp.

4.6 Relationship between nucleosome positioning and gene expression

To begin to understand nucleosome structure and gene expression, transcriptional activity was compared to nucleosome structure, correlating levels of accumulated mRNA from RNA-seq data to nucleosomal maps generated from the same cells. Genes were ranked by expression level (FPKM) in a heatmap from most highly expressed to zero expression (Figure 4.11). Genes were also split into five groups and CFD graphs produced for each quintile (Figure 4.11). The most highly expressed (top quintile) gene class exhibited the most prominent nucleosome peak heights with a well defined +1 nucleosome, whereas the gene class with very low or no detectable (bottom quintile) expression had less defined peak-height reads and a weaker +1 nucleosome. Similar transcriptional/chromatin linkages have been described in other organisms (Lantermann *et al.*, 2010, Teves and Henikoff, 2011). All the expression classes had similarly spaced nucleosomal repeat lengths of ~165bp. This increased nucleosome pattern in expressed genes against unexpressed genes suggests during gene expression nucleosomes are required to be well positioned and although nucleosomes are present on unexpressed genes their exact positioning relative to the start of the gene is unimportant.

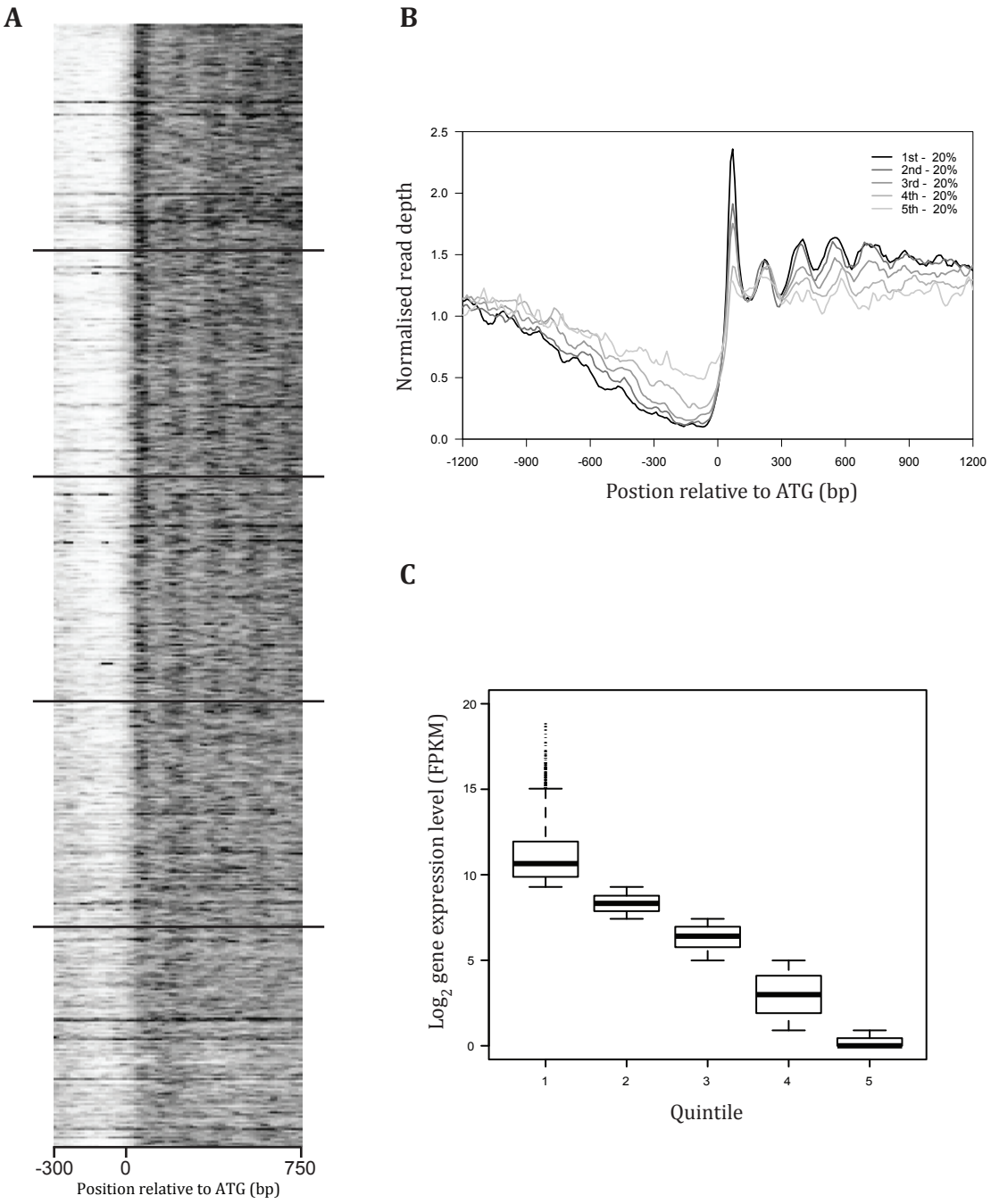


Figure 4.11: Highly expressed genes have a stronger nucleosome phasing than unexpressed genes

A. Genes were sorted by their expression level (FPKM) as determined by RNA-seq on growing Ax2 cells and nucleosome positions displayed in a heatmap format. **B.** Sorted genes from A. were divided into quintiles. A CFD graph for each quintile is shown. **C.** Box plot showing the range of $\log_2$ FPKM values in each quintile in B.

4.7 Nucleosome changes during development

Chromatin is important for regulation of gene expression. Further to mapping nucleosome positions in growing cells nucleosomes were mapped at the loose mound stage (Figure 4.12). From growing cells to loose-mound stage, greater than 60% of all genes change expression by >2 fold (Figure 4.13). Globally nucleosome patterns at the start and end of genes are similar to that of growing cells (Figure 4.12). Although there was a small increase in repeat length from 165bp in growing cells to 169bp in loose mound cells, similar to previous observation in *Dictyostelium* (Chang *et al*., 2012). Although nucleosome differences are observed in nucleosome structure at individual genes, little change in nucleosome structure is observed in genes that have >2-fold change in gene expression (Figure 4.13).

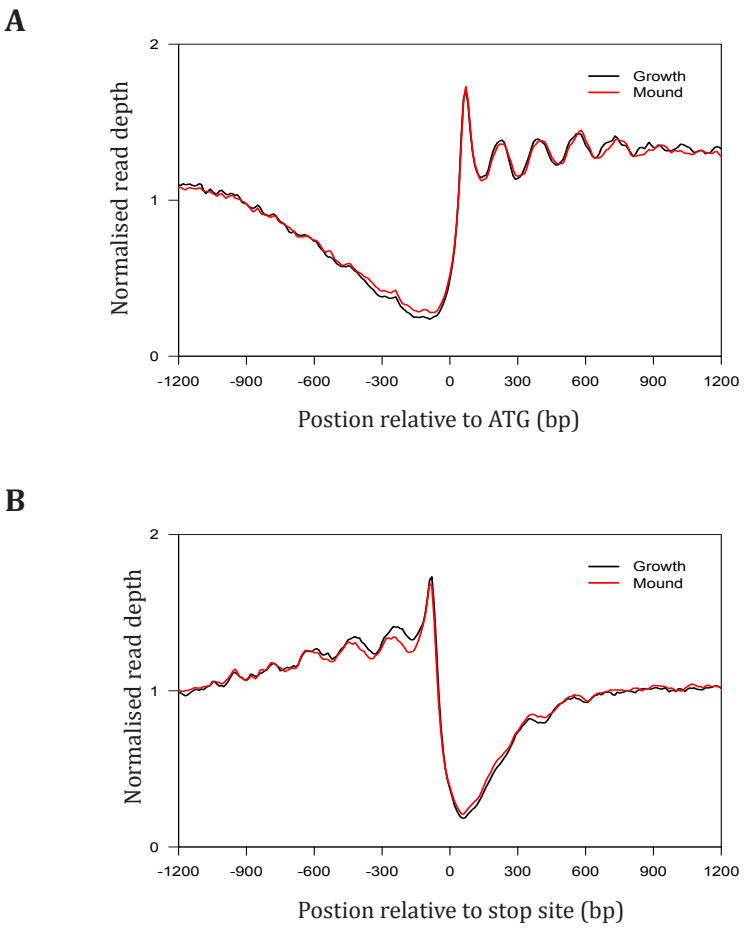


Figure 4.12: Nucleosome positions at loose-mound stage of development

CFD graph comparing nucleosome positions in growing cells and loose-mound stage cells in relation to **A.** ATG start site and **B.** translational stop site.

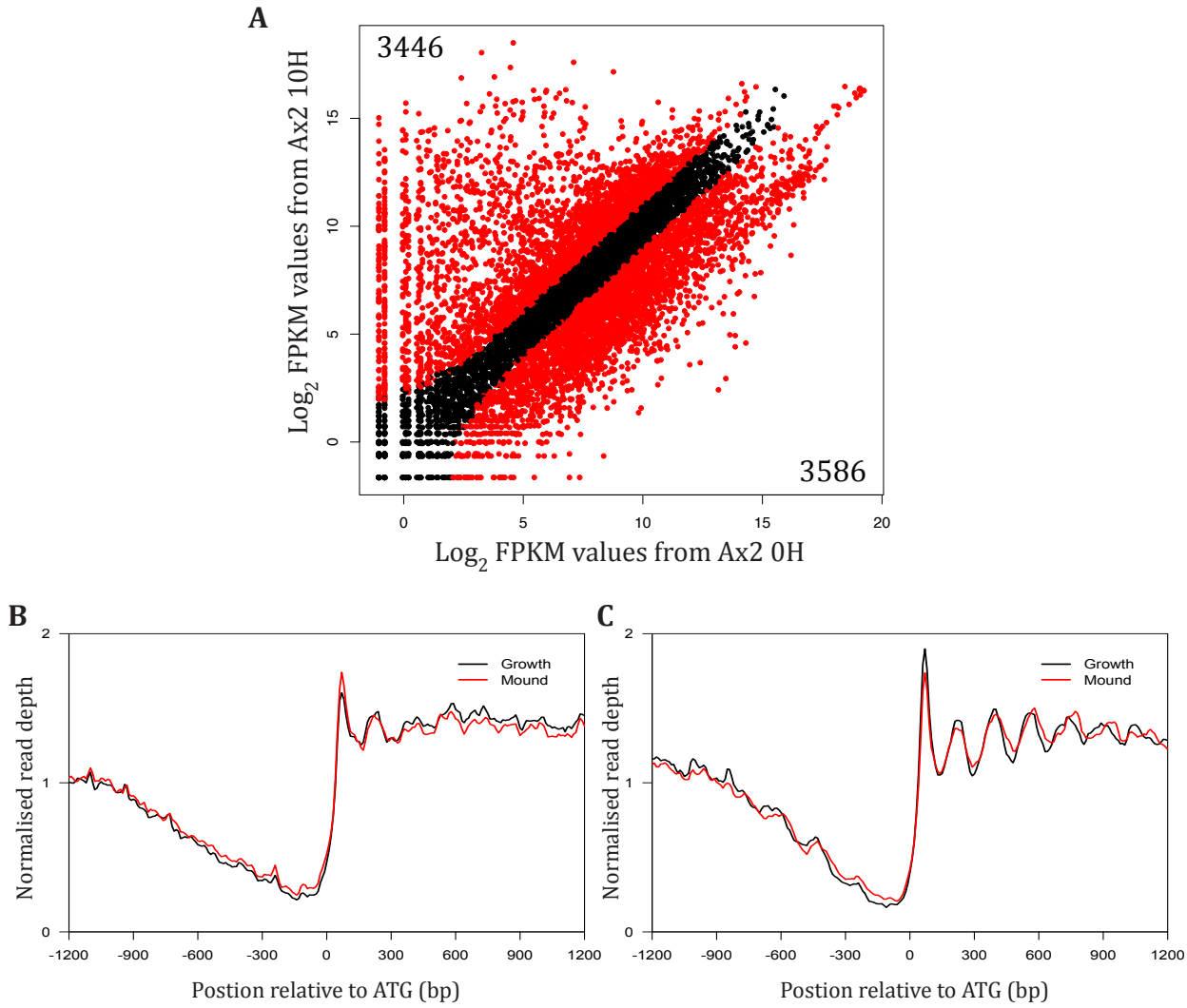


Figure 4.13: Changes in gene expression through development have little affect on nucleosome structure

A. Scatter plot of log_2 FPKM expression values calculated for growing Ax2 cells against loose mound stage (10 hours post starvation) cells. Black points are genes with identical expression values between the two cell lines, <2-fold change or p>0.05. Red points are genes that are differentially expressed between the two cell lines, ≥ 2-fold change and p≤0.05. CFD graph of nucleosomes from growing cells (black) and loose-mound stage cells (red) of **B.** 4,059 genes that are upregulated ≥2 fold (p ≤ 0.05) in loose-mound stage cells relative to growing cells and **C.** 3,703 genes that are downregulated ≥2 fold (p ≤0.05) in loose-mound stage cells relative to growing cells.

4.8 Discussion: Nucleosomes are positioned in *Dictyostelium*

Utilising modern nucleosome mapping techniques, I was able to recapitulate early work on the ribosomal (rDNA) palindrome (Ness *et al*., 1983) and observed many comparable findings to Chang *et al*. (2012) using a different sequencing approach demonstrating the utility and reproducibility of these high-throughput approaches to mapping nucleosomes. Nucleosome mapping was also highly reproducible across the two experimental replicates.

Nucleosomes are well phased relative to the start of genes. Such phasing has been observed in most eukaryotic organisms that have been studied (Albert *et al*., 2007, Lantermann *et al*., 2010, Schones *et al*., 2008, Mavrich *et al*., 2008, Li *et al*., 2011, Field *et al*., 2009) and recently in archaea (Ammar *et al*., 2012). Normalising nucleosome dyads to 'dyads' from a naked DNA MNase control revealed a classic nucleosome-free region and upstream nucleosomes as observed in other species (Lee *et al*., 2007, Mavrich *et al*., 2008). When genes were grouped by gene expression, highly expressed gene classes displayed better nucleosome phasing and a more prominent +1 nucleosome than genes that were not expressed, although repeat length was invariable. This linkage of nucleosome structure and gene expression has been observed in other organisms (Lantermann *et al*., 2010, Teves and Henikoff, 2011). Although perhaps counterintuitive, the data suggests the most actively transcribed genes generally are better phased than untranscribed genes despite the polymerase passing over the gene. The question: is the nucleosome phasing at transcribed genes a consequence of transcription or a requirement? There have been

suggestions that transcription is required to place the +1 nucleosome and that the other nucleosomes position themselves relative to that (Hughes *et al*., 2012) although nucleosome arrays have been reconstituted in vitro with whole cell extract and ATP in the absence of transcription (Zhang *et al*., 2011).

Growing cells have an average repeat length between dyad centres of 165bp, with the nucleosome core being ~147bp. This therefore assumes an average linker DNA length of ~18bp. This repeat length increases to 169bp in 10-hour developed cells, with an assumed average linker length of 22bp. It is as yet unclear why the linker length increases during development. The repeat length has been shown to vary widely between different organisms (Felsenfeld and Groudine, 2003), and even in different tissues of the same organism (Valouev *et al*., 2011).

Work of others using salt gradient dialysis to reconstitute nucleosome positions in vitro showed nucleosome positions are not due to intrinsic properties encoded by the DNA sequence (although NFRs are to some extent formed). Whole cell extract and ATP were required to recapitulate much of the nucleosome pattern (Zhang *et al*., 2011). In fact, recent work (Hughes *et al*., 2012) demonstrated in vivo that repeat length is dependent on species-specific cellular factors (remodellers, etc.). By inserting a yeast artificial chromosome (YAC) of *Kluyveromyces lactis* DNA into *S.cerevisiae*, nucleosomes on the YAC adopted the *S.cerevisiae* repeat length (~165bp) rather than the *K.lactis* repeat length (~178bp) (Heus *et al*., 1993).

Hence, the increase in repeat length in *Dictyostelium* cells as they progress though development must be linked to changes in chromatin remodellers, linker histones or other chromatin binding factors in the nucleus. In the previous chapter Chd proteins were identified as being developmentally regulated and developmental regulators of gene expression. In the next chapter developmental regulation of nucleosome positioning by one chromatin remodeller will be investigated.

Chapter 5:

Genome-wide nucleosome alterations in *chdC*-null

5.1 Introduction

Previous chapters demonstrated Chd proteins regulate distinct stages of development in *Dictyostelium* and are required for the expression of large and separate subsets of genes in both growing and developed cells. I then developed a technique to map nucleosome positions on a global scale, showing that nucleosomes were well organised in *Dictyostelium*. Growing cells had a 165bp nucleosome repeat length that increased at loose mound stage to 169bp.

Work from others has demonstrated a role for CHD1 in replacement and stabilisation of nucleosomes after transcription preventing cryptic transcription during elongation (Radman-Livaja *et al*., 2012, Smolle *et al*., 2012). A ChIP-seq experiment in *S.cerevisiae* found Chd1 binding to NFR and gene bodies. There was no relationship with steady-state gene expression levels but there was a positive correlation of binding with histone turnover and transcription rate (Zentner *et al*., 2013).

Much work has focused on CHD1 due to its ubiquitous presence in eukaryotes. Here the nucleosome remodelling properties of ChdC, a member of the third subfamily of CHD proteins, a family absent from fungi, is investigated. When knocked out it caused wide-ranging phenotypic effects and gene expression changes (Chapter 3). ChdC has closest homology to CHD7. Mutations in CHD7 cause CHARGE syndrome, a severe developmental disorder (Vissers *et al*., 2004). CHD7 has been shown in vitro to have nucleosome sliding activity that requires external DNA outside of the nucleosome, suggestive of a nucleosome spacing role

(Bouazoune and Kingston, 2012). ChdC's role in nucleosome positioning was investigated using the global nucleosome mapping method demonstrated in the previous chapter in both growing and developed cells.

Presented in this chapter is to my knowledge the first investigation into the nucleosome remodelling activity of a CHD subfamily III member in vivo using a global nucleosome mapping method.

5.2 Mapping nucleosomes in *chdC*-null

Nucleosomes were mapped in *chdC*-null cells in both growing cells and loose mound-stage cells. To mitigate any differences in nucleosome structure due to the developmental delay in *chdC*-nulls, cells were morphologically matched 10 hours post-starvation in Ax2 cells and 12 hours in *chdC*-null cells. Chromatin was digested, sequenced and plotted identically to Ax2 in Chapter 4. There was no difference in the nucleosome ladder produced from digesting chromatin from Ax2 and *chdC*-nulls, demonstrating globally nucleosomes are still present (Figure 5.1). Having sequenced and mapped all of the genome in both Ax2 and *chdC*-null cells, the data could be used to demonstrate the absence of the region previously targeted to produce a knockout mutant for the *chdC* gene (Figure 5.2).

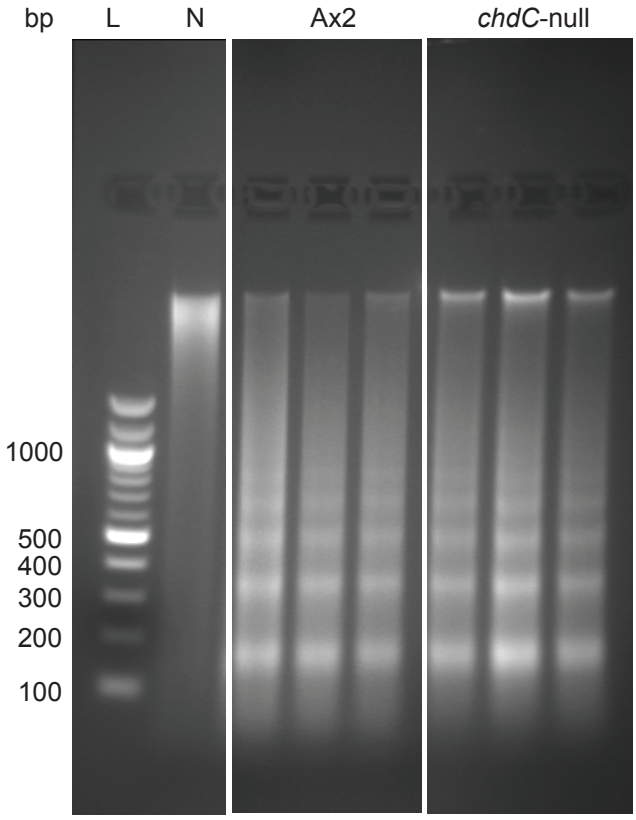


Figure 5.1: Micrococcal nuclease ladders in Ax2 and *chdC*-null cells

Micrococcal nuclease digestions from Ax2 and *chdC*-null cells produce identical nucleosome ladders. L - ladder. N -naked DNA digest.

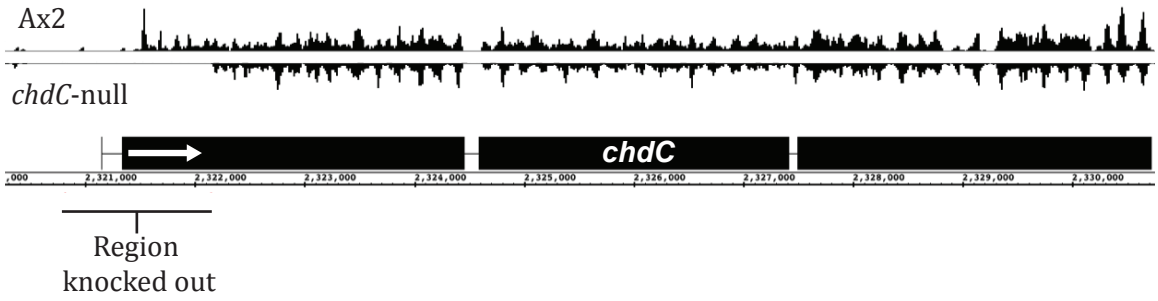


Figure 5.2: Sequencing demonstrates loss of 1kb region in *chdC* gene.

Nucleosome dyads across the *chdC* locus, in *chdC*-null cells reads originating from the region knocked out are absent.

5.3 Global nucleosome patterns in *chdC*-null are similar to Ax2

Only modest differences in overall nucleosome organisation were observed in *chdC*-null cells when all 12,750 genes were aligned to the ATG start site and stop site (Figure 5.3) in both growing and loose mound stage cells. Nonetheless, there were small, reproducible differences in the phasing of the averaged global nucleosome positions at the start of genes within the coding regions.

5.4 A subset of genes have altered nucleosome structure in *chdC*-nulls

To identify potential nucleosomal modelling changes in specific loci that resulted from the loss of ChdC, a custom algorithm was used to interrogate nucleosome position differences in both growing and loose mound stage cells. A threshold of ≥3 nucleosome position differences between Ax2 and *chdC*-nulls that occur within 1Kb of the ATG start site was selected, as nucleosome differences appeared to be limited to the 5' end rather than the 3' end of genes (Figure 5.3). This stringent threshold was set to reduce false positive nucleosome differences caused by noise. Using this threshold a conservative estimate of the number of differentially remodelled genes was calculated.

Using these criteria, 1,685 genes in growing cells were identified as reproducibly differentially organised, in the biological and technical replicates, between Ax2 and *chdC*-null cells. These genes correspond to ~13% of the total number of genes. When this subset of genes was aligned to the ATG start site and visualised using a CFD graph (Figure 5.4A), a bigger difference than all genes (Figure 5.3)

was observed. A larger difference was observed when genes were aligned based on the mid-point of the +1 nucleosome of each gene in Ax2 cells (Figure 5.4B). Nucleosomes in this subset of genes remain well phased in the *chdC*-null although the distance between nucleosome peaks increases from 165bp in Ax2 cells to 180bp in *chdC*-nulls. This results in a progressive shift in nucleosome positions for each nucleosome. Comparison to the genes that were below the threshold for remodelling, i.e. not altered in *chdC*-null, demonstrates a clear difference (Figure 5.4C). Using the same analysis method with loose mound-stage nucleosomes, 1,964 genes were identified as reproducibly differentially organised in *chdC*-null cells. An increase in repeat length was again observed, on top of the increased repeat length of Ax2 loose mound-stage nucleosomes (Figure 5.4A and B).

When the genes with altered nucleosome structure in growing cells were compared with those from loose mound-stage cells, a small but significant overlap of 578 genes was observed ($p<2.2x10^{-16}$, hypergeometric test). Although some genes have altered nucleosome organisation, there is clearly a large number of genes that are altered in only one stage of development. Chromatin from these subsets of genes (altered in growth; altered in loose mound; altered in growth and loose mound) were plotted in CFD graphs using nucleosomes from both growing and loose mound-stage cells, this clearly demonstrated that ChdC is spacing nucleosomes in different subsets of genes throughout development (Figure 5.5). This ChdC-dependent chromatin remodelling was found not to be enriched on specific chromosomes (Figure 5.6).

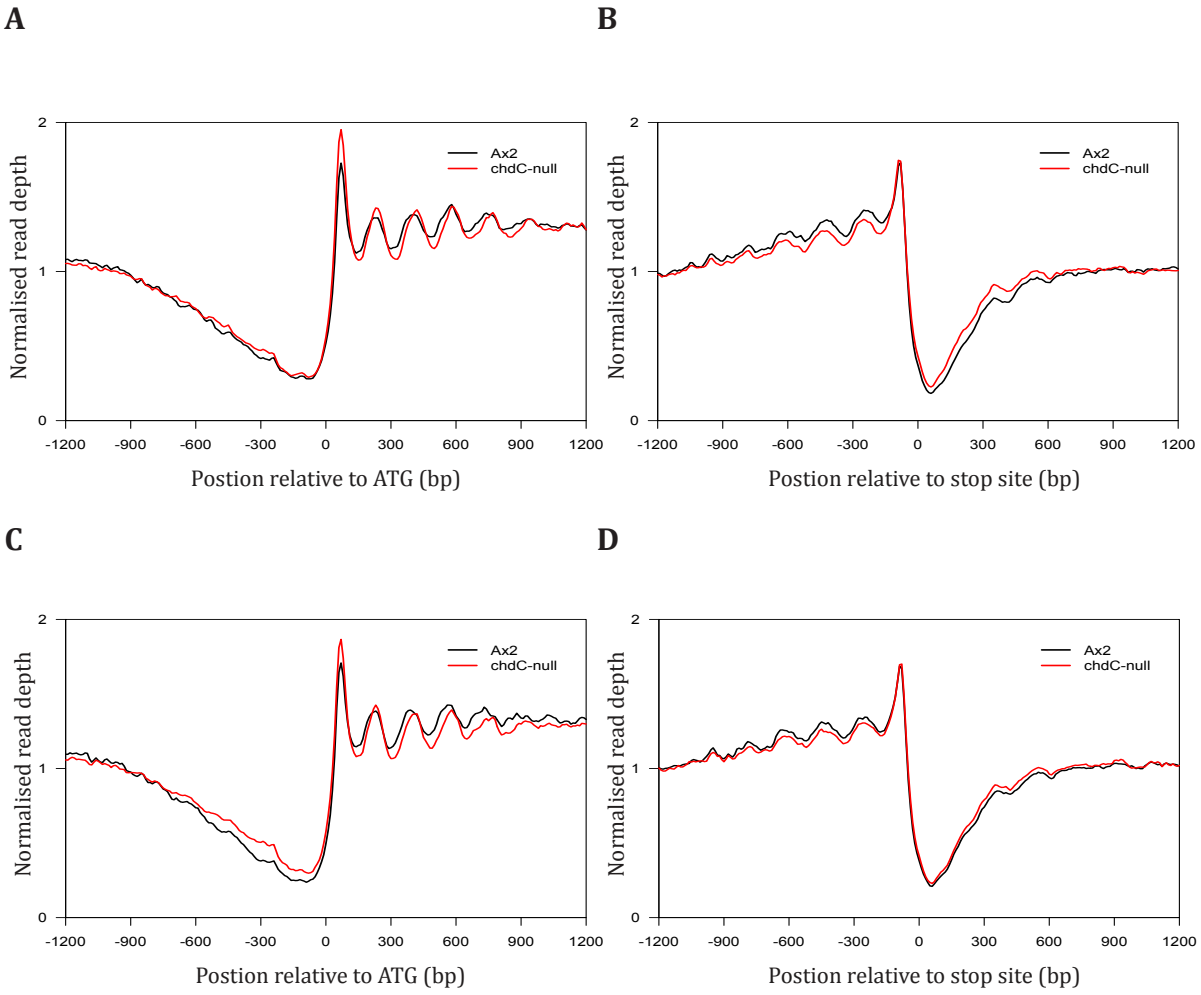


Figure 5.3: Nucleosome patterns are globally similar in *chdC*-nulls.

Cumulative frequency distribution graph for nucleosome dyads in Ax2 (black) and *chdC*-null (red) cells around the **A.** ATG start site and **B.** translation stop site of 12,750 protein-coding genes in growing cells. Cumulative frequency distribution graph for nucleosome dyads around the **C.** ATG start site and **D.** translation stop site of 12,750 protein-coding genes in loose-mound-stage cells.

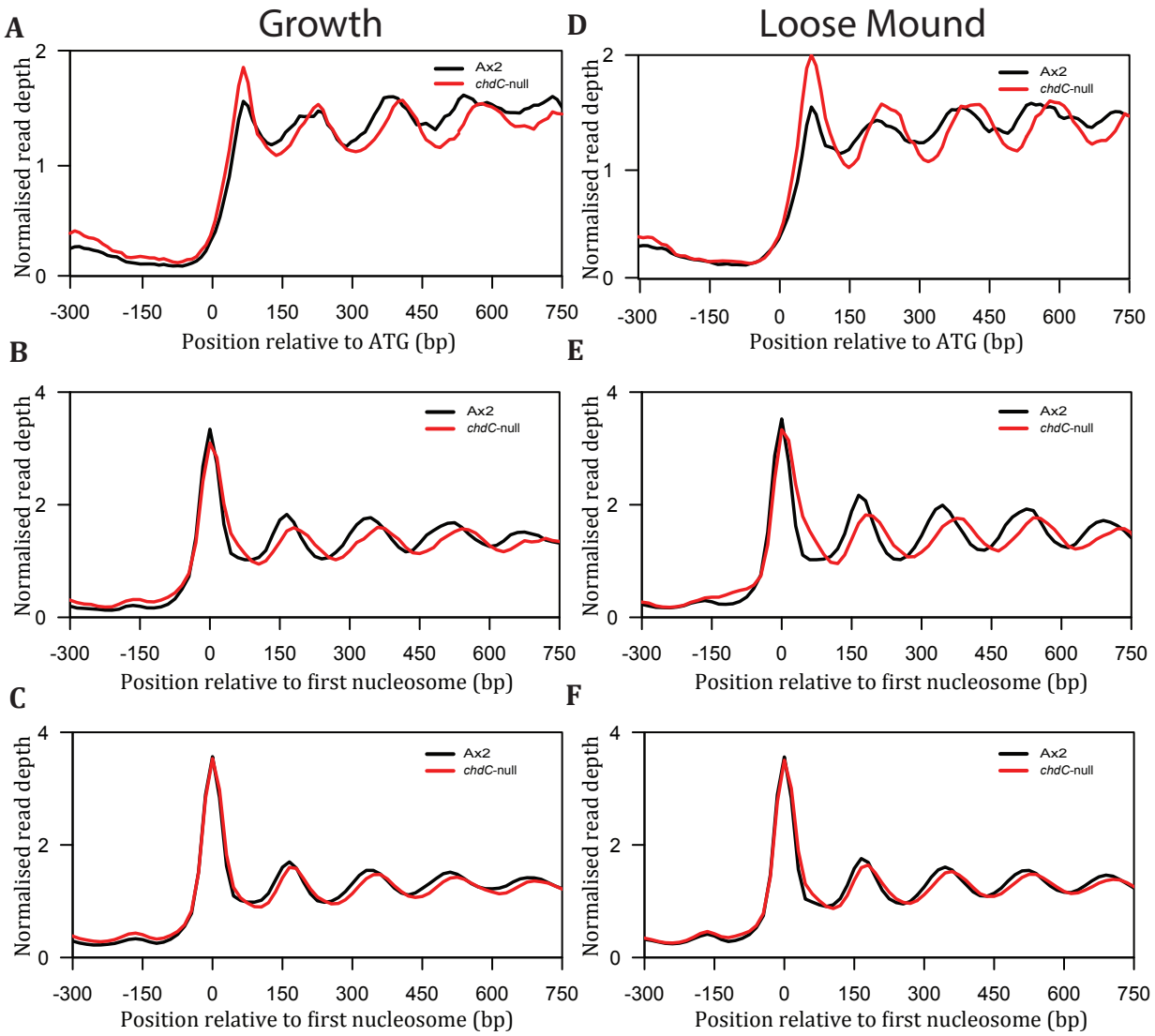


Figure 5.4: Increased nucleosome repeat length in a subset of genes in *chdC*-nulls

Cumulative frequency distribution graphs comparing nucleosome dyads from Ax2 (black) and *chdC*-null (red) cells. **A.** 1,685 genes that were identified as remodelled in growing cells aligned to the ATG start site. **B.** 1,685 genes that were identified as remodelled in growing cells aligned to the mid-point of their first defined nucleosome in Ax2. **C.** The genes that were identified as not remodelled in growing cells aligned to the mid-point of their first defined nucleosome in Ax2. **D.** 1,964 genes that were identified as remodelled in loose mound-stage cells aligned to the ATG start site. **E.** 1,964 genes that were identified as remodelled in loose mound-stage cells aligned to the mid-point of their first defined nucleosome in Ax2. **F.** The genes that were identified as not remodelled in loose mound-stage cells aligned to the mid-point of their first defined nucleosome in Ax2.

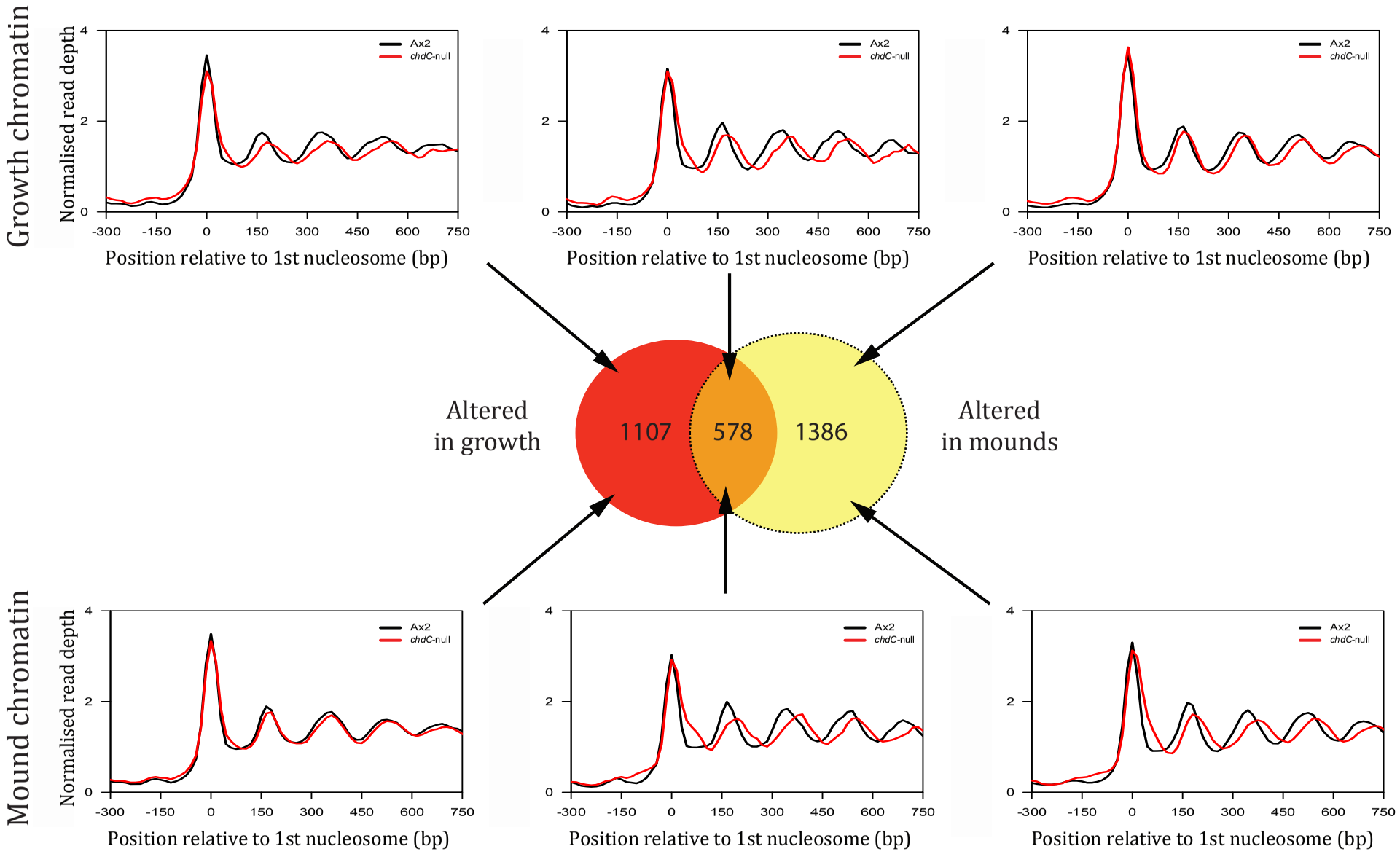


Figure 5.5: ChdC control of nucleosome structure is developmentally regulated - See legend on next page.

Figure 5.5: ChdC control of nucleosome structure is developmentally regulated

A Venn diagram displaying the overlap of genes that were identified as having altered nucleosome structure in growing and loose mound-stage cells. Cumulative frequency distribution graphs comparing nucleosome dyads from Ax2 (black) and *chdC*-null (red) cells from each subset of genes. Top, nucleosome dyads from growing cells. Bottom, nucleosome dyads from loose mound-stage cells. The overlap is significant ($p<2.2x10^{-16}$). The hypergeometric test for significance was used.

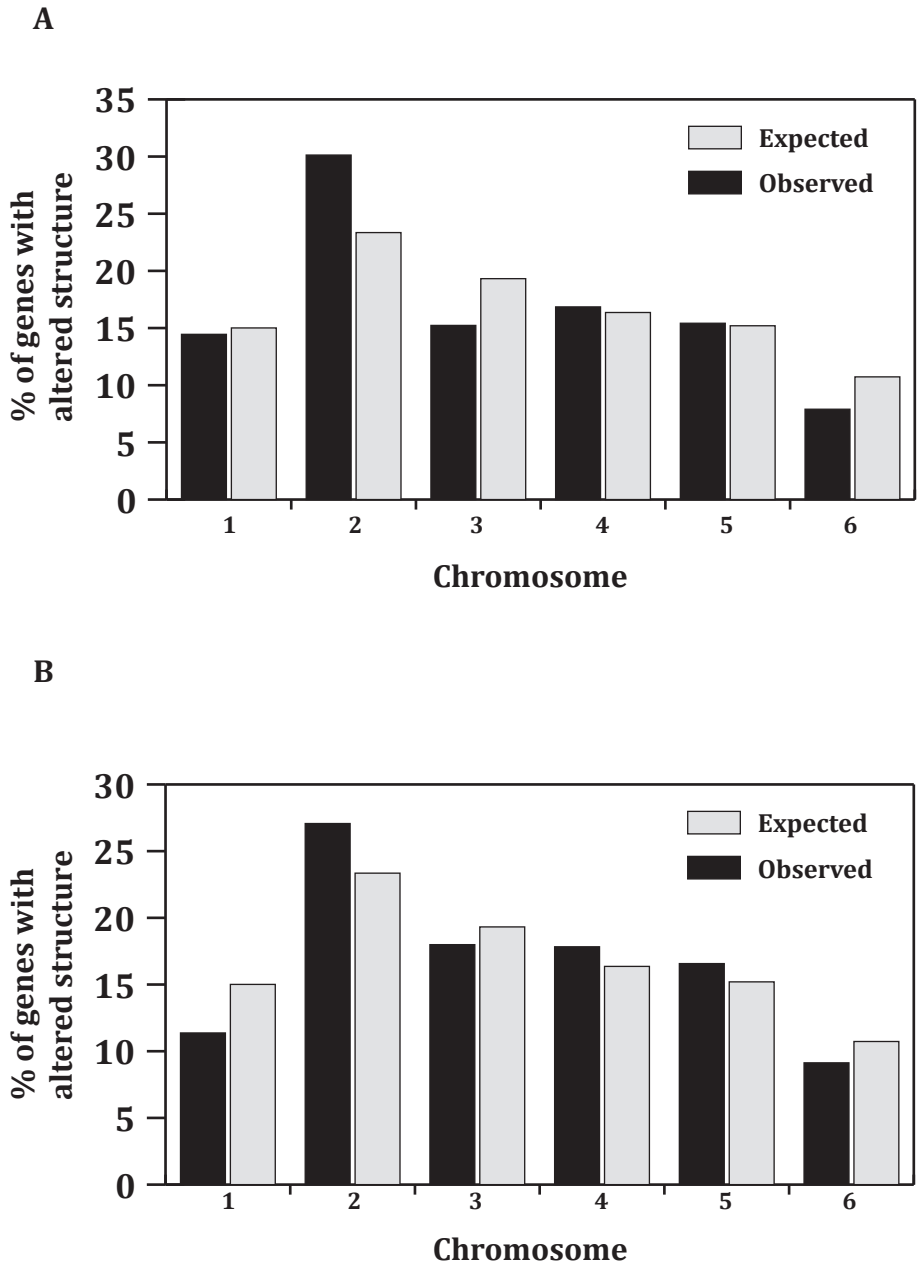


Figure 5.6: Genomic locations of genes with altered nucleosome structure

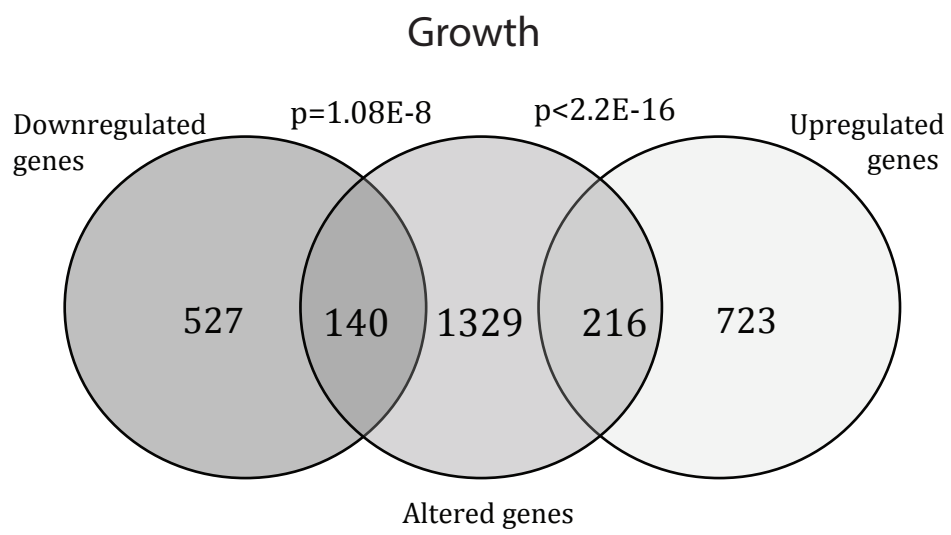
The number of genes that were identified as having altered nucleosome structure in *chdC*-nulls present on each chromosome, showing observed against expected values in **A.** growing cells and **B.** loose mound-stage cells.

5.5 Genes with altered chromatin structure are enriched for gene expression changes

Since aberrant chromatin remodelling and therefore misplaced nucleosomes may alter gene expression patterns, it was of interest to determine if altered genes in *chdC*-nulls exhibited specific differences in transcription during growth and development. To examine this, comparative high-throughput transcriptome sequencing (RNA-seq) of growth and loose mound-stage cells from Ax2 and *chdC*-null strains from Chapter 3 was used. To investigate the relationship between chromatin organisation and gene expression in the absence of ChdC, genes that had altered nucleosome structure and were misexpressed (≥2 fold, $p \leq 0.05$) in *chdC*-nulls were compared (Figure 5.7). In growing cells, the ~80% of misregulated genes did not have altered nucleosome structure, and similarly ~80% of genes with altered nucleosome structure are not misregulated. Although altered chromatin organisation is not a definitive predictor of altered gene expression, there is a highly statistically significant correlation with both upregulated and downregulated genes (Figure 5.7). A similar relationship was observed at the loose mound stage, although only genes that were upregulated were enriched for altered chromatin organisation (Figure 5.7). Gene expression changes at this stage are complicated by developmentally dependent gene expression changes. Misregulated genes that don't display altered chromatin organisation might be explained by secondary gene expression changes. However, there were also a substantial number of genes with altered chromatin organisation without changes in gene expression.

To ensure cumulative changes in repeat length observed are representative of individual genes, individual genes were studied. Previously, the inositol synthesis gene *ino1* that is upregulated at mound stage was shown by chromatin immunoprecipitation (ChIP) to be bound by ChdC (Rogers, 2010). Encouragingly, *ino1* was in the subset of genes at loose mound stage with altered chromatin structure and nucleosome shifting was observed (Figure 5.8), although shifting was also observed in genes with no expression changes and no expression (Figure 5.8). These results indicate that extended nucleosome spacing due to loss of ChdC can impact gene regulation, although increased nucleosome spacing is not sufficient to alter expression and may not require transcription.

A



B

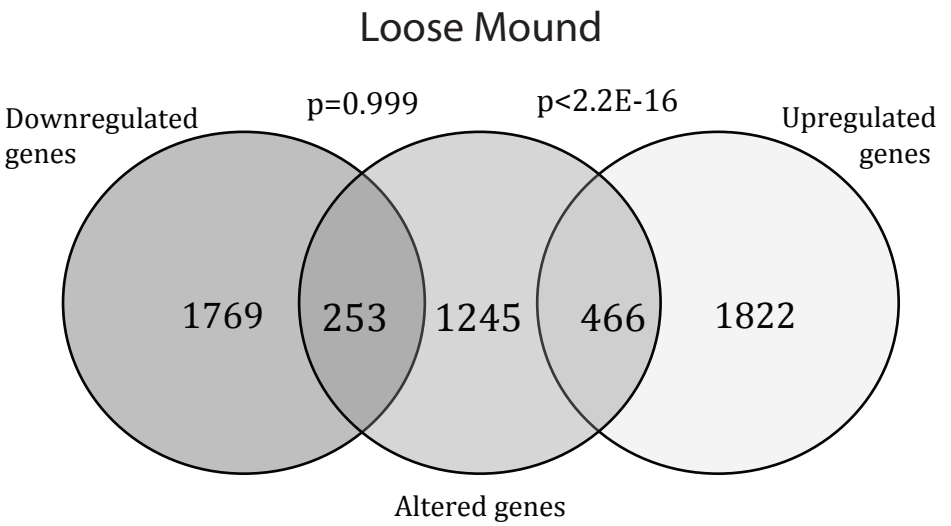


Figure 5.7: Correlation of altered nucleosome structure and gene expression changes

Venn diagram representation of the overlap between genes that are remodelled in *chdC*-nulls and genes that are upregulated and downregulated relative to Ax2 cells. **A.** Growing cells. **B.** Loose mound-stage cells. Hypergeometric test for significance was used.

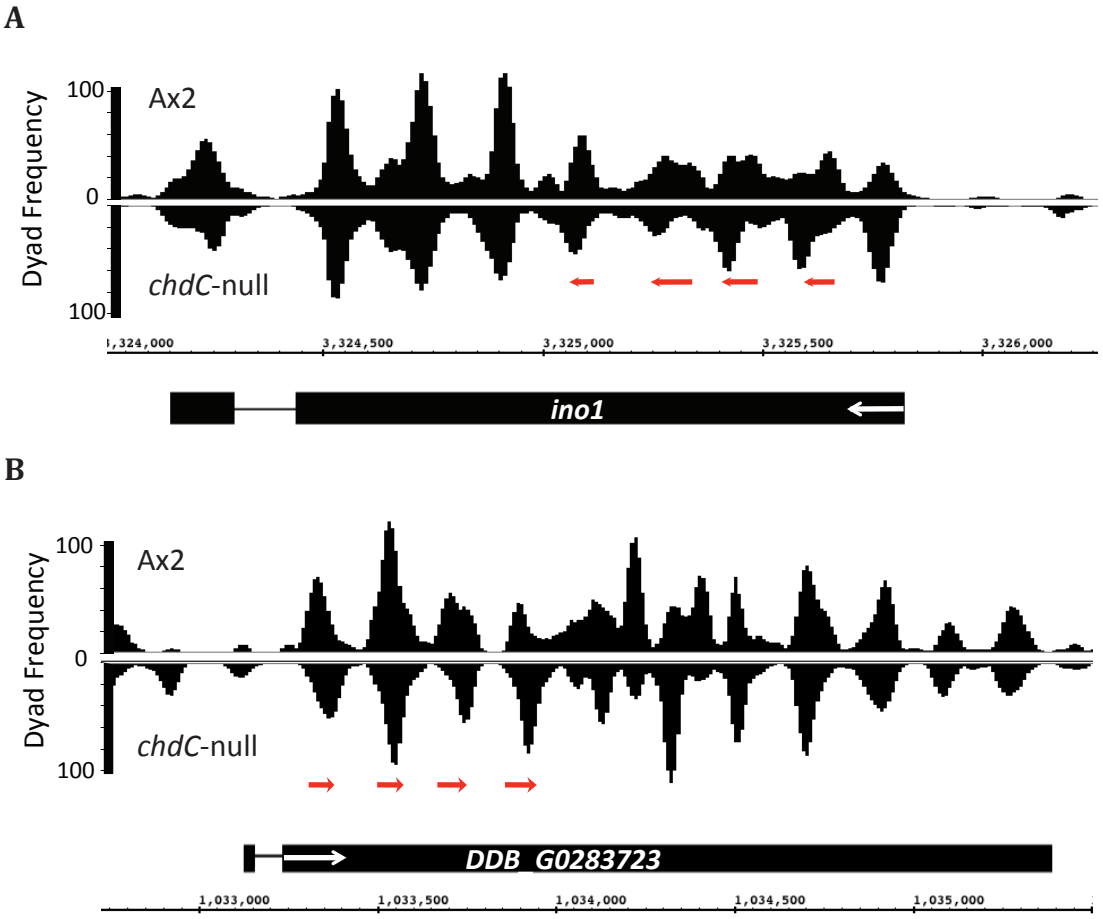


Figure 5.8: Increased nucleosome spacing is observed in individual genes

A. Genome browser trace of nucleosome dyad frequency from both Ax2 and *chdC*-null cells at loose mound stage around in the *ino1* gene, which is upregulated in *chdC*-null cells. **B.** Genome browser trace of nucleosome dyad frequency from both Ax2 and *chdC*-null cells in growing cells around the *DDB_G0283723* gene, which is not expressed in either Ax2 or *chdC*-nulls. *chdC*-null trace is plotted as a mirror image for comparison with Ax2. Arrows on gene bodies indicate direction of transcription. Red arrows indicate the direction a nucleosome has moved.

5.6 Remodelling in normal development

During the course of normal wild type (Ax2) development large changes in gene expression are observed as cells switch from vegetative growth to differentiation stages (Figure 4.13). Repeat length was observed to increase globally slightly from 165 to 169 from growth to loose mound stage (Figure 4.12), and only minor differences were observed in subsets of genes with expression changes. The algorithm used to identify differences in *chdC*-nulls was utilised, and this identified 2,856 genes with repeatable different chromatin structure in loose mound-stage cells compared to growing cells. Cumulative representation of these genes shows a flattening of all peaks and a marginal shift of nucleosomes in line with the global changes in repeat length (Figure 5.9). Flattening of the peaks during development suggests genes are being remodelled but not all in the same directional manner, in contrast to *chdC*-null, in which peak heights are maintained and the nucleosomes positioned at +2 onwards shift further into the gene (Figure 5.4).

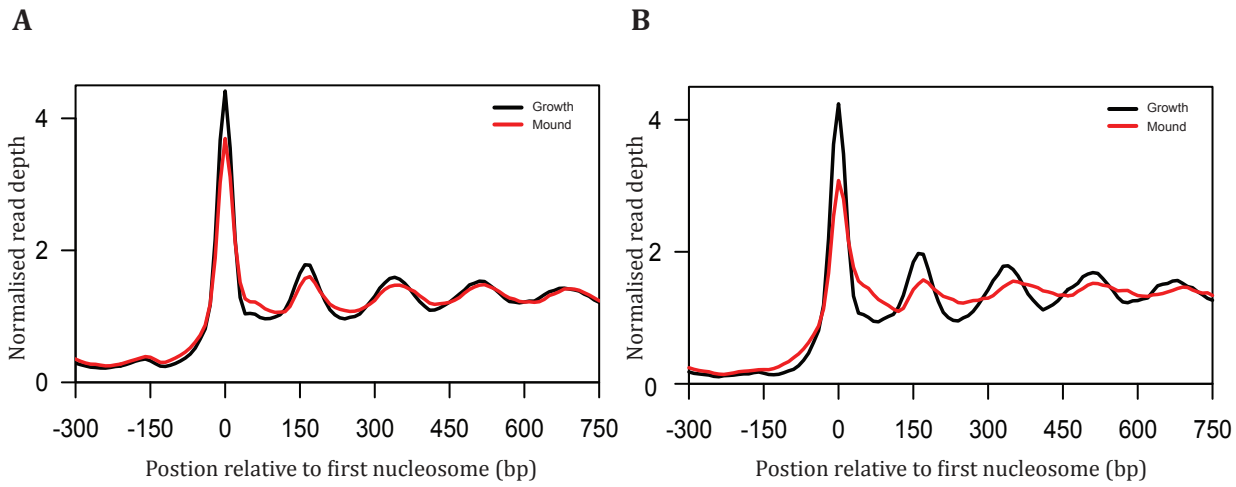


Figure 5.9: Normal chromatin remodelling from growing to developed cells

Cumulative frequency distribution graphs comparing nucleosome dyads from Ax2 cells during growth (black) and loose mound stage (red). **A.** All genes aligned to the first nucleosome in growing cells. **B.** 2,856 genes identified as remodelled at the loose mound stage relative to growing cells.

5.7 Discussion: ChdC is a developmental regulator of nucleosome repeat length in a subset of genes

This chapter described how nucleosome positions were mapped globally in *Dictyostelium* cells that lack a CHD subfamily III member, ChdC, a homologue of human CHD7, which is mutated in CHARGE syndrome (Vissers *et al*., 2004). This work demonstrates that in the absence of ChdC a subset of genes display an increased nucleosome repeat length, on average 15bp greater than Ax2. Two separate but overlapping groups (~15% of the genome each) display increased repeat lengths only in growing and loose mound stage. ChdC as well as being developmentally regulated itself (Figure 3.2) appears to be regulating nucleosome repeat length in a developmentally regulated way, responsible for maintaining 'normal' (165/169bp) repeat length in a subset of genes. To determine if the action of ChdC was restricted to distinct regions of the genome, the possibility of chromosomal enrichment of genes with altered nucleosomes was studied. However, ChdC-dependent remodelling was found not to be enriched on specific chromosomes, although the possibility these genes come together in 3D space within the nucleus cannot be excluded.

Genes that are misexpressed in *chdC*-null were enriched for altered nucleosome structure, and satisfyingly the *ino1* gene, which was previously shown to be bound by ChdC by ChIP and is upregulated in the mound stage, was in this subset of genes with increased nucleosome repeat length. Although ~80% of genes that were misexpressed had normal chromatin, some of these genes that were misexpressed but didn't display increased repeat length might be explained by

secondary gene expression changes. For example, transcription factors *stkA* (down), *bzpM* and *bzpJ* (up) were misregulated in growing cells, and *srfV*, *cinB*, *cinC* (up) and *rblA* (down) were misregulated in loose mound-stage cells. Alterations in steady state levels of transcription factors will lead to further gene expression changes, although ~80% of the genes that had nucleosome repeat length changes had no accompanying change in gene expression. Therefore, loss of normal nucleosome repeat can result in expression changes but on its own is not sufficient.

This is the first study that has investigated the in vivo genome-wide role of a CHD subfamily III member. The data here fits with evidence showing human CHD7 can slide nucleosomes in vitro (Bouazoune and Kingston, 2012). The nucleosome-sliding activity of CHD7 requires external DNA outside of the nucleosome core. It has been proposed, but not demonstrated, that CHD7 might have a nucleosome spacing role (Bouazoune and Kingston, 2012).

A wealth of evidence implicates CHD1 in the elongation complex (Stokes *et al*., 1996, Simic *et al*., 2003, McDaniel *et al*., 2008) and studies have found it to be important in the formation of the nucleosomal arrays past the +1 nucleosome (Gkikopoulos *et al*., 2011), replacement and stabilisation of nucleosomes after transcription preventing cryptic transcription (Radman-Livaja *et al*., 2012, Smolle *et al*., 2012). As in this study, alterations in nucleosome structure do not correlate well with gene expression changes (Gkikopoulos *et al*., 2011), although Zentner *et al*. (2013) did find a positive correlation of CHD1 binding (determined

by ChIP-seq) with histone turnover and transcription rate, suggesting the high histone turnover is a consequence of transcriptional elongation.

It might be presumed that all CHD proteins work in the same way to CHD1, traversing with the elongation machinery involved in replacing nucleosomes after the polymerase has passed through. It appears ChdC may act differently as nucleosomes are still present in the absence of ChdC. The data suggests ChdC may be involved in spacing nucleosomes in a subset of genes rather than replacing nucleosomes. Increased nucleosome spacing was also observed in genes that were not expressed in either wild type or mutant cells, suggesting this mechanism of nucleosome spacing by ChdC might be independent of transcription.

Chapter 6:

Discussion

6.1 Aims of the project

The aims of my project were as follows:

- To produce knockouts of the complete family of CHD proteins and study their roles in *Dictyostelium* development.
- To investigate the role of the Chd family in regulating gene expression.
- To develop a novel method to map nucleosome positions globally in *Dictyostelium*.
- To investigate the role of nucleosome positioning and gene expression.
- To investigate the role of chromatin remodellers in nucleosome positioning.

6.2 The roles of Chd proteins in *Dictyostelium* development

In this thesis I have shown that the three *Dictyostelium* Chd proteins are developmentally regulated. Chds are expressed in growth and with different development profiles. I produced knockout cell lines for each of the *chd* genes using homologous recombination. Each of the mutants displayed distinct phenotypes suggesting the Chd proteins have non-redundant function. *chdA*-null displayed a delay entering development, a chemotaxis defect and a culmination defect, persisting in the migratory slug stage for an extended period of time. *chdB*-null cells displayed a heterochrony phenotype, entering and finishing development ~3-4 hours later than wild type. *chdC*-null displayed the most severe phenotype with multiple defects. *chdC*-null cells grow poorly with a cell cycle double the length of Ax2 cells. *chdC*-nulls chemotax towards cAMP and

enter development slightly late. During aggregation, *chdC*-nulls fail to form clear streams and produce large mounds, which arrest and never proceed to the fruiting body stage. Other work showed *chdC*-nulls to have a reduced efficiency in differentiating into stalk and spore cells (Platt *et al*., 2013).

ChdC is a subfamily III chromatin remodeller with homology to human CHD7. CHD7 is mutated in 60–80% of patients diagnosed with CHARGE syndrome, a severe developmental disorder (Lalani *et al*., 2006, Janssen *et al*., 2012). Many characteristics of the syndrome have been recapitulated in $CHD7^{+/-}$ mice (Hurd *et al*., 2007, Bosman *et al*., 2005), Xenopus (Bajpai *et al*., 2010) and zebrafish (Patten *et al*., 2012), with $CHD7^{-/-}$ animals being embryonic lethal. CHD7 was shown to be important for neural crest cell formation and migration in vitro (Bajpai *et al*., 2010). CHD7 is critical for olfactory neural stem cell proliferation and regeneration of olfactory sensory neurons (Layman *et al*., 2009). Remarkably, *chdC*-nulls display strikingly parallel phenotypes to CHD7 mutants: aberrant cell movement, defective differentiation/development and slow proliferation rates. Complete knockouts and knockdowns of CHD7 are lethal in higher eukaryotes (Hurd *et al*., 2007, Bajpai *et al*., 2010). Without the unique uncoupling of growth and development cycles in *Dictyostelium*, *chdC*-nulls would also be lethal as they do not produce fruiting bodies or spore cells. In addition to being a molecular and genetic model, *chdC*-null may also provide a biological model for CHARGE syndrome in *Dictyostelium*.

Overall I demonstrated that Chd proteins are important regulators of development in *Dictyostelium*. This is in line with the developmental disorders

associated with mutations in CHD genes (Okawa *et al*., 2008, Vissers *et al*., 2004) and observations in higher eukaryotes that show that CHD proteins are essential for developmental programmes and that complete knockouts often result in embryonic lethality (Marfella *et al*., 2006, Khattak *et al*., 2002, Hurd *et al*., 2007). These experiments demonstrate *Dictyostelium* is a good model to study CHD function in a multicellular environment.

6.3 The roles of Chd proteins in regulating gene expression

CHD proteins are well established as regulators of gene expression both as activators and repressors (Murawska and Brehm, 2011) and are involved in various stages of the transcription cycle (McDaniel *et al*., 2008, Alen *et al*., 2002). Therefore it was a key aim of the project to begin to investigate the regulation that Chd proteins in *Dictysotelium* have on gene expression.

I demonstrated using high-throughput mRNA sequencing (RNA-seq) that each Chd member regulates the expression, directly or indirectly, of up to 15% of genes in growing cells and 30% in 5-hour cAMP pulsed cells. Large numbers of genes were both repressed and derepressed upon loss of Chd proteins, suggesting that Chd chromatin remodellers have complex roles in both activation and repression of gene expression. Interestingly, I discovered that the genes misexpressed in each of the *chd*-nulls were distinctly separate; few genes are regulated by more than one Chd remodeller, indicating that Chds display specificity for the genes they regulate. Remarkably no compensation, at the RNA level, by the other *chd* genes was observed upon the loss of another *chd* gene.

This is contrary to compensation observed between human CHD3 and CHD4 (Williams *et al*., 2004). This evidence further re-enforces the distinct roles of Chds in regulating different genes, as there appears to be no regulatory system that enables them to compensate for each other.

chdC-null has decreased expression of nuclear genes that encode mitochondrial proteins, which may cause the slow growth observed due to low mitochondrial function and therefore energy production. In chemotactically competent cells, misexpression of genes that encode components of cAMP signalling pathways was observed. In *chdC*-nulls, genes are misexpressed in all of the signalling pathways that signal cAMP. These signalling pathways are altered to a lesser extent in *chdA*-nulls. These alterations prove their chemotaxis defects are due to the inability to signal the response to cAMP. Further fitting with their inability to chemotax towards cAMP, almost all of the genes that were previously identified as being regulated by cAMP pulses (Iranfar *et al*., 2006) are downregulated both in *chdA* and *chdC*-nulls.

Despite presenting a minimal phenotype *chdB*-nulls still misexpress 12–15% of the genome at the stages of development that were tested. More phenotypic effects may appear in more specific conditions or assays, although GO term enrichment didn't suggest further possible phenotypes.

To the best of my knowledge this study is the first time that the gene specificity of an entire family of CHD proteins within an organism has been investigated. The *S.cerevisiae* genome encodes only one gene, yet the human and mouse

genomes contain nine. Due to the unknown function of many of the mammalian CHD proteins, the large disparity in CHD gene number remains unexplained. I have demonstrated that individual Chd proteins have both developmental and gene specificity, regulating different gene sets during development. This fits with tissue-specific expression of CHDs observed in other organisms (Kim *et al*., 1999, Lathrop *et al*., 2010, Thompson *et al*., 2003, Sanlaville *et al*., 2006, Shur and Benayahu, 2005).

6.4 A global nucleosome map in *Dictyostelium*

I developed a method to globally map nucleosomes in *Dictyostelium* (Platt *et al*., 2013). The method uses a combination of micrococcal nuclease digestion, an endonuclease, and high-throughput sequencing to infer the positions of nucleosomes (Kent *et al*., 2010). Using this approach I was able to recapitulate early work on the ribosomal (rDNA) palindrome (Ness *et al*., 1983). The method was highly reproducible and was confirmed by independent findings (Chang *et al*., 2012) in which a different sequencing approach was utilised. This demonstrated the utility and reproducibility of using high-throughput methods to map nucleosomes in *Dictyostelium*.

Using this approach I demonstrated that nucleosomes in *Dictyostelium* are globally positioned relative to the starts of genes, which is akin to positioning in other higher eukaryotes. Here the ATG start site was used as a proxy for the transcriptional start site (TSS). A previous approach focused on estimated TSS from only 5,468 genes, which is less than 43% of the genome (Chang *et al*.,

2012). In this study nucleosomes were analysed for all 12,750 protein-coding genes, producing similar results, although the di-nucleosome organisation within gene bodies that had previously been suggested was not observed (Chang *et al*., 2012). These observations are consistent with the view that *Dictyostelium* has animal-like chromatin structure, placing its transcription start site within the nucleosome-free region (Chang *et al*., 2012), whereas transcriptional start sites in fungi are within the +1 nucleosome.

Highly expressed gene classes display better average nucleosome phasing relative to the ATG and a more prominent +1 nucleosome than genes that are not expressed, although repeat length is invariable. This linkage of nucleosome structure and gene expression has been observed in other organisms (Lantermann *et al*., 2010, Teves and Henikoff, 2011). Although counterintuitive, the data shows the most actively transcribed genes have more not less phasing, despite the polymerase passing over the gene. It is thought that transcription is required to position the +1 nucleosome and the other nucleosomes are positioned relative to that as the elongation complex passes along the gene and replaces the nucleosomes after transcription (Hughes *et al*., 2012). However, nucleosome arrays have been reproduced in vitro with whole cell extract and ATP in the absence of transcription, also demonstrating the importance of ATP-dependent chromatin remodellers in forming nucleosome arrays (Zhang *et al*., 2011).

I determined growing cells have an average repeat length between dyad centres of 165bp, with the nucleosome core being ~147bp. This therefore assumes an

average linker DNA length of ~18bp. This repeat length increases to 169bp in 10-hour developed cells, with an assumed linker length of 22bp. It is as yet unclear why the linker length increases during development. The repeat length has been shown to vary widely between different organisms (Felsenfeld and Groudine, 2003) and in different tissues of the same organism (Valouev *et al*., 2011). Differences in DNA sequence between species have the potential to explain repeat length differences, but do not explain tissue specific changes. Recent work (Hughes *et al*., 2012) conclusively demonstrated in vivo that repeat length is dependent on species-specific cellular factors and not DNA sequence. By inserting a yeast artificial chromosome (YAC) of *Kluyveromyces lactis* DNA into *S.cerevisiae*, nucleosomes on the YAC adopted the *S.cerevisiae* repeat length rather than the *K.lactis* repeat length.

Although small, a 5bp difference in repeat length is predicted to have a significant impact on chromatin fibre folding (Correll *et al*., 2012). The reason for the increase in repeat length during *Dictyostelium* development is unclear, though it seems possible the increase is due to changes in chromatin remodellers and chromatin associated proteins during development. At the mound stage cells have begun cell fate commitment, becoming prespore and prestalk cells. Further experiments on separate prestalk and prespore populations will provide interesting insights.

6.5 ChdC is required to maintain normal nucleosome spacing

ChdC is a homologue of human CHD7, which is mutated in CHARGE syndrome (Vissers *et al*., 2004). *chdC*-null presented the most severe phenotype of the three *chd*-nulls with parallel phenotypes to CHD7 mutants, hence its selection for nucleosome mapping genome wide. Unlike Chd1 (Gkikopoulos *et al*., 2011), there was little genome-wide effect on nucleosome positioning in *chdC*-nulls. I confirmed this using comparison of individual genes. Only ~15% of genes are affected in growing cells. These altered genes display an increased average repeat length up 15bp from 165bp to 180bp. Further, another ~15% of genes were altered in loose mound-stage cells with an increased repeat length, although this was not the same gene set. This suggested ChdC maintains the normal repeat length in a subset of the genome in a developmentally regulated way.

Genes that are misexpressed in ChdC were statistically enriched for nucleosome repeat length changes, demonstrating repeat length impacts on gene expression. Although ~80% of genes that were misexpressed had normal chromatin structure. Clearly some of these genes that were misexpressed but didn't display increased repeat length can be explained by secondary gene expression changes. Various transcription factors were found to have altered expression, although ~80% of the genes that have altered structure had no accompanying change in gene expression. Further, increased nucleosome spacing was observed in genes that were not expressed in either wild type or mutant cells suggesting this mechanism of nucleosome spacing by ChdC might be independent of

transcription. Increased nucleosome spacing clearly impacts upon transcription regulation, although it remains unclear how. There may be involvement of a secondary factor at the genes with expression changes.

ChdC appears to be regulating normal nucleosome repeat length in a subset of genes but why nucleosomes revert to a longer repeat length in the absence of ChdC remains unclear. Is this due to nucleosomes moving to the most thermodynamically favorable position? Or does another chromatin remodeller have an antagonistic action to ChdC? Work in *S.cerevisiae* has shown remodellers may work together to organise nucleosome arrays by shifting nucleosomes in opposing directions (Yen *et al*., 2012).

This is the first study, to my knowledge, that has investigated the in vivo genome-wide role of a CHD subfamily III member. The data here for ChdC fits with evidence showing in vitro that the human homologue CHD7 can slide nucleosomes (Bouazoune and Kingston, 2012). CHD7 requires external DNA outside of the nucleosome core for in vitro sliding activity, which is suggestive of a nucleosome spacing role (Bouazoune and Kingston, 2012). The data I have presented suggest ChdC function is different to the described role of Chd1. Chd1 is involved in the formation of the nucleosomal arrays past the +1 nucleosome (Gkikopoulos *et al*., 2011). It replaces and stabilises nucleosomes after transcription and prevents cryptic transcription (Radman-Livaja *et al*., 2012, Smolle *et al*., 2012). ChdC appears to be acting on fewer genes than Chd1. Whether or not transcription is involved, nucleosomal arrays are still present but

are spaced further apart. Hence, ChdC doesn't appear to be involved in replacing nucleosomes but instead spacing them.

6.6 Summary

In this study, I have demonstrated the three Chd proteins in *Dictyostelium* are important regulators of development. The Chd proteins appear to regulate distinct stages of development and the expression of distinct gene sets. Further adding to their non-redundant function, the chds do not compensate for each other at the transcript level.

I also developed a method using high-throughput sequencing and micrococcal nuclease digestion to map nucleosome positions on a global scale in *Dictyostelium*. Despite high AT content, *Dictyostelium* positions nucleosomes within genic regions like other higher eukaryotes. I applied this method of mapping nucleosomes to *chdC*-null cells, which is the first, to my knowledge, in vivo investigation of chromatin remodelling in a CHD subfamily III member. ChdC appears to be developmentally regulating nucleosome spacing within a subset of genes in a regulated way, which impacts on gene expression.

Dictyostelium fills a gap in the complexity scale, with a multicellular development stage and greater level of complexity than *S.cerevisiae* whilst still possessing a compact manipulatable genome. Unfortunately chromatin regulation in *Dictyostelium* has been understudied. In this thesis I have made significant advances in these studies in *Dictyostelium*: I have provided a new tool to study

nucleosome positioning and dynamics genome wide and begun dissecting the roles of Chd proteins in regulating development, gene expression and finally nucleosome positioning.

Chapter 7:

References

Albert, I., Mavrich, T. N., Tomsho, L. P., Qi, J., Zanton, S. J., Schuster, S. C. & Pugh, B. F. (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomycescerevisiae* genome. *Nature,* 446**,** 572-576.

Alen, C., Kent, N. A., Jones, H. S., O'Sullivan, J., Aranda, A. & Proudfoot, N. J. (2002) A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol Cell,* 10**,** 1441-1452.

Allan, J., Cowling, G. J., Harborne, N., Cattini, P., Craigie, R. & Gould, H. (1981) Regulation of the higher-order structure of chromatin by histones H1 and H5. *J Cell Biol,* 90**,** 279-288.

Allen, M. D., Religa, T. L., Freund, S. M. & Bycroft, M. (2007) Solution structure of the BRK domains from CHD7. *J Mol Biol,* 371**,** 1135-1140.

Allfrey, V. G., Faulkner, R. & Mirsky, A. E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A,* 51**,** 786-794.

Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G. D., Giaever, G. & Nislow, C. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife,* 1**,** e00078.

Anders, S. & Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol,* 11, R106.

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res,* 19**,** 1732-1741.

Arents, G. & Moudrianakis, E. N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A,* 92**,** 11170-11174.

Axel, R. (1975) Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. *Biochemistry,* 14**,** 2921-2925.

Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H. & Mills, A. A. (2007) CHD5 Is a Tumor Suppressor at Human 1p36. *Cell,* 128**,** 459-475.

Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C. P., Zhao, Y., Swigut, T. & Wysocka, J. (2010) CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature,* 463**,** 958-962.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science,* 290**,** 972-977.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. & Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* 410**,** 120-124.

Bartels, C. F., Scacheri, C., White, L., Scacheri, P. C. & Bale, S. (2010) Mutations in the CHD7 gene: the experience of a commercial laboratory. *Genet Test Mol Bioma,* 14**,** 881-891.

Basu, S., Fey, P., Pandit, Y., Dodson, R., Kibbe, W. A. & Chisholm, R. L. (2013) dictyBase 2013: integrating multiple Dictyostelid species. *Nucleic Acids Res,* 41**,** D676-683.

Batsukh, T., Pieper, L., Koszucka, A. M., von Velsen, N., Hoyer-Fender, S., Elbracht, M., Bergman, J. E., Hoefsloot, L. H. & Pauli, S. (2010) CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Hum Mol Genet,* 19**,** 2858-2866.

Batsukh, T., Schulz, Y., Wolf, S., Rabe, T. I., Oellerich, T., Urlaub, H., Schaefer, I. M. & Pauli, S. (2012) Identification and characterization of FAM124B as a novel component of a CHD7 and CHD8 containing complex. *PLoS ONE,* 7**,** e52640.

Baxevanis, A. D., Arents, G., Moudrianakis, E. N. & Landsman, D. (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res,* 23**,** 2685-2691.

Baxevanis, A. D. & Landsman, D. (1996) Histone Sequence Database: A compilation of highly-conserved nucleoprotein sequences. *Nucleic Acids Res,* 24**,** 245-247.

Becker, P. B. & Hörz, W. (2002) ATP-dependent nucleosome remodeling. *Annu Rev Biochem,* 71**,** 247-273.

Berger, E. A. & Clark, J. M. (1983) Specific cell-cell contact serves as the developmental signal to deactivate discoidin I gene expression in *Dictyostelium discoideum*. *Proc Natl Acad Sci USA,* 80**,** 4983-4987.

Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. & Snyder, M. (2012) An intergrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.

Bhaumik, S. R., Smith, E. & Shilatifard, A. (2007) Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol,* 14**,** 1008-1016.

Blumberg, D. D., Agarwal, A. K., Sloger, M. S. & Yoder, B. K. (1991) Gene expression and chromatin structure in the cellular slime mold, *Dictyostelium discoideum*. *Dev Genet,* 12**,** 65-77.

Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. (2004) Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Mol Cell,* 14**,** 667-673.

Bosman, E. A., Penn, A. C., Ambrose, J. C., Kettleborough, R., Stemple, D. L. & Steel, K. P. (2005) Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet,* 14**,** 3463-3476.

Bouazoune, K. & Kingston, R. E. (2012) Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc Natl Acad Sci U S A*, 109, 19238-19243

Bouazoune, K., Mitterweger, A., Langst, G., Imhof, A., Akhtar, A., Becker, P. B. & Brehm, A. (2002) The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J,* 21**,** 2430-2440.

Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. (2004) Mi-2/NuRD: multiple complexes for many purposes. *Biochim Biophys Acta,* 1677**,** 52-57.

Brehm, A., Langst, G., Kehle, J., Clapier, C. R., Imhof, A., Eberharter, A., Muller, J. & Becker, P. B. (2000) dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *EMBO J,* 19**,** 4332-4341.

Brock, D. A. & Gomer, R. H. (1999) A cell-counting factor regulating structure size in *Dictyostelium*. *Gene Dev,* 13**,** 1960-1969.

Brock, D. A., Hatton, R. D., Giurgiutiu, D. V., Scott, B., Jang, W., Ammann, R. & Gomer, R. H. (2003) CF45-1, a secreted protein which participates in *Dictyostelium* group size regulation. *Euk Cell,* 2**,** 788-797.

Brock, D. A., van Egmond, W. N., Shamoo, Y., Hatton, R. D. & Gomer, R. H. (2006) A 60-kilodalton protein component of the counting factor complex regulates group size in *Dictyostelium discoideum*. *Euk Cell,* 5**,** 1532-1538.

Brodeur, G. M., Green, A. A., Hayes, F. A., Williams, K. J., Williams, D. L. & Tsiatis, A. A. (1981) Cytogenetic features of human neuroblastomas and cell lines. *Cancer Res,* 41**,** 4678-4686.

Brodeur, G. M., Sekhon, G. S. & Goldstein, M. N. (1977) Chromosomal-aberrations in human neuroblastomas. *Cancer,* 40**,** 2256-2263.

Burkhardt, L., Fuchs, S., Krohn, A., Masser, S., Mader, M., Kluth, M., Bachmann, F., Huland, H., Steuber, T., Graefen, M., Schlomm, T., Minner, S., Sauter, G., Sirma, H. & Simon, R. (2013) CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. *Cancer Res*, 73, 2795-2805.

Cairns, B. R., Kim, Y. J., Sayre, M. H., Laurent, B. C. & Kornberg, R. D. (1994) A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proc Natl Acad Sci U S A,* 91**,** 1950-1954.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

Carlson, M. & Laurent, B. C. (1994) The SNF/SWI family of global transcriptional activators. *Curr Opin Cell Biol,* 6**,** 396-402.

Carruthers, L. M., Bednar, J., Woodcock, C. L. & Hansen, J. C. (1998) Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. *Biochemistry,* 37**,** 14776-14787.

Carvill, G. L., Heavin, S. B., Yendle, S. C., McMahon, J. M., O'Roak, B. J., Cook, J., Khan, A., Dorschner, M. O., Weaver, M., Calvert, S., Malone, S., Wallace, G., Stanley, T., Bye, A. M. E., Bleasel, A., Howell, K. B., Kivity, S., Mackay, M. T., Rodriguez-Casero, V., Webster, R., Korczyn, A., Afawi, Z., Zelnick, N., Lerman-Sagie, T., Lev, D., Moller, R. S., Gill, D., Andrade, D. M., Freeman, J. L., Sadleir, L. G., Shendure, J., Berkovic, S. F., Scheffer, I. E. & Mefford, H. C. (2013) Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet,* 45, 825-830.

Chang, G. S., Noegel, A. A., Mavrich, T. N., Muller, R., Tomsho, L., Ward, E., Felder, M., Jiang, C., Eichinger, L., Glockner, G., Schuster, S. C. & Pugh, B. F. (2012) Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in *Dictyostelium*. *Genome Res,* 22**,** 1098-1106.

Chen, G. K., Shaulsky, G. & Kuspa, A. (2004) Tissue-specific G1-phase cell-cycle arrest prior to terminal differentiation in *Dictyostelium*. *Development,* 131**,** 2619-2630.

Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R. & Winston, F. (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol,* 6**,** e277.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J. A., Kuo, F., Kim, J., Cokus, S. J., Casero, D., Bernal, M., Huijser, P., Clark, A. T., Kramer, U., Merchant, S. S., Zhang, X., Jacobsen, S. E. & Pellegrini, M. (2010) Relationship between nucleosome positioning and DNA methylation. *Nature,* 466**,** 388-392.

Chubb, J. R., Bloomfield, G., Xu, Q. K., Kaller, M., Ivens, A., Skelton, J., Turner, B. M., Nellen, W., Shaulsky, G., Kay, R. R., Bickmore, W. A. & Singer, R. H. (2006a) Developmental timing in *Dictyostelium* is regulated by the Set1 histone methyltransferase. *Dev Biol,* 292**,** 519-532.

Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. (2006b) Transcriptional pulsing of a developmental gene. *Curr Biol,* 16**,** 1018-1025.

Chung, H. R., Dunkel, I., Heise, F., Linke, C., Krobitsch, S., Ehrenhofer-Murray, A. E., Sperling, S. R. & Vingron, M. (2010) The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE,* 5**,** e15754.

Clapier, C. R. & Cairns, B. R. (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem,* 78**,** 273-304.

Clarke, M., Schatten, G., Mazia, D. & Spudich, J. A. (1975) Visualization of actin fibers associated with the cell membrane in amoebae of *Dictyostelium discoideum*. *Proc Natl Acad Sci USA,* 72**,** 1758-1762.

Coates, J. C. & Harwood, A. J. (2001) Cell-cell adhesion and signal transduction during *Dictyostelium* development. *J Cell Sci,* 114**,** 4349-4358.

Cockburn, A. F., Taylor, W. C. & Firtel, R. A. (1978) *Dictyostelium* rDNA consists of non-chromosomal palindromic dimers containing 5S and 36S coding regions. *Chromosoma,* 70**,** 19-29.

Corona, D. F. V. & Tamkun, J. W. (2004) Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta,* 1677**,** 113-119.

Correll, S. J., Schubert, M. H. & Grigoryev, S. A. (2012) Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J,* 31**,** 2416-2426.

Curk, T., Demsar, J., Xu, Q., Leban, G., Petrovic, U., Bratko, I., Shaulsky, G. & Zupan, B. (2005) Microarray data mining with visual programming. *Bioinformatics,* 21**,** 396-398.

Dawson, M. A. & Kouzarides, T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell,* 150**,** 12-27.

de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. (2004) Open source clustering software. *Bioinformatics,* 20**,** 1453-1454.

De Lozanne, A. (1987) Homologous recombination in *Dictyostelium* as a tool for the study of developmental genes. *Meth Cell Biol,* 28**,** 489-495.

De Lozanne, A. & Spudich, J. A. (1987) Disruption of the *Dictyostelium* myosin heavy chain gene by homologous recombination. *Science,* 236**,** 1086-1091.

Delmas, V., Stokes, D. G. & Perry, R. P. (1993) A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc Natl Acad Sci U S A,* 90**,** 2414-2418.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M. & Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res,* 36**,** W465-W469.

Dion, M. F., Altschuler, S. J., Wu, L. F. & Rando, O. J. (2005) Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci U S A,* 102**,** 5501-5506.

Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res,* 36**,** e105.

Dubin, M., Fuchs, J. r., Gräf, R., Schubert, I. & Nellen, W. (2010) Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res,* 38**,** 7526-7537.

Durr, H., Flaus, A., Owen-Hughes, T. & Hopfner, K. P. (2006) Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Res,* 34**,** 4160-4167.

Ebbert, R., Birkmann, A. & Schüller, H.-J. (1999) The product of the SNF2/SWI2 paralogue INO80 of *Saccharomyces cerevisiae* required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol Microbiol,* 32**,** 741-751.

Edwards, C. A. & Firtel, R. A. (1984) Site-specific phasing in the chromatin of the rDNA in *Dictyostelium discoideum*. *J Mol Biol,* 180**,** 73-90.

Ehrenhofer-Murray, A. E. (2004) Chromatin dynamics at DNA replication, transcription and repair. *Eur J Biochem,* 271**,** 2335-2349.

Ehrlich, M. & Wang, R. Y. (1981) 5-Methylcytosine in eukaryotic DNA. *Science,* 212**,** 1350-1357.

Eichinger, L., Pachebat, J. A., Glockner, G., Rajandream, M. A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B. A., Rivero, F., Bankier, A. T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Madan Babu, M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M. A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E. C., Chisholm, R. L., Gibbs, R., Loomis, W. F., Platzer, M., Kay, R. R., Williams, J., Dear, P. H., Noegel, A. A., Barrell, B. & Kuspa, A. (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature,* 435**,** 43-57.

Engelen, E., Akinci, U., Bryne, J. C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D. H., Demmers, J., Rijkers, E. J., Bhattacharya, S., Philipsen, S., Pevny, L. H., Grosveld, F. G., Rottier, R. J., Lenhard, B. & Poot, R. A. (2012) Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet,* 43**,** 607-611.

Escalante, R., Moreno, N. & Sastre, L. (2003) *Dictyostelium discoideum* developmentally regulated genes whose expression is dependent on MADS box transcription factor SrfA. *Euk Cell,* 2**,** 1327-1335.

Faix, J., Kreppel, L., Shaulsky, G., Schleicher, M. & Kimmel, A. R. (2004) A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res,* 32**,** E143.

Faix, J., Linkner, J., Nordholz, B., Platt, J. L., Liao, X. H. & Kimmel, A. R. (2013) The application of the Cre-loxP system for generating multiple knock-out and knock-in targeted loci. *Methods Mol Biol,* 983**,** 249-267.

Felsenfeld, G. & Groudine, M. (2003) Controlling the double helix. *Nature,* 421**,** 448-453.

Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C. & Liu, H.-K. (2013) The Chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription factors. *Cell stem cell,* 13**,** 62-72.

Fertey, J., Ammermann, I., Winkler, M., Stoger, R., Iftner, T. & Stubenrauch, F. (2010) Interaction of the papillomavirus E8--E2C protein with the cellular CHD6 protein contributes to transcriptional repression. *J Virol,* 84**,** 9505-9515.

Field, Y., Fondufe-Mittendorf, Y., Moore, I. K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J. D., Widom, J. & Segal, E. (2009) Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat Genet,* 41**,** 438-445.

Flanagan, J. F., Blus, B. J., Kim, D., Clines, K. L., Rastinejad, F. & Khorasanizadeh, S. (2007) Molecular implications of evolutionary differences in CHD double chromodomains. *J Mol Biol,* 369**,** 334-342.

Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F. & Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature,* 438**,** 1181-1185.

Fury, W., Batliwalla, F., Gregersen, P. K. & Li, W. (2006) Overlapping probabilities of top ranking gene lists, hypergeometric distribution, and stringency of gene selection criterion. *Conf Proc IEEE Eng Med Biol Soc,* 1**,** 5531-5534.

Gao, X., Gordon, D., Zhang, D., Browne, R., Helms, C., Gillum, J., Weber, S., Devroy, S., Swaney, S., Dobbs, M., Morcuende, J., Sheffield, V., Lovett, M., Bowcock, A., Herring, J. & Wise, C. (2007) CHD7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. *Am J Hum Genet,* 80**,** 957-865.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., Ramalho-Santos, J., McManus, M. T., Plath, K., Meshorer, E. & Ramalho-Santos, M. (2009) Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature,* 460**,** 863-868.

Gaudet, P., Fey, P., Basu, S., Bushmanova, Y. A., Dodson, R., Sheppard, K. A., Just, E. M., Kibbe, W. A. & Chisholm, R. L. (2011) dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. *Nucleic Acids Res*, 39, D620-624.

Gaudet, P., Fey, P. & Chisholm, R. (2008) *Dictyostelium discoideum*: The social ameba. *CSH Protoc,* 2008**,** pdb.emo109.

Gaudet, P., Pilcher, K. E., Fey, P. & Chisholm, R. L. (2007) Transformation of *Dictyostelium discoideum* with plasmid DNA. *Nature Protoc,* 2**,** 1317-1324.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J. L., Barton, G. J. & Owen-Hughes, T. (2011) A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science,* 333**,** 1758-1760.

Gorbalenya, A. E. & Koonin, E. V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struc Biol,* 3**,** 419-429.

Gorringe, K. L., Choong, D. Y., Williams, L. H., Ramakrishna, M., Sridhar, A., Qiu, W., Bearfoot, J. L. & Campbell, I. G. (2008) Mutation and methylation analysis of the chromodomain-helicase-DNA binding 5 gene in ovarian cancer. *Neoplasia,* 10**,** 1253-1258.

Gowher, H., Leismann, O. & Jeltsch, A. (2000) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J,* 19**,** 6918-6923.

Grewal, S. I. S. & Jia, S. (2007) Heterochromatin revisited. *Nat Rev Genet,* 8**,** 35-46.

Grskovic, M., Chaivorapol, C., Gaspar-Maia, A., Li, H. & Ramalho-Santos, M. (2007) Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. *PLoS Genet,* 3**,** e145.

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F. V., Becker, P. B. & Müller, C. W. (2003) Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell,* 12**,** 449-460.

Guschin, D., Wade, P. A., Kikyo, N. & Wolffe, A. P. (2000) ATP-Dependent histone octamer mobilization and histone deacetylation mediated by the Mi-2 chromatin remodeling complex. *Biochemistry,* 39**,** 5238-5245.

Hall, J. A. & Georgel, P. T. (2007) CHD proteins: a diverse family with strong ties. *Biochem Cell Biol,* 85**,** 463-476.

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J. & Workman, J. L. (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell,* 111**,** 369-379.

Hauk, G., McKnight, J. N., Nodelman, I. M. & Bowman, G. D. (2010) The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Mol Cell,* 39**,** 711-723.

Heid, P. J., Geiger, J., Wessels, D., Voss, E. & Soll, D. R. (2005) Computer-assisted analysis of filopod formation and the role of myosin II heavy chain phosphorylation in *Dictyostelium*. *J Cell Sci,* 118**,** 2225-2237.

Heidel, A. J., Lawal, H. M., Felder, M., Schilde, C., Helps, N. R., Tunggal, B., Rivero, F., John, U., Schleicher, M., Eichinger, L., Platzer, M., Noegel, A. A., Schaap, P. & Glockner, G. (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res,* 21**,** 1882-1891.

Heus, J. J., Zonneveld, B. J., Bloom, K. S., de Steensma, H. Y. & van den Berg, J. A. (1993) The nucleosome repeat length of Kluyveromyces lactis is 16 bp longer than that of *Saccharomyces cerevisiae*. *Nucleic Acids Res,* 21**,** 2247-2248.

Hewish, D. R. & Burgoyne, L. A. (1973) Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Bioph Res Co,* 52**,** 504-510.

Hirschhorn, J. N., Brown, S. A., Clark, C. D. & Winston, F. (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Gene Dev,* 6**,** 2288-2298.

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. & Bradbury, E. M. (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem,* 268**,** 305-314.

Huang, S., Gulzar, Z. G., Salari, K., Lapointe, J., Brooks, J. D. & Pollack, J. R. (2011) Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. *Oncogene,* 31**,** 4164-4170.

Hughes, A. L., Jin, Y., Rando, O. J. & Struhl, K. (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell,* 48**,** 5-15.

Hurd, E. A., Capers, P. L., Blauwkamp, M. N., Adams, M. E., Raphael, Y., Poucher, H. K. & Martin, D. M. (2007) Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mamm Genome,* 18**,** 94-104.

Iranfar, N., Fuller, D. & Loomis, W. F. (2006) Transcriptional regulation of post-aggregation genes in *Dictyostelium* by a feed-forward loop involving GBF and LagC. *Dev Bio,* 290**,** 460-469.

Ishihara, K., Oshimura, M. & Nakao, M. (2006) CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell,* 23**,** 733-742.

Jakovcevski, M. & Akbarian, S. (2012) Epigenetic mechanisms in neurological disease. *Nat Med,* 18**,** 1194-1204.

Janssen, N., Bergman, J. E., Swertz, M. A., Tranebjaerg, L., Lodahl, M., Schoots, J., Hofstra, R. M., van Ravenswaaij-Arts, C. M. & Hoefsloot, L. H. (2012) Mutation update on the CHD7 gene involved in CHARGE syndrome. *Hum Mutat,* 33**,** 1149-1160.

Jones, P. A. & Baylin, S. B. (2007) The epigenomics of cancer. *Cell,* 128**,** 683-692.

Jongmans, M. C., van Ravenswaaij-Arts, C. M., Pitteloud, N., Ogata, T., Sato, N., Claahsen-van der Grinten, H. L., van der Donk, K., Seminara, S., Bergman, J. E., Brunner, H. G., Crowley, W. F., Jr. & Hoefsloot, L. H. (2009) CHD7 mutations in patients initially diagnosed with Kallmann syndrome--the clinical overlap with CHARGE syndrome. *Clin Genet,* 75**,** 65-71.

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J. & Segal, E. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature,* 458**,** 362-366.

Katoh, M., Curk, T., Xu, Q., Zupan, B., Kuspa, A. & Shaulsky, G. (2006) Developmentally regulated DNA methylation in *Dictyostelium discoideum*. *Euk Cell,* 5**,** 18-25.

Kay, R. R., Langridge, P., Traynor, D. & Hoeller, O. (2008) Changing directions in the study of chemotaxis. *Nat Rev Mol Cell Biol,* 9**,** 455-463.

Keim-Reder, M. (2006) Regulation of Dictyostelium gene expression and chemotaxis by inositol signalling. *MRC Laboratory of Molecular Cell Biology.* London, University College London.

Kent, N. A., Adams, S., Moorhouse, A. & Paszkiewicz, K. (2010) Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucleic Acids Res,* 39**,** e26.

Kent, N. A. & Mellor, J. (1995) Chromatin structure snap-shots: rapid nuclease digestion of chromatin in yeast. *Nucleic Acids Res,* 23**,** 3786-3787.

Khattak, S., Lee, B. R., Cho, S. H., Ahnn, J. & Spoerel, N. A. (2002) Genetic characterization of *Drosophila* Mi-2 ATPase. *Gene,* 293**,** 107-114.

Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. (1993) BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature,* 366**,** 170-174.

Kim, H. G., Kurth, I., Lan, F., Meliciani, I., Wenzel, W., Eom, S. H., Kang, G. B., Rosenberger, G., Tekin, M., Ozata, M., Bick, D. P., Sherins, R. J., Walker, S. L., Shi, Y., Gusella, J. F. & Layman, L. C. (2008) Mutations in CHD7, encoding a chromatin-remodeling protein, cause idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. *Am J Hum Genet,* 83**,** 511-519.

Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., Kingston, R. & Georgopoulos, K. (1999) Ikaros DNA-binding

proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity,* 10**,** 345-355.

Kim, M. S., Chung, N. G., Kang, M. R., Yoo, N. J. & Lee, S. H. (2011) Genetic and expressional alterations of CHD genes in gastric and colorectal cancers. *Histopathology,* 58**,** 660-668.

Knezetic, J. A. & Luse, D. S. (1986) The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell,* 45**,** 95-104.

Konijn, T. M., van de Meene, J. G. C., Bonner, J. T. & Barkley, D. S. (1967) The acrasin activity of adenosine-3',5'-cyclic phosphate. *Proc Natl Acad Sci USA,* 58**,** 1152-1154.

Kornberg, R. D. (1974) Chromatin structure: A repeating unit of histones and DNA. *Science,* 184**,** 868-871.

Kortholt, A. & van Haastert, P. J. (2008) Highlighting the role of Ras and Rap during *Dictyostelium* chemotaxis. *Cell Signal,* 20**,** 1415-1422.

Kouzarides, T. (2007) Chromatin modifications and their function. *Cell,* 128**,** 693-705.

Kruskal, W. H. & Wallis, W. A. (1952) Use of ranks in one-criterion variance analysis. *J Am Stat Assoc,* 47**,** 583-621.

Kuhlmann, M., Borisova, B. E., Kaller, M., Larsson, P., Stach, D., Na, J. B., Eichinger, L., Lyko, F., Ambros, V., Soderbom, F., Hammann, C. & Nellen, W. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res,* 33**,** 6405-6417.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature,* 410**,** 116-120.

Lalani, S. R., Safiullah, A. M., Fernbach, S. D., Harutyunyan, K. G., Thaller, C., Peterson, L. E., McPherson, J. D., Gibbs, R. A., White, L. D., Hefner, M., Davenport, S. L., Graham, J. M., Bacino, C. A., Glass, N. L., Towbin, J. A., Craigen, W. J., Neish, S. R., Lin, A. E. & Belmont, J. W. (2006) Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am J Hum Genet,* 78**,** 303-314.

Lang, J., Tobias, E. S. & MacKie, R. (2011) Preliminary evidence for involvement of the tumour suppressor gene CHD5 in a family with cutaneous melanoma. *Brit J Dermatol,* 164**,** 1010-1016.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10, R25.

Lantermann, A. B., Straub, T., Stralfors, A., Yuan, G.-C., Ekwall, K. & Korber, P. (2010) *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. *Nat Struct Mol Biol,* 17**,** 251-257.

Lathrop, M. J., Chakrabarti, L., Eng, J., Rhodes, C. H., Lutz, T., Nieto, A., Liggitt, H. D., Warner, S., Fields, J., Stoger, R. & Fiering, S. (2010) Deletion of the Chd6 exon 12 affects motor coordination. *Mamm Genome,* 21**,** 130-142.

Laurent, B. C., Treich, I. & Carlson, M. (1993) The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Gene Dev,* 7**,** 583-591.

Layman, W. S., McEwen, D. P., Beyer, L. A., Lalani, S. R., Fernbach, S. D., Oh, E., Swaroop, A., Hegg, C. C., Raphael, Y., Martens, J. R. & Martin, D. M. (2009) Defects in neural stem cell proliferation and olfaction in Chd7 deficient mice indicate a mechanism for hyposmia in human CHARGE syndrome. *Hum Mol Genet,* 18**,** 1909-1923.

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R. & Nislow, C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet,* 39**,** 1235-1244.

Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K. H. (2011) The nucleosome map of the mammalian liver. *Nat Struct Mol Biol,* 18**,** 742-746.

Liu, W., Lindberg, J., Sui, G., Luo, J., Egevad, L., Li, T., Xie, C., Wan, M., Kim, S. T., Wang, Z., Turner, A. R., Zhang, Z., Feng, J., Yan, Y., Sun, J., Bova, G. S., Ewing, C. M., Yan, G., Gielzak, M., Cramer, S. D., Vessella, R. L., Zheng, S. L., Gronberg, H., Isaacs, W. B. & Xu, J. (2011) Identification of novel CHD1-associated collaborative alterations of genomic structure and functional assessment of CHD1 in prostate cancer. *Oncogene,* 31**,** 3939-3948.

Lorch, Y., LaPointe, J. W. & Kornberg, R. D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell,* 49**,** 203-210.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8A resolution. *Nature*, 389, 251-260.

Lusser, A., Urwin, D. L. & Kadonaga, J. T. (2005) Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol,* 12**,** 160-166.

Lutz, T., Stöger, R. & Nieto, A. (2006) CHD6 is a DNA-dependent ATPase and localizes at nuclear sites of mRNA synthesis. *FEBS Lett,* 580**,** 5851-5857.

Maere, S., Heymans, K. & Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics,* 21**,** 3448-3449.

Maizels, N. (1976) *Dictyostelium* 17S, 25S, and 5S rDNAs lie within a 38,000 base pair repeated unit. *Cell,* 9**,** 431-438.

Mann, H. B. & Whitney D. R. (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat,* 18**,** 50-60.

Marfella, C. G. & Imbalzano, A. N. (2007) The Chd family of chromatin remodelers. *Mutat Res,* 618**,** 30-40.

Marfella, C. G., Ohkawa, Y., Coles, A. H., Garlick, D. S., Jones, S. N. & Imbalzano, A. N. (2006) Mutation of the SNF2 family member Chd2 affects mouse development and survival. *J Cell Physiol,* 209**,** 162-171.

Marom, R., Shur, I., Hager, G. L. & Benayahu, D. (2006) Expression and regulation of CReMM, a chromodomain helicase-DNA-binding (CHD), in marrow stroma derived osteoprogenitors. *J Cell Physiol,* 207**,** 628-635.

Martinez-Campa, C., Politis, P., Moreau, J. L., Kent, N., Goodall, J., Mellor, J. & Goding, C. R. (2004) Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Mol Cell,* 15**,** 69-81.

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C., Gilmour, D. S., Albert, I. & Pugh, B. F. (2008) Nucleosome organization in the *Drosophila* genome. *Nature,* 453**,** 358-362.

McDaniel, I. E., Lee, J. M., Berger, M. S., Hanagami, C. K. & Armstrong, J. A. (2008) Investigations of CHD1 function in transcription and development of *Drosophila* melanogaster. *Genetics,* 178**,** 583-587.

McKnight, J. N., Jenkins, K. R., Nodelman, I. M., Escobar, T. & Bowman, G. D. (2011) Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Mol Cell Biol,* 31**,** 4746-4759.

McMains, V. C., Liao, X. H. & Kimmel, A. R. (2008) Oscillatory signaling and network responses during the development of *Dictyostelium discoideum*. *Ageing Res Rev,* 7**,** 234-248.

McPherson, C. E. & Singleton, C. K. (1992) V4, a gene required for the transition from growth to development in *Dictyostelium discoideum*. *Dev Biol,* 150**,** 231-242.

Minucci, S. & Pelicci, P. G. (2006) Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer,* 6**,** 38-51.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S. & Wu, C. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling Complex. *Science,* 303**,** 343-348.

Morettini, S., Tribus, M., Zeilner, A., Sebald, J., Campo-Fernandez, B., Scheran, G., Worle, H., Podhraski, V., Fyodorov, D. V. & Lusser, A. (2011) The chromodomains

of CHD1 are critical for enzymatic activity but less important for chromatin localization. *Nucleic Acids Res,* 39**,** 3103-3115.

Müller, K. & Gerisch, G. (1978) A specific glycoprotein as the target site of adhesion blocking Fab in aggregating *Dictyostelium* cells. *Nature,* 274**,** 445-449.

Müller-Taubenberger, A., Bonisch, C., Furbringer, M., Wittek, F. & Hake, S. B. (2010) The histone methyltransferase Dot1 is required for DNA damage repair and proper development in *Dictyostelium*. *Biochem Bioph Res Co,* 404**,** 1016-1022.

Muramoto, T. & Chubb, J. R. (2008) Live imaging of the *Dictyostelium* cell cycle reveals widespread S phase during development, a G2 bias in spore differentiation and a premitotic checkpoint. *Development,* 135**,** 1647-1657.

Muramoto, T., Müller, I., Thomas, G., Melvin, A. & Chubb, J. R. (2010) Methylation of H3K4 Is required for inheritance of active transcriptional states. *Curr Biol,* 20**,** 397-406.

Murawska, M. & Brehm, A. (2011) CHD chromatin remodelers and the transcription cycle. *Transcription,* 2**,** 244-253.

Musselman, C. A., Ramirez, J., Sims, J. K., Mansfield, R. E., Oliver, S. S., Denu, J. M., Mackay, J. P., Wade, P. A., Hagman, J. & Kutateladze, T. G. (2011) Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc Natl Acad Sci U S A,* 109**,** 787-792.

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L. S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E. L., Campbell, N. G., Geller, E. T., Valladares, O., Schafer, C., Liu, H., Zhao, T., Cai, G., Lihm, J., Dannenfelser, R., Jabado, O., Peralta, Z., Nagaswamy, U., Muzny, D., Reid, J. G., Newsham, I., Wu, Y., Lewis, L., Han, Y., Voight, B. F., Lim, E., Rossin, E., Kirby, A., Flannick, J., Fromer, M., Shakir, K., Fennell, T., Garimella, K., Banks, E., Poplin, R., Gabriel, S., DePristo, M., Wimbish, J. R., Boone, B. E., Levy, S. E., Betancur, C., Sunyaev, S., Boerwinkle, E., Buxbaum, J. D., Cook Jr, E. H., Devlin, B., Gibbs, R. A., Roeder, K., Schellenberg, G. D., Sutcliffe, J. S. & Daly, M. J. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature,* 485**,** 242-245.

Neigeborn, L. & Carlson, M. (1984) Genes affecting the regulation of Suc2 gene expression by glucose repression in *saccharomyces cerevisiae*. *Genetics,* 108**,** 845-858.

Nelson, H. C., Finch, J. T., Luisi, B. F. & Klug, A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature,* 330**,** 221-226.

Ness, P. J., Labhart, P., Banz, E., Koller, T. & Parish, R. W. (1983) Chromatin structure along the ribosomal DNA of *Dictyostelium*. Regional differences and changes accompanying cell differentiation. *J. Mol. Biol.,* 166**,** 361-381.

Nguyen, A. T. & Zhang, Y. (2011) The diverse functions of Dot1 and H3K79 methylation. *Gene Dev,* 25**,** 1345-1358.

Nicol, J. W., Helt, G. A., Blanchard, S. G., Raja, A. & Loraine, A. E. (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics,* 25**,** 2730-2731.

Nioi, P., Nguyen, T., Sherratt, P. J. & Pickett, C. B. (2005) The carboxy-terminal Neh3 domain of Nrf2 is required for transcriptional activation. *Mol Cell Biol,* 25**,** 10895-10906.

Nishiyama, M., Nakayama, K., Tsunematsu, R., Tsukiyama, T., Kikuchi, A. & Nakayama, K. I. (2004) Early embryonic death in mice lacking the beta-catenin-binding protein Duplin. *Mol Cell Biol,* 24**,** 8386-8394.

Nishiyama, M., Skoultchi, A. I. & Nakayama, K. I. (2012) Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-beta-catenin signaling pathway. *Mol Cell Biol,* 32**,** 501-512.

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J. B., Turner, E. H., Levy, R., O,'Day, D. R., Krumm, N., Coe, B. P., Martin, B. K., Borenstein, E., Nickerson, D. A., Mefford, H. C., Doherty, D., Akey, J. M., Bernier, R., Eichler, E. E. & Shendure, J. (2012) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science,* 338**,** 1619-1622.

Okada, M., Okawa, K., Isobe, T. & Fukagawa, T. (2009) CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. *Mol Biol Cell,* 20**,** 3986-3995.

Okawa, E. R., Gotoh, T., Manne, J., Igarashi, J., Fujita, T., Silverman, K. A., Xhao, H., Mosse, Y. P., White, P. S. & Brodeur, G. M. (2008) Expression and sequence analysis of candidates for the 1p36.31 tumor suppressor gene deleted in neuroblastomas. *Oncogene,* 27**,** 803-810.

Olins, A. L. & Olins, D. E. (1974) Spheroid chromatin units (v Bodies). *Science,* 183**,** 330-332.

Oliver, S. S., Musselman, C. A., Srinivasan, R., Svaren, J. P., Kutateladze, T. G. & Denu, J. M. (2012) Multivalent recognition of histone tails by the PHD fingers of CHD5. *Biochemistry,* 51**,** 6534-6544.

Oudet, P., Gross-Bellard, M. & Chambon, P. (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell,* 4**,** 281-300.

Ozsolak, F., Song, J. S., Liu, X. S. & Fisher, D. E. (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat Biotech,* 25**,** 244-248.

Pagon, R. A., Graham, J. M., Jr., Zonana, J. & Yong, S. L. (1981) Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *J Pediatr,* 99**,** 223-227.

Parikh, A., Miranda, E. R., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., Curk, T., Sucgang, R., Chen, R., Zupan, B., Loomis, W. F., Kuspa, A. & Shaulsky, G. (2010) Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol,* 11**,** R35.

Parish, R. W., Stalder, J. & Schmidlin, S. (1977) Biochemical evidence for a DNA repeat length in the chromatin of *Dictyostelium discoideum*. *FEBS Lett,* 84**,** 63-66.

Paro, R. & Hogness, D. S. (1991) The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci U S A,* 88**,** 263-267.

Patel, A., McKnight, J. N., Genzor, P. & Bowman, G. D. (2011) Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding. *J Biol Chem,* 286**,** 43984-43993.

Patten, S. A., Jacobs-McDaniels, N. L., Zaouter, C., Drapeau, P., Albertson, R. C. & Moldovan, F. (2012) Role of Chd7 in Zebrafish: a model for CHARGE syndrome. *PLoS ONE,* 7**,** e31650.

Paul, S., Kuo, A., Schalch, T., Vogel, H., Joshua-Tor, L., McCombie, W. R., Gozani, O., Hammell, M. & Mills, A. A. (2013) Chd5 requires PHD-mediated histone 3 binding for tumor suppression. *Cell Reports,* 3**,** 92-102.

Pavlovic, J., Banz, E. & Parish, R. W. (1989) The effects of transcription on the nucleosome structure of four *Dictyostelium* genes. *Nucleic Acids Res,* 17**,** 2315-2332.

Peterson, C. L., Dingwall, A. & Scott, M. P. (1994) Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. *Proc Natl Acad Sci U S A,* 91**,** 2905-2908.

Peterson, C. L. & Herskowitz, I. (1992) Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell,* 68**,** 573-583.

Platt, J. L., Kent, N. A., Harwood, A. J. & Kimmel, A. R. (2013) Analysis of chromatin organization by deep sequencing technologies. *Methods Mol Biol,* 983**,** 173-83.

Platt, J. L., Rogers, B. J., Rogers, K. C., Harwood, A. J. & Kimmel, A. R. (2013) Different CHD chromatin remodelers are required for expression of distinct gene sets and specific stages during development of *Dictyostelium discoideum*. *Development*, 120, 4926-4936.

Pleasance, E. D., Stephens, P. J., O'Meara, S., McBride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C., Varela, I., Nik-Zainal, S., Davies, H. R., Ordonez, G. R., Mudie, L. J., Latimer, C., Edkins, S., Stebbings, L., Chen, L., Jia, M.,

Leroy, C., Marshall, J., Menzies, A., Butler, A., Teague, J. W., Mangion, J., Sun, Y. A., McLaughlin, S. F., Peckham, H. E., Tsung, E. F., Costa, G. L., Lee, C. C., Minna, J. D., Gazdar, A., Birney, E., Rhodes, M. D., McKernan, K. J., Stratton, M. R., Futreal, P. A. & Campbell, P. J. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature,* 463**,** 184-190.

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Stralfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J. P., Ekwall, K. & Korber, P. (2012) CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in *S. pombe*. *EMBO J,* 31**,** 4388-4403.

Potts, R. C., Zhang, P., Wurster, A. L., Precht, P., Mughal, M. R., Wood, W. H., III, Zhang, Y., Becker, K. G., Mattson, M. P. & Pazin, M. J. (2011) CHD5, a Brain-Specific Paralog of Mi2 Chromatin Remodeling Enzymes, Regulates Expression of Neuronal Genes. *PLoS ONE,* 6**,** e24515.

R Development Core Team (2012) R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Radman-Livaja, M., Quan, T. K., Valenzuela, L., Armstrong, J. A., van Welsem, T., Kim, T., Lee, L. J., Buratowski, S., van Leeuwen, F., Rando, O. J. & Hartzog, G. A. (2012) A key role for Chd1 in histone H3 dynamics at the 3' ends of long genes in yeast. *PLoS Genet,* 8**,** e1002811.

Raper, K. B. (1935) *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves. *J Agr Res,* 50**,** 135-147.

Razin, A. & Riggs, A. D. (1980) DNA methylation and gene function. *Science,* 210**,** 604-610.

Robbins, C. M., Tembe, W. A., Baker, A., Sinari, S., Moses, T. Y., Beckstrom-Sternberg, S., Beckstrom-Sternberg, J., Barrett, M., Long, J., Chinnaiyan, A., Lowey, J., Suh, E., Pearson, J. V., Craig, D. W., Agus, D. B., Pienta, K. J. & Carpten, J. D. (2011) Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors. *Genome Res,* 21**,** 47-55.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem,* 273**,** 5858-5868.

Rogers, B. J. (2010) ATP dependent chromatin remodeling and inositol phosphate signaling in *Dictyostelium discoideum*. *School of Biosciences.* Cardiff, Cardiff University.

Rot, G., Parikh, A., Curk, T., Kuspa, A., Shaulsky, G. & Zupan, B. (2009) dictyExpress: a *Dictyostelium discoideum* gene expression database with an explorative data analysis web-based interface. *BMC Bioinformatics,* 10**,** 265.

Ryan, D. P. & Owen-Hughes, T. (2011) Snf2-family proteins: chromatin remodellers for any occasion. *Curr Opin Chem Biol,* 15**,** 649-656.

Ryan, D. P., Sundaramoorthy, R., Martin, D., Singh, V. & Owen-Hughes, T. (2011) The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *EMBO J,* 30**,** 2596-2609.

Saldanha, A. J. (2004) Java Treeview – extensible visualization of microarray data. *Bioinformatics,* 20**,** 3246-3248.

Sanlaville, D., Etchevers, H. C., Gonzales, M., Martinovic, J., Clement-Ziza, M., Delezoide, A. L., Aubry, M. C., Pelet, A., Chemouny, S., Cruaud, C., Audollent, S., Esculpavit, C., Goudefroye, G., Ozilou, C., Fredouille, C., Joye, N., Morichon-Delvallez, N., Dumez, Y., Weissenbach, J., Munnich, A., Amiel, J., Encha-Razavi, F., Lyonnet, S., Vekemans, M. & Attie-Bitach, T. (2006) Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development. *J Med Genet,* 43**,** 211-217.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. & Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-411.

Sawarkar, R., Visweswariah, S. S., Nellen, W. & Nanjundiah, V. (2009) Histone deacetylases regulate multicellular development in the social amoeba *Dictyostelium discoideum*. *J Mol Biol,* 391**,** 833-848.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Meth,* 9**,** 671-675.

Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., Adams, D. J., Flicek, P., Crawford, G. E., Laframboise, T., Tesar, P., Wei, C. L. & Scacheri, P. C. (2010) CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet,* 6**,** e1001023.

Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G. & Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell,* 132**,** 887-898.

Schuster, E. F. & Stöger, R. (2002) CHD5 defines a new subfamily of chromodomain-SWI2/SNF2-like helicases. *Mamm Genome,* 13**,** 117-119.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-342.

Schwartz, S., Meshorer, E. & Ast, G. (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol,* 16**,** 990-995.

Seelig, H. P., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M. & Genth, E. (1995) The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis Rheum,* 38**,** 1389-1399.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J. Z. & Widom, J. (2006) A genomic code for nucleosome positioning. *Nature,* 442**,** 772-778.

Segal, E. & Widom, J. (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol,* 19**,** 65-71.

Sharma, A., Jenkins, K. R., Heroux, A. & Bowman, G. D. (2011) Crystal structure of the chromodomain helicase DNA-binding protein 1 (Chd1) DNA-binding domain in complex with DNA. *J Biol Chem,* 286**,** 42099-42104.

Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. (2000) A chromatin remodelling complex involved in transcription and DNA processing. *Nature,* 406**,** 541-544.

Shim, Y. S., Choi, Y., Kang, K., Cho, K., Oh, S., Lee, J., Grewal, S. I. & Lee, D. (2012) Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *EMBO J,* 31**,** 4375-4387.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M. & Iyer, V. R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol,* 6**,** e65.

Shur, I. & Benayahu, D. (2005) Characterization and functional analysis of CReMM, a novel chromodomain helicase DNA-binding protein. *J Mol Biol,* 352**,** 646-655.

Shur, I., Socher, R. & Benayahu, D. (2006) In vivo association of CReMM/CHD9 with promoters in osteogenic cells. *J Cell Physiol,* 207**,** 374-378.

Simic, R., Lindstrom, D. L., Tran, H. G., Roinick, K. L., Costa, P. J., Johnson, A. D., Hartzog, G. A. & Arndt, K. M. (2003) Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J,* 22**,** 1846-1856.

Sims, R. J., 3rd, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S. & Reinberg, D. (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem,* 280**,** 41789-41792.

Smith, S. S. & Ratner, D. I. (1991) Lack of 5-methylcytosine in *Dictyostelium discoideum* DNA. *Biochem J,* 277**,** 273-275.

Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., Washburn, M. P. & Workman, J. L. (2012) Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol,* 19**,** 884-892.

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics,* 27**,** 431-432.

Stern, M., Jensen, R. & Herskowitz, I. (1984) Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol,* 178**,** 853-868.

Stevense, M., Chubb, J. R. & Muramoto, T. (2011) Nuclear organization and transcriptional dynamics in *Dictyostelium*. *Dev Growth Differ,* 53**,** 576-586.

Stockdale, C., Flaus, A., Ferreira, H. & Owen-Hughes, T. (2006) Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. *J Biol Chem,* 281**,** 16279-16288.

Stokes, D. G. & Perry, R. P. (1995) DNA-binding and chromatin localization properties of CHD1. *Mol Cell Biol,* 15**,** 2745-2753.

Stokes, D. G., Tartof, K. D. & Perry, R. P. (1996) CHD1 is concentrated in interbands and puffed regions of *Drosophila* polytene chromosomes. *Proc Natl Acad Sci U S A,* 93**,** 7137-7142.

Strahl, B. D. & Allis, C. D. (2000) The language of covalent histone modifications. *Nature*, 403, 41-45.

Strasser, K., Bloomfield, G., MacWilliams, A., Ceccarelli, A., MacWilliams, H. & Tsang, A. (2012) A retinoblastoma orthologue is a major regulator of S-phase, mitotic, and developmental gene expression in *Dictyostelium*. *PLoS ONE,* 7**,** e39914.

Sucgang, R., Kuo, A., Tian, X., Salerno, W., Parikh, A., Feasley, C. L., Dalin, E., Tu, H., Huang, E., Barry, K., Lindquist, E., Shapiro, H., Bruce, D., Schmutz, J., Salamov, A., Fey, P., Gaudet, P., Anjard, C., Babu, M. M., Basu, S., Bushmanova, Y., van der Wel, H., Katoh-Kurasawa, M., Dinh, C., Coutinho, P. M., Saito, T., Elias, M., Schaap, P., Kay, R. R., Henrissat, B., Eichinger, L., Rivero, F., Putnam, N. H., West, C. M., Loomis, W. F., Chisholm, R. L., Shaulsky, G., Strassmann, J. E., Queller, D. C., Kuspa, A. & Grigoriev, I. V. (2011) Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome Biol,* 12**,** R20.

Sukumaran, S., Brown, J. M., Firtel, R. A. & McNally, J. G. (1998) lagC-null and gbf-null cells define key steps in the morphogenesis of *Dictyostelium* mounds. *Dev Biol,* 200**,** 16-26.

Sullivan, K. F., Hechenberger, M. & Masri, K. (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol,* 127**,** 581-592.

Surapureddi, S., Viswakarma, N., Yu, S., Guo, D., Rao, M. S. & Reddy, J. K. (2006) PRIC320, a transcription coactivator, isolated from peroxisome proliferator-binding protein complex. *Biochem Bioph Res Co,* 343**,** 535-543.

Sussman, R. & Sussman, M. (1967) Cultivation of *Dictyostelium discoideum* in axenic culture. *Biochem Bioph Res Co,* 29**,** 53-55.

Tai, H. H., Geisterfer, M., Bell, J. C., Moniwa, M., Davie, J. R., Boucher, L. & McBurney, M. W. (2003) CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins. *Biochem Bioph Res Co,* 308**,** 170-176.

Teves, S. S. & Henikoff, S. (2011) Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Gene Dev,* 25**,** 2387-2397.

Thompson, B. A., Tremblay, V., Lin, G. & Bochar, D. A. (2008) CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol,* 28**,** 3894-3904.

Thompson, P. M., Gotoh, T., Kok, M., White, P. S. & Brodeur, G. M. (2003) CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. *Oncogene,* 22**,** 1002-1011.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J. & Guigo, R. (2009) Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol,* 16**,** 996-1001.

Tirosh, I., Sigal, N. & Barkai, N. (2010) Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biol,* 11**,** R49.

Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature,* 395**,** 917-921.

Tran, H. G., Steger, D. J., Iyer, V. R. & Johnson, A. D. (2000) The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *EMBO J,* 19**,** 2323-2331.

Trapnell, C., Pachter, L. & Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-1111.

Tsukiyama, T., Palmer, J., Landel, C. C., Shiloach, J. & Wu, C. (1999) Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in *Saccharomyces cerevisiae*. *Gene Dev,* 13**,** 686-697.

Tsukiyama, T. & Wu, C. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell,* 83**,** 1011-1020.

Urushihara, H. (2009) The cellular slime mold: eukaryotic model microorganism. *Exp Anim,* 58**,** 97-104.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A. & Johnson, S. M. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res,* 18**,** 1051-1063.

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. & Sidow, A. (2011) Determinants of nucleosome organization in primary human cells. *Nature,* 474**,** 516-520.

Varga-Weisz, P. (2001) ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. *Oncogene,* 20**,** 3076-3085.

Vissers, L. E., van Ravenswaaij, C. M., Admiraal, R., Hurst, J. A., de Vries, B. B., Janssen, I. M., van der Vliet, W. A., Huys, E. H., de Jong, P. J., Hamel, B. C., Schoenmakers, E. F., Brunner, H. G., Veltman, J. A. & van Kessel, A. G. (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet,* 36**,** 955-957.

von Zelewsky, T., Palladino, F., Brunschwig, K., Tobler, H., Hajnal, A. & Muller, F. (2000) The *C. elegans* Mi-2 chromatin-remodelling proteins function in vulval cell fate determination. *Development,* 127**,** 5277-5284.

Wade, P. A., Jones, P. L., Vermaak, D. & Wolffe, A. P. (1998) A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol,* 8**,** 843-846.

Walfridsson, J., Bjerling, P., Thalen, M., Yoo, E.-J., Park, S. D. & Ekwall, K. (2005) The CHD remodeling factor Hrp1 stimulates CENP-A loading to centromeres. *Nucleic Acids Res,* 33**,** 2868-2879.

Walfridsson, J., Khorosjutina, O., Matikainen, P., Gustafsson, C. M. & Ekwall, K. (2007) A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly. *EMBO J,* 26**,** 2868-2879.

Wang, G. G., Allis, C. D. & Chi, P. (2007a) Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol Med,* 13**,** 363-372.

Wang, G. G., Allis, C. D. & Chi, P. (2007b) Chromatin remodeling and cancer, part II: ATP-dependent chromatin remodeling. *Trends Mol Med,* 13**,** 373-380.

Wang, H. B. & Zhang, Y. (2001) Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor. *Nucleic Acids Res,* 29**,** 2517-2521.

Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. & Crabtree, G. R. (1996) Diversity and specialization of mammalian SWI/SNF complexes. *Gene Dev,* 10**,** 2117-2130.

Wang, X., Lau, K. K., So, L. K. & Lam, Y. W. (2009) CHD5 is down-regulated through promoter hypermethylation in gastric cancer. *J Biomed Sci,* 16**,** 95.

Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q. & Zhao, K. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet,* 40**,** 897-903.

Warnes, G. R. (2012) gplots: Various R programming tools for plotting data.

Watson, A. A., Mahajan, P., Mertens, H. D., Deery, M. J., Zhang, W., Pham, P., Du, X., Bartke, T., Edlich, C., Berridge, G., Chen, Y., Burgess-Brown, N. A., Kouzarides, T., Wiechens, N., Owen-Hughes, T., Svergun, D. I., Gileadi, O. & Laue, E. D. (2012) The

PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J Mol Biol,* 422**,** 3-17.

Watson, J. D. & Crick, F. C. (1953) A structure for deoxyribose nucleic acid. *Nature,* 171**,** 737-738.

Watts, D. J. & Ashworth, J. M. (1970) Growth of myxameobae of the cellular slime mould *Dictyostelium discoideum* in axenic culture. *Biochem J,* 119**,** 171-174.

Wessels, D., Kuhl, S. & Soll, D. R. (2006) Application of 2D and 3D DIAS to motion analysis of live cells in transmission and confocal microscopy imaging. *Methods Mol Biol,* 346**,** 261-279.

Westenberger, S. J., Cui, L., Dharia, N. & Winzeler, E. (2009) Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics,* 10**,** 610.

Whitehouse, I., Rando, O. J., Delrow, J. & Tsukiyama, T. (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature,* 450**,** 1031-1035.

Williams, C. J., Naito, T., Arco, P. G.-d., Seavitt, J. R., Cashman, S. M., Souza, B. D., Qi, X., Keables, P., Andrian, U. H. V. & Georgopoulos, K. (2004) The chromatin remodeler Mi-2β Is required for CD4 expression and T cell development. *Immunity,* 20**,** 719-733.

Williams, H. P. & Harwood, A. J. (2003) Cell polarity and *Dictyostelium* development. *Curr Opin Microbiol,* 6**,** 621-627.

Williams, J. G. (2010) *Dictyostelium* finds new roles to model. *Genetics,* 185**,** 717-726.

Winston, F. & Carlson, M. (1992) Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet,* 8**,** 387-391.

Wolffe, A. P. & Hayes, J. J. (1999) Chromatin disruption and modification. *Nucleic Acids Res,* 27**,** 711-720.

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. & Collins, F. S. (1997) Characterization of the CHD family of proteins. *Proc Natl Acad Sci U S A,* 94**,** 11472-11477.

Woodcock, C. L. & Dimitrov, S. (2001) Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev,* 11**,** 130-135.

Woodcock, C. L. F., Safer, J. P. & Stanchfield, J. E. (1976) Structural repeating units in chromatin: I. Evidence for their general occurrence. *Exp Cell Res,* 97**,** 101-110.

Wu, L. & Winston, F. (1997) Evidence that Snf-Swi controls chromatin structure over both the TATA and UAS regions of the SUC2 promoter in *Saccharomyces cerevisiae*. *Nucleic Acids Res,* 25**,** 4230-4234.

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J. & Wang, W. (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell,* 2**,** 851-861.

Yap, K. L. & Zhou, M.-M. (2011) Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry,* 50**,** 1966-1980.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. & Pugh, B. F. (2012) Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell,* 149**,** 1461-1473.

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J. & Rando, O. J. (2005) Genome-scale identification of nucleosome positions in *S.cerevisiae*. *Science,* 309**,** 626-630.

Zentner, G. E., Layman, W. S., Martin, D. M. & Scacheri, P. C. (2010) Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *Am J Med Genet A,* 152A**,** 674-686.

Zentner, G. E., Tsukiyama, T. & Henikoff, S. (2013) ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet,* 9**,** e1003317.

Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell,* 95**,** 279-289.

Zhang, Z., Wippo, C. J., Wal, M., Ward, E., Korber, P. & Pugh, B. F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science,* 332**,** 977-980.

Zigmond, S. H. (1988) Orientation chamber in chemotaxis. *Methods Enzymol,* 162**,** 65-72.

Appendix A

Supplementary results

A.1-10: Enriched GO terms in RNA-seq comparisons

Diagrams show gene ontology (GO) enrichment terms for up and downregulated genes. Size of circle defines the number of genes annotated with a term. Significance is indicated by yellow, orange and dark orange. Dark orange is the most significant; white is not significant.

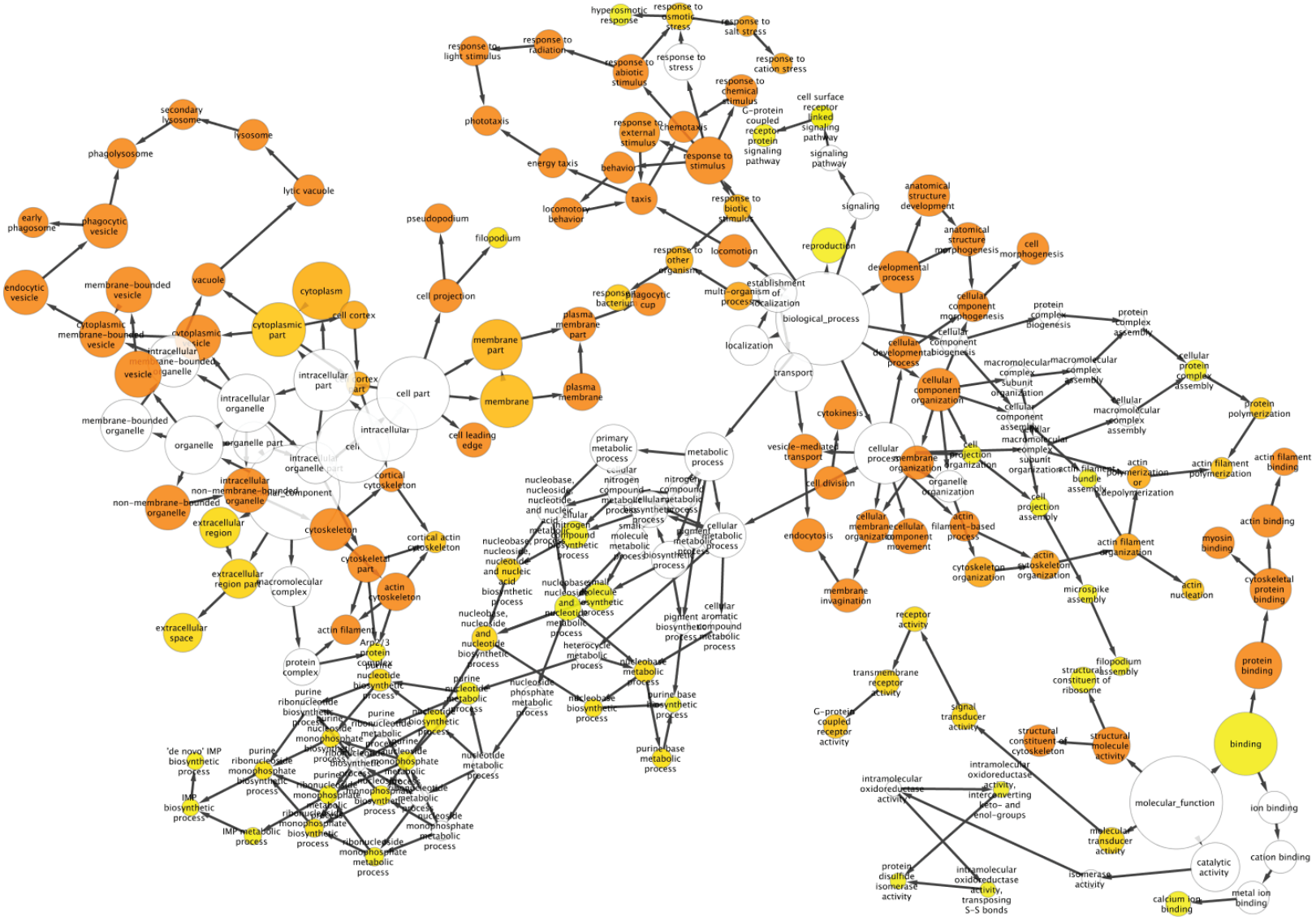


Figure A.1: GO enrichment in downregulated genes in growing *chdA*-null cells

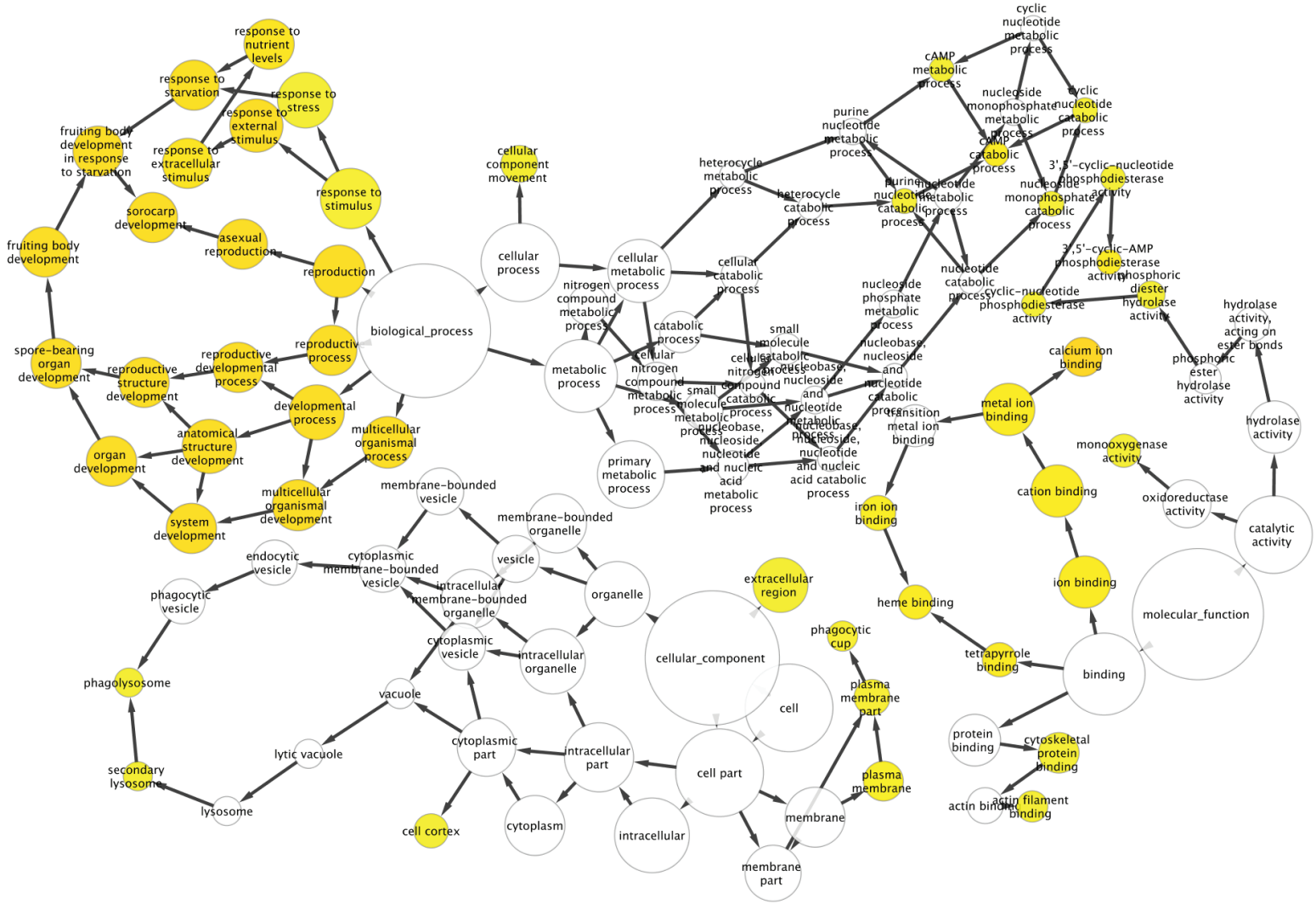


Figure A.2: GO enrichment in downregulated genes in five-hour pulsed *chdA*-null cells

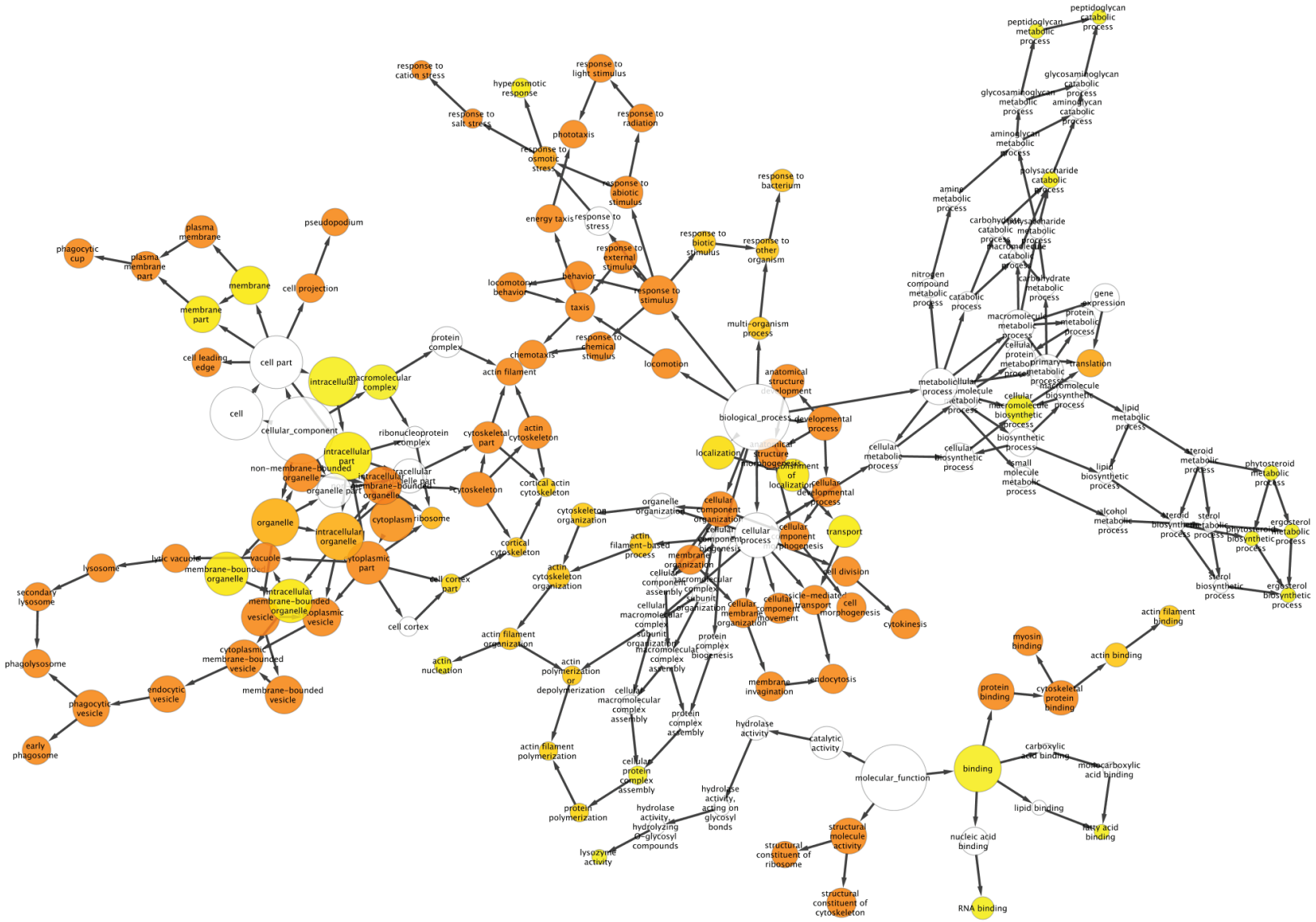


Figure A.3: GO enrichment in downregulated genes in growing *chdB*-null cells

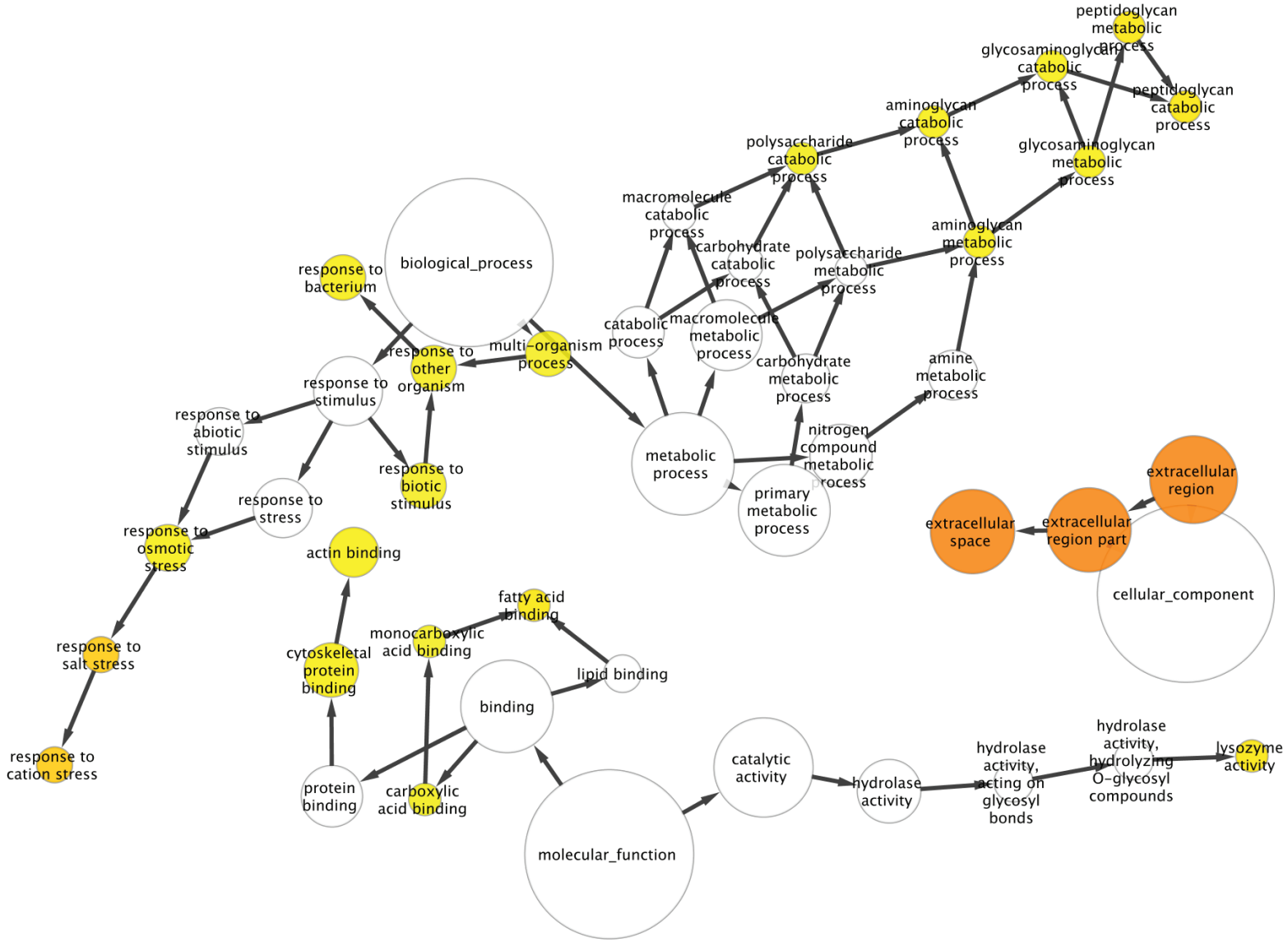


Figure A.4: GO enrichment in downregulated genes in growing *chdC*-null cells

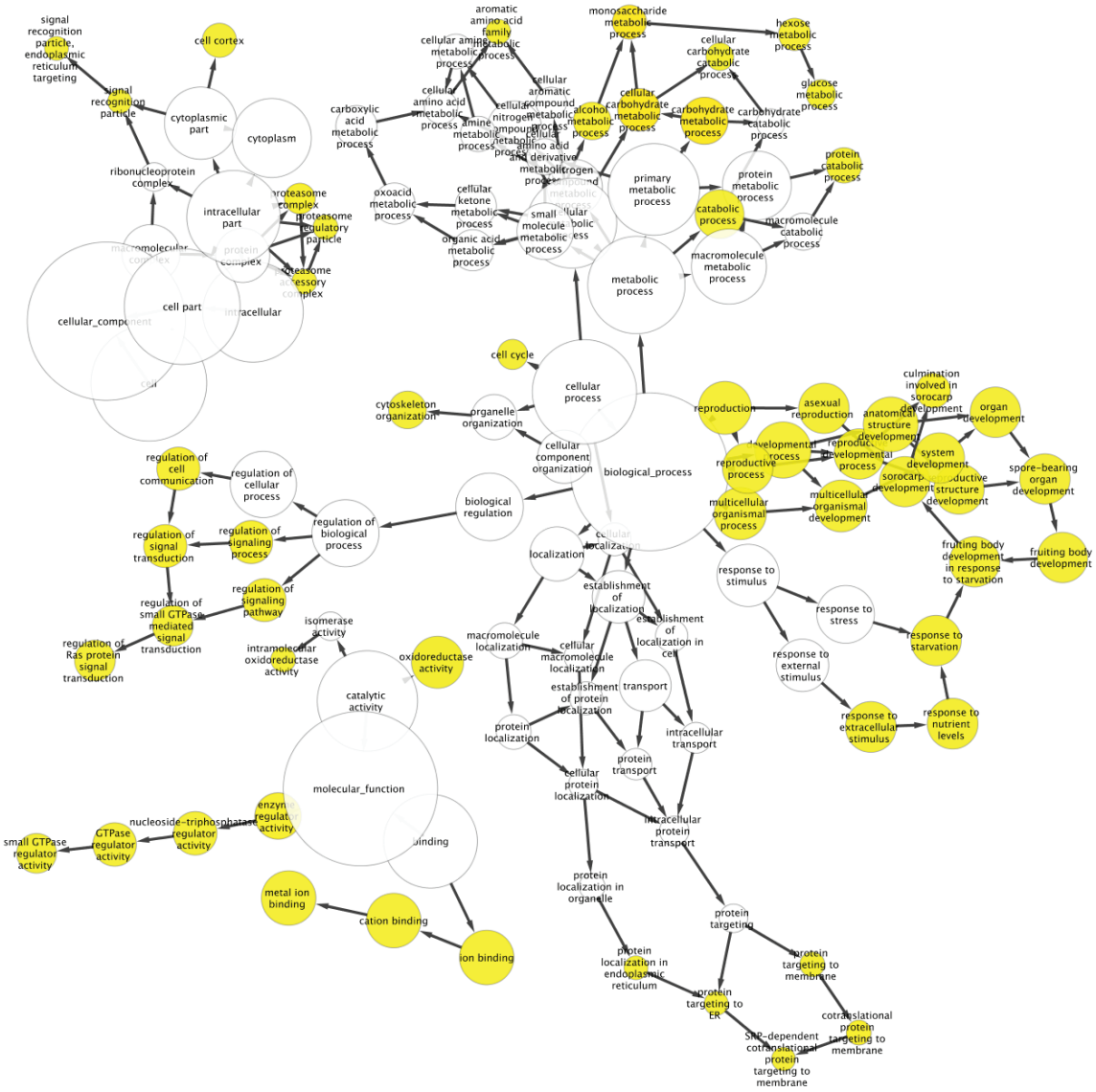


Figure A.5: GO enrichment in downregulated genes in five-hour pulsed *chdC*-null cells

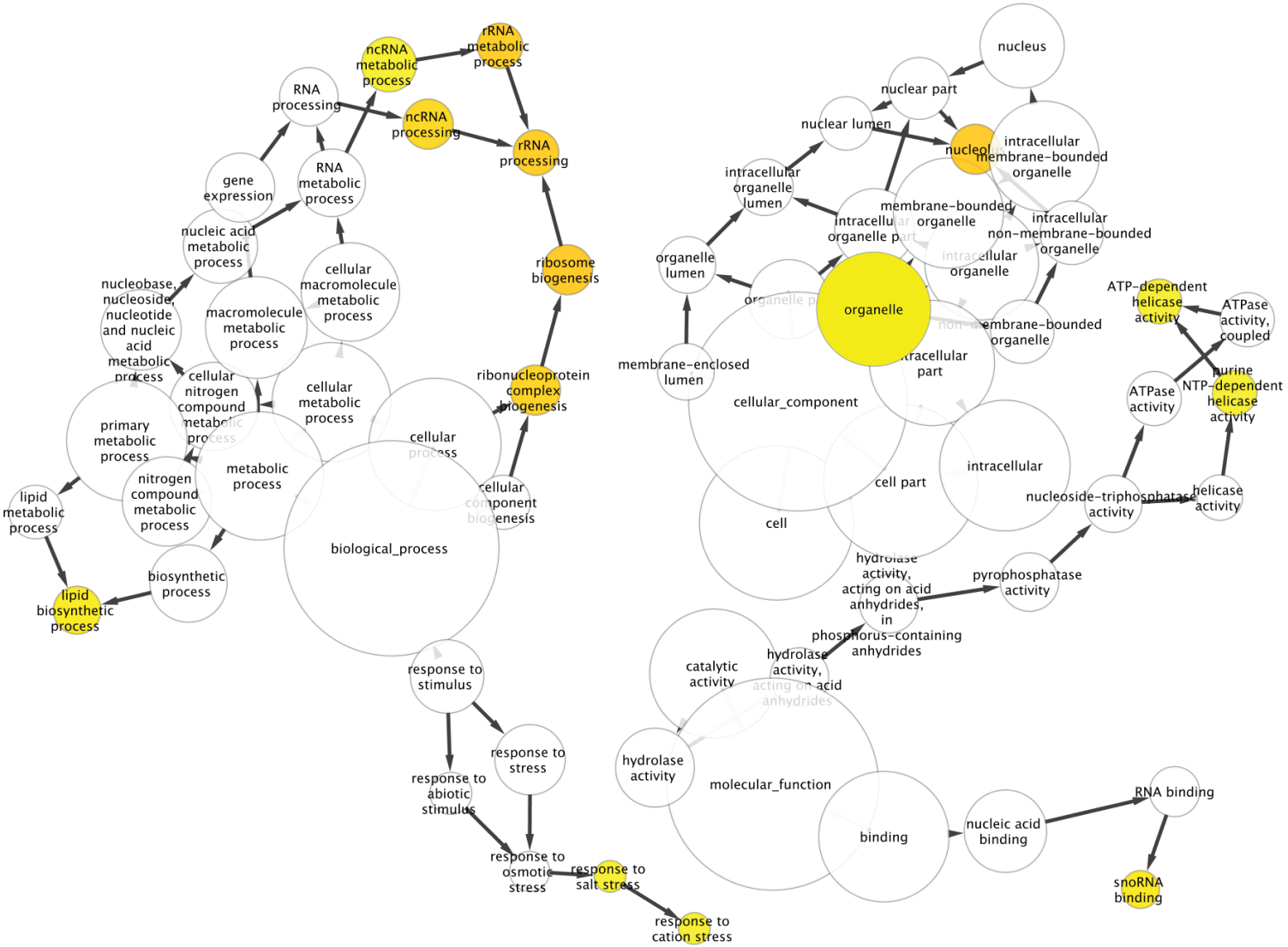


Figure A.6: GO enrichment in upregulated genes in five-hour pulsed *chdC*-null cells

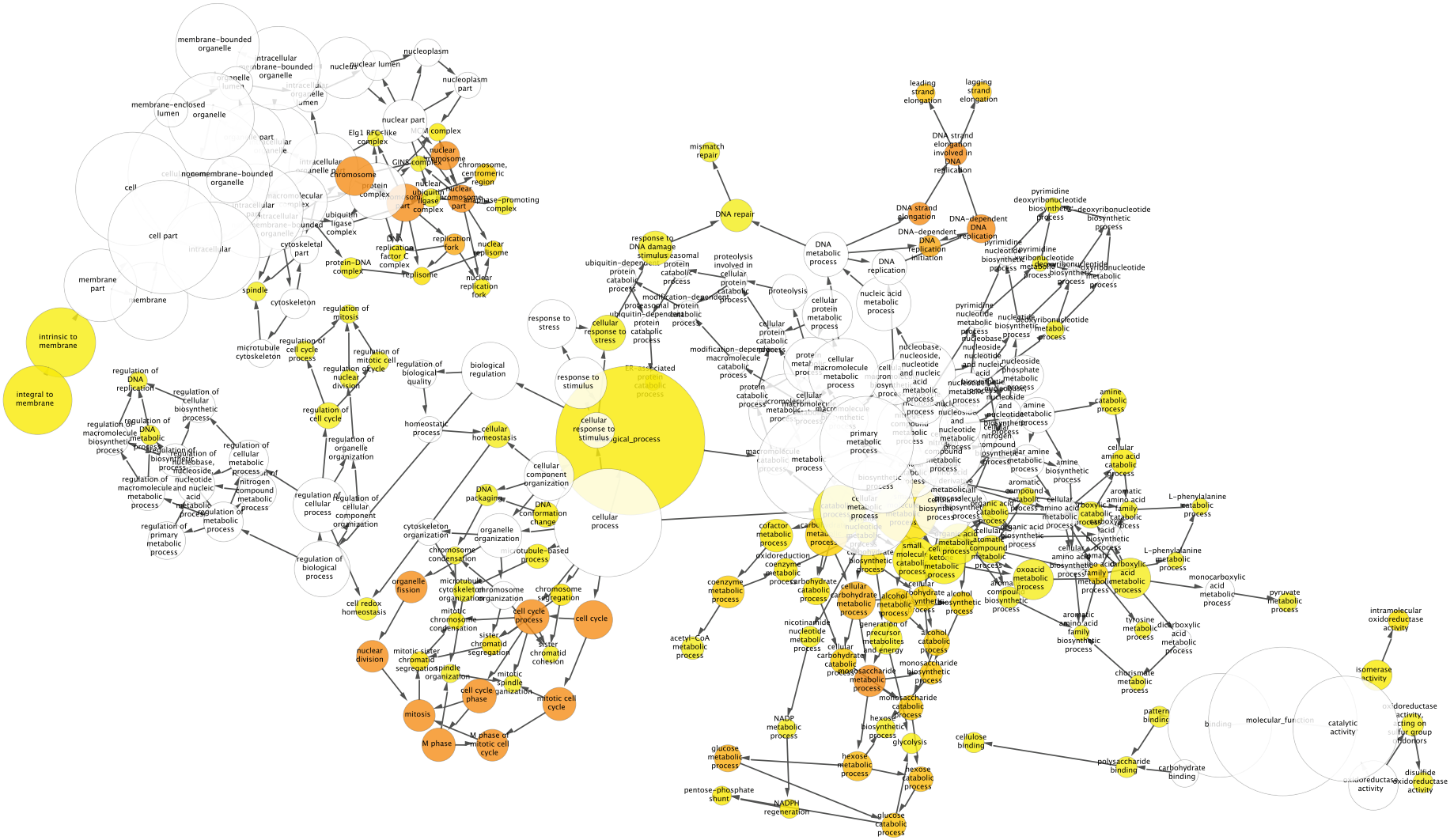


Figure A.7: GO enrichment in downregulated genes in loose-mound-stage *chdC*-null cells

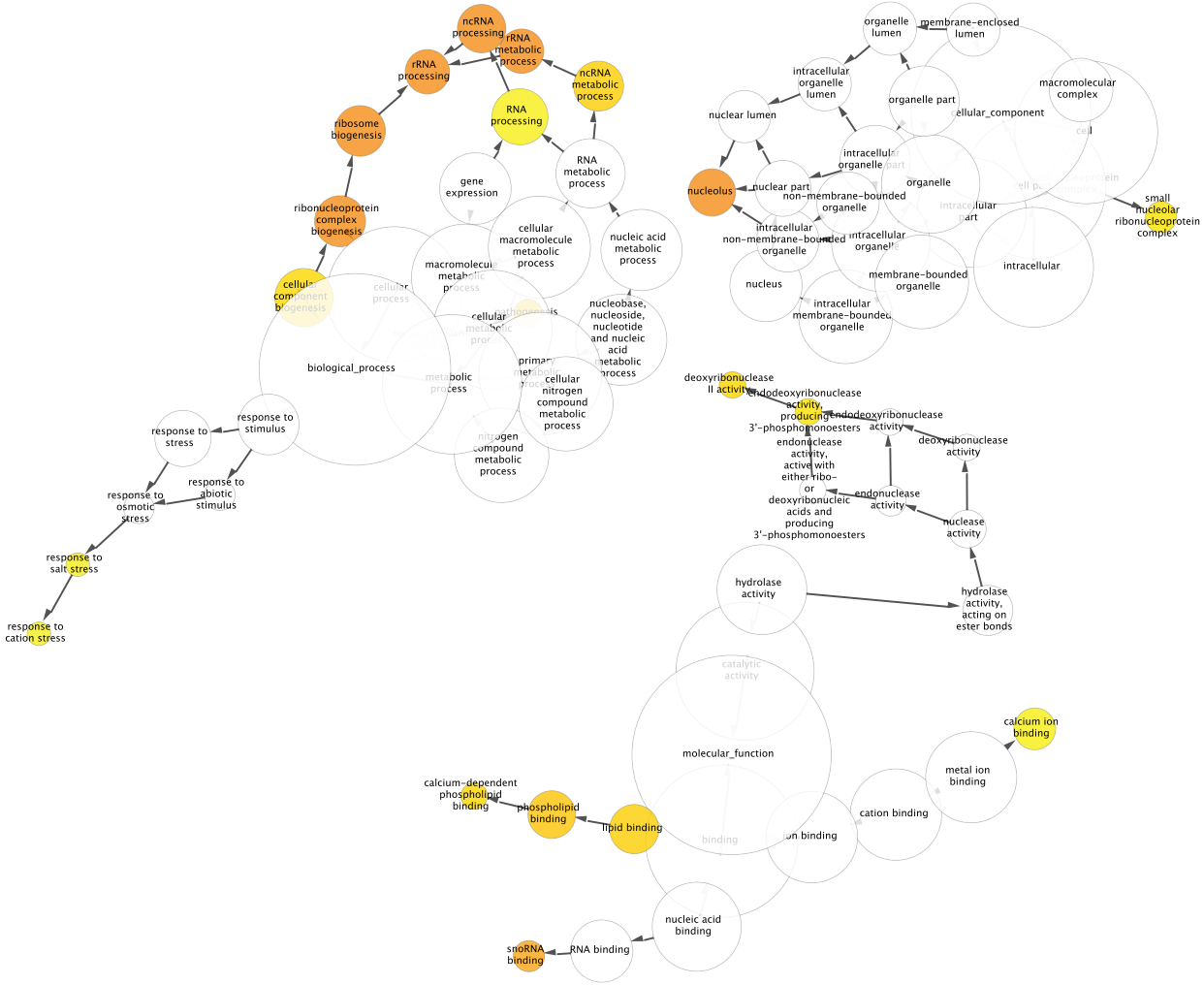


Figure A.8: GO enrichment in upregulated genes in loose-mound-stage *chdC*-null cells

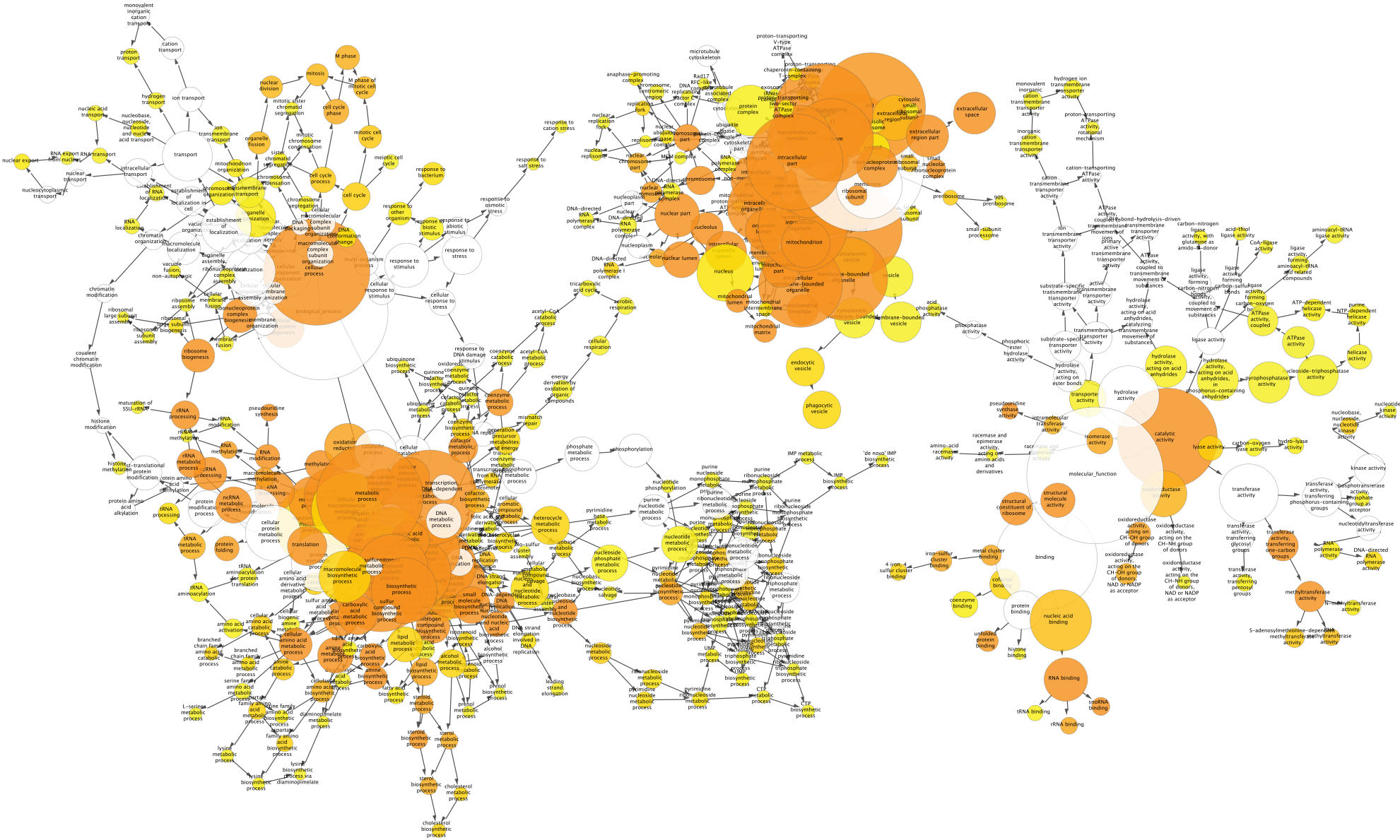


Figure A.9: GO enrichment in downregulated genes in five-hour pulsed Ax2 cells relative to growing Ax2 cells.

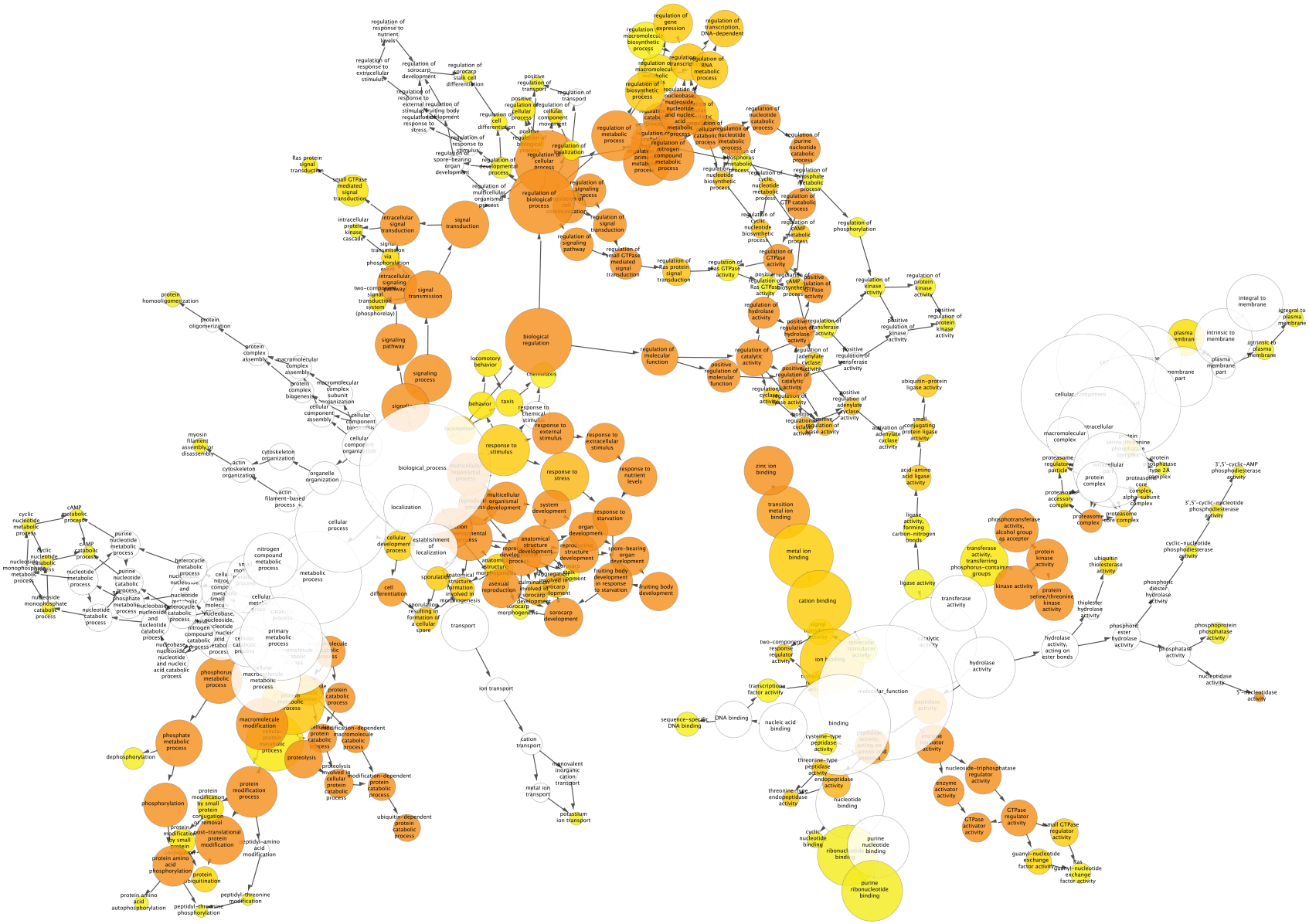


Figure A.10: GO enrichment in upregulated genes in five-hour pulsed Ax2 cells relative to growing Ax2 cells.

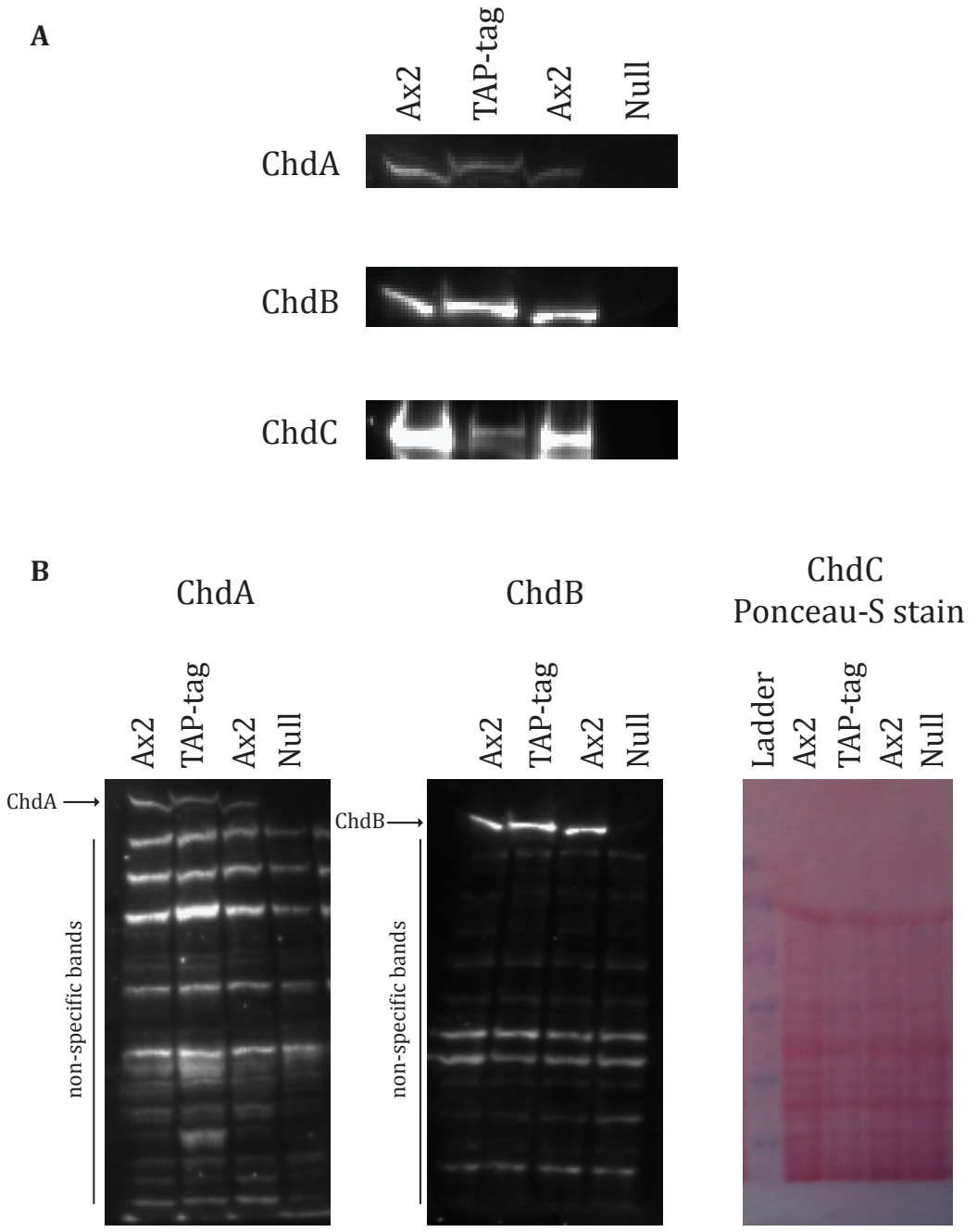


Figure A.11: Quality control of Chd specific antibodies

A. Western blotting on whole cell extracts from growing cells. Ax2, cell lines with TAP (Tandem Affinity Purification)-tagged Chd proteins and cells lacking Chd protein. **B.** Full blots from A. showing non-specific bands are unchanged in *chdA*-null and *chdB*-null cells. Ponceau-S staining of the membrane for the ChdC blot demonstrates equal loading.

Appendix B

Published work

© 2013. Published by The Company of Biologists Ltd | Development (2013) 140, 4926-4936 doi:10.1242/dev.099879

RESEARCH ARTICLE

Different CHD chromatin remodelers are required for expression of distinct gene sets and specific stages during development of *Dictyostelium discoideum*

James L. Platt[1,2], Benjamin J. Rogers[2], Kelley C. Rogers[1], Adrian J. Harwood[2,*] and Alan R. Kimmel[1,*]

ABSTRACT

Control of chromatin structure is crucial for multicellular development and regulation of cell differentiation. The CHD (chromodomain-helicase-DNA binding) protein family is one of the major ATP-dependent, chromatin remodeling factors that regulate nucleosome positioning and access of transcription factors and RNA polymerase to the eukaryotic genome. There are three mammalian CHD subfamilies and their impaired functions are associated with several human diseases. Here, we identify three CHD orthologs (ChdA, ChdB and ChdC) in *Dictyostelium discoideum*. These CHDs are expressed throughout development, but with unique patterns. Null mutants lacking each CHD have distinct phenotypes that reflect their expression patterns and suggest functional specificity. Accordingly, using genome-wide (RNA-seq) transcriptome profiling for each null strain, we show that the different CHDs regulate distinct gene sets during both growth and development. ChdC is an apparent ortholog of the mammalian Class III CHD group that is associated with the human CHARGE syndrome, and GO analyses of aberrant gene expression in *chdC* nulls suggest defects in both cell-autonomous and non-autonomous signaling, which have been confirmed through analyses of *chdC* nulls developed in pure populations or with low levels of wild-type cells. This study provides novel insight into the broad function of CHDs in the regulation development and disease, through chromatin-mediated changes in directed gene expression.

KEY WORDS: SNF2, Transcriptome profiling, RNA-seq, Chemotaxis, Growth, Differentiation

INTRODUCTION

The packaging of DNA into chromatin physically limits access of transcription factors and other elements of the transcription machinery to their target genes. Throughout development, changes in chromatin organization are exploited to control gene expression. In particular, alterations in nucleosome positioning along genomic DNA, termed chromatin remodeling, impact transcriptional activity, both positively and negatively (Ho and Crabtree, 2010; Euskirchen et al., 2012). Importantly, as nucleosome positions are preserved during DNA replication, inherited differences in positioning can alter multicellular development or cellular properties during disease states.

ATP-dependent chromatin remodeling proteins use the energy of ATP hydrolysis to physically change the interaction between DNA and nucleosomes, which can promote or limit DNA access to a variety of regulators (Hopfner et al., 2012). The archetypal ATP-dependent chromatin remodeler was discovered in yeast where mutations in a single yeast gene, designated *SWI2/SNF2*, lead to aberrant mating-type switching (*swi2*) (Breeden and Nasmyth, 1987) and glucose repression (*snf2*) (Abrams et al., 1986). *SWI2/SNF2* has a DNA-dependent (helicase-like) ATPase that is part of a large multiprotein complex. The *Drosophila* ortholog of SWI2/SNF2, Brahma (Brm), regulates cell fate by acting as a transcriptional co-activator of homeotic gene expression in the Antennapedia and Bithorax complexes (Tamkun et al., 1992), and the mammalian ortholog BRG1 (SMARCA4 – Mouse Genome Informatics), Brahma-related gene 1, is implicated in T-cell development, ES cell differentiation and tumor suppression (Hargreaves and Crabtree, 2011).

Other chromatin remodeling families have been identified that share the ATPase motif, but differ in other protein domains (Clapier and Cairns, 2009; Eisen et al., 1995; Ryan and Owen-Hughes, 2011). These include ISWI (imitation switch), INO80 (inositol requiring 80) and CHD (chromodomain, helicase, DNA binding) proteins (Ho and Crabtree, 2010). The first member of the CHD family was identified in mice (Delmas et al., 1993), but CHD family members were subsequently described in a variety of eukaryotic organisms, including humans, *Drosophila*, *Saccharomyces cerevisiae*, *C. elegans* and *Arabidopsis* (Flanagan et al., 2007). The human CHD family comprises nine members, CHD1-9, divided into three subfamilies – CHD1-2 (class I), CHD3-5 (class II) and CHD6-9 (class III) – based upon their chromodomain types and additional motif features (Yap and Zhou, 2011).

Functionally, the CHD proteins play important roles in the development of multicellular organisms. CHD1 is essential for maintaining pluripopotency of mouse ES cells (Gaspar-Maia et al., 2009), and *Drosophila Chd1* nulls are both male and female sterile (McDaniel et al., 2008). Loss of *Chd2* in mice causes embryonic lethality (Marfella et al., 2006a) and conditional inactivation of *CHD4* in hematopoietic cells impairs their differentiation and development (Williams et al., 2004). Haploinsufficient mutations of *CHD7* are observed in 60-80% of cases of CHARGE syndrome (Janssen et al., 2012; Lalani et al., 2006; Layman et al., 2010; Zentner et al., 2010), a severe developmental disorder in humans characterized by coloboma of the eye, heart defects, atresia of the choanae, retardation of growth, and genital and ear abnormalities (Pagon et al., 1981). *Chd7*$^{+/-}$ mice recapitulate some of the human symptoms, and *Chd7*$^{-/-}$ mice are embryonic lethal (E10.5) (Bosman

[1]Laboratory of Cellular and Developmental Biology, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [2]School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AX, UK.

*Authors for correspondence (harwoodaj@cf.ac.uk; alank@helix.nih.gov)

Received 6 June 2013; Accepted 20 September 2013

et al., 2005; Hurd et al., 2007). Wild-type CHD7 can direct nucleosome sliding *in vitro*, and CHD7 proteins that recapitulate CHARGE mutations are significantly impaired in this activity (Bouazoune and Kingston, 2012). CHD7 is suggested to function on tissue-specific genes that regulate development and differentiation (Schnetz et al., 2009; Schnetz et al., 2010). CHD8 also has an important role in development, is highly expressed during early embryogenesis and appears to modulate WNT/β-catenin signaling through interaction and downregulation of response genes (Nishiyama et al., 2012; Thompson et al., 2008). Other CHD functions are associated with human diseases, including some neuroblastomas (CHD5), Hodgkin's lymphoma (CHD3) (Lemos et al., 2003) and small-cell lung cancers (CHD7) (Pleasance et al., 2010).

It is not yet clear how the different CHDs may cause these multiple, diverse and distinct patterns of developmental deficiencies. Broad investigations of the developmental targets of the individual CHD proteins and their effect on cell behavior are required to establish the connectivity of CHD function with gene expression, developmental phenotypes and disease states. We have now identified the CHD orthologs of *Dictyostelium* and associate loss-of-function mutations with distinct developmental phenotypes and altered gene expression profiles.

Dictyostelium grow as single-celled organisms, in which cellular growth and division are dependent on an adequate food supply (McMains et al., 2008). Nutrient depletion triggers a developmental program to ultimately form a terminally differentiated, multicellular structure: the fruiting body. Development occurs in distinct stages as cells first sense starvation and high population density, and then aggregate via a chemotactic response to secreted cAMP signals to form the initial multicellular organism. These multicellular mounds then differentiate into patterned prespore and prestalk cell populations, before terminal differentiation into structures comprising mature spore and stalk cells (Kessin, 2001).

Using genome-wide transcript profiling (RNA-seq) during growth and development to assay gene expression in *Dictyostelium* (Loomis and Shaulsky, 2011; Parikh et al., 2010) that individually lack the different CHD family members, we correlate altered transcription profiles with the developmental phenotypes of the different *chd* nulls. We conclude that the different CHDs act to control diverse aspects of *Dictyostelium* development, perhaps by targeting distinct sets of genetic loci.

ChdC in *Dictyostelium* is an apparent ortholog of the mammalian Class III CHD group that is associated with the human CHARGE syndrome. Similar to *CHD7* mutations in humans (Zentner et al., 2010), *chdC* nulls have the strongest phenotype in *Dictyostelium*, with impaired growth, chemotaxis, cell differentiation and developmental arrest. Analyses following transcriptome profiling indicate defects in both cell-autonomous and non-autonomous functions, phenotypes confirmed by chimeric development of *chdC* nulls with wild-type cells. These data yield novel insight into the broad function of CHDs, with particular relevance to the Class III CHD group and their association with highly severe genetic syndromes in humans.

RESULTS

Dictyostelium possess three CHD proteins

We initiated separate insertional (REMI) mutagenic screens (Kuspa, 2006; Artemenko et al., 2011) to identify *Dictyostelium* strains that exhibited defects in chemotaxis to cAMP or that arrested development. An identical CHD-encoding gene was identified in these distinct screens as a novel genetic modifier of both chemotaxis and development. Extended searches for related sequences in the *Dictyostelium* genome identified two additional CHD proteins (supplementary material Fig. S1). We have termed the *Dictyostelium* CHD genes *ChdA* (DDB_G0284171), *ChdB* (DDB_G0280705) and *ChdC* (DDB_G0293012); *ChdC* was the gene found in the original mutagenic screens. Our sequence analyses also identified related SWI2/SNF2-type genes (Ryan and Owen-Hughes, 2011) in *Dictyostelium*, encoding members of the bona fide SWI2/SNF2 (*snf2*; DDB_G0271052 and DDB_G0285205), ISWI (*isw*; DDB_G0292948), SWR (*swr1*; DDB_G0267638) and INO80 (*ino80*; DDB_G0292358) chromatin remodeling families (see supplementary material Figs S1, S2). All of the encoded proteins possess the DNA-dependent ATPase domain and a DNA-binding motif, but differ in overall organization (supplementary material Fig. S2); the tandem chromodomains are a defining moiety for the CHDs (Blus et al., 2011). More distantly related SWI2/SNF2 members in *Dictyostelium* were also noted, including the Mot1 (modifier of transcription) and Rad54 families (supplementary material Fig. S1).

Dictyostelium ChdA shows the strongest similarity to human CHD subfamily I, clustering with human CHD1/2 and other species Chd1-subtypes through the entire protein sequence (Ryan et al., 2011), including the C-terminal DUF4208 (supplementary material Figs S1, S2). *Dictyostelium* ChdB possesses weak C-terminal CHDCT2 domains, which are generally characteristic of CHD subfamily II (Ryan and Owen-Hughes, 2011); however, ChdB lacks the N-terminal PHD finger present in most subfamily II members and globally aligns better with human CHD6-9 proteins (supplementary material Fig. S1). ChdC (DDB_G0293012) is the largest *Dictyostelium* CHD, and clusters most strongly with the CHD6-9, subfamily III group (supplementary material Figs S1, S2). Although ChdC has C-terminal SANT and SLIDE DBD motifs characteristic of other CHD III members (Ryan and Owen-Hughes, 2011), it lacks the conserved BRK site. Motif prediction in ChdC is complicated by the large number of homopolymeric residues embedded in its long C-terminal region. In addition, although BRK domains are of unknown function, they appear to be unique to the metazoa; thus, although the yeast SWI2/SNF2 and metazoan Brm/BRG are considered to be orthologs and functionally equivalent, only Brm/BRG proteins possess BRK domains (Ryan and Owen-Hughes, 2011). ChdC may uniquely possess a SUMO-like amino acid sequence of unknown function. Collectively, the individual CHDs of *Dictyostelium* would seem to embody different CHD types.

Developmental expression of ChdA, ChdB and ChdC

We examined developmental protein expression patterns of each CHD by immunoblotting, using protein-specific antibodies (Fig. 1A). Each CHD is detected during growth and the early stages of development, but they exhibit large expression differences following multicellular aggregation. ChdA has low relative expression during aggregation and early stages of multicellular development, but has increased abundance during terminal differentiation, after 16 hours. By contrast, ChdB shows peak abundance at 4 hours of development, the mid-stage of cellular aggregation (Fig. 1A). ChdB abundance then declines and is undetectable during late stages of development. Finally, ChdC shows peak abundance between 8 and 10 hours of development, as cells complete aggregate formation and initiate multicellular development. Again, as seen for ChdB, ChdC is undetected during late stages of development; thus, ChdA predominates at terminal differentiation.

Development (2013) doi:10.1242/dev.099879

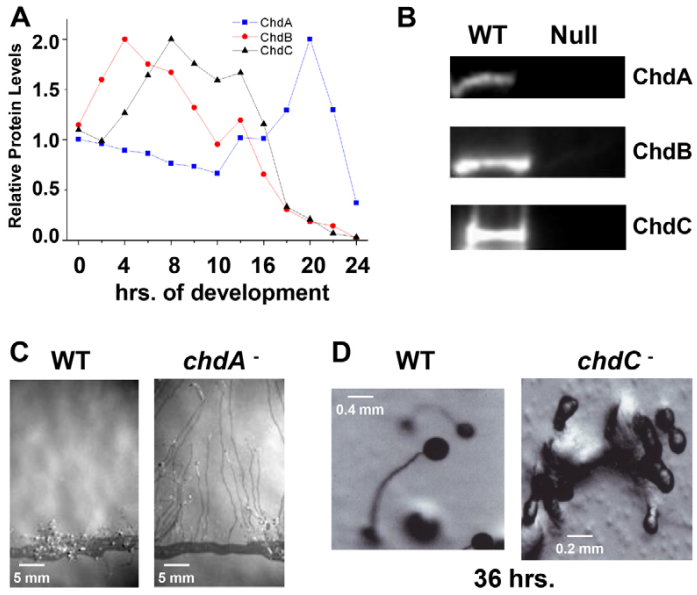


Fig. 1. The three CHD proteins of *Dictyostelium* have distinct developmental roles. (A) Whole-cell extracts were prepared from developing *Dictyostelium* at 2-hour intervals and ChdA, ChdB and ChdC proteins levels measured by immunoblot assay. Quantified data are expressed as normalized protein levels relative to expression at 0 hour. (B) Whole-cell extracts were prepared from growing wild-type, *chdA*-null, *chdB*-null and *chdC*-null cells, and ChdA, ChdB and ChdC proteins expression measured by immunoblot assay. (C) Wild-type and *chdA*-null cells were plated for development at the edge of nitrocellulose membranes and pseudoplasmodia migration assayed by exposure to a directional light source. (D) Wild-type and *chdC*-null cells were terminally developed. *chdC* nulls arrest development at multi-cell aggregation.

Major phenotypic developmental defects in cells lacking *chdA*, *chdB* or *chdC*

To investigate CHD functions in *Dictyostelium*, we made deletion strains for each of the Chd genes; immunoblotting confirmed the absence of the targeted protein in each mutant strain (Fig. 1B). Mutants lacking each gene were viable, but possessed different developmental phenotypes that largely correlated with their peak expression patterns. Disruption of *chdA* causes a slight (~1 hour) delay in streaming maxima, but no apparent effect on aggregate, tip or slug formation during early *Dictyostelium* development. However, the *chdA* nulls exhibited a significant delay at the onset of terminal differentiation (i.e. culmination), ~16 hours post-starvation. This delay is best illustrated when *Dictyostelium* are developed on nitrocellulose filters where wild-type cells rarely develop migratory pseudoplasmodia, but instead rapidly progress to terminal differentiation. Under the same developmental conditions, *chdA*-null mutants form migratory pseudoplasmodia at high frequency (Fig. 1C) and persist at this stage for extended times, when compared with wild type.

ChdB is maximally expressed during early development and *chdB*-null mutants take ~2-4 hours longer to form multicellular mounds than do wild type. The timing defect does not appear to be stage specific; initiation of streaming, loose mound formation and tight aggregation are proportionally delayed. It is currently unclear how this phenotype originates. Beyond this developmental delay, no prominent phenotypic effects were apparent.

The strongest developmentally defective *chd*-null phenotype was observed with cells lacking *chdC*. *chdC*-null cells exhibited a delay during aggregation of ~2 hours and a developmental arrest at the multicellular mound stage, at ~12 hours; development failed to progress to the tipped mound stage, but instead formed bulbous extensions when development terminated (Fig. 1D). The developmental arrest in *chdC*-null mutant cells was completely penetrant, with no migratory pseudoplasmodia or mature terminal wild-type structures formed.

Differential growth effects of the CHDs

During cell culture, we observed a slower growth rate for *chdC* nulls, than for the other strains in axenic suspension. To quantify growth rates, we identically diluted wild type, and *chdA*, *chdB* and *chdC* nulls from log phase growth (~2×10^6 cells/ml) into fresh growth media, and monitored growth over several days (Fig. 2A). The growth rate for *chdC* nulls was consistently halved compared with all other strains. The growth defect of *chdC* nulls is not due to deficiencies in cytokinesis, and *chdC* nulls are not multi-nucleate.

We considered whether *chdC* nulls might be aberrantly regulated during the transition from growth to development in response to nutrient depletion. Prior to the onset of starvation, *Dictyostelium* monitor the accumulation of various secreted factors to sense an increase in cell density (Burdine and Clarke, 1995). Inappropriate cell density sensing by *chdC* nulls could lead to growth defects or delayed entry into development and subsequent asynchrony. Using RNA and protein samples taken from growing wild-type cultures at ranges of ~5×10^4 to 5×10^6 cells/ml, we show that the relative expression of both *ChdC* mRNA (Fig. 2B) and ChdC protein increases as cell density rises from ~5×10^5 to 1×10^6 cells/ml. Thus, *ChdC* regulation appears linked to a pre-starvation response.

To determine whether ChdC function might additionally participate in the pre-starvation response, we examined cell density-dependent expression of the prototypic pre-starvation response marker Discoidin I (Burdine and Clarke, 1995) in wild-type and *chdC*-null cells. Discoidin I mRNA expression was similarly enhanced in wild-type and *chdC*-null cells at a cell density transition from ~3×10^5 to 1×10^6 cells/ml (Fig. 2C). Although the data indicate that ChdC expression is induced by the pre-starvation response, as with several other early developmental markers, they do not support a requirement for ChdC to mediate this process.

CHD effects on chemotaxis

To further assess CHD function during early development, we examined wild-type and *chd*-null cells that had been induced to differentiate in shaking culture in the presence of exogenous pulses of cAMP. These cAMP conditions mimic early stage developmental signaling and bypass the potential for defects in endogenous cAMP synthesis (McMains et al., 2008). Cells were then assayed for their ability to move by chemotaxis towards an exogenous cAMP source. Chemotactic behavior of *chdB* nulls is highly similar to wild type

Development (2013) doi:10.1242/dev.099879

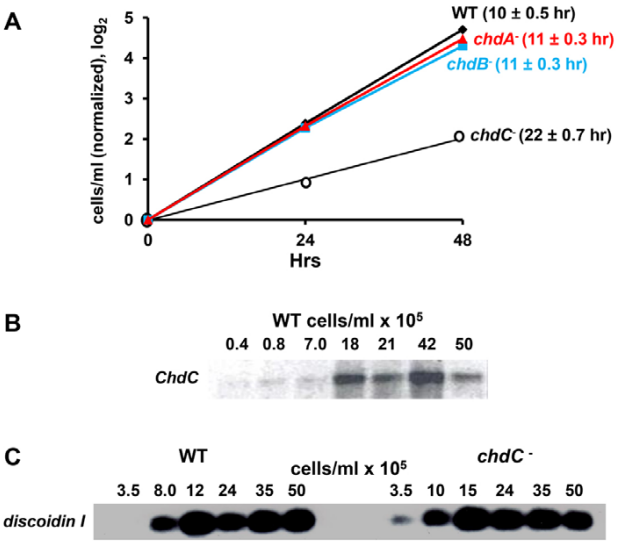


Fig. 2. ChdC and cell growth. (A) Wild-type, *chdA*-, *chdB*- and *chdC*-null cells were grown to log phase (<2×10^6 cells/ml) and diluted ~20× into fresh media and cell aliquots taken at various times to evaluate relative growth rate. Growth increase for each strain was normalized to the initial cell density and plotted as a $\log_2$ scale; doubling times for each strain are indicated. (B) RNA was prepared from growing wild-type cells at the indicated cell densities and hybridized by RNA gel blot with a *ChdC* probe. (C) RNA was prepared from growing wild-type and *chdC*$^-$ cells at the indicated cell densities and hybridized by RNA gel blot to a probe specific for the coding region of discoidin I.

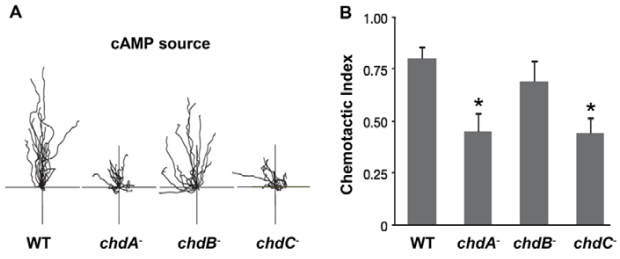


Fig. 3. Differential CHD functions during chemotaxis. (A) Individual wild-type, *chdA*-null, *chdB*-null and *chdC*-null cells were imaged for chemotaxis toward a directed cAMP point source during a 15-minute period. Tracings reflect the migration paths for different cells of each strain. (B) The chemotactic index was plotted for each strain as mean±s.e.m. The Mann-Whitney test was used to evaluate significance. *P<0.001.

(Fig. 3). By contrast, both *chdA*- and *chdC*-null cells had significant chemotaxis defects, including reduced cell speed, increased cell turning and lower chemotactic indices. Nonetheless, chemotaxis to cAMP was not absent in either *chdA* or *chdC* nulls, consistent with their ability to form multicellular aggregates during development. The discrete phenotypes of each strain suggest that the individual CHDs have distinct actions, with only limited functional redundancy.

Aberrant and unique transcriptional profiles of *chd* nulls during growth and early development

To examine directly differential effects on gene expression upon loss of ChdA, ChdB or ChdC, we applied high-throughput transcriptome sequencing (RNA-seq) for the three mutant strains, and compared relative mRNA transcript levels of each strain with that of wild type and with each other. Transcriptome profiles were determined for cells during growth and following 5 hours of stimulation with exogenous 100 nM pulses of cAMP. These latter conditions promote early differentiation and bypass the requirement for endogenous extracellular cAMP signaling (McMains et al., 2008).

During growth, ~10-15% of all genes are mis-expressed by over twofold (P<0.05) in each strain compared with wild type (Fig. 4A). Yet most wild-type gene expression is not altered upon loss of ChdA, ChdB or ChdC. When expression profile differences are compared between any two strains (Fig. 4A) or if all mis-expressed genes are collectively analyzed by comparative heat maps (Fig. 4B), they cluster as distinct gene sets. Although the gene sets for each strain are not fully unique, their observed intersections are no greater than would occur by the random distribution of independent variables. We conclude that ChdA, ChdB and ChdC predominantly regulate distinct gene sets during growth. We did not detect compensatory expression profile changes of *chdA*, *chdB* or *chdC* mRNAs in the various mutant strains.

We next surveyed for functional gene classes that might be mis-expressed during growth of the various *chd* nulls, using Gene Ontology (GO) annotation. Despite the large number of genes that are downregulated in *chdA* and *chdB* nulls, little group biases could be discerned during growth. This is consistent with a lack of a significant growth phenotype for these strains. By contrast, *chdC* nulls displayed a poor growth phenotype and a large number of genes involved in mitochondrial and metabolic processes were poorly expressed exclusively in *chdC* nulls (Fig. 5A and supplementary material Fig. S3A).

Many regulators of aggregation also showed altered expression in growing *chdC* nulls. *ctnA*, *cf45-1* and *cf60*, three gene members of the secreted countin complex that potentiate inter-cell signaling and chemotactic movement (Brock and Gomer, 1999; Brock et al., 2003; Brock et al., 2006), were downregulated in *chdC* nulls. Expression of *lmcA*, *lmcB* and *srsA*, which function similarly to countin, was also suppressed, while expression of other modulators of aggregate size (e.g. *cnrJ*, *cnrG* and *cnrK*) was upregulated.

We also mapped gene expression patterns in *chdA*, *chdB* and *chdC* nulls that had been induced to synchronously differentiate by cAMP pulsing, to a time point akin to ~8 hours of development, when cells are chemotactically competent (Fig. 4C). As wild-type cells initiate development, there are large changes in the *Dictyostelium* transcriptome (Van Driessche et al., 2002; Loomis and Shaulsky, 2011). These changes are further augmented in the *chd* nulls, where individually 15-30% of all genes are mis-expressed by more than twofold (P<0.05). All of the mis-expressed genes were clustered in order to compare expression differences among all the cell lines, and displayed as heat maps (Fig. 4D). Although there is a trend for each strain to similarly exhibit more gene de-repression (i.e. overexpression) than repression (i.e. under-expression), the expression patterns for the *chdA*, *chdB* and *chdC* nulls display distinct gene expression profiles, underscoring their unique developmental phenotypes.

A complex series of interacting circuitries are shown to direct chemotactic movement and signal response to extracellular cAMP in *Dictyostelium* (Kortholt and van Haastert, 2008; McMains et al., 2008; Swaney et al., 2010). Although they primarily function downstream of the CAR1/Gα2βγ (cAMP receptor-G protein) complex, they diverge into several broad paths, including the cGMP-dependent and -independent actions of guanylyl cyclases

Development (2013) doi:10.1242/dev.099879

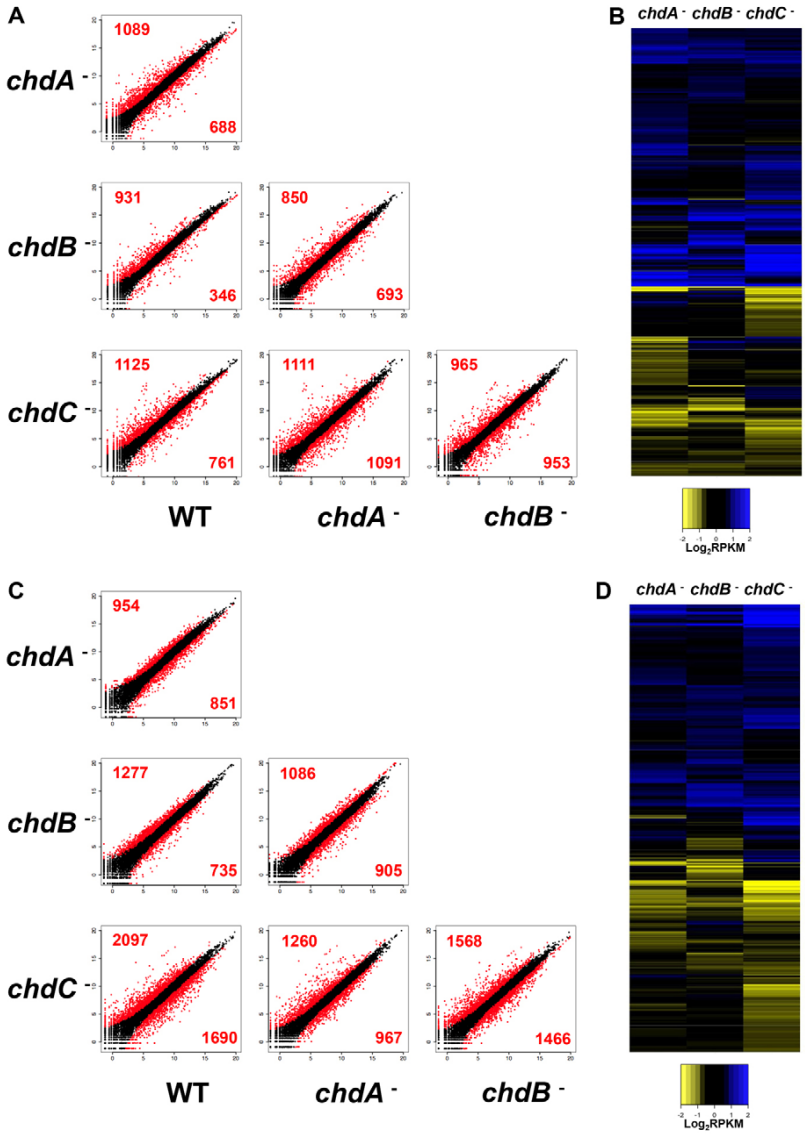


Fig. 4. Large transcriptome differences among the *chd* nulls during growth or differentiation. Scatter plots of normalized (RPKM, $\log_2$) expression values for genes in wild-type, *chdA*-, *chdB*- and *chdC*-null cells during growth (A) or after 5 hours of cAMP-pulsing (C). Data points with more than twofold (P<0.05) expression differences between indicated strains are in red, and less than twofold (P<0.05) expression differences are black. Heat maps of $\log_2$-fold changes in RPKM values in all genes with expression differences [≥2-fold (P<0.05)] in *chdA*, *chdB* or *chdC* relative to wild type, during growth (B) or after 5 hours of cAMP pulsing (D).

sGC and GCA, the PIP_3-dependent and -independent functions mediated via mTORC2, inter- and intracellular cAMP accumulation and signaling, and the interactions of Ca^{+2} and phospholipase A_2 pathways (Kortholt and van Haastert, 2008). These networks have been defined biochemically and genetically and include ~50 genes. Of these, 50% are mis-expressed in *chdC* nulls (Fig. 5B), suggesting defects in chemotactic response, but also in extracellular cAMP signal relay (e.g. Gα2, PKBR1, CRAC, ERK2). Other gene classes that influence chemotaxis via cAMP wave propagation and cytoskeletal organization (supplementary material Fig. S3B,C) are also preferentially mis-expressed in *chdC* nulls, and to a lesser degree in *chdA* nulls. Thus, although the ChdC and ChdA gene sets are distinct, there are some shared elements. Yet as many more chemotactic genes are mis-expressed in *chdC*-nulls than in *chdA*-nulls, the data also explain the more minor streaming defects of *chdA* nulls during development. In addition, as several independent signaling pathways regulate chemotaxis during *Dictyostelium* development, their collective and compensatory activities can dampen effects caused by deficiencies in selective elements (Van Haastert and Veltman, 2007; Kortholt and van Haastert, 2008). Accordingly, chemotaxis is compromised, but is not eliminated, in *chdA* and *chdC* nulls.

The culmination defect of *chdA* nulls is a post-slug formation, late developmental event. The regulatory pathways for this stage are complex (Harwood et al., 1992; Guo et al., 1999) and we do not observe gene expression defects at early stages that are predictive of the phenotype. Although *chdB* nulls show large numbers of mis-expressed genes, the cells do not exhibit a strong early phenotype and the mis-expressed genes do not group into simple GO collections.

cAMP signaling also directs developmentally regulated gene expression changes. Cells respond differentially to low (nM) level pulses of cAMP during early development and to non-varying, higher (μM) cAMP levels at the multicellular stage. Although only 13 genes are currently classified as exclusively pulse-dependent (Iranfar et al., 2003), 11 of these are under-expressed in *chdC*-nulls

Development (2013) doi:10.1242/dev.099879

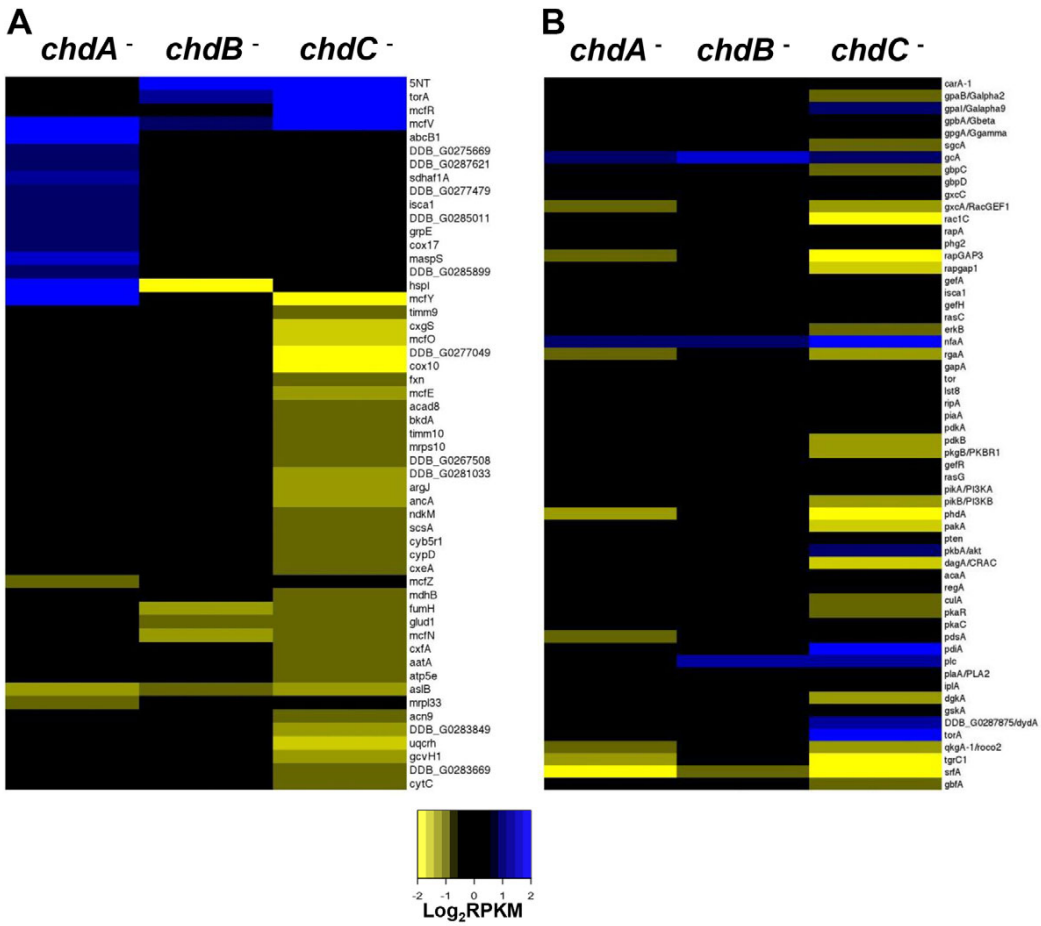


Fig. 5. Expression changes in GO clustered genes for growth or development. (A) Heat maps show $\log_2$-fold changes in RPKM values for 53 genes with GO annotations associated with mitochondrial functions in the *chdA*, *chdB* and *chdC* nulls relative to wild type during growth. (B) Fifty-nine gene functions have been mapped to a cAMP chemotaxis network that acts downstream of CAR1 (carA). Heat maps show $\log_2$-fold change in RPKM values for these genes in the *chdA*, *chdB* and *chdC* nulls relative to wild type, during differentiation in the presence of cAMP pulses for 5 hours. Genes are sorted approximately into separate signaling pathways downstream of the cAMP receptor.

and 10 are under-expressed *chdA* nulls (supplementary material Fig. S3C); only 1 gene is affected in *chdB* nulls. In contrast to pulse-regulated gene expression, aggregation and cell-type specific gene expression is mediated, in part, by response to a non-varying cAMP stimulus, involving regulatory genes such as *tgrC1*/*lagC*, *gbfA* and *srfA* (Fig. 5B). Only the *chdC* nulls show developmental arrest during aggregation and only *chdC* nulls show impaired expression of all these essential regulatory factors.

ChdC-mediated regulation of multicellular development and pattern formation

The observed chemotaxis (see Fig. 3) and morphology (see Fig. 1) defects of the *chdC* nulls fit precisely with those identified during our initial genetic screens for aberrant chemotaxis and multicellular development. The RNA-seq data from the 5-hour pulsed set indicate poor expression of genes (e.g. *tgrC1*/*lagC*, *gbfA*, *srfA*) that are required for multi-cell differentiation. Indeed, the strongest phenotype of *chdC*-null mutant cells arises as cells begin multicellular development, where wild-type cells in the aggregation mound differentiate into the precursors of the terminally spore and stalk cells. Prespore cells comprise ~80% of mound cells, with the remaining prestalk cells that give rise to the stalk, basal disc and supporting cup structures (Williams, 2006). We, thus, examined multi-cellular differentiation of the *chdC* nulls.

Endogenous prespore *pspA* and *cotB* and prestalk *ecmA* and *ecmB* mRNA (Williams, 2006) levels are diminished in *chdC* nulls compared with wild type (Fig. 6A). *In vivo* reporter expression using *pspA* promoter-*lacZ* fusions shows decreased reporter activity and restricted expression patterning compared with wild type (Fig. 6B). In addition, the expression of 53% of genes annotated as prespore-specific (including *pspA* and *cotB*) is halved in *chdC*-nulls as assayed by RNA-seq at the early mound stage, fully consistent with the defects described for prespore/spore differentiation (see Fig. 6).

Prestalk cell differentiation is more complex, marked by multiple sub-compartments of defined expression termed the pstA, pstO and pstB regions, which are primarily defined by subpromoter regions of prestalk genes *ecmA* and *ecmB* (Jermyn et al., 1987). In *chdC*-null mutants, *ecmA*/*lacZ* is expressed with a weaker and scattered pattern, without defined sorting to the tip of

Development (2013) doi:10.1242/dev.099879

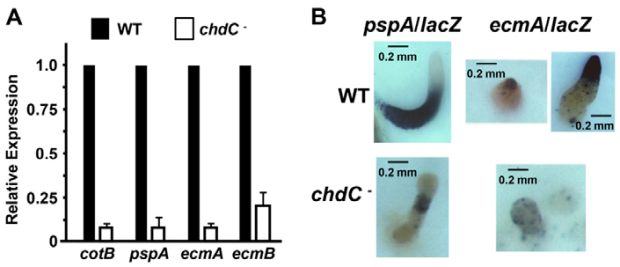


Fig. 6. Regulation of prespore/spore and prestalk/stalk differentiation by ChdC. (A) RNA was prepared from 12-hour developed cells, and relative expression levels of the prespore genes *pspA* and *cotB* and the prestalk genes *ecmA* and *ecmB* were quantified by RT-qPCR. (B) Wild-type and *chdC*-null cells were transformed with either the prespore-specific expression construct *pspA/lacZ* or the prestalk-specific expression construct *ecmA/lacZ*, plated for development on nitrocellulose filters and stained for β-galactosidase activity.

the mound (Fig. 6B). Global transcriptional profiling also gave a clear view of the observed broad prestalk defects, where the expression of 56% of genes annotated as prestalk specific is halved in *chdC* nulls.

Other GO terms that were enriched in the downregulated population of *chdC* nulls at the loose mound stage broadly segregate as regulators of cell division or metabolism. The former group includes cell cycle genes (*cdk1*, *cdc45*, *cycB*), DNA replication genes (*polA1*, *polA3*, *polD2*, *pcna*, *rcf3*, *rcf4*, *rcf5*) and mitotic genes (*nek2*, *kif2*, *kif12*). Although controversy surrounds the requirement of cell cycle progression during *Dictyostelium* development (Chen et al., 2004; Muramoto and Chubb, 2008), this gene set is significantly upregulated during wild-type development (Strasser et al., 2012).

Autonomous and non-autonomous defects in *chdC*-nulls

The RNA-seq data indicate that *chdC* nulls have defects in both cell-cell signaling (e.g. cAMP production) and cell-autonomous (e.g. the TgrC1/LagC-GbfA) regulatory pathways. These data suggest that although certain developmental aspects of *chdC*-null development might be rescued by exogenous wild-type signals, they would be insufficient to bypass defects in cell-autonomous functions.

To examine possible defects in both cell-autonomous and non-autonomous pathways in *chdC* nulls, we investigated cell pattern specification in a series of chimeric, mixed cell developments. *chdC* nulls were mixed at various cell ratios with the parental wild type. Remarkably, although 100% *chdC* nulls are unable to form terminal fruiting body structures, chimeric organisms with as few as 2.5% wild-type cells will progress to fruiting bodies, albeit with a reduced spore mass and a significantly larger basal disk (Fig. 7A).

In chimeras with unmarked 90% *chdC* nulls, the 10% ubiquitously marked *rZIP/lacZ* (Balint-Kurti et al., 1997) wild-type cells preferentially, but not exclusively, populate the anterior prestalk region (Fig. 7B). In a reverse experiment, where the 90% *chdC* nulls are marked with *ecmA/lacZ* and mixed with 10% unmarked wild-type cells, stained *chdC* nulls can localize to the prestalk/stalk regions (Fig. 7C). These data suggest that the failure of tip formation in *chdC* nulls may result from defects in signal propagation and not simply from a loss of an autonomous intracellular function.

To investigate non-autonomous parameters further, we examined terminal cell formation under monolayer induction conditions (Kay, 1987). Stalk cell differentiation is induced by the sequential exposure of cells to first cAMP and then to DIF-1. Under these conditions ~55% of wild-type cells were induced to form vacuolated stalk cells in the presence of DIF-1, a process that was highly suppressed by simultaneous exposure to cAMP (Fig. 7D). By contrast, only ~10-12% of *chdC* nulls could form vacuolated cells in the presence of exogenous DIF-1, regardless of the presence of cAMP (Fig. 7D).

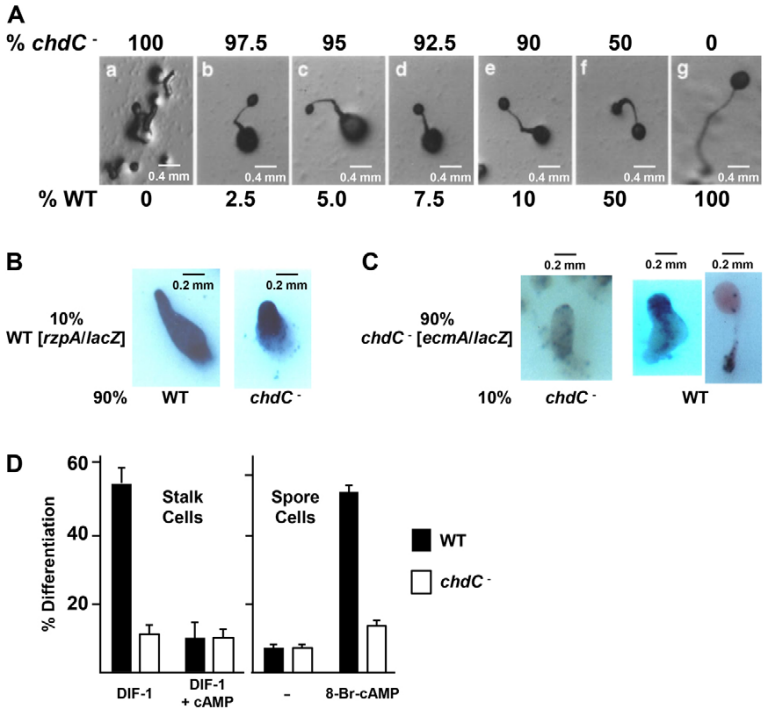


Fig. 7. Developmental defects of *chdC* nulls are partially rescued by chimeric development with wild-type cells. (A) The indicated percentages of wild-type and *chdC*$^{-}$ cells were mixed prior to plating for development. Images were obtained after 36 hours of development. (B) Wild-type cells were transformed with the ubiquitously expressed *rzpA/lacZ* reporter construct, mixed and developed as a 10% ratio with 90% unmarked wild-type cells or with 90% unmarked *chdC*$^{-}$ cells, and stained for β-galactosidase activity. (C) *chdC*$^{-}$ cells were transformed with the prestalk-specific reporter construct *ecmA/lacZ*, mixed and developed as a 90% ratio with 10% unmarked wild-type cells or with 10% unmarked *chdC*$^{-}$ cells, and stained for β-galactosidase activity. (D) Stalk cell assays: wild-type and *chdC*-null cells were plated at low cell density in monolayer with DIF-1 or with DIF-1 + cAMP. The percentage of cells that were fully vacuolated (mature stalk cells) was determined using phase-contrast microscopy. Spore cell assays: wild-type and *chdC*-null cells were plated at low cell density in monolayer without or with 8Br-cAMP to promote sporulation. Assays were scored by phase-contrast microscopy for the percentage of phase-bright fully mature spore cells.

In low cell density monolayers, spore formation can be directly induced with the membrane-permeant cAMP analog 8Br-cAMP, which directly activates cAMP receptors and the cAMP-dependent protein kinase (PKA) and bypasses other signaling requirements (Kay, 1987). Under these conditions ~60% of wild-type cells differentiated into spores, whereas only 15% of *chdC* null cells formed spores, efficiencies similar to that of uninduced wild-type cells (Fig. 7D), indicating an intrinsic loss in spore-forming ability of *chdC* nulls.

We also examined spore differentiation following chimeric development of 10% wild type with 90% *chdC* nulls. Spores were harvested and germinated in the absence and presence of blasticidin to differentiate and quantify wild-type and *chdC* nulls. Here, fewer than 10% of total spores from the mix were blasticidin-resistant *chdC* nulls. Although spore differentiation is not totally blocked in *chdC* nulls, it was substantially reduced to a frequency of ~10% that of WT, confirming a cell-autonomous defect for spore differentiation in *chdC* nulls. Although exogenous wild-type cells are able to influence developmental patterning in *chdC* nulls, they are unable to rescue terminal cell differentiation of *chdC* nulls.

DISCUSSION

Of the multiple SWI2/SNF2, ATPase-type chromatin remodeling proteins in eukaryotes, the CHD family is among the largest and most diverse (Ryan and Owen-Hughes, 2011). Humans have nine distinct CHDs that group into three structural subfamilies, CHD I, CHD II and CHD III. The CHD I subfamily is ubiquitous, found in the yeasts, plants and metazoa, whereas CHD II and CHD III members are generally restricted to multicellular organisms, with the latter seemingly absent in plants (Ryan and Owen-Hughes, 2011). It is, thus, striking that *Dictyostelium*, a facultative metazoan, encodes two proteins that cluster with the CHD III subfamily. Accordingly, *Dictyostelium* provides a unique model system for analyzing multiple and diverse regulatory aspects of different CHD family members; *Dictyostelium* does not seem to encode CHD II members, which are defined by the presence of PHD fingers (Ryan and Owen-Hughes, 2011).

Although CHD I members are suggested to facilitate an open chromatin organization at active promoters and to regulate nucleosome spacing more distally (Hennig et al., 2012; Persson and Ekwall, 2010; Zentner et al., 2013), their interactions at these promoter sites do not seem to be functionally essential at a global level. *S. cerevisiae* are viable without CHD1 (Tsukiyama et al., 1999) and *D. melanogaster* that lack CHD1 develop to maturity, although both males and females are sterile (McDaniel et al., 2008). Nonetheless, mouse ES cells depleted of CHD1 exhibit defects in pluripotency (Gaspar-Maia et al., 2009). Although *Chd1*$^{-/-}$ mice have not been reported, *Chd2*$^{-/-}$ mice are embryonically lethal (Marfella et al., 2006b). Therefore, CHD I family members can have crucial functions in the control of developmentally essential gene sets.

chdA-null *Dictyostelium* are viable, without apparent impact upon growth on bacteria or in axenic culture. However, *chdA* nulls exhibit a highly defined early developmental defect, involving poor transcriptional response to the extracellular activating signal cAMP. Furthermore, the morphological transitions and cell fate determinations required for terminal differentiation, which occur co-temporaneously with maximal levels of *ChdA* expression in wild-type cells, are effectively arrested in cells lacking ChdA. To some extent, the phenotypic effects of ChdA (i.e. a CHD1/2 ortholog) loss in *Dictyostelium* parallels the essential developmental requirement for CHD I members in mouse.

Consistent with the observed phenotypes, we also show that loss of *ChdA* in *Dictyostelium* causes a restricted, rather than global, impact on gene expression. Fewer than 15% of the *Dictyostelium* genome shows altered patterns of gene expression during growth or early development when parental and *chdA*-null cells are compared. We observe a similar distribution of genes that exhibit up- or downregulation. The data indicate that ChdA in *Dictyostelium*, and potentially CHD1 in other regulatory networks, is not essential for global chromatin organization and gene expression induction. ChdA is likely to have more confined operational targets and integrate with other chromatin organizing components.

Studies of the CHD III class in *Drosophila* (e.g. Kismet) or in humans (e.g. CHD7) suggest unique developmental functions compared with other CHD members. Kismet regulates cell fate determination of *Drosophila* and can antagonize the repressive effects of Polycomb in homeotic patterning (Srinivasan et al., 2008). Similarly, loss-of-function mutations or haploinsufficiencies of *CHD7* in mouse, humans and zebrafish (Layman et al., 2010; Patten et al., 2012) have complex developmental defects. In humans, the eye, heart, ear and genitalia are specifically affected, and many phenotypes are recapitulated in mouse and zebrafish. Although CHD7 is not essential for self renewal, pluripotency or somatic reprogramming of mouse ES cells, it has been suggested that a significant defect in human CHARGE syndrome may involve lineage effects in the proliferation, cell migration and specification of neural crest cells (Zentner et al., 2010).

As in the metazoa, *Dictyostelium* has multiple (ChdB and ChdC) CHD III subfamily members, with ChdC having the most essential and complex regulatory functions of all CHD members. At the phenotypic level, *chdC*-null cells are viable, but have reduced growth rates. They respond to nutrient depletion for induction of multi-cell development, but rapidly display abnormalities that temporally parallel the upregulation of ChdC in wild-type cells. Response to cAMP, chemotaxis and aggregate formation are impaired, as are cell-fate commitments.

Many of the defects in *chdC* nulls can be directly attributed to specific abnormalities in gene expression patterns compared with parentals. Although only a small population of genes are mis-regulated during growth of *chdC* nulls, they include a large number of cellular metabolic and energetic regulators. By contrast, the mis-regulated genes in *chdB* nulls do not simply cluster in a few functionally related groups, but are distributed broadly among GO classes, and, under the conditions studied, ChdB seems to be only minimally required.

During aggregation, *Dictyostelium* organize a complex inter-cell signaling pathway, which directs chemotaxis and multi-cell aggregation, but also cytodifferentiation and morphogenetics. Central to this pathway is the production and secretion (signal-relay) of the chemoattractant cAMP (Kortholt and van Haastert, 2008; McMains et al., 2008; Swaney et al., 2010). Receptor-mediated response to cAMP involves a highly integrated pathway that collectively promotes signal amplification, cytoskeletal reprogramming, cell polarization and non-terminal, cell-fate specification; many individual components of the cAMP signaling pathway are required for all cellular responses (Kortholt and van Haastert, 2008; McMains et al., 2008; Swaney et al., 2010). We show that ~50% of the genes, which lie most immediate to receptor sensing, are mis-regulated in *chdC* nulls. Their mis-regulations are predicted to compromise cell-autonomous requirements of intracellular signaling that directs chemotaxis and differentiation, but also the non-autonomous, extracellular elements involved in cAMP signal relay. Thus, although certain non-autonomous defects of *chdC*

nulls may be rescued by co-development with a small population of wild-type cells, the cell-autonomous pathways that are regulated by intracellular signaling remain compromised even in the presence of exogenous cAMP or wild-type factors.

Gene expression patterns have also been studied in *Chd7*$^{+/+}$and *Chd7*$^{-/-}$ mouse ES cells (Schnetz et al., 2010). Data indicate an impact on only a restricted gene set, with significant enrichment of genes primarily expressed in the ES population (Schnetz et al., 2010). In addition, more CHD7-binding sites map to enhancers of these genes than to non-ES class genes. Interestingly, these ES-enriched and CHD7-bound gene classes are generally upregulated in the absence of CHD7, suggesting a modulating repressive function on this ES gene set. Nonetheless, CHD7 is not entirely repressive, as other gene sets are downregulated upon the loss of CHD7. Regardless of the impact of CHD7 on ES cell transcriptional patterns as both a positive and negative effector, processes of self renewal, pluripotency and somatic cell reprogramming are not impaired in *CHD7*-deficient ES cells (Schnetz et al., 2010).

The directed and interactive roles of the different CHDs in multicellular differentiation are complex and still largely undefined. *Dictyostelium* provides a unique system to separately dissect these functions. We have shown that phenotypes resulting from deficiencies of ChdA, ChdB or ChdC in *Dictyostelium* are not global and, thus, parallel those in the more complex metazoa. It is particularly striking that proliferative, migration and cell specification defects of *chdC*-null *Dictyostelium* are mirrored in *Chd7*$^{-/-}$ neural crest cells. We further speculate that defective inter- and intra-cell signaling, as seen in *chdC*-null *Dictyostelium*, may also contribute to the developmental abnormalities seen in CHARGE syndrome.

Our data extend conclusions on the role for CHDs in gene expression. Each of the CHDs in *Dictyostelium* regulates different developmental patterns and different gene sets. The CHDs do not appear to act globally and can have both repressive and activating functions. We suggest that the mechanistic functions of CHD in *Dictyostelium* are similarly shared in the metazoa and that more-complete analyses will contribute significantly to the understanding for the ultimate treatment of genetic diseases, such as CHARGE syndrome, caused by loss-of-function mutations in *CHD* loci.

MATERIALS AND METHODS

Bioinformatics of CHD and SWI2/SNF2-related proteins

Identification of genes in *Dictyostelium* was performed by BLAST search of the *Dictyostelium* genome (http://dictybase.org/tools/blast). Three CHD proteins were identified: ChdA is DDB_G0284171, ChdB is DDB_G0280705 and ChdC is DDB_G0293012 (Basu et al., 2013). Additional domains were identified using EMBL SMART software (Letunic et al., 2012) and NCBI conserved domain software (Marchler-Bauer et al., 2011). The phylogenic tree was created using http://phylogeny.lirmm.fr/phylo_cgi/index.cgi (Dereeper et al., 2008) and full-length proteins. Similar clusterings were generated using only the chromodomains or helicase motifs.

Dictyostelium strains

Dictyostelium Ax2 (wild-type) cells were grown axenically in HL5 medium at 20°C. Gene deletion constructs were produced using standard PCR-based methods. Gene regions were disrupted by the floxed blasticidin resistance cassette from pLPBLP (Faix et al., 2004) using TOPO-Blunt II (Invitrogen). Gene deletion strains were generated by transformation with 15 µg of linearized plasmid using electroporation and selection in HL5 media supplemented with 10 µg/ml Blasticidin S. Transformants were screened by PCR and gene disruptions confirmed by immunoblotting.

Dictyostelium development and phenotype analyses

To analyze development, growing cells in log phase (1-3×10^6 cells/ml) were washed twice in KK2 buffer (15 mM KH_2PO_4, 3 mM K_2HPO_4) and developed on 0.45 µm nitrocellulose filters. Developmental expression of the individual CHD proteins were analyzed every 2 hours by immunoblotting. Slug migration assays were carried out in a 'slug can'. Washed cells were plated at one edge of a nitrocellulose filter, developed to the slug stage, and subjected to unidirectional light; images were taken after 48 hours. For chemotaxis, 5×10^7 cells/ml were shaken in KK2 buffer for 5 hours while being pulsed every 6 minutes with cAMP to a final concentration of 10^{-7} M. The cells were then placed in the Zigmond chamber (Z02; Neuro Probe, Gaithersburg, MD) in a cAMP gradient (source at 10^{-6} M cAMP, sink with no cAMP). Differential interference contrast (DIC) images of cells were captured at 6-second intervals for 15 minutes. Cell movement was analyzed using Dynamic Image Analysis System (DIAS) software (version 3.4.1) (Soll Technologies, Iowa City, IA; (Wessels et al., 2009). Stalk and spore cell differentiation in monoloayer assays were performed at a cell density of 1×10^4 cells/cm^2 (Kay, 1987) and scored by phase-contrast microscopy.

Antibodies and immunoblotting

Antibodies specific to ChdA, ChdB and ChdC were raised in rabbits by immunizing with peptides (~20 amino acids) specific for each protein (Openbiosystems, 70 day protocol). Immunoblotting was by standard methodologies, using whole-cell, SDS lysates, 3-8% Tris-acetate gels and blotting overnight at low voltage.

RNA-seq; sequencing and bioinformatics

RNA was isolated from all strains using TRIzol (Life Technologies). RNA quality was analyzed using the Agilent 2100 Bioanlyzer. 5 µg of total RNA was used for the Illumina libraries. RNA was polyA enriched and libraries were single-end, sequenced for 50 bases, using the Illumina HiSeq2000. Filtered reads from the manufacturers recommended pipeline (version 1.7) were aligned to the *Dictyostelium* genome with Tophat (version 1.3.0) (Trapnell et al., 2009) while masking the duplicated region of chromosome 2 (nucleotides 3,015,984 to 3,768,555) using the gene models from dictybase (http://dictybase.org/) (Basu et al., 2013). Reads were counted and normalized, and differential expression was calculated with HTSeq and DESeq (Anders and Huber, 2010). Heat maps were generated with the heatmap.2 function in the gplots package (Warnes, 2012) scatterplots were also produced in R (R Development Core Team, 2012). GO term enrichment used Orange (Curk et al., 2005). The RNA-seq data have been deposited in the GEO database (Accession Number, GSE47222).

Acknowledgements

We are indebted to Dr Harold Smith and the NIDDK genomics core.

Competing interests

The authors declare no competing financial interests.

Author contributions

J.L.P., B.J.R., K.C.R., A.J.H. and A.R.K. developed the concepts and approaches. J.L.P., B.J.R., K.C.R. and A.R.K. performed experiments. J.L.P., B.J.R., K.C.R., A.J.H. and A.R.K. analyzed the data. J.L.P., A.J.H. and A.R.K. prepared the manuscript prior to submission.

Funding

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), the National Institute of Diabetes and Digestive and Kidney Diseases, and a Wellcome Trust/NIH Programme Studentship to J.L.P. Deposited in PMC for immediate release.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.099879/-/DC1

References

Abrams, E., Neigeborn, L. and Carlson, M. (1986). Molecular analysis of SNF2 and SNF5, genes required for expression of glucose-repressible genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **6**, 3643-3651.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106.
Artemenko, Y., Swaney, K. F. and Devreotes, P. N. (2011). Assessment of development and chemotaxis in Dictyostelium discoideum mutants. *Methods Mol. Biol.* **769**, 287-309.
Balint-Kurti, P., Ginsburg, G., Rivero-Lezcano, O. and Kimmel, A. R. (1997). rZIP, a RING-leucine zipper protein that regulates cell fate determination during Dictyostelium development. *Development* **124**, 1203-1213.
Basu, S., Fey, P., Pandit, Y., Dodson, R. J., Kibbe, W. A. and Chisholm, R. L. (2013). DictyBase 2013: integrating multiple Dictyostelid species. *Nucleic Acids Res.* **41 Database issue**, D676-D683.
Blus, B. J., Wiggins, K. and Khorasanizadeh, S. (2011). Epigenetic virtues of chromodomains. *Crit. Rev. Biochem. Mol. Biol.* **46**, 507-526.
Bosman, E. A., Penn, A. C., Ambrose, J. C., Kettleborough, R., Stemple, D. L. and Steel, K. P. (2005). Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum. Mol. Genet.* **14**, 3463-3476.
Bouazoune, K. and Kingston, R. E. (2012). Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc. Natl. Acad. Sci. USA* **149**, 19238-19243.
Breeden, L. and Nasmyth, K. (1987). Cell cycle control of the yeast HO gene: cis- and trans-acting regulators. *Cell* **48**, 389-397.
Brock, D. A. and Gomer, R. H. (1999). A cell-counting factor regulating structure size in Dictyostelium. *Genes Dev.* **13**, 1960-1969.
Brock, D. A., Hatton, R. D., Giurgiutiu, D. V., Scott, B., Jang, W., Ammann, R. and Gomer, R. H. (2003). CF45-1, a secreted protein which participates in Dictyostelium group size regulation. *Eukaryot Cell* **2**, 788-797.
Brock, D. A., van Egmond, W. N., Shamoo, Y., Hatton, R. D. and Gomer, R. H. (2006). A 60-kilodalton protein component of the counting factor complex regulates group size in Dictyostelium discoideum. *Eukaryot Cell* **5**, 1532-1538.
Burdine, V. and Clarke, M. (1995). Genetic and physiologic modulation of the prestarvation response in Dictyostelium discoideum. *Mol. Biol. Cell* **6**, 311-325.
Chen, G. K., Shaulsky, G. and Kuspa, A. (2004). Tissue-specific G1-phase cell-cycle arrest prior to terminal differentiation in Dictyostelium. *Development* **131**, 2619-2630.
Clapier, C. R. and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273-304.
Curk, T., Demsar, J., Xu, Q., Leban, G., Petrovic, U., Bratko, I., Shaulsky, G. and Zupan, B. (2005). Microarray data mining with visual programming. *Bioinformatics* **21**, 396-398.
Delmas, V., Stokes, D. G. and Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc. Natl. Acad. Sci. USA* **90**, 2414-2418.
Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M. et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36 Web Server issue**, W465-W469.
Eisen, J. A., Sweder, K. S. and Hanawalt, P. C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* **23**, 2715-2723.
Euskirchen, G., Auerbach, R. K. and Snyder, M. (2012). SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *J. Biol. Chem.* **287**, 30897-30905.
Faix, J., Kreppel, L., Shaulsky, G., Schleicher, M. and Kimmel, A. R. (2004). A rapid and efficient method to generate multiple gene disruptions in Dictyostelium discoideum using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res.* **32**, e143.
Flanagan, J. F., Blus, B. J., Kim, D., Clines, K. L., Rastinejad, F. and Khorasanizadeh, S. (2007). Molecular implications of evolutionary differences in CHD double chromodomains. *J. Mol. Biol.* **369**, 334-342.
Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., Ramalho-Santos, J., McManus, M. T., Plath, K., Meshorer, E. et al. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863-868.
Guo, K., Anjard, C., Harwood, A., Kim, H. J., Newell, P. C. and Gross, J. D. (1999). A myb-related protein required for culmination in Dictyostelium. *Development* **126**, 2813-2822.
Hargreaves, D. C. and Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21**, 396-420.
Harwood, A. J., Hopper, N. A., Simon, M. N., Driscoll, D. M., Veron, M. and Williams, J. G. (1992). Culmination in Dictyostelium is regulated by the cAMP-dependent protein kinase. *Cell* **69**, 615-624.
Hennig, B. P., Bendrin, K., Zhou, Y. and Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Rep.* **13**, 997-1003.
Ho, L. and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature* **463**, 474-484.
Hopfner, K. P., Gerhold, C. B., Lakomek, K. and Wollmann, P. (2012). Swi2/Snf2 remodelers: hybrid views on hybrid molecular machines. *Curr. Opin. Struct. Biol.* **22**, 225-233.
Hurd, E. A., Capers, P. L., Blauwkamp, M. N., Adams, M. E., Raphael, Y., Poucher, H. K. and Martin, D. M. (2007). Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mamm. Genome* **18**, 94-104.
Iranfar, N., Fuller, D. and Loomis, W. F. (2003). Genome-wide expression analyses of gene regulation during early development of Dictyostelium discoideum. *Eukaryot. Cell* **2**, 664-670.
Janssen, N., Bergman, J. E., Swertz, M. A., Tranebjaerg, L., Lodahl, M., Schoots, J., Hofstra, R. M., van Ravenswaaij-Arts, C. M. and Hoefsloot, L. H. (2012). Mutation update on the CHD7 gene involved in CHARGE syndrome. *Hum. Mutat.* **33**, 1149-1160.
Jermyn, K. A., Berks, M., Kay, R. R. and Williams, J. G. (1987). Two distinct classes of prestalk-enriched mRNA sequences in Dictyostelium discoideum. *Development* **100**, 745-755.
Kay, R. R. (1987). Cell differentiation in monolayers and the investigation of slime mold morphogens. *Methods Cell Biol.* **28**, 433-448.
Kessin, R. H. (2001). *Dictyostelium – Evolution, Cell Biology and the Development of Multicellularity*. Cambridge, UK: Cambridge University Press.
Kortholt, A. and van Haastert, P. J. (2008). Highlighting the role of Ras and Rap during *Dictyostelium* chemotaxis. *Cell. Signal.* **20**, 1415-1422.
Kuspa, A. (2006). Restriction enzyme-mediated integration (REMI) mutagenesis. *Methods Mol. Biol.* **346**, 201-209.
Lalani, S. R., Safiullah, A. M., Fernbach, S. D., Harutyunyan, K. G., Thaller, C., Peterson, L. E., McPherson, J. D., Gibbs, R. A., White, L. D., Hefner, M. et al. (2006). Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am. J. Hum. Genet.* **78**, 303-314.
Layman, W. S., Hurd, E. A. and Martin, D. M. (2010). Chromodomain proteins in development: lessons from CHARGE syndrome. *Clin. Genet.* **78**, 11-20.
Lemos, T. Ã. A., Passos, D. O., Nery, F. C. and Kobarg, J. (2003). Characterization of a new family of proteins that interact with the C-terminal region of the chromatin-remodeling factor CHD-3. *FEBS Lett.* **533**, 14-20.
Letunic, I., Doerks, T. and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40 Database issue**, D302-D305.
Loomis, W. F. and Shaulsky, G. (2011). Developmental changes in transcriptional profiles. *Dev. Growth Differ.* **53**, 567-575.
Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R. et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39 Database issue**, D225-D229.
Marfella, C. G., Ohkawa, Y., Coles, A. H., Garlick, D. S., Jones, S. N. and Imbalzano, A. N. (2006a). Mutation of the SNF2 family member Chd2 affects mouse development and survival. *J. Cell. Physiol.* **209**, 162-171.
McDaniel, I. E., Lee, J. M., Berger, M. S., Hanagami, C. K. and Armstrong, J. A. (2008). Investigations of CHD1 function in transcription and development of Drosophila melanogaster. *Genetics* **178**, 583-587.
McMains, V. C., Liao, X. H. and Kimmel, A. R. (2008). Oscillatory signaling and network responses during the development of Dictyostelium discoideum. *Ageing Res. Rev.* **7**, 234-248.
Muramoto, T. and Chubb, J. R. (2008). Live imaging of the Dictyostelium cell cycle reveals widespread S phase during development, a G2 bias in spore differentiation and a premitotic checkpoint. *Development* **135**, 1647-1657.
Nishiyama, M., Skoultchi, A. I. and Nakayama, K. I. (2012). Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-β-catenin signaling pathway. *Mol. Cell. Biol.* **32**, 501-512.
Pagon, R. A., Graham, J. M., Jr, Zonana, J. and Yong, S. L. (1981). Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *J. Pediatr.* **99**, 223-227.
Parikh, A., Miranda, E. R., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., Curk, T., Sucgang, R., Chen, R., Zupan, B. et al. (2010). Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol.* **11**, R35.
Patten, S. A., Jacobs-McDaniels, N. L., Zaouter, C., Drapeau, P., Albertson, R. C. and Moldovan, F. (2012). Role of Chd7 in zebrafish: a model for CHARGE syndrome. *PLoS ONE* **7**, e31650.
Persson, J. and Ekwall, K. (2010). Chd1 remodelers maintain open chromatin and regulate the epigenetics of differentiation. *Exp. Cell Res.* **316**, 1316-1323.
Pleasance, E. D., Stephens, P. J., O'Meara, S., McBride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C. et al. (2010). A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190.
R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Ryan, D. P. and Owen-Hughes, T. (2011). Snf2-family proteins: chromatin remodellers for any occasion. *Curr. Opin. Chem. Biol.* **15**, 649-656.
Ryan, D. P., Sundaramoorthy, R., Martin, D., Singh, V. and Owen-Hughes, T. (2011). The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *EMBO J.* **30**, 2596-2609.
Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., Zhang, X., Song, L., Wang, Z., Laframboise, T. et al. (2009). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res.* **19**, 590-601.
Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., Adams, D. J., Flicek, P., Crawford, G. E., Laframboise, T. et al. (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* **6**, e1001023.
Srinivasan, S., Dorighi, K. M. and Tamkun, J. W. (2008). Drosophila Kismet regulates histone H3 lysine 27 methylation and early elongation by RNA polymerase II. *PLoS Genet.* **4**, e1000217.
Strasser, K., Bloomfield, G., MacWilliams, A., Ceccarelli, A., MacWilliams, H. and Tsang, A. (2012). A retinoblastoma orthologue is a major regulator of S-phase, mitotic, and developmental gene expression in Dictyostelium. *PLoS ONE* **7**, e39914.

Swaney, K. F., Huang, C. H. and Devreotes, P. N. (2010). Eukaryotic chemotaxis: a network of signaling pathways controls motility, directional sensing, and polarity. *Annu Rev Biophys* **39**, 265-289.
Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C. and Kennison, J. A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561-572.
Thompson, B. A., Tremblay, V., Lin, G. and Bochar, D. A. (2008). CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol. Cell. Biol.* **28**, 3894-3904.
Trapnell, C., Pachter, L. and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111.
Tsukiyama, T., Palmer, J., Landel, C. C., Shiloach, J. and Wu, C. (1999). Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. *Genes Dev.* **13**, 686-697.
Van Driessche, N., Shaw, C., Katoh, M., Morio, T., Sucgang, R., Ibarra, M., Kuwayama, H., Saito, T., Urushihara, H., Maeda, M. et al. (2002). A transcriptional profile of multicellular development in Dictyostelium discoideum. *Development* **129**, 1543-1552.
Van Haastert, P. J. and Veltman, D. M. (2007). Chemotaxis: navigating by multiple signaling pathways. *Sci. STKE* **2007**, pe40.
Warnes, G. R. (2012). *gplots: Various R Programming Tools for Plotting Data*.
Wessels, D. J., Kuhl, S. and Soll, D. R. (2009). Light microscopy to image and quantify cell movement. *Methods Mol. Biol.* **571**, 455-471.
Williams, J. G. (2006). Transcriptional regulation of Dictyostelium pattern formation. *EMBO Rep.* **7**, 694-698.
Williams, C. J., Naito, T., Arco, P. G., Seavitt, J. R., Cashman, S. M., De Souza, B., Qi, X., Keables, P., Von Andrian, U. H. and Georgopoulos, K. (2004). The chromatin remodeler Mi-2β is required for CD4 expression and T cell development. *Immunity* **20**, 719-733.
Yap, K. L. and Zhou, M.-M. (2011). Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* **50**, 1966-1980.
Zentner, G. E., Layman, W. S., Martin, D. M. and Scacheri, P. C. (2010). Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *Am. J. Med. Genet. A.* **152A**, 674-686.
Zentner, G. E., Tsukiyama, T. and Henikoff, S. (2013). ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet.* **9**, e1003317.

Chapter 9

Analysis of Chromatin Organization by Deep Sequencing Technologies

James L. Platt, Nick A. Kent, Adrian J. Harwood, and Alan R. Kimmel

Abstract

Micrococcal nuclease (MNase) is an endonuclease that cleaves native DNA at high frequency, but is blocked in chromatin by sites of intimate DNA–protein interaction, including nucleosomal regions. Protection from MNase cleavage has often been used to map transcription factor binding sites and nucleosomal positions on a single-gene basis; however, by combining MNase digestion with high-throughput, paired-end DNA sequencing, it is now possible to simultaneously map DNA-protein interaction regions across the entire genome. Biochemical and bioinformatic protocols are detailed for global mono-nucleosome positioning at ~160 bp spacing coverage, but are applicable to mapping more broadly or for site-specific binding of transcription factors at ~50 bp resolution.

Key words Nucleosome mapping, MNase, High-throughput, Genome-wide, *Dictyostelium*

1 Introduction

The positioning of nucleosomes along the genomic DNA backbone can significantly impact the accessibility of factors that regulate transcription, replication, repair, and recombination. Although micrococcal nuclease (MNase) will digest naked DNA along its entire axis, MNase action is largely protected in chromatin regions that are intimately associated with proteins. MNase has, thus, proven a particularly effective probe for the precise mapping of nucleosome positions on chromosomes, in both gene-specific and, more recently, genome-wide global assays (1). Historically, studies on chromatin organization in *Dictyostelium* were particularly limited (2, 3). However, with the availability of a fully annotated genome sequence (4), deep mass spectrometric analyses of the histone complement and modifications (5), and a complete developmental transcription profile (6), we have initiated studies to globally map nucleosomal positions throughout the genomes of wild-type and mutant cells. Methods described are optimized from those first

Ludwig Eichinger and Francisco Rivero (eds.), *Dictyostelium discoideum Protocols*, Methods in Molecular Biology 983,
DOI 10.1007/978-1-62703-302-2_9,

applied in *Saccharomyces cerevisiae* (7, 8); chromatin is predigested by MNase in situ and the released nucleosome-protected DNA is then sequenced by a high-throughput, paired-end approach, which allows direct alignment mapping of nucleosome positions on a genome-wide scale. We present biochemical methods for generating and analyzing high-throughput, nucleosomal sequence reads at a global level. We also discuss bioinformatic approaches that are required to analyze large data sets and to comparatively assign nucleosomal positions at single-gene and genome-wide bases.

2 Materials

All solutions are prepared in deionized water unless indicated otherwise.

2.1 Cell Culture

Dictyostelium cells are grown axenically at 22°C in shaking culture in HL-5 medium including glucose (Formedium #HLG0102) supplemented with 5 μg/mL vitamin B_{12} and 200 μg/mL folic acid.

2.2 Chromatin Digestion

1. Wash buffer: 100 mM sorbitol.
2. Digestion buffer: 100 mM sorbitol, 50 mM NaCl, 10 mM Tris–HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM 2-mercaptoethanol, 0.5 mM spermidine, 0.1% Nonidet P-40 (NP-40). Store at room temperature; make fresh every month.
3. Micrococcal nuclease (Affymetrix, USB #70196Y): 15 U/μL in 10 mM Tris–HCl, pH 7.5, 10 mM NaCl, 100 μg/mL bovine serum albumin (BSA). Store at −20°C.
4. Stop solution: 5% SDS, 250 mM EDTA, pH 8.4. Make fresh and incubate at 37°C.

2.3 DNA Cleanup

1. Phenol solution: Buffer-saturated phenol in 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.
2. Chloroform.
3. RNaseA: 10 μg/μL DNase-free ribonuclease A (Sigma). Store at −20°C (see Note 1).
4. 3.5 M sodium acetate, pH 5.2.
5. 100% ethanol.
6. 70% ethanol.
7. TE buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA.

2.4 DNA Isolation

1. DNA loading dye (Fermentas).
2. TAE buffer: 10 mM Tris–HCl, pH 8.0, 20 mM acetic acid, 1 mM EDTA.
3. 1.5% agarose in TAE buffer.

4. Standard agarose gel electrophoresis equipment.
5. Costar Spin-X 0.45 μm cellulose acetate or polysulfone (Corning Inc.).
6. NanoDrop spectrophotometer (Thermo Scientific).

2.5 Sequencing

1. Library preparation kit: Illumina paired-end sample prep kit (#PE-102-1001).
2. Agilent 2100 Bioanalyzer (Agilent Technologies).
3. Agilent High Sensitivity DNA chip (Agilent Technologies #5067-4626).
4. Sequencer: Illumina HiSeq2000.
5. Software: Bowtie (9), Integrated Genome Browser (10), Cluster 3 (11), and TreeView (12).

3 Methods

3.1 Optimized and Preparative Chromatin Digestions

Prior to preparing sequencing-scale quantities of nucleosomal DNA (see Subheadings 3.4 and 3.5), the conditions for MNase cleavage of chromatin must first be optimized to obtain reproducible digests. Conditions can vary with the organism or even cell type. Variables to consider include NP-40 concentration in the digestion buffer, activity units of MNase, digestion temperature, and digestion time (see Note 2 and Subheading 3.2, steps 12–13).

1. A total of 5×10^8 cells are required, with 1×10^8 cells used for each MNase treatment (see steps 2 and 3). Cells are routinely grown axenically to a density of $1–2 \times 10^6$ cells/mL. However, cells can be taken at other densities or after growth on bacteria. Alternatively, cells may be taken at any stage of development (see Note 3).
2. Set up five 1.5-mL spin tubes with increasing amounts of MNase, ranging from 0 to 300 U/tube (see Note 2).
3. Split cells into five separate tubes (~10^8 cells/tube) and wash each, twice with 25 mL of wash buffer at 4°C, 4,000 × *g*.
4. Gently resuspend the pellet of one tube (~10^8 cells) in 400 μL of digest buffer and immediately, as cells begin to lyse, transfer to one of the MNase tubes (see step 2) and mix gently; incubate at 37°C.
5. Repeat step 4 for each of the remaining four tubes.
6. After 2 min, add 40 μL of stop solution and shake vigorously for 10 s to lyse the cells and halt the digestion reaction (see Note 4); stop times for each of the 5 tubes are staggered according to their start times.
7. Samples can be stored at −20°C or directly processed for DNA purification (see Subheading 3.2).

3.2 DNA Purification and Analyses

1. Add 200 μL of phenol and 200 μL of chloroform to each chromatin digestion tube (from Subheading 3.1, step 7), vortex for 10 s. If samples were previously frozen, thaw first at 4°C.
2. Microcentrifuge at 14,500 × *g* for 5 min at room temperature.
3. Transfer the aqueous phase containing nucleic acids to a fresh tube with 14 μL of RNase A and mix.
4. Incubate samples at 37°C for 30 min; remix each tube after 15 min.
5. Repeat steps 1 and 2 and transfer the aqueous phase to a fresh tube.
6. Add 1/10 volume of sodium acetate and 2 volumes of 100% ethanol (see Note 5) to the tube with the aqueous phase.
7. Incubate at −20°C for at least 15 min.
8. Microcentrifuge at 14,500 × *g*, 4°C for 15 min to pellet DNA.
9. Remove and discard the supernatant. Wash the DNA pellet with 70% ethanol and microcentrifuge at 14,500 × *g*, 4°C for 5 min to pellet DNA.
10. Remove and discard the supernatant and air-dry the pellet.
11. Resuspend the DNA pellet in 100 μL of TE buffer. DNA recovery is estimated using a NanoDrop spectrophotometer.
12. Take 10 μL of each sample and analyze DNA size by electrophoresis in a 1.5% agarose gel in comparison to DNA markers with a size range of 50–10,000 bp. It is suggested to use a tracking dye that migrates at <50 bp (see Note 6). Limited MNase digestion will create a range of partially digested, protected DNA sizes (see Fig. 1a).
13. Once chromatin digestion conditions have been optimized (see step 12) to generate maximum DNA amounts as a ~50–1,000 bp array (see Fig. 1a), the digestions in Subheading 3.1 should be repeated to produce 3 identical preparations that will be pooled in Subheading 3.4, step 1. Analyze a small aliquot of this preparative digest to verify cleavage reproducibility, before fragment size isolation (see Subheading 3.4). Alternatively, narrower size ranges, covering only mono-nucleosomes, sub-nucleosomes, di-nucleosomes, etc., can also be selected (see Subheading 3.6, step 3).

3.3 Naked DNA Control Digestion

We suggest the digestion of purified, "naked" DNA, devoid of protein, to control for sequence-specific cleavage by MNase, but also for sequencing and analytic biases (see Subheadings 3.5 and 3.6). As described for chromatin in Subheading 3.2, steps 12–13, MNase cleavage conditions for naked DNA must be first optimized and verified.

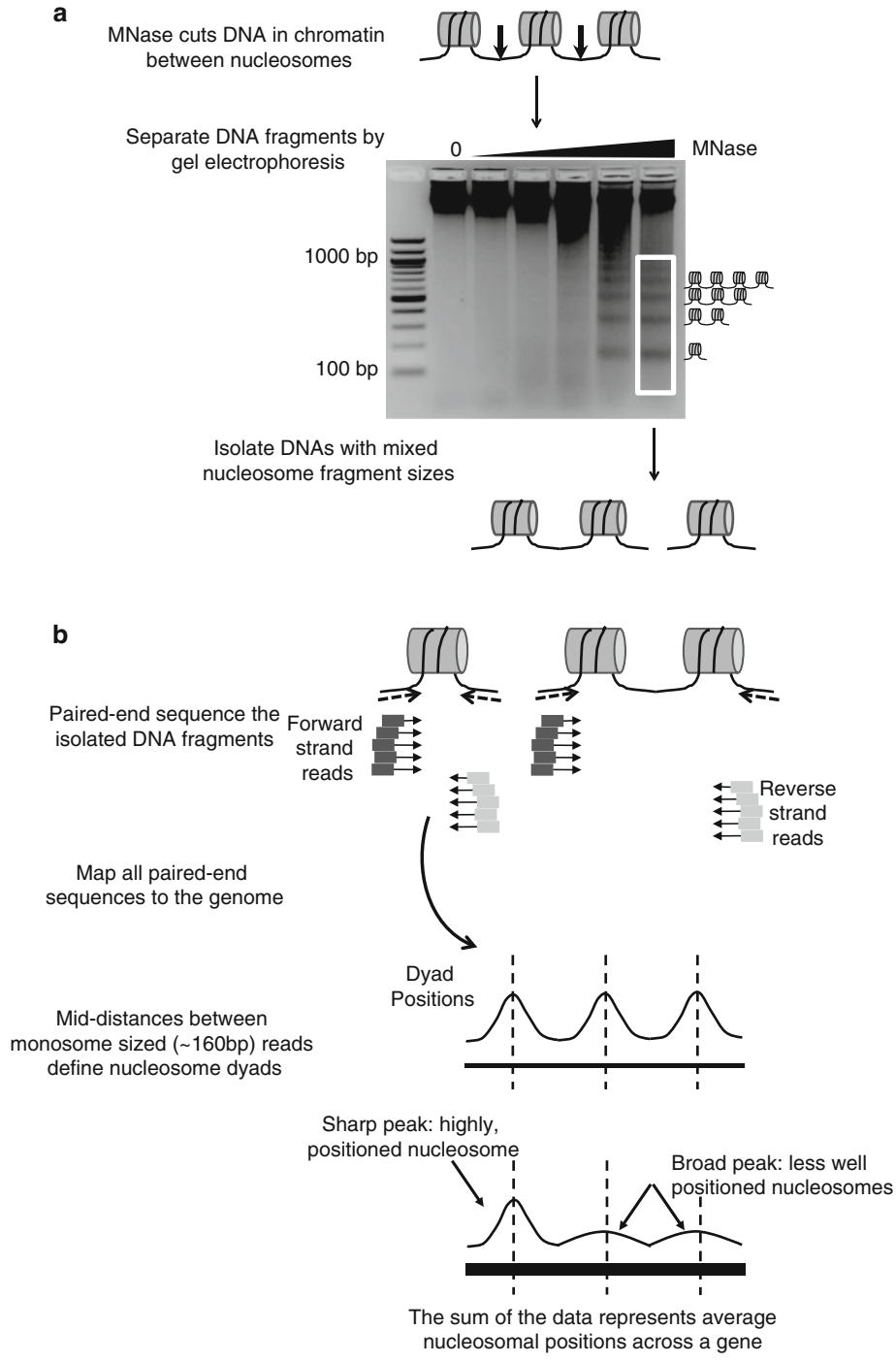


Fig. 1 Analysis of chromatin organization by deep sequencing. (**a**) Chromatin is digested in situ with varying amounts of micrococcal nuclease (MNase). The DNA is extracted and separated on an agarose gel. Digestion by MNase produces a characteristic ladder, with ~160 bp nucleosome repeat length. The DNA (~50–1,000 bp) is size selected by excision from the gel. (**b**) The excised DNA is purified and paired-end sequenced. The resulting sequences are mapped to the genome. Distances between mapped reads correspond generally to nucleosome repeat-length spacings. The midpoint between two ~160 bp, paired reads defines the dyad of a nucleosome; the sum of all read peaks assigns an average position of every nucleosome. Nucleosomes with highly similar positioning in most cells will exhibit a very sharp coverage peak, whereas more flexibly positioned nucleosomes result in broader peaks

1. Set up five 1.5-mL spin tubes with a concentration series of MNase, at 0–5 U/tube (see Note 7).
2. Add 10 μg of purified, undigested DNA (see Note 8) in 400 μL of digestion buffer to each tube, mix gently, and incubate at room temperature for 30 s.
3. Add 40 μL of stop solution and shake vigorously for 10 s to stop the digestion reaction.
4. Samples can be stored at −20°C or directly processed for DNA purification and size analyses (see Subheading 3.2, steps 12–13).

3.4 DNA Fragment Isolation

1. Pool separately the three chromatin and the three naked DNA digestions (see Subheading 3.2, steps 12–13 and Subheading 3.3, step 4) that give maximum DNA amounts in a ~50–1,000 bp (or narrower size) array (see Fig. 1a).
2. Separate the pooled DNA digests by electrophoreses using 4 lanes in a 1.5-mm 1.5% agarose gel. DNA in the chosen size range is isolated by gel excision.
3. Add the excised agarose gel slice to a Costar column and freeze at −20°C for 10 min; thaw at 4°C for 5 min and repeat freeze and thaw one time.
4. Microcentrifuge, with the hinge of the tube facing outwards, at 14,500 × *g*, 4°C for 5 min to release liquid and DNA from the gel.
5. Remove the released liquid and reserve the DNA.
6. Microcentrifuge again, but now with the hinge of the tube facing inwards, at 14,500 × *g*, 4°C for 5 min to release the remaining gel liquid.
7. Pool the released DNA with the previous tube (see step 5).
8. Phenol:chloroform extract the DNA and precipitate as previously described (see Subheading 3.2, steps 1–11), using a 2:1:1 ratio of sample:phenol:chloroform. Estimate DNA recovery using a NanoDrop spectrophotometer.

3.5 Library Generation and Sequencing

Illumina paired-end DNA libraries require ~5 μg of genomic DNA, regardless if prepared from chromatin or naked DNA controls. Library preparations primarily follow manufacturer instructions (see Note 9). DNA ends are first repaired using a mixture of T4 polynucleotide kinase + Klenow fragment of DNA polymerase I + T4 DNA polymerase, which is included in the Illumina paired-end sample prep kit. Adaptors are ligated to the DNA ends and DNA >100 bp is purified by gel electrophoresis (see Subheading 3.4, steps 2–8) to remove the self-ligated adaptors, ~80 bp.

1. The recovered DNA is amplified with primers complementary to the adaptor sequences, to enrich for DNA fragments with

adaptors on both ends. While the Illumina instructions suggest an additional gel purification step, we have found that this is not necessary (see Note 10).

2. Libraries are quality checked on the Agilent Bioanalyzer 2100 using a high-sensitivity DNA chip to confirm that the libraries are not largely contaminated with adaptor dimers or primer pairs (see Note 10).
3. Libraries are sequenced at 76 bp, paired-end on the Illumina HiSeq 2000 at a relatively low density of ~350,000 clusters/mm^2. Recent protocols expect 100–200 million reads per lane (see Note 11). Paired-end reads allow both forward and reverse template sequencing of each DNA cluster (see Fig. 1b). Alignment of the paired sequence reads to the genome contains positional information that, when queried by varying the distances across the genome map, can reveal nucleosome repeat lengths or transcription factor spacings (see Subheading 3.6). Sequencing data for naked DNA is processed similarly to that of chromatin-derived DNA (see Subheading 3.6, step 6).

3.6 Data Analyses

1. Illumina reads (see Subheading 3.5, step 3) can be aligned to the genome using Bowtie (9) or another aligner designed to handle large numbers of short reads (see Note 12). Bowtie accepts Illumina reads in a fastq format, as well as the assembled *Dictyostelium* genome [see http://www.dictybase.org/Downloads/, 4).
2. Bowtie will output the genome coordinates of each aligned paired-end read and indicate distances between them, in a resulting SAM (Sequence Alignment Map) file (13).
3. If heterogeneously sized chromatin DNA libraries were used (see Subheading 3.2, steps 12–13 and Subheading 3.4, steps 1–2), the distances between each read end will vary (see Fig. 1b). However, the Bowtie SAM file can be queried to extract read alignments into files of specific sized spacings, e.g., mono-nucleosomes, di-nucleosomes, etc., which can be further processed separately. If the excised DNA lengths at the outset were more restricted (see Subheading 3.2, steps 12–13 and Subheading 3.4, steps 1–2), size selection within the SAM file is unnecessary. An alternative approach using long-read (e.g., 454-type) sequencing can essentially cover DNA lengths from purified mono-nucleosomes (14).
4. Each read endpoint defines an accessible site within the chromatin for MNase cleavage, and the mid-distance between two paired reads corresponds to the axis of symmetry that separates the two MNase-sensitive sites. For mono-nucleosome-sized selected spacings (~160 bp), the midpoint also represents the dyad for an individual nucleosome (see Fig. 1b). By processing all reads, one can determine the frequency distribution for

Fig. 2 Nucleosome positioning viewed in Integrated Genome Browser. A subregion of chromosome 4 is displayed, with two divergently expressed genes. Highly positioned (i.e., *sharp peaks*) nucleosomes are seen for DDB0191460, whereas nucleosomal positioning in DDB0233398 is less defined. Introns and noncoding regions tend to be underrepresented in nucleosomal reads, for both biological and technical reasons (e.g., see Note 12)

nucleosome (or other) positions across the entire mapped genome. This derived frequency distribution file [.sgr or .bed 8) can be loaded into a genome browser, such as Integrated Genome Browser (10), for visualization (see Fig. 2 and Notes 12 and 13).

5. Genome-wide comparisons often yield globally applicable information that is missed when data are analyzed at the single-gene level. Indeed, all (or defined subset) of the *Dictyostelium* genes can be first aligned by a defined, universal gene feature (e.g., start codon), with further interrogation for common nucleosome positioning within the entire gene set (see Fig. 3). Alternative ways to further examine the aligned data include clustering based upon shared characteristics of a nucleosomal repeat pattern, display as heatmaps, and comparative analyses of wild-type and mutant strains or during a developmental time course. Clustering can be achieved with Cluster 3 (11) and visualization with TreeView (12), or similar software packages.
6. Peaks can be normalized or compared to the naked DNA controls to ensure peaks represent actual nucleosome presence and not biases introduced by MNase digestion.

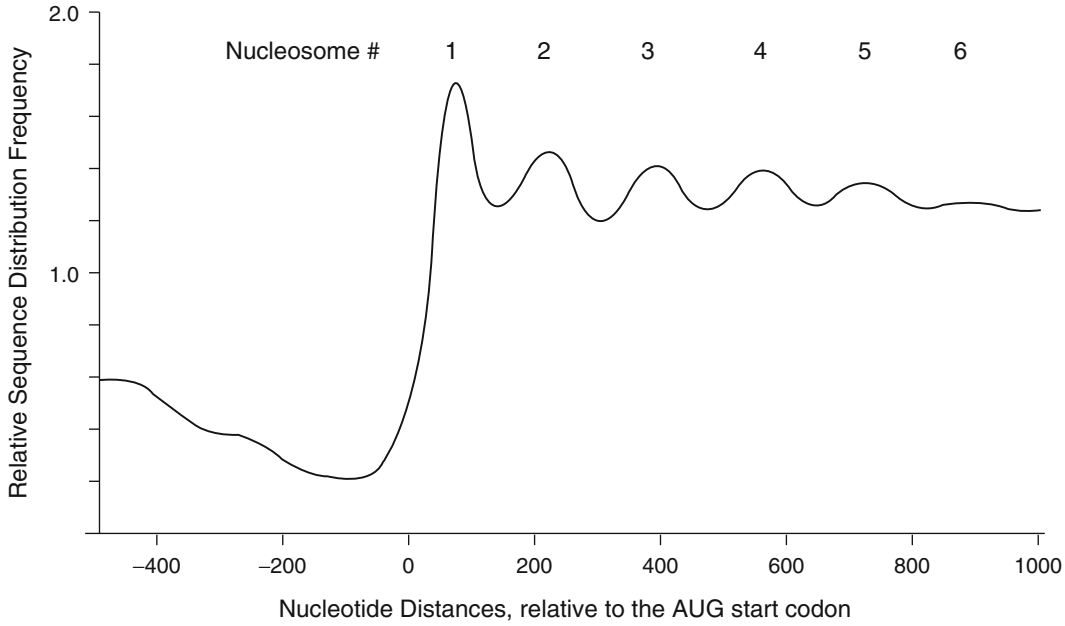


Fig. 3 Genome-wide alignment of all nucleosomal reads relative to translational start sites. The ~12,500 protein-coding genes were aligned at their translational start sites. Nucleosome coverage for each gene was then placed relative to their AUG start. Data represent the sum of the genome-wide aligned data. *The numbered peaks* represent globally shared nucleosome positions with ~160 bp spacings; 5′-promoter regions are relatively underrepresented

4 Notes

1. If the ribonuclease A is not DNase free, the solution should be boiled for 5 min prior to freezing to inactivate the contaminating DNase.
2. NP-40 is used to permeabilize the plasma membrane to allow MNase to enter the cell. Increasing the NP40 concentration will increase MNase exposure, but also decrease the integrity of the cell. Ideally, digestion should occur quickly to avoid nucleosomal movement. MNase works optimally at 37°C, but we have observed efficient digestions at lower temperatures (22–30°C), but with more units of MNase. It should also be noted that MNase activity can vary from batch to batch. *Dictyostelium* at different developmental time points require different amounts of MNase to create equivalent digests. This may be due to differences in sensitivity to NP-40 for cell permeabilization.
3. For stages other than growth, cell numbers should be increased two- to threefold.
4. At the end of the digestion, prior to adding the stop solution, the cell suspension can be spun at 14,500 × *g* for 5 sec, and the

supernatant quickly transferred to a fresh tube containing 40 µL of stop solution. This centrifugation step serves to pellet cell debris and high molecular weight chromatin fragments (DNA species >1 kb) whilst releasing lower molecular weight chromatin particles of mono- to hexa-nucleosome sizes into the supernatant.

5. Other salts may be used to precipitate the digested DNA. However, ammonium acetate should be avoided as it inhibits T4 polynucleotide kinase, which is required during library preparation.
6. Using a loading buffer with a rapidly migrating dye, such as orange G, is suggested to avoid interfering with visualization of DNA in the gel.
7. MNase-digested naked DNA serves as a control for sequence digestion biases. Pure DNA is digested very quickly by MNase and so requires a lower enzyme concentration and a shorter digestion time than chromatin.
8. Purified control DNA is obtained by resuspending ~10^8 cells in 400 µL of TE buffer and following Subheading 3.2, steps 1–11.
9. The exact protocol is included with the paired-end kit, but is routinely updated and optimized; the most recent editions may be downloaded at http://www.illumina.com/support/sequencing/sequencing_kits/pe_dna_sample_prep_kit.ilmn, and therein.
10. Contamination with adaptor dimers will reduce the overall yield of sequencing reads. If library preparation has succeeded, but is still contaminated with adaptor dimers, another gel purification may be required.
11. Other high-throughput "paired-end" sequencing platforms should also be effective, although we have not assessed them.
12. Reads with highly redundant (e.g., extreme A+T bias) or extensively repeated [e.g., (AAC)n] sequences may not align uniquely to the genome.
13. Perl scripts to handle data manipulation are provided as supplemental data by Kent and colleagues (8).

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health, the National Institute of Diabetes and Digestive and Kidney Diseases, and a Wellcome Trust/NIH Programme Studentship to J.L.P. There are no conflicts or competing interests.

References

1. Clark DJ (2010) Nucleosome positioning, nucleosome spacing and the nucleosome code. J Biomol Struct Dyn 27:713–894
2. Pavlovic J, Banz E, Parish RW (1989) The effects of transcription on the nucleosome structure of four *Dictyostelium* genes. Nucleic Acids Res 17:2315–2332
3. Edwards CA, Firtel RA (1984) Site-specific phasing in the chromatin of the rDNA in *Dictyostelium discoideum*. J Mol Biol 180: 73–90
4. Gaudet P, Fey P, Basu S, Bushmanova YA, Dodson R, Sheppard KA, Just EM, Kibbe WA, Chisholm RL (2011) dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. Nucleic Acids Res 39:D620–D624
5. Stevense M, Chubb JR, Muramoto T (2011) Nuclear organization and transcriptional dynamics in *Dictyostelium*. Develop Growth Diff 53:576–586
6. Rot G, Parikh A, Curk T, Kuspa A, Shaulsky G, Zupan B (2009) dictyExpress: a *Dictyostelium discoideum* gene expression database with an explorative data analysis web-based interface. BMC Bioinformatics 10:265
7. Kent NA, Mellor J (1995) Chromatin structure snap-shots: rapid nuclease digestion of chromatin in yeast. Nucleic Acids Res 23: 3786–3787
8. Kent NA, Adams S, Moorhouse A, Paszkiewicz K (2010) Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. Nucleic Acids Res 39:e26
9. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
10. Nicol JW, Helt GA, Blanchard SG, Raja A, Loraine AE (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25:2730–2731
11. de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20:1453–1454
12. Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. Bioinformatics 20:3246–3248
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
14. Chang GS, Noegel AA, Mavrich TN, Muller R, Tomsho LP, Ward E, Felder M, Jiang C, Eichinger L, Glockner G, Schuster SC, Pugh BF (2012) Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in *Dictyostelium*. Genome Res 22:1098–1106

Chapter 13

The Application of the Cre-*loxP* System for Generating Multiple Knock-out and Knock-in Targeted Loci

Jan Faix, Joern Linkner, Benjamin Nordholz, James L. Platt, Xin-Hua Liao, and Alan R. Kimmel

Abstract

Dictyostelium discoideum is an exceptionally powerful eukaryotic model to study many aspects of growth, development, and fundamental cellular processes. Its small-sized, haploid genome allows highly efficient targeted homologous recombination for gene disruption and knock-in epitope tagging. We previously described a robust system for the generation of multiple gene mutations in *Dictyostelium* by recycling the Blasticidin S selectable marker after transient expression of the Cre recombinase. We have now further optimized the system for higher efficiency and, additionally, coupled it to both, knock-out and knock-in gene targeting, allowing the characterization of multiple and cooperative gene functions in a single cell line.

Key words Cre recombinase, *Dictyostelium*, *loxP*, Homologous recombination, Gene replacement, Knock-outs, Epitope tags

1 Introduction

The genome of *Dictyostelium discoideum* is relatively small (~34 Mb) in comparison to most eukaryotes, but has a highly compact and haploid organization with short (<1 kb, average) intergenic distances (1, 2). Thus, *Dictyostelium* has ~12,500 protein-coding genes (1–3), yet frequencies for targeted homologous recombination at an individual locus can exceed 20%. In addition, genome-wide restriction enzyme-mediated insertional (REMI) mutagenic approaches allow the targeting and recovery of multiple loci (4). Nonetheless, there remain only few, highly effective selectable markers for *Dictyostelium*, which had limited the ability to create multiple mutations within a single cell. In addition, genetic crosses between strains are still limited (5). To overcome these limitations, we developed a very efficient system for the creation of multiple gene mutations within an individual *Dictyostelium*

Ludwig Eichinger and Francisco Rivero (eds.), *Dictyostelium discoideum Protocols*, Methods in Molecular Biology 983, DOI 10.1007/978-1-62703-302-2_13,

cell (6–8) that recycles a single selectable marker, Blasticidin S-resistance (9), using the Cre-*loxP* system (10). This has allowed the study of epistatic relationships among many genes in the large *Dictyostelium* transcriptome and to evaluate potential redundancies among various pathways (11, 12).

We created a universal gene-targeting vector backbone with a *Bsr* cassette flanked (floxed) by *loxP* recombination sites; translational stop codons in all six reading frames were placed outside one *loxP* site (6–8). Transient expression of the Cre recombinase removes the Blasticidin S-resistance expression cassette from the disrupted gene, but leaves translational stop codons in all the reading frames. This intramolecular recombination event creates a nonsense mutation within the targeted gene. Since the resulting cells are Blasticidin S-sensitive, they can be utilized for additional rounds of gene disruption or for REMI mutagenic screening. A new extrachromosomal, Cre-expressing vector makes the process even more efficient (8).

We have additionally engineered a new knock-in targeting vector that will add an epitope tag to the C-terminus of any targeted protein (13, 14). This ensures that any protein can now be tagged and expressed under control of their endogenous regulatory elements, for purification and cellular localization studies. Although we describe a tandem affinity purification (TAP) tag, the approach is adaptable to many other epitopes and for N-terminal tagging.

Further, many transformation vectors exist that direct temporal- or spatial-specific expression of mutated or tagged proteins (e.g., GFP, RFP) in *Dictyostelium*, and cells deleted of Bsr remain G418-sensitive and can still be engineered for the regulated expression of specific protein variants, for global cDNA screening, for complementation expression, or for gene interference by RNAi or antisense methodologies (1, 15–19).

Although we have not observed any abnormal phenotypes upon expression of Cre in *Dictyostelium*, it is prudent to compare the behavior of parental and Cre-recombined cells to ensure that *Bsr* deletion (or Cre-transformation) has not created a dominant (or secondary) phenotype. For example, we have frequently detected normal levels of mRNA expression after the floxed-*Bsr* is removed via Cre recombination. Although an in-frame nonsense codon has been incorporated into the gene, there is the potential for production of a truncated protein that may create a phenotype that is distinct from a loss-of-function mutation. Unless one is specifically interested in the function of proteins with carboxyl-terminal truncations, floxed-*Bsr* insertions should generally be designed near the 5′-end of the gene.

The regulated expression of Cre, using cell-specific promoters or the tetracycline-responsive system (16), may permit conditional gene disruption. Thus, it may be possible to study the effects of gene loss at specific developmental stages or in specific cell-types or the function of essential genes. The floxed-*Bsr* cassette has other advantages.

Once *loxP* sites have been inserted into a gene of interest, they are efficient recombination targets for the creation of gene knock-ins to express mutant or tagged protein variants or for novel promoter/reporter fusions (20).

Finally, it should be noted that recombination between *loxP* sites has been used effectively in mammalian systems to induce gene expression by fusing a promoter and gene target that had been separated by a floxed-inactivating sequence, or conversely, to repress expression by deleting an element that is essential for transcription.

2 Materials

2.1 Generation of Targeting Vectors and Transformation of Escherichia coli

1. *E. coli* cells (e.g., DH5α or any other suitable *E. coli* strain).
2. LB broth: 10.0 g of bacteriological tryptone (Difco), 5.0 g of yeast extract, 5.0 g of NaCl. Bring to 1 L with deionized water and adjust pH to 7.4 using 1 N NaOH.
3. Transformation buffer I (TFB I): 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$, 30 mM potassium acetate, 15% glycerol; adjust pH to 5.8 with saturated acetic acid. Do not autoclave, sterile filter. Store at –20°C.
4. Transformation buffer II (TFB II): 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% glycerol; adjust pH to 7.0 with 1 N NaOH, sterile filter. Store at –20°C.
5. Plasmid vector DNA (pLPBLP), available from the Dicty Stock Center (http://dictybase.org/StockCenter/StockCenter.html) (1).
6. PCR custom primers.
7. *Taq* polymerase, 5 U/μL (Roche).
8. 10× PCR buffer: 500 mM KCl, 100 mM Tris–HCl, pH 8.3.
9. 10× dNTPs 2 mM each dATP, dTTP, dGTP, and dCTP.
10. Thermal cycler (VWR Doppio or other suitable).
11. TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.
12. DNA loading buffer: 40.0 g of sucrose, 0.5 g of SDS, 0.25 g of bromophenol blue. Bring to 100 mL with TE-buffer.
13. Ethidium bromide (EtBr) solution, 10 mg/mL.
14. 10× Tris-phosphate buffer: 355 mM Tris, 302 mM NaH_2PO_4, 10 mM EDTA-disodium dihydrate. The pH of the 1× solution will self adjust to pH 7.8.
15. Agarose in 1× Tris-phosphate buffer, pH 7.8.
16. Agarose gel electrophoresis system (Biorad).
17. Restriction nucleases.
18. Gel extraction kit (Qiagen).

19. PCR purification kit (Qiagen).
20. Alkaline phosphatase, 1 U/μL (Roche).
21. T4 DNA ligase, 1 U/μL (Roche).
22. 10× Ligation buffer: 660 mM Tris–HCl, pH 7.5, 100 mM $MgCl_2$, 10 mM spermidine, 10 mM ATP, 20 mM DTT, 1.5 mg/mL BSA. Commercially available 10× ligase buffers also work well.
23. TCM solution: 10 mM Tris–HCl, pH 8.0, 100 mM $CaCl_2$, 100 mM $MgCl_2$.
24. Ampicillin stock solution (1,000×): 0.5 g of ampicillin; bring to 10 mL with deionized water, sterile filter. Store at 4°C.
25. LB-ampicillin-plates: 15.0 g of Bacto-agar (Difco); bring to 1 L with LB broth, autoclave, cool in a water bath to 50°C, add 1 mL of ampicillin stock solution.
26. Quiaprep Spin Miniprep Kit (Qiagen).
27. Plasmid DNA Maxi Kit (Qiagen).

2.2 Transformation of Dictyostelium Cells

1. *D. discoideum* cells (e.g., Ax2 or Ax3; see http://dictybase.org/StockCenter/StockCenter.html) (1).
2. Ax medium: 14.3 g of bacteriological peptone (L-34, Oxoid), 7.15 g of yeast extract, 18.0 g of maltose, 0.616 g of $Na_2HPO_4 \cdot 2H_2O$, 0.486 g of KH_2PO_4; bring to 1 L with deionized water and adjust pH to 6.7 with 1 N NaOH. Autoclave. Store at 4°C. Ax medium is also commercially available from Formedium.
3. B12/folic acid mix: 5 mg of B12 (cyanocobalamine), 200 mg of folic acid; add 95 mL of deionized water, adjust pH to 6.5 with 5 N NaOH, and bring to 100 mL. Filter sterilize and store at −20°C protected from light.
4. HL5 medium: 17.8 g of bacteriological peptone (L-85, Oxoid), 7.2 g of yeast extract, 0.54 g of Na_2HPO_4, 0.4 g of KH_2PO_4, 130 μL of B12/folic acid mix; bring to 1 L with deionized water and adjust pH to 6.5. Autoclave and add 20 mL of 50% (w/v) glucose. Store at 4°C.
5. HL5-C medium with (or without glucose) is commercially available from Formedium.
6. 17 mM Na-K-phosphate buffer, pH 6.0: 0.356 g of Na_2HPO_4, 1.99 g of KH_2PO_4. Bring to 1 L with deionized water, autoclave. Store at 4°C.
7. Electroporation buffer: 10 mM K-PO_4, pH 6.1, 50 mM glucose, Filter sterilize, and store at 4°C.
8. Electroporation cuvettes: 4-mm gap.
9. Electroporator XCell (Biorad).

10. Healing solution: 100 mM $CaCl_2$, 100 mM $MgCl_2$.
11. 1,000× Blasticidin S stock solution: 10 mg/mL in deionized water, sterile filtered. Store at 4°C or freeze in aliquots.
12. *Klebsiella aerogenes.*
13. SM agar plates: 15.0 g of Bacto-agar (Difco), 10.0 g of bacteriological peptone (L-34, Oxoid), 1.0 g of yeast extract, 10.0 g of glucose, 1.0 g of $MgSO_4 \cdot 7H_2O$, 2.2 g of KH_2PO_4, 1.0 g of K_2HPO_4, bring to 1 L and autoclave. Store at 4°C.
14. SM agar plates (Formedium): Suspend 41.7 g of powdered medium in deionized water, bring to 1 L, and autoclave. Store at 4°C.
15. Sterile toothpicks.
16. 100× Ampicillin/streptomycin stock solution: 5 mg/mL of ampicillin, 4 mg/mL of streptomycin sulfate, sterile filtered. Store at 4°C or freeze at –20°C. Commercially available 100× penicillin/streptomycin stock solutions also work well.

2.3 Validation of Knock-out Mutants

1. High Pure PCR Template kit (Roche).
2. Dimethylsulfoxide (DMSO).

2.4 Removal of the Floxed-Bsr Cassette

1. Plasmid vector DNA pDEX-NLS-Cre and pTX-NLS-Cre, available from the Dicty Stock Center (http://dictybase.org/StockCenter/StockCenter.html) (1).
2. Phosphate agar plates: 15.0 g of Bacto-agar (Difco); bring to 1 L with 17 mM Na-K-phosphate buffer, autoclave. Store at 4°C.
3. G418 stock solution: 10 mg/mL in deionized water, sterile filtered. Store at 4°C or freeze in aliquots.

3 Methods

3.1 Preparation and Transformation of Chemically Competent E. coli Cells

3.1.1 Preparation of Competent E. coli Cells

1. Inoculate *E. coli* strain DH5α (or any other suitable *E. coli* host used in your laboratory) in 10 mL of LB broth and allow to grow at 37°C overnight.
2. Inoculate 5 mL of the overnight culture into 200 mL of fresh LB broth in a 1-L Erlenmeyer flask.
3. Grow the cells at 37°C at 220 rpm to an OD_{600} of approximately 0.3–0.6. The best results are obtained when the cells are harvested during early log phase.
4. Chill the cells in an ice/water bath for ~15 min. For all subsequent steps, keep the cells as close to 0°C as possible and chill all containers and centrifuges before adding cells. Transfer the cells to four sterile 50-mL centrifugation tubes and centrifuge at 4,000 × *g* for 15 min at 4°C.

5. Carefully pour off and discard supernatant and resuspend the pellets in 50 mL of ice cold TFB I buffer and keep cells on ice for 1 h.
6. Centrifuge the cells at 4,000 × *g* for 15 min at 4°C.
7. Carefully pour off and discard supernatant and resuspend the pellet in 8 mL of ice-cold TFB II buffer and keep cells on ice for 1 h.
8. Prepare 200-μL aliquots of the cells in prechilled 1.5-mL microfuge tubes and immediately shock freeze in liquid nitrogen.
9. Store competent cells at −80°C.

It is also possible to purchase *E. coli* cells that are competent for transformation from various companies (e.g., Promega, Thermo Scientific, Stratagene).

3.1.2 Transformation of E. coli Cells

1. Place frozen chemically competent *E. coli* cells on ice and allow the cells to thaw slowly.
2. Add 200 μL of the cells to a ligation mix (see Subheading 3.2, step 14) and incubate on ice for 1 h.
3. Heat shock cells for 1 min at 42°C.
4. Chill cells for 5 min on ice.
5. Add 800 μL of LB broth and incubate for 1 h at 37°C.
6. Plate cells on LB agar plates containing 50 μg/mL ampicillin and incubate at 37°C overnight.

3.2 Construction of Targeting Vectors for Gene Disruption

1. Synthesize two primer pairs for the amplification of a 5′ and 3′ fragment of your gene of interest (see Note 1) that will generate PCR products of approximately 500 bp for each fragment. At their 5′ ends the primers should carry recognition sites for restriction endonucleases to facilitate cloning in targeting vector pLPBLP (see Fig. 1). The unique cloning sites of vector pLPBLP are *Kpn*I-*Sal*I-*Acc*I-*Cla*I-*Hin*dIII-*Sma*I-(floxed-*Bsr* cassette)-*Pst*I-*Sma*I-*Bam*H-*Spe*I-*Not*I.
2. Amplify the 5′ and 3′ fragments from genomic DNA or plasmid DNA by PCR. Set up a 100-μL reaction consisting of: 0.01 up to 0.1 μg of plasmid DNA or genomic template, 0.1 μM of each primer, 0.2 mM of each dNTP, and 10 μL of 10× PCR buffer. Bring to 99 μL with deionized water, add 1 μL of *Taq* DNA polymerase (see Note 2), and mix.
3. Confirm size and quantity of PCR products by analytical agarose gel electrophoresis and staining with EtBr.
4. Digest the PCR products with appropriate restriction enzymes overnight at 37°C in a final volume of 200 μL (see Note 3).
5. Add 30 μL of DNA loading buffer and run the entire sample on preparative agarose gel for each fragment.

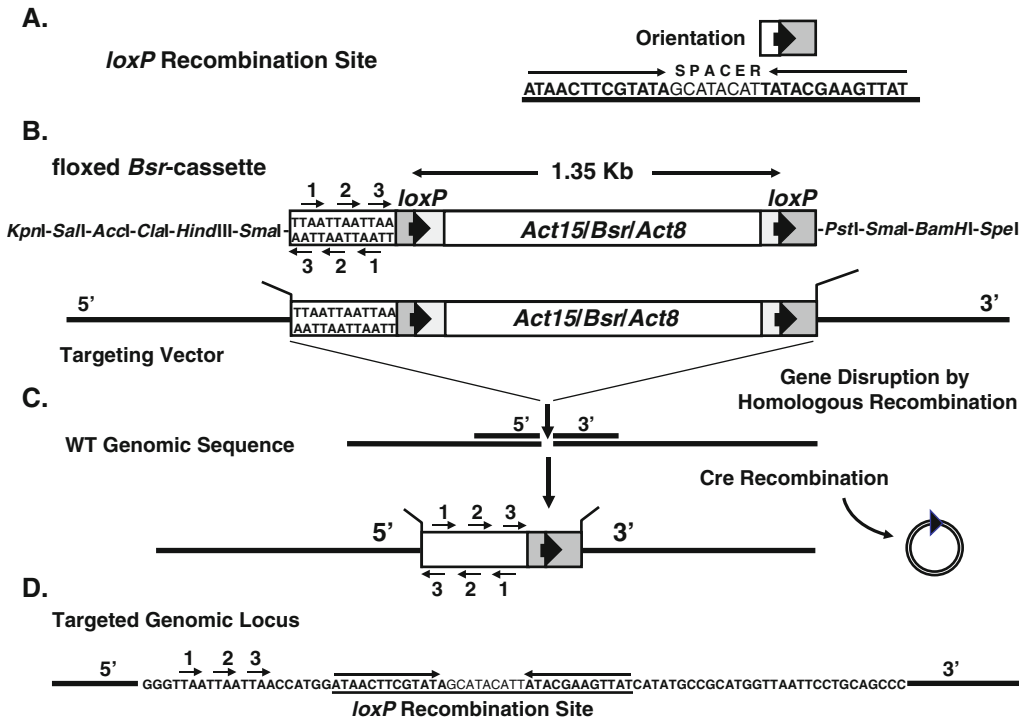


Fig. 1 The strategy for Cre-*loxP* recycling of the *Bsr* selectable marker. (**a**) The *loxP* recombination site includes a 13 bp inverted repeat (*Bold*), separated by a short spacer sequence. (**b**) The floxed-*Bsr* cassette was constructed with *loxP* sites in the same orientation flanking both sides of the Blasticidin-resistance (*Bsr*) expression cassette (9, 10); the *Bsr* gene is flanked by the *Act15* promoter and the *Act8* terminator. An oligonucleotide cassette is added upstream, with translational stop codons in all six reading frames. Restriction enzyme sites outside of the floxed-*Bsr* cassette permit the cloning of 5′ and 3′ gene sequences for targeted disruption. The entire floxed-*Bsr* cassette is part of the pLPBLP vector (~4.5 kb), which additionally contains a bacterial origin of replication and the ampicillin resistance gene (6–8). (**c**) The floxed-*Bsr* cassette is inserted between 5′ and 3′ gene-targeting fragments (>500 bp). Wild-type cells are transformed for gene disruption by homologous recombination, selected for resistance to Blasticidin S, and screened. Transient expression of Cre promotes recombination between the two *loxP* sites, leaving only a short (<100 bp) sequence that includes the translational stop cassette and a single *loxP* site. Since translational stops interrupt all six reading frames, the orientation of the floxed-*Bsr* cassette within the targeting vector is generally not significant. (**d**) The resulting sequence organization within the targeted gene is shown with the single *loxP* site and in-frame stop codons indicated

6. Cut out the bands of interest with a razor blade and purify the appropriate DNA fragments using a gel extraction kit (Qiagen).
7. Confirm quantity of purified PCR products by analytical agarose gel electrophoresis and staining with EtBr. Store the fragments at −20°C for later use.
8. Digest 10 μg of vector pLPBLP with two restriction enzymes compatible for the insertion of the first PCR fragment for 2 h at 37°C in a final volume of 200 μL. Place the mix on ice and confirm completion of the digestion by running a small aliquot of the reaction in an agarose gel.

9. Stop the reaction by addition of fivefold volume (1 mL) of PB buffer (included in the PCR purification kit) and purify the DNA fragments using the PCR purification kit.
10. Dephosphorylate cleaved vector pLPBLP with 1 μL of alkaline phosphatase for 1 h at 37°C in a final volume of 200 μL.
11. Stop the reaction by addition of 1 mL of PB buffer and purify the DNA fragments using the PCR purification kit.
12. Ligate 50–100 ng of cleaved and dephosphorylated vector pLPBLP with approximately fivefold molar excess of the first PCR fragment with 1 μL of T4 DNA ligase in 1× ligation buffer and a final volume of 20 μL overnight at 15°C.
13. Add 30 μL of TCM solution to the ligation buffer and chill on ice.
14. Transform ligation reaction as described in Subheading 3.1.2.
15. Inoculate colonies in 4 mL of LB broth containing 50 μg/mL of ampicillin and grow the cells to the stationary phase overnight at 37°C.
16. Prepare plasmid DNA with Qiaprep Spin Miniprep kit or other standard procedure and isolate derivatives of plasmid pLPBLP containing the first fragment.
17. Confirm insertion of appropriate fragment by restriction digest.
18. Digest confirmed plasmid with the second combination of restriction enzymes and perform treatment with alkaline phosphatase as described in steps 8–11.
19. Ligate second PCR fragment into this vector as described in step 12 and repeat steps 13–17 in order to isolate gene disruption plasmid pLPBLP containing both fragments flanking the *Bsr* cassette.
20. Inoculate cells containing the final gene disruption plasmid in 300 mL of LB broth containing 50 μg/mL ampicillin and incubate at 37°C overnight.
21. Prepare plasmid DNA with Maxiprep kit.
22. Digest approximately 100 μg (this amount of plasmid DNA will allow one to perform three *Dictyostelium* transformations) of plasmid DNA for 2 h at 37°C in a final volume of 250 μL with a combination of restriction enzymes that liberates two DNA fragments: one plasmid fragment will contain the 5′-genomic fragment-floxed-*Bsr* cassette-3′-genomic fragment and the other plasmid fragment will contain the remainder of the vector backbone (~2.7 kbp).
23. Confirm completion of the digestion by running a small aliquot of the reaction in an agarose gel (see Note 4).
24. Stop the reaction by addition of 1 mL of PB buffer and purify the DNA fragments using the PCR purification kit (Qiagen).

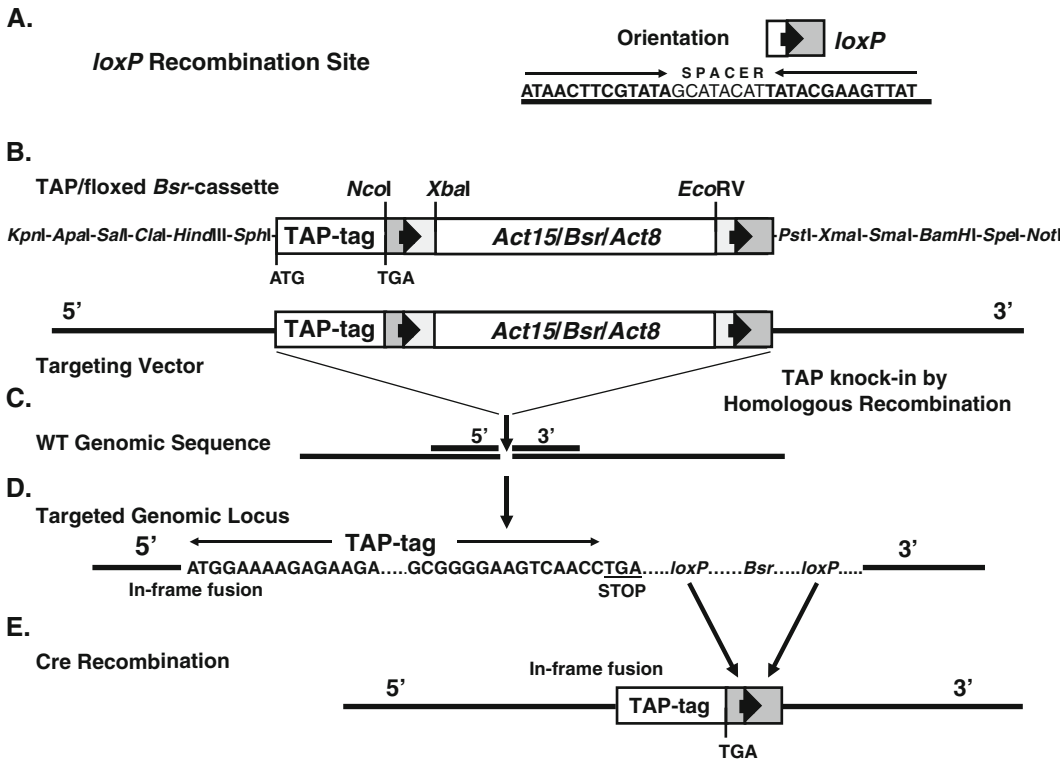


Fig. 2 TAP-tag knock-in strategy. (**a**) Vector organization. The *loxP* recombination site includes a 13 bp inverted repeat (*Bold*), separated by a short spacer sequence. (**b**) The TAP/floxed-*Bsr* cassette was constructed by fusing a TAP construct (13, 14) to the floxed-*Bsr* cassette (see Fig. 1). Restriction enzyme sites, upstream of the ATG translation start codon in the TAP-tag, are used to create C-terminal protein fusions. The entire TAP/floxed-*Bsr* cassette is part of the pLPBLP-TAP vector (~5.0 kb), which additionally contains a bacterial origin of replication and the ampicillin resistance gene. (**c**) The TAP/floxed-*Bsr* cassette is inserted between 5′ and 3′ gene-targeting fragments (>500 bp). The 5′ targeting fragment, would encode the C-terminus of the protein to be "tagged"; the endogenous translational STOP must be removed and the fusion to TAP must create an in-frame read-through (see Fig. 3). Wild-type cells are transformed for gene disruption by homologous recombination, selected for resistance to Blasticidin S, and screened. (**d**) The resulting sequence organization within the targeted in-frame, gene fusion is shown with the entire floxed-*Bsr* cassette still present. (**e**) Transient expression of Cre promotes recombination between the two *loxP* sites, leaving only the TAP-tag and a single *loxP* site

It is not necessary to isolate and separate the *Bsr* cassette flanked by the 5′ and 3′-fragments from the vector backbone.

25. Elute the DNA fragments with 60 µL of 10 mM Tris–HCl, pH 8.0 (see Note 5).
26. Run 1 µL of the sample on an analytical 0.7% agarose gel to validate quality and quantity of the DNA fragments.
27. Store eluted DNA at −20°C for later use.

3.3 Construction of Targeting Vectors for Epitope Tagging

The strategy for targeted in-frame fusion of an epitope within a gene locus is conceptually similar to that described for gene disruption (see Subheading 3.2). However, four fragment cloning aspects are fundamentally different (Fig. 2). First, the orientation of the 5′

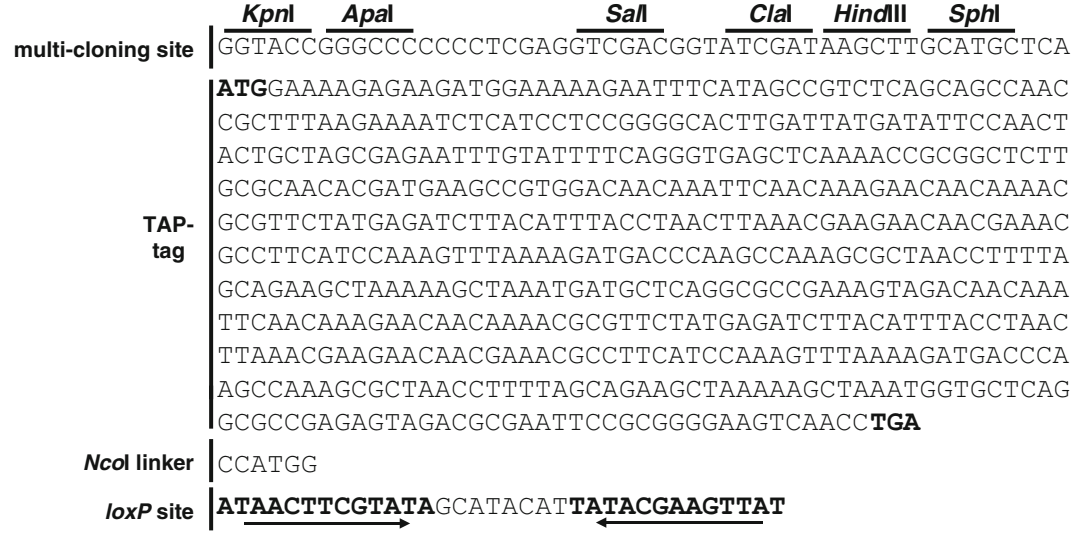


Fig. 3 In-frame TAP sequence organization. The TAP-tag sequence is shown, including the translational start/stop codons. Upstream cloning sites are indicated, as is downstream linkage with *loxP*

and 3′ fragments is absolutely critical. The 5′ fragment must be placed upstream of the epitope tag. Second, the 5′ fragment must encode the C-terminal domain of the selected gene, with no additional genomic sequences. Third, the STOP codon must be removed from the C-terminal coding fragment. Lastly, the C-terminal coding fragment must have an in-frame read-through when fused to the epitope tag (Fig. 3). This may include amino acids that derive from the multi-cloning site. We also often place three glycines between the endogenous gene and the TAP-tag. All other aspects of DNA cloning, transfection into *Dictyostelium*, selection, screening, etc., follow procedures outlined in Subheadings 3.1, 3.2, and 3.4 (see Note 6).

3.4 Transformation of Dictyostelium Cells by Electroporation

3.4.1 Preparation of Electrocompetent Dictyostelium Cells

1. Inoculate *Dictyostelium* cells at a concentration of $5–7 \times 10^5$ cells/mL into 200 mL of fresh axenic growth medium (Ax, HL5, or HL5-C) in a 1,000-mL flask. The cells may be washed off a plastic petri dish or transferred from liquid medium. The doubling time of *Dictyostelium* in shaken suspension is approximately 8–10 h at 21°C.
2. Incubate the culture at 21°C for about 24 h with shaking at 150 rpm and harvest the cells at a density of not more than 5×10^6 cells/mL.
3. Transfer 100 mL of the cells into two sterile, disposable 50 mL centrifugation tubes and incubate on ice for 15 min.
4. Pellet the cells by centrifugation at $500 \times g$ for 2 min at 4°C.
5. Carefully pour off and discard the supernatant and place the centrifugation tubes with the cell pellets on ice.

6. Pool the pellets and resuspend in 50 mL of ice cold 17 mM Na-K-phosphate buffer, pH 6.0.
7. Pellet the cells by centrifugation at $500 \times g$ for 2 min at 4°C, again pour off and discard the supernatant and resuspend the pellet in 50 mL of ice cold electroporation buffer.
8. Repeat step 7 but resuspend the cells at a concentration of 1×10^7 cells/mL in ice cold electroporation buffer. Keep the cells on ice and use as soon as possible for electroporation.

3.4.2 Electroporation of Dictyostelium Cells

1. Pipette the DNA samples to be electroporated (~35 μg) into sterile 4-mm electroporation cuvettes and place them on ice.
2. Add 700 μL of the competent cells to each DNA sample, mix gently, and incubate on ice.
3. Electroporate the cells using Biorad Xcell gene pulser preset protocol for *Dictyostelium* (these conditions are square wave, $V = 1.0$ kV, 10 μF, 1.0 ms pulse length, 2 pulses, and 5 s pulse interval). The time constant should be ~1 ms. The voltage should be ~1.0 kV. This setup routinely yields hundreds of transformants.
4. Remove the cuvette from the chamber and plate the cells in the middle of a petri dish. Place the dish on a laboratory shaker and gently shake at ~40 rpm for 15 min at room temperature.
5. Adjust the suspension to 2 mM $CaCl_2$ and 2 mM $MgCl_2$ with healing solution and continue shaking for another 15 min at room temperature.
6. Add 12 mL of axenic growth medium and allow the cells to recover over night at 21°C.
7. Add Blasticidin S at a final concentration of 10 μg/mL and incubate at 21°C.
8. Gently change medium-containing Blasticidin S every 3 days and select the transformants for approximately 10–14 days at 21°C until colonies with a diameter of approximately 1 mm are clearly visible.
9. Isolate clonal *Dictyostelium* cell lines by spreader dilution. For this prepare a dense suspension of *K. aerogenes* that were grown on SM (or SM-Formedium) agar plates overnight at 37°C with sterile filtered 17 mM Na-K-phosphate buffer, pH 6.0.
10. Pre-dry approximately five SM agar plates per transformation in a laminar flow bench for ~30 min at room temperature and add 250 μL of the *K. aerogenes* suspension into the center of each agar plate.
11. Wash off the transformants from the plastic surface by gently pipetting up and down the medium and bring one to three drops in a 1.5-mL microfuge tube filled with 1 mL of medium. Briefly vortex the cells to disrupt cell aggregates.

12. Add increasing amounts (1, 2.5, 5, 10, and 20 μL) of the cell suspension into the corresponding drop of *K. aerogenes* and immediately spread the drop over the entire SM plate using a sterile Drygalski spatula.
13. Allow the plates to dry and subsequently incubate at 21°C.
14. After ~3 days individual plaque forming colonies will appear in the bacterial lawn.
15. After the colonies have reached a diameter of 1–2 mm, pick individual clonal cell lines with sterile toothpicks and transfer them into sterile 24-well plates containing axenic medium supplemented with 10 μg/mL Blasticidin S and ampicillin/streptomycin solution to prevent bacterial growth.

3.5 Validation of Knock-out Mutants

Two PCRs are performed to rapidly validate targeted gene disruption. One PCR examines the presence of the wild-type or disrupted gene using primers that flank the floxed-*Bsr* insertion sites (see Fig. 4; Note 7). A second PCR design uses a primer from within the *Bsr* cassette and another primer from outside of the targeting

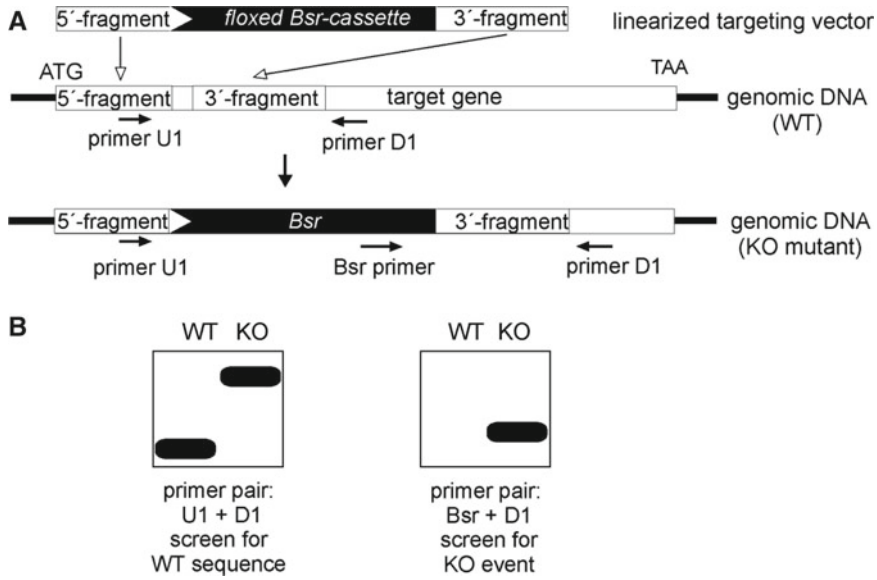


Fig. 4 Validation of targeted gene disruption. (**a**) Generation of a null cell line for a target gene. 5′ and 3′ specific sequences of the target gene are cloned into pLPBLP. The linear-targeting vector is then used to disrupt the gene by homologous recombination. Dependent on the spacing between the 5′ and 3′ fragments, targeted integration may cause a small deletion in the gene. Upstream primer U1 is within the 5′ targeting fragment, primer Bsr is within the *Bsr* gene, and primer D1 is within the endogenous genomic sequence, but downstream of the 3′ targeting fragment. (**b**) Schematic validation of targeted gene disruption by PCR. A wild type and a Blasticidin-resistant KO mutant are examined by PCR amplification employing the two different sets of primers indicated. *Left Panel*—The primer combination of U1 and D1 amplifies both the WT and homologously target gene; the target gene PCR product is approximately 1.3 kbp larger than that of WT. *Right Panel*—The primer combination of Bsr and D1 specifically amplifies only a fragment after the homologous recombination event and, hence, is seen only in the bona fide disrupted mutant. This figure is reprinted from ref. 8, with permission from Elsevier

Table 1
Reliable *Bsr*-specific primers that are used in different laboratories

Primer sequence	Length (bp)	Orientation	Position in *Bsr* cDNA (bp), relative to ATG	Laboratory
CAGTTACTCGTCCTATATACG	21	Antisense	152–172	Jan Faix
CATTGTAATCTTCTCTGTCG CTACTTCTAC	30	Antisense	50–79	David Traynor
GTGTAGGGAGTTGATTTCAG ACTATGCACC	30	Sense	306–335	David Traynor
AGTATTCGAGTGGTAAGTCCTTG	23	Sense	276–298	Alan Kimmel
GGTGCATAGTCTGAAATCAACTC	23	Antisense	313–335	Alan Kimmel

fragment. A number of reliable *Bsr*-specific primers used in different laboratories are listed (Table 1). A similar strategy is used to identify successful knock-in targeting. A different primer is required in the second PCR, replacing the primer from the *Bsr* cassette with one from the TAP cassette (see Figs. 2 and 3). A TAP in-frame knock-in should also be confirmed by immunoblot with an antibody to the TAP epitope (see Note 6).

1. Inoculate 10-cm petri dish plates containing 12 mL of axenic medium supplemented with 10 μg/mL Blasticidin S with individual clones from the 24-well plates and allow the cells to colonize the entire plate.
2. Prepare genomic DNA using the High Pure PCR Template Preparation kit as described by the manufacturer following the protocol for cultured cells (see Note 8).
3. Set up a 100-μL reaction consisting of 0.1 μg of genomic template DNA, 0.1 μM of each primer, 0.2 mM of each dNTP, and 10 μL of 10× PCR buffer. Bring to 99 μL with deionized water, add 1 μL of *Taq* DNA polymerase, and mix (see Note 9).
4. For a final product size of approximately 1 kbp perform the reaction using the following general protocol: 94°C for 30 s, 49°C for 60 s, and 70°C for 90 s for 30 cycles. Allow a longer extension time in the last cycle.
5. For a final product size of approximately about 2.5 kbp perform the reaction using the following general protocol: 94°C for 30 s, 49°C for 60 s, and 70°C for 160 s for 30 cycles. Allow a longer extension time in the last cycle.
6. Validate targeted gene disruption by examination of aliquots of the two PCR reactions in a 0.7% analytical agarose gel and EtBr staining.

7. Store the knock-out mutants by slowly freezing the cells in growth medium supplemented with 7–10% DMSO. For this, harvest ~5×10^7 cells during exponential growth by pelleting at $500 \times g$ for 2 min at 4°C. Discard supernatant and resuspend the pellet in ice cold DMSO-containing growth medium and transfer the suspension to prechilled liquid nitrogen storage vials. Place the vials into the gaseous phase above the liquid nitrogen in a nitrogen tank for 2 h to allow for slow cooling (~1°C/min). Subsequently transfer vials to liquid nitrogen for long-term storage, and if liquid nitrogen is unavailable, storage frozen cells in a −80°C freezer. Alternatively prepare spores and store them at −80°C.

3.6 Removal of the Floxed-Bsr Cassette by Transient Expression of Cre

1. Bsr knock-out cells are electroporated with 35 μg of either pDEX-NLS-Cre or the further developed extrachrosomally replicating pTX-NLS-Cre (see Fig. 5) as described in Subheading 3.4.
2. After the 1–3 h recovery period, add G418 to a final concentration of 10–20 μg/mL.
3. Selection is continued for 3–10 days in case of pDEX-NLS-Cre or up to 14 days or more in case of pTX-NLS-Cre (see Note 10).
4. After appearing of colonies the cells are spreader diluted for clonal selection on SM agar plates containing *K. aerogenes* as described in Subheading 3.4.2, steps 9–14.
5. After the colonies have reached a diameter of 1–2 mm, pick individual clonal cell lines with sterile toothpicks and transfer them into a 24-well plate containing 1 mL axenic medium in the absence of G418 but supplemented with 50 μg/mL ampicillin and 40 μg/mL streptomycin to prevent bacterial growth.
6. Aspirate off the medium, replace with fresh medium every second day to remove remaining bacteria, and allow the cells to reach cell densities of ~1×10^6 cells/mL.
7. Transfer 400-μL aliquots of each well into two separate 24-well plates containing growth medium supplemented with either G418 or Blasticidin S at 15 μg/mL. Cell lines sensitive to G418 and Blasticidin S can be easily identified by comparing growth of the cells in the three tested conditions within approximately 3 days.
8. Cells not growing in Blasticidin S- and G418-containing media are of potential interest. Usually, >95% of cells selected by these growth criteria will show appropriate Cre recombination. Take these cells from the first 24-well plate and inoculate fresh 10-cm petri dishes containing 12 mL of axenic medium and allow the cells to colonize the entire plate.

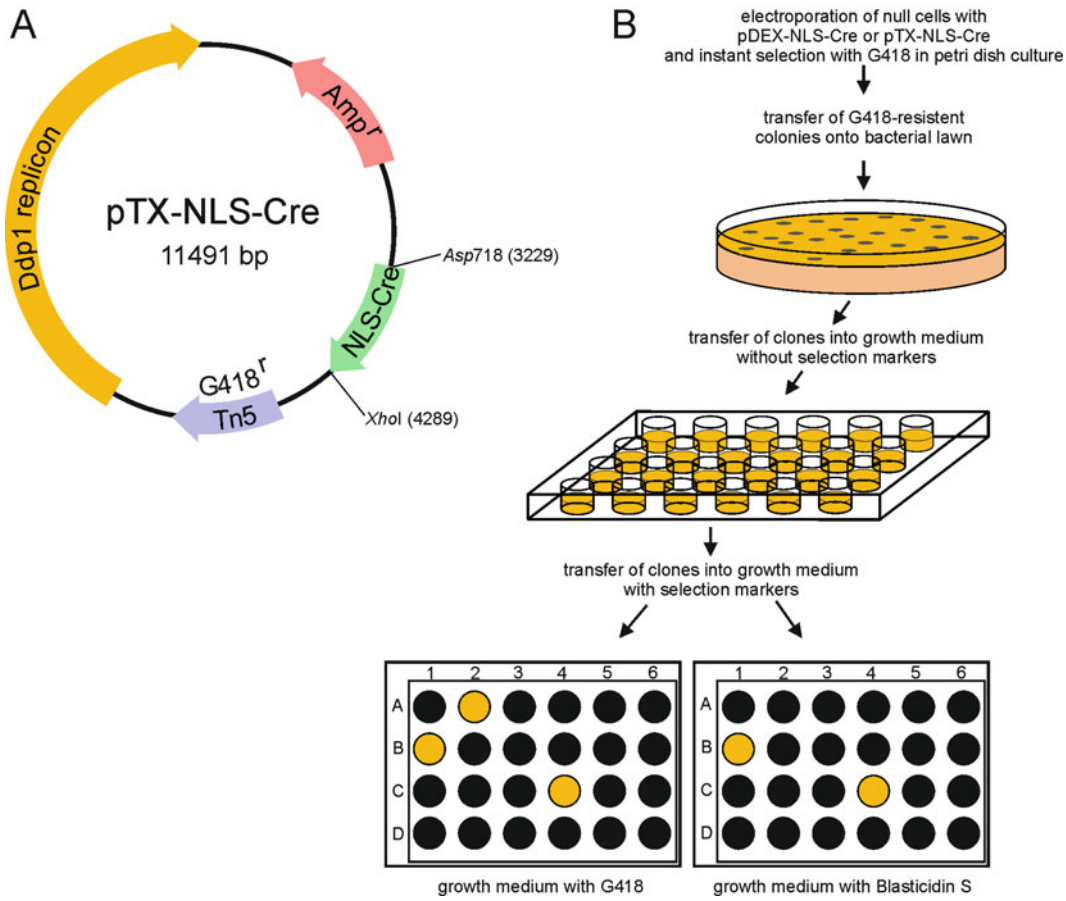


Fig. 5 Screening of G418 and Blasticidin S sensitive null-mutants cells after transient expression of NLS-Cre. (**a**) Schematic view of the extrachromosomal pTX-NLS-Cre-expression plasmid. The actin15 promotor drives expression of the Cre recombinase fused to an N-terminal signal peptide for nuclear localization (NLS-Cre). An ampicillin resistance cassette (Amp^r) enables selection in *E. coli* and the neomycin resistance cassette (G418^r) allows selection of transfected *D. discoideum* cells. Extrachromosomal replication of the plasmid is conferred by the Ddp1 replicon. (**b**) After transfection (of any selected mutants containing a floxed-*Bsr* cassette; see Fig. 4) with a Cre-expression plasmid such as pDEX-NLS-Cre or pTX-NLS-Cre, the cells are selected with G418 in liquid medium in petri dish cultures for 5–14 days (see Note 10). G418 selection is then removed, and 3–7 days later, the cells are plated in low density with *K. aerogenes* onto SM agar plates. Clonal cells are subsequently transferred into 24-well plates, replica plated, and separately screened for sensitivity towards both G418 and Blasticidin S. Growing cells are indicated by *ochre spots*, whereas *black spots* display clones that are unable to grow under selection. Clones that are unable to grow in both G418 and Blasticidin S are expanded in 24-well plates grown without either antibiotic and tested for excision of the *Bsr* cassette by diagnostic PCR (see Fig. 6)

9. Isolate genomic DNA from the Blasticidin S- and G418-sensitive cell lines and validate Cre-mediated recombination by PCR as described in Subheading 3.5, steps 1–6, using the same two primer pairs used to validate the knock-out event (see Fig. 6).
10. The cells devoid of the *Bsr* cassette can now be used for the next round of *Bsr*-mediated targeted gene disruption.

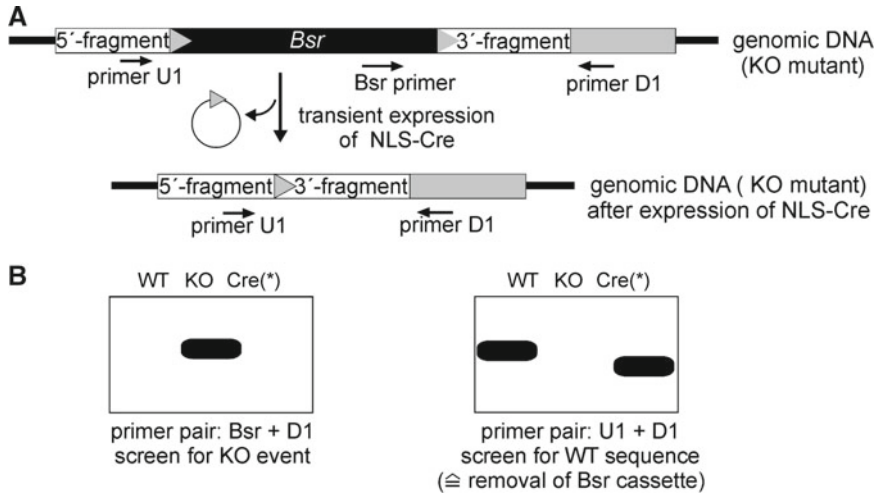


Fig. 6 Validation for removal of the floxed-*Bsr* cassette by expression of the Cre recombinase. (**a**) Strategy for removal of the *Bsr* cassette by transient expression of NLS-Cre. Deletion of the floxed-*Bsr* cassette leaves a sequence of <100 bp (see Fig. 1). Primer U1 is within the 5′ targeting fragment, primer Bsr is within the *Bsr* gene, and primer D1 is within the endogenous genomic sequence that is downstream of the 3′ targeting fragment. (**b**) PCR analysis of null cells following transient expression of NLS-Cre (*). Most of the clonal cell lines that are sensitive to both Blasticidin and G418 for growth also lack the floxed-*Bsr* cassette

4 Notes

1. The orientation of the *Bsr* cassette relative to the 5′ and 3′ genomic fragments is generally not important for targeted gene disruption. However, the 5′ and 3′ fragments must be inserted in the same orientation in the targeting vector as they are in the genome.
2. There are numerous thermostable polymerases currently available. This protocol has been adapted for *Taq* polymerase, since this enzyme is reliable and relatively inexpensive; however several other enzymes can be used as well. The "proofreading" activities of thermostable polymerases such as *Phusion*, *Vent* and *Deep Vent* (New England Biolabs), *Pfu* (Stratagene), or *KOD* (Novagene) are not required.
3. Depending on the restriction nuclease used for cloning, the primers and the length of the primer overhang, it might be very important to digest the amplified DNA fragments overnight. This procedure significantly increases the number of appropriately digested primer ends and hence the number of *E. coli* transformants. Digestions of DNA fragments at 37°C overnight should contain between 0.1 and 0.25 mM EDTA to avoid degradation of DNA by nonspecific nucleases.

4. The restriction digest of the final gene-targeting vector must be complete. Otherwise the undigested gene-targeting vector might integrate by a single crossover event into the *Dictyostelium* genome. This event may considerably complicate the interpretation of the PCR data and in most cases does not lead to disruption of the target gene.
5. It is recommended not to use EDTA-containing buffers for this step. Millimolar concentrations of EDTA considerably inhibit the transformation efficiency of *Dictyostelium* cells. Thus, TE buffers should be avoided. Use instead 10 mM Tris–HCl, pH 8.0, or deionized water for final elution of the DNA prior to transformation of *Dictyostelium* cells.
6. The targeting construct must be sequenced to ensure that the coding region is in-frame with the epitope tag and that no inadvertent mutations have been incorporated. Immunoblots with an antibody to the TAP epitope (Open Biosystems #CAB1001) must be used to confirm the successful incorporation of the epitope tag as a gene knock-in. The construct outlined (see Fig. 2) uses a TAP-tag, but as designed many other epitope-coding sequences can be easily inserted, including FLAG, myc, GFP, etc. Since the 5′ targeting fragment is derived from a very C-terminal coding region, the downstream 3′ targeting fragment is very likely to be comprised of A+T rich, intergenic elements, as well as sequences from the downstream gene. We have not observed deleterious effects in targeting efficiency. N-terminal tagging is also possible with appropriate reengineering to make certain there is in-frame read through following removal of the floxed-*Bsr* cassette. Finally, knock-in cell lines must be evaluated to ensure that the epitope tag does not cause a gain-of-function phenotype.
7. Unless the knock-out is confirmed by immunoblot analysis using a specific antibody, a control PCR for the WT allele is highly recommended. In some cases we observed the presence of the WT allele in addition to the disrupted locus. This might be due to local gene duplications or polyploidization in the chosen strain. In some instances, confirmation by Southern blot hybridization may be required. Integration of the *Bsr* cassette into a single-copy genomic locus is detected as a defined restriction enzyme fragment length change by hybridization with a gene-specific probe.
8. In some instances, unpurified DNA in cell lysates can be used for PCR screening. ~10 μL of cells from confluent plates are added to 10 μL of freshly made lysis buffer (50 mM KCl, 10 mM Tris–HCl, pH 8.3, 2.5 mM $MgCl_2$, 0.45% Tween 20, and 0.4 mg/mL proteinase K). The cell suspension is incubated at 22°C for 5 min, and then at 95°C for 1–5 min, to inactivate proteinase K. Use 1 μL for PCR screening (Subheading 3.5).

9. We have been able to detect gene targeting using Bsr/D1 (and TAP/D1) primers pairs (see Fig. 4) in mixed cell populations, at a sensitivity of 0.01%. Using this strategy, one can screen thousands of transformants in mixed pools of cells to identify those with the rare (e.g., poorly growing) positives. Cells in the appropriate pool are then aliquoted and rescreened. Generally only several rounds of such pool-aliquot/dilution screening are required to select pure single colony positives.
10. The duration of G418 selection is dependent upon a number of parameters, including the Cre-expression plasmid used, the concentration of G418 used in the selection, the medium, the temperature, and probably the strain or cell line used. The most important aspect of this step is the removal of background cells that had not been transfected with Cre recombinase expressing plasmids. This must be optimized for each laboratory. If selection with G418 is too short, knock-out cells will persist that still harbor the *Bsr* cassette; if selection is too long, cells may lack the *Bsr* cassette, but now contain pDEX-NLS-Cre as a stably integrated plasmid into the genome. To avoid this, we developed the extrachromosomal replicating plasmid pTX-NLS-Cre (8). This type of vectors not only yield higher transformation efficiencies when compared to integrating vectors, but most importantly, are known to be genetically unstable (cured) upon removal of positive selection. Thus, G418 selection can be considerably extended (>14 days) without the risk to encounter genomic integration. After longer selection periods, it is recommended to grow the cells for 5–7 days in the absence of G418 prior to PCR screening (see Fig. 6) to promote curing of the pTX-NLS-Cre plasmid.

Acknowledgements

We thank Alexander Junemann and Moritz Winterhoff for thoroughly testing plasmid pTX-NLS-Cre and Drs. Katrin Koch and Ralf Gräf for their TAP-tag plasmid. This research was supported by Deutsche Forschungsgemeinschaft (Fa 330/4-2, Fa 330/5-1), the Intramural Research Program of the National Institutes of Health, the National Institute of Diabetes and Digestive and Kidney Diseases, and the WellcomeTrust/NIH Program Studentship to J.L.P.; J.L.P. is joint student with Dr. Adrian Harwood (Cardiff University) and A.R.K. (NIH). There are no conflicts or competing interests.

References

1. Gaudet P, Fey P, Basu S, Bushmanova YA, Dodson R, Sheppard KA, Just EM, Kibbe WA, Chisholm RL (2011) dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa. Nucleic Acids Res 39:D620–D624
2. Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L, Platzer M, Noegel AA, Schaap P, Glöckner G (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res 21:1882–1891
3. Rot G, Parikh A, Curk T, Kuspa A, Shaulsky G, Zupan B (2009) dictyExpress: a *Dictyostelium discoideum* gene expression database with an explorative data analysis web-based interface. BMC Bioinformatics 10:265
4. Kuspa A (2006) Restriction enzyme-mediated integration (REMI) mutagenesis. Methods Mol Biol 346:201–209
5. King J, Insall R (2006) Parasexual genetics using axenic cells. Methods Mol Biol 346:125–135
6. Faix J, Kreppel L, Shaulsky G, Schleicher M, Kimmel AR (2004) A rapid and efficient method to generate multiple gene disruptions in *Dictyostelium discoideum* using a single selectable marker and the Cre-*loxP* system. Nucleic Acids Res 32:e143
7. Kimmel AR, Faix J (2006) Generation of multiple knockout mutants using the Cre-*loxP* system. Methods Mol Biol 346:187–199
8. Linkner J, Nordholz B, Junemann A, Winterhoff M, Faix J (2012) Highly effective removal of floxed Blasticidin S resistance cassettes from *Dictyostelium discoideum* mutants by extrachromosomal expression of Cre. Eur J Cell Biol 91:156–160
9. Sutoh K (1993) A transformation vector for *Dictyostelium discoideum* with a new selectable marker bsr. Plasmid 30:150–154
10. Sauer B (2002) Cre/lox: one more step in the taming of the genome. Endocrine 19:221–228
11. McMains VC, Myre M, Kreppel L, Kimmel AR (2010) *Dictyostelium* possesses highly diverged presenilin/gamma-secretase that regulates growth and cell-fate specification and can accurately process human APP: a system for functional studies of the presenilin/gamma-secretase complex. Dis Model Mech 3:581–594
12. Hoeller O, Kay RR (2007) Chemotaxis in the absence of PIP3 gradients. Curr Biol 17:813–817
13. Koch KV, Reinders Y, Ho TH, Sickmann A, Gräf R (2006) Identification and isolation of *Dictyostelium* microtubule-associated protein interactors by tandem affinity purification. Eur J Cell Biol 85:1079–1090
14. Liao X-H, Buggey J, Kimmel AR (2010) Chemotactic activation of *Dictyostelium* AGC-family kinases AKT and PKBR1 requires separate but coordinated functions of PDK1 and TORC2. J Cell Sci 123:983–992
15. Rosel D, Khurana T, Majithia A, Huang X, Bhandari R, Kimmel AR (2012) TOR complex 2 (TORC2) in *Dictyostelium* suppresses phagocytic nutrient capture independently of TORC1-mediated nutrient sensing. J Cell Sci 125:37–48
16. Veltman DM, Keizer-Gunnink I, van Haastert PJ (2009) An extrachromosomal, inducible expression system for *Dictyostelium discoideum*. Plasmid 61:119–125
17. Kuhlmann M, Popova B, Nellen WB (2006) RNA interference and antisense-mediated gene silencing in *Dictyostelium*. Methods Mol Biol 346:211–226
18. Robinson DN, Spudich JA (2000) Dynacortin, a genetic link between equatorial contractility and global shape control discovered by library complementation of a *Dictyostelium discoideum* cytokinesis mutant. J Cell Biol 150:823–838
19. Spann TP, Brock DA, Lindsey DF, Wood SA, Gomer RH (1996) Mutagenesis and gene identification in *Dictyostelium* by shotgun antisense. Proc Natl Acad Sci U S A 93:5003–5007
20. Bretsche MS, Clotworthy M (2007) Using single *loxP* sites to enhance homologous recombination: ts mutants in Sec1 of *Dictyostelium discoideum*. PLoS One 2:e724