KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree project, in Biotechnology
Second cycle 30 credits

Epigenetic Profiling of Canine Brain

Towards understanding the evolution of wolf to dog

ROXANNE CARLSSON NORLIN

Stockholm, Sweden 2022
www.kth.se

Table of contents

Abstract

The domestication of dog to wolf started around 35000 BP and is believed to be the oldest domestication event among both plants and animals. During this event, dogs have developed differences in morphological and behavioural traits to their ancestors, such as wider snouts, higher skulls, and lower tendencies to taking risks. It is now suggested that many of these differences can be explained by differences in active regulatory regions. The main objective of this thesis is to map chromatin interactions in the genome of wolf and dog brain tissues to annotate regulatory variants between the canine species. This will hopefully provide novel information regarding genomic changes mediating traits gained through domestication. We will perform Capture Hi-C (HiCap) on tissue samples of hypothalamus and prefrontal cortex of wolf and dog. HiCap is a method derived from the chromosome capture method 3C. In HiCap, the interacting regions of the DNA are crosslinked, ensuring that promoter-enhancer interactions will not be lost. These interacting regions are then ligated together, followed by sequencing library preparation. Subsequently, sequencing of these libraries will provide information of which promoters are actively regulated by enhancers in the nucleus. We successfully prepared Hi-C libraries for all tissues and animals. However, there were longer fragments in some libraries which can be removed. Due to lack of necessary probes for sequence capture, the laboratory work was cut short and no major results were obtained. By continuing the laboratory work, hopefully these libraries will result in novel insights in the domestication of the dog.

Keywords: Dog, wolf, domestication, evolution, gene regulation, HiCap, chromosome conformation capture, enhancers

Sammanfattning

Hundens domesticering är en av de äldsta och tros ha startats 35000 f.n. Under domesticeringen från varg till hund, har hunden utvecklat både morfologiska och beteendemässiga skillnader så som bredare snot, högre skallar och minskat risktagande. Många av dessa skillnader tros bero på skillnader i aktiva regulatoriska regioner mellan arternas genom. Syftet med detta examensarbete är att kartlägga genomiska interaktioner för hjärnvävnader hos både hund och varg för att identifiera regulatoriska skillnader mellan arterna. Förhoppningsvis kan detta leda till nya insikter i genomiska skillnader som utvecklats under hundens domesticering. För att jämföra arternas regulatoriska regioner användes metoden Capture Hi-C (HiCap) på vävnadsprover av både hypotalamus och prefrontala cortex för varg och hund. HiCap är en metod utvecklad från chromosome conformation capture-metoden 3C. I HiCap så fixeras interagerande delar av DNA:t så att promotor-enhancerinteraktioner förblir. Dessa interagerande regioner ligeras sedan samman och biblioteksbereds för sekvensering. Genom sekvensering fastställs vilka promotorer som aktivt regleras av enhancers i cellkärnan. Hi-C bibliotek förberedes för alla vävnader för båda arterna. I vissa av biblioteken upptäcktes längre DNA-fragment som kan renas bort.
På grund av avsaknad av probes för sequence capture så kunde laborationen ej fullföljas och därmed inga särskilda resultat erhållas. Om laborationen fullföljs kan resultaten förhoppningsvis ge nya epigenetiska insikter i hundens domesticering.

Nyckelord: Hund, varg, domesticering, evolution, genreglering, HiCap, chromosome conformation capture, enhancers

Acknowledgements

I would like to express my gratitude and appreciation for my supervisor professor Pelin Sahlén who has guided me through this project with her expertise and passion for genetics. I would also like to thank my co-supervisor Jörg Bachmann for all his support, help and encouragement. Furthermore, I would like to thank my family and friends for their support throughout this project.

1. Introduction

1.1 Background

1.1.1 Domestication of the dog

The dog (*Canis familiaris)* is believed to be the first species to be domesticated among both plants and animals with the domestication event starting 35000 BP in the Middle East (Galibert, Quignon, Christophe, & Catherine, 2011). Dogs today present a vast variety in morphological and physiological differences; however, they all share the same ancestor, the Gray wolf, *Canis lupus* (Perri, 2016). Dogs, in comparison to wolves, are smaller, have wider snouts, higher skull and are less prone to taking risks, among many other differences (Janssens, Perri, Crombe, & Lawler, 2019) (Pescini-Marshall, Besserdich, Kratz, & Range, 2016). Despite their differences, only 2% of the mitochondrial DNA differ between dogs and wolves (Vila & Savolainen, 1997). There are many theories for the evolutionary mechanism that led to the domestication of the dog in their coexistence with humans, such as population-based selection and human-dog co-evolution (Miklósi, Ádám, 2015). The theory of population-based selection states that humans' food waste provided an easy food source which wolves exploited. These wolves scavenging of human's food waste, started isolating themselves from hunting wolves which by genetic drift led to their genetic divergence (Miklósi, Ádám, 2015) (Serpell, 1995). Unlike Population-based selection, the theory of human-dog co-evolution suggests that both species has impacted each other's evolvement. This theory states that due to wolves' superior smell, humans have relied on canines' smell which has enabled their skulls to develop more advanced speech (Paxton, 2000). Independent of theory, wolves have coexisted with humans and their evolution can provide insight into the human evolution.

A variety of genetical studies have addressed the genetical differences that has resulted from the evolution from wolf to dog. The main focus of these studies has been on the differences in coding sequences of the genomes of the species (Caniglia, et al., 2018) (Krzeminska, Nowacka-Woszuk, & Switonski, 2022) (Saetre, et al., 2004). In humans, less than 2% of the genome consists of coding regions (Alexander, Fang, Snyder, & Gerstein, 2012), however, it is believed that 80% of the genome has some sort of biochemical function (Anshul, Patrick, & Khatun, 2012). The noncoding regions of the genome provides functions such as transcription regulation, chromosome structure and replication (Palazzo & Ryan, 2014). There are different classes of regulatory noncoding DNA, two of them being enhancers and promoters, which are the main focus of this research project (Strachan, Goodship, & Chinnery, 2015).

It is believed that epigenetic differences in regulatory regions of wolf and dog may play a big role in their morphological and physiological differences. The study of these differences along with their functional properties could provide further insight into the domestication of

the dog (Sahlén, et al., 2021). One previous study by P. Sahlén et al. found enriched pathways for dogs, such as in oxytocin signaling, carbohydrate digestion, cancer risk as well as other phenotypical differences (Sahlén, et al., 2021). These findings support a role of noncoding regulatory elements in the domestication of the dog.

For this research project which aims at comparing the DNA regulation of dog and wolf brain, prefrontal cortex and hypothalamus are in focus, as both play a big role in the behaviour of the canine. The hypothalamus is one of the oldest parts of the brain and controls basic life functions such as energy metabolism, digestion, energy expenditure, sleeping patterns, metabolic control, emergency responses to stressors and reproduction (Clifford & Bradford, 2014). The prefrontal cortex controls cognitive functions such as "attention, impulse inhibition, prospective memory and cognitive flexibility" (Murray & Nowicki, 2020).

The brain tissues collected from wolf are taken from Swedish zoos, the dog tissues are mongrels collected from Hungary. Mongrels are free-breeding dogs who have not been subjected to breeding which has resulted in their genetical structure being formed by natural selection. Mongrels are therefore well suited to study early domestication events of the dog which is why tissue samples from mongrels are used for this research project. However, the term mongrel is also used for mixed-breed dogs that can have pure-breed ancestors (Pilot & Malewski, 2015). Mongrel dogs defined in this way are not as suited for evolutionary studies since their genetic structure has actively been influenced by breeding.

1.1.2 Epigenetics

Enhancers upregulate gene transcription and are able to interact with promoters regardless of their orientation. Enhancers can be located up to 1 Mbp away from the gene that they regulate which can make them hard to locate. Due to this distance, the DNA will create a loop with the enhancer to reach spatial proximity to the transcription starting site. This allows for interaction of the two regions and their bound proteins. An insulator called CTFT helps ensure interaction between the correct enhancer and promoter by blocking wrongful interactions (Strachan, Goodship, & Chinnery, 2015, pp. 152-153)

Enhancers are cis-acting regulatory elements, meaning that they exclusively act within the same DNA molecule which they reside in (Strachan, Goodship, & Chinnery, 2015, pp. 42, 150). Cis-acting regulatory elements are more prone to rapid evolution, making them play a big role in evolution and interesting targets when studying evolution from a genetic perspective (Strachan, Goodship, & Chinnery, 2015, p. 42). A mutation in an enhancer will only affect the level of transcription of the gene it upregulates and not the actual function of the gene, which often makes for a more subtle variation than a mutation in a coding region. Mutations in enhancers will therefore more likely not generate a disruptive change but more gradually modify the organism (Klein, Keith, Agarwal, Durham, & Shendure, 2018).

Since enhancers are cis-regulatory elements which interact with promoters through genome looping, they can successfully be identified by chromosome conformation capture methods (Belton, et al., 2012). Chromosome conformation capture is an effective method to identify differences in regulatory regions in tissues and cells (Mifsud, Tavares-Cadete, & Young, 2015).

1.1.3 Chromosome conformation capture technique

Chromosome conformation capture is a technique used to study the three-dimensional arrangement of chromatin in the nucleus and with that, interactions within the chromosome. One common chromosome conformation capture method is 3C which has been the precursor of many other capture methods such as 4C, 5C, Hi-C and HiCap (de Laat & Dekker, 2013) (Sahlén, et al., 2021).

1.1.3.1 3C: Chromosome conformation capture

In 3C, the structural arrangement of the chromatin is fixated by formaldehyde. Formaldehyde crosslinks proteins with both proteins and DNA, fixating DNA regions through covalently bonded protein bridges (Gavrilov, et al., 2009) (Han, Zhiliang, & Wang, 2018). This will ensure that the chromatin-interaction pattern stays intact (Sahlén, et al., 2021). The chromatin is then digested by restriction enzymes followed by intramolecular ligation of crosslinked interacting regions. This proximity ligation is ensured by low DNA concentrations, favoring interaction of crosslinked, and thus spatially close, regions (Gavrilov, et al., 2009). Crosslinking is then reversed, and the interaction regions are analysed through PCR and/or sequencing (Gavrilov, et al., 2009).

Because 3C relies on PCR, the sequences to be analysed must be previously known for the design of PCR primers. This limits the amount of information which can be provided by this method. Since primers are designed for the regions of interest, only interactions between these regions of interest are detected, resulting in a “one vs one” detection. Another shortcoming of the method is that 3C has a limited range of a few hundred kb of where it can analyse interactions (Han, Zhiliang, & Wang, 2018).

1.1.3.2 4C: Circular chromosome conformation capture

4C is a chromosome conformation capture method based of 3C which generates genome-wide information at a high resolution (Matelot & Noordermeer, 2016). Unlike 3C, the method uses a microarray or next -generation sequencing and is capable of detecting interactions between the regions of interest and all other interaction regions. This makes 4C a “one vs all” detection method. After reversal of crosslinking, the ligated regions are further digested with a secondary digestion enzyme and ligated further to create circular DNA consisting of the two interaction regions (Han, Zhiliang, & Wang, 2018). Since 4C is using two rounds of

digestion enzymes, it will generate shorter sequences resulting in a higher resolution (Matelot & Noordermeer, 2016). During PCR, 4C uses bait-specific primers which enable this "one vs all" approach. Lastly, the resulting fragments, which are interacting with the region of interest are analysed using PCR or microarrays (Han, Zhiliang, & Wang, 2018).

4C can identify tens of thousands of interacting regions which is much higher than other 3C-derived methods (Matelot & Noordermeer, 2016). However, 4C experiences some amplification biases due to ineffective amplification of GC-rich fragments which will decrease the validity of the results (Han, Zhiliang, & Wang, 2018).

1.1.3.3 5C: Chromosome capture carbon copy

5C is capable of detecting interaction regions with no preselection of regions of interest and is therefore an "all vs all" detection method. This is due to 5C using a large library of primers containing a universal sequence. All interactions can therefore be screened for within one single assay, which eliminated the bias selection of primers both 3C and 4C exhibits. However, 5C is still limited in the size of the region it can analyse (Han, Zhiliang, & Wang, 2018).

1.1.3.4 Hi-C: High-throughput chromosome conformation capture

Hi-C, like 5C is an "all vs all" method that relies on high throughput DNA sequencing. In Hi-C, the sticky ends generated from the restriction enzymes are repaired using a biotinylated nucleotide which is followed by blunt-end ligation. The biotin in end-repaired, non-ligated fragments are removed using an exonuclease which prevents these fragments from being misinterpreted as consisting of interacting regions (Belaghzal, Dekker, & Gibcus, 2017). The DNA fragments are then sheared and purified (Han, Zhiliang, & Wang, 2018). Biotin enables purification and enrichment of the ligated crosslinked regions from the unwanted digested chromatin fragments using streptavidin coated magnetic beads (Belton, et al., 2012). This selection for biotinylated fragments will significantly reduce the number of unwanted regions for sequencing and data analysis. Finally, the DNA fragments will go through sequencing and sequence mapping to identify which regions are interacting (Belton, et al., 2012). However, in Hi-C the mapping resolution is dependent on the sequencing depth. This need for a high sequencing depth can be economically limiting (Han, Zhiliang, & Wang, 2018).

1.1.3.5 HiCap: Capture Hi-C

HiCap is based of Hi-C but has the addition of sequence capture of promoter regions and the substitute of a 4-cutter enzyme (Sahlén, et al., 2021).These changes to the Hi-C protocol results in a higher sensitivity and resolution. By using sequence capture, it is possible to select for the targets of interest, for example promotor sequences by hybridisation of the sequencing library to DNA or RNA probes. This enrichment step limits the number of

sequences needed for sequencing and with that, the sequencing depth required (Pradhananga, 2020).

By using a 4-cutter enzyme, the enhancer region is more precisely identified than with using a 6-cutter enzyme due to shorter fragments being generated (on average 422 bp). This increases the resolution of the regulatory regions being identified (Pradhananga, 2020) (Sahlén, et al., 2021).

1.2 Objective

The main objective of this degree project is to create a detailed sequence, epigenetic and regulatory map from wolf and dog brain tissues. More specifically, the enhancer and promoter regions of DNA from prefrontal cortex and hypothalamus will be mapped using Capture Hi-C (HiCap) along with high throughput sequencing. From this, regulatory epigenetic maps will be created as a tool to annotate regulatory variants between canine species. The biological function of these genetic regulatory differences will also be investigated in regard to the domestication and evolution of the dog.

2. Materials and methods

2.1 Sample collection

All four wolf individuals (*Canis lupus)* used in this project were euthanised for population control at Swedish zoos in 2019. Two of the samples (W1, female and W2, male) were collected from Orsa zoo, and two were obtained from Lycksele zoo (W21, male and W22, female). The tissue samples were dissected by a veterinarian surgeon and snap-frozen after dissection. From this sample collection, hypothalamus (Hyp) and prefrontal cortex (PFC) were used for this research project. These tissue samples were provided in larger pieces and then cut into smaller pieces at SciLifeLab in a cryostat.

The four Mongrel dog individuals used were street dogs collected and dissected in Hungary and provided by the Department of Ethology at Eötvös Loránd University. Two of the dogs were female (E and F) and two male (H and G), all were euthanised due to medical reasons in 2019-2021. The samples were dissected by the same surgeon and delivered precut in adequate sizes.

2.2 Nuclei extraction and fixation

The tissue samples were stored in Eppendorf tubes in freezers at -80°C in pieces of approximately 0.5 cm^3. On a given day, two samples of the same species and tissue type, from two different individuals, one female and one male, were processed in parallel. The tissues were thawed in 7 ml 1xPBS containing cOmplete Protease Inhibitor tablets (Roche) and homogenised using a dounce homogeniser with pestle A (~ 8 strokes). For wolf, two pieces of tissue were used for prefrontal cortex and one piece for hypothalamus to ensure adequate cell numbers. For dog, only one piece was used for all samples. After the debris of the homogenate (fat and bigger pieces of tissue) had sunk to the bottom of the homogeniser, the upper cell suspension (6 ml) was pelleted by centrifugation (600 g, 10 min). During these steps, the cell concentration of the homogenate, cell suspension and supernatant was checked using a cell counter (Countess 3 FL Automated Cell Counter).

The pellet was crosslinked by resuspension in 1% formaldehyde, 1xPBS solution (1 ml per million cells) at room temperature. The reaction was quenched after 10 min by incubation with glycine (0.125 M, 5 min) and the cells were collected by centrifugation (400 g, 5 min). Cell lysis buffer (10 mM NaCl, 10 mM Tris-HCl pH 8.0 and 0.20% Triton-X) was used to lyse the cells (1ml per million cells, 10 min incubation on ice) followed by centrifugation (~5 million cells per aliquot, 600 g, 5 min). The pellets were snap-frozen on dry ice and stored at -80°C.

2.3 Chromosome conformation capture

1X FastDigest Mbol digestion buffer was added to nuclei extracted pellets of ~5 million cells (550 μl). An aliquot of 50 μl was taken as an undigested control. For solubilisation of the chromatin, the sample was incubated in 20% SDS (final concentration 0.3%, 37°C, 950 rpm, 60 min). SDS was then quenched by addition of 20% Triton-X (to final concentration 2%) to ensure that it does not disturb the activity of the digestion enzyme in later steps. Additional FastDigest Mbol digestion buffer (10X) was added to compensate for the change in volume, followed by incubation (37°C, 950 rpm, 60 min). FastDigest Mbol enzyme (cut motif GATC) (6 μl) was added and digested the enzyme for 2 h (37°C, 450 rpm) after which the enzyme was deactivated by heat denaturation (75°C, 10 min). Two aliquots of the reaction were taken as digested control (50 μl) and 3C control (75 μl). The resulting 5' sticky ends of the enzymatic digestion were end-repaired using Klenow fragment (10U/μl), dTTP, dGTP, dCTP (1.2 μl, 10 mM) and biotin-14-dATP (30 μl, 10.4 mM). By using a biotinylated nucleotide, target fragments can be extracted using streptavidin-coated beads in later steps. After incubation (23°C, 450 rpm, 4 h), the end-repair reaction was quenched by addition of EDTA (9.6 μl, 0.5 M) and incubation (75°C, 10 min). The fragments which have been in spatial proximity were then ligated using T4 DNA ligase (50 Weiss units) by adding ligation buffer (150 μl), ATP (15 μl, 100 mM) and reaching a final volume of 500 μl by addition of ddH20. To the 3C control, 6 Weiss units of T4 DNA ligase was added along with ligation buffer (50 μl), ATP (5 μl, 100 mM) and ddH20 (396 μl). For both the digested and undigested control, ligation buffer (50 μl) was added along with the 400 μl ddH_20 to reach a final volume of 500 μl. The samples and controls were incubated overnight at 16°C.

After the spatially close fragments had been ligated, the crosslinking was removed by incubation with Proteinase K at a high temperature (final conc. 3.58 mU/ μl, 65°C, 6 h). The DNA was then purified by two cycles of purification using phenol-chloroformisoamyl alcohol (PCI) (25:24:1) (1:1 ratio of PCI to sample, centrifugation at 16000 g, 5 min). The purified samples were then precipitated by overnight incubation (-20°C) in a mixture of 100% Ethanol (2.5X v/v), NaOAc (0.1X (v/v), pH 5.2) and glycogen (1 μl).

The precipitate was pelleted by centrifugation (4°C, 16 000 g, 20 min) resuspended in 70% ethanol (200 μl) and pelleted (4°C, 16 000 g, 5 min). To remove any remaining ethanol, the samples were air dried (10 min) before resuspension in 100 μl of ddH2O (30 μl was used for controls). This was followed by purification from RNA using RNAse A (1 μl for Hi-C, 0.2 μl for controls, 37°C, 1h).

The quantity of the samples and controls were measured using a Qubit fluorometric quantification (Invitrogen) with the Broad Range dsDNA kit and the quality of the samples and controls were analysed using 2100 Bioanalyzer systems (Agilent).

2.3.1 Mock Hi-C

Since biotin-14-dATP is costly, "mock" Hi-Cs were performed on wolf hypothalamus and prefrontal cortex using dATP to see if the protocol was suitable for these tissues before using biotin-14-dATP.

2.4 Library preparation

Library preparation was performed on four samples in parallel. When possible, the starting material was 5 μg of chromosome captured DNA. For the samples containing less than 5 μg, all DNA was used, at a minimum amount of 3 μg, see Appendix B, Table 5.1 for initial DNA amounts used. To reach a sample volume of 78.25 μl with a 5 μg DNA content, some samples were dried using a speedvac and resuspended, and to some samples, ddH_20 was added.

2.4.1 Biotin removal from unligated ends

Firstly, biotin was removed from unligated, end-repaired ends by using T4 DNA polymerase (2.5 U). T4 DNA polymerase has a strong 3'-5' exonuclease activity which was used for this purpose. To favour this activity, dATP and dGTP (5 nMol) was added along with BSA (5 ug) and 5X T4 DNA polymerase Buffer (20 μl) followed by incubation (12°C, 15 min). The reaction was quenched by addition of EDTA (1 μMol) and the DNA was purified by two cycles of PCI extraction. The purified samples were then precipitated by overnight incubation (-20°C) in a mixture of 100% Ethanol (2.5X v/v), NaOAc (0.1X (v/v), pH 5.2) and glycogen (0.4 μl). The precipitate was pelleted (4°C, 16000 g, 20 min), washed in 70% ethanol and air dried (10 min) followed by resuspension in ddH_20 (100 μl). The DNA concentrations were measured using a Qubit fluorometric quantification (Invitrogen) with a Broad Range dsDNA kit.

2.4.2 Sonication

3 μg of DNA was taken for sonication (if total DNA count was lower, all DNA was taken) see Appendix C, Table 5.2 for initial DNA amounts used for sonication. The sonication was performed using Covaris Ultrasonicator S2 along with the SonoLab software (Duty cycle:10%, Intensity:5, Cycles per burst: 200, Time: 5 cycles of 40 s, Set mode: frequency sweeping, Temp: 4-7°C).

Purification of the DNA fragments was performed on the sonicated fragments using AMPure XP beads (Beckman Coulter). AMPure XP beads (180 μl) were mixed with each sample of 130 μl in tubes and incubated (room temperature, 5 min) to let the fragments bind the beads. The tubes were then placed on a magnetic rack (5 min) and the fragments bound to the beads were washed in two cycles of 80% ethanol (200 μl per sample). After purification, the samples were dried while still on the beads (5 min) and then released from the beads by

addition of ddH_20 (50 μl per sample). After a two-minute incubation on a magnetic rack in room temperature, the supernatants containing the samples were collected (approximately 48 μl per sample).

The samples were then quantified using Qubit fluorometric quantification (Invitrogen) with a High Sensitivity dsDNA kit and then analysed using 2100 Bioanalyzer systems (Agilent) see Table 5.3 in Appendix D and Figure 3.4 in Results for results.

2.4.3 Selection of biotin-labelled fragments

The remaining steps of the library preparation were performed on eight samples in parallel. Dynabeads T1 (200 μl per sample) were washed twice in 400 μl of Tween 1X Wash Buffer (5.04 mM Tris-HCl pH 8.0, 1.008 M NaCl, 0.504 mM EDTA, 0.5 vol‰ Tween). The sonicated and size selected DNA fragments (48 μl per sample) were diluted in 102 μl ddH_20 each and then bound to 1.7 ml of streptavidin beads (Dynabeads T1) by 15 min incubation in 1X No Tween Wash Buffer (5 mM Tris-HCl pH 8.0, 1M NaCl, 0.5 mM EDTA). This was followed by two wash cycles of 1X No Tween Buffer (400 μl) to wash away unbound fragments and other unwanted components. To prepare for ligation, the beads were then washed in 1X T4 DNA ligation buffer (200 μl) and resuspended in 52 μl of ddH_20.

2.4.4 End-repair and dA-tailing

16 μl of End Repair-A Tailing Buffer mixed with 4 μl of End Repair-A Tailing Enzyme Mix was added to the Dynabeads T1 and the samples were run in a thermocycler (20°C for 15 min, 72°C for 15 min, 4°C hold). The end-repair will create blunt ends from the sticky ends generated from the biotin removal of unligated ends. This allows for ligation of the dA-tails to the DNA fragments.

2.4.5 Adaptor and primer ligation

Adaptors, for primers to be added later, were ligated to the fragments by addition of 5μl SureSelect XT HS2 Adaptor Oligo Mix along with 25 μl of Ligation master Mix (23 μl 10X T4 DNA Ligation Buffer, 2 μl T4 DNA Ligase). Prior to this, the ligation master mix had been incubated at room temperature for 30 to 40 minutes to active the T4 DNA ligase. The reaction was performed in a thermocycler (20°C for 30 min, 4°C hold). The beads were then washed in two cycles of 1X No Tween Buffer (150 μl) and one cycle of 1X low EDTA buffer (150 μl) and resuspended in 34 μl of ddH_20.

To be able to distinguish between the samples after pooling in later steps such as sequencing, the samples were indexed using unique dual-index primers. These unique primers also contain universal primers for sequencing and PCR amplification. These primers (SureSelect XT HS2 Index Primer Pair) were added to the samples along with PCR reaction mixture (1:10 of Herculase II Fusion DNA Polymerase : 5X Herculase II Buffer with dNTPs) and run in a thermocycler (1 cycle of 98°C 2 min, 8-9 cycles of 98°C 30 s, 60°C 30 s, 72°C 1 min, 1 cycle of 72°C 5 min followed by 4°C hold).

2.4.6 PCR clean up

The supernatant (50 μl), containing the amplified DNA was taken from the bead suspension and mixed with AMPure XP beads (50 μl) for a 1:1 ratio of sample volume to bead volume. The DNA bound to the beads during a 5 min incubation. Any unwanted reagents and fragments were removed through two cycles of wash using 80% ethanol. To release the DNA from the beads, the beads were dried (37°C 1-2 min) and resuspended in 30 μl ddH_20 and the supernatant was taken for further steps. The products were quantified using Qubit fluorometric quantification (Invitrogen) with the High Sensitivity dsDNA kit and the ssDNA kit as well as analysed using 2100 Bioanalyzer systems (Agilent) for results see Table 3.2 and Figure 3.5.

2.4.7 Remaining steps of library preparation

After PCR clean up, the DNA libraries would be hybridised to probes. However, due to incoherent annotations in the dog genome used a reference, these probes were not able to be properly designed within the time scope of this project. The following steps of the capture Hi-C which were not performed are mentioned in 4.3 Future perspectives.

3. Results

3.1 Results from nuclei extraction and Hi-C

3.1.1 Bioanalyzer results from chromosome conformation capture

3.1.1.1 Successful Hi-C

Figure 3.1 shows the bioanalyzer results from a successful Hi-C performed on prefrontal cortex from Dog E. The Hi-C sample in graph A shows a clear shift in comparison to the digested sample in graph C. This increase in fragment length indicates that the end-repair and blunt-ligation has been successful and thus, that the Hi-C has worked properly. By comparing the undigested and digested controls in graphs B and C, it is evident that the restriction digestion also has been effective. In graph C, there is a high peak at 1627 bp of digested fragments, while in graph B, due to high molecular-weight genomic DNA, no clear peak is visible. For the 3C control, no end-repair has been performed, making the ligation a sticky-end ligation. Since sticky-ends are more prone to ligation, more fragments will have been ligated together, making the fragments longer than for Hi-C. This shift is clear when comparing graphs A and D with fragment lengths peaking around 7000bp and 13000bp respectively. For successful Hi-C bioanalyzer results for all samples, see Figure 5.1 in Appendix A.

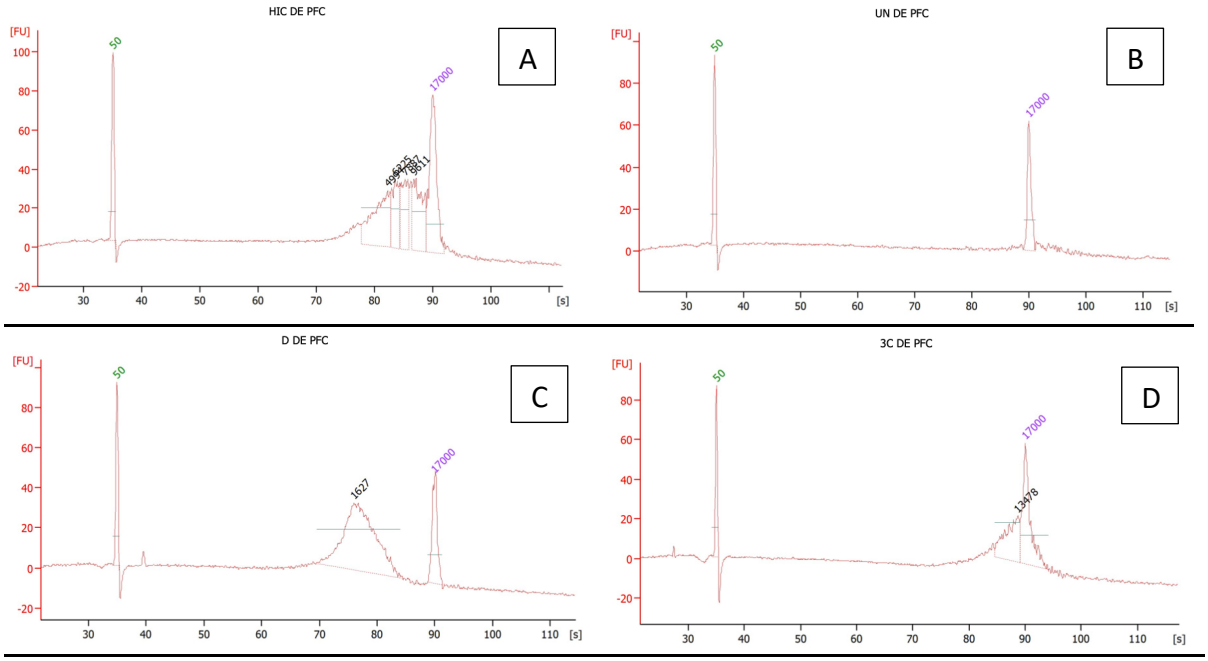


Figure 3.1: The four graphs are obtained from 2100 Bioanalyzer systems (Agilent) run on sample Dog E prefrontal cortex. Graph A depict the Hi-C sample and graph B, C and D are the three controls of undigested DNA, digested DNA and 3C control. The upper and lower markers indicate at 50 bp and 17000 bp.

3.1.1.2 Hi-C with smaller shift

For Wolf 1 hypothalamus, two Hi-Cs were performed to reach approximately 3 μg of DNA as starting material for library prep, see Table 5.1 in Appendix B. The results from the bioanalyzer for the second run can be seen in Figure 3.2. This was the run with the smallest shift between Hi-C and the digested control with peaks around 2000 bp and 1081 bp, as can be seen in graph E and G. This decrease in ligation efficiency might be due to a smaller volume of ligase being added or some other human error. Due to time constrains, this sample was not rerun. It was pooled together with the first run of Wolf 1 hypothalamus which had a better shift (see Appendices A Figure 5.1). Both Hi-Cs on Wolf 1 hypothalamus were performed on the same nuclei extracted pellet which further emphasises that human error may have cause this lower shift. This run was performed in parallel with other successful Hi-Cs which indicates that it was an isolated error.

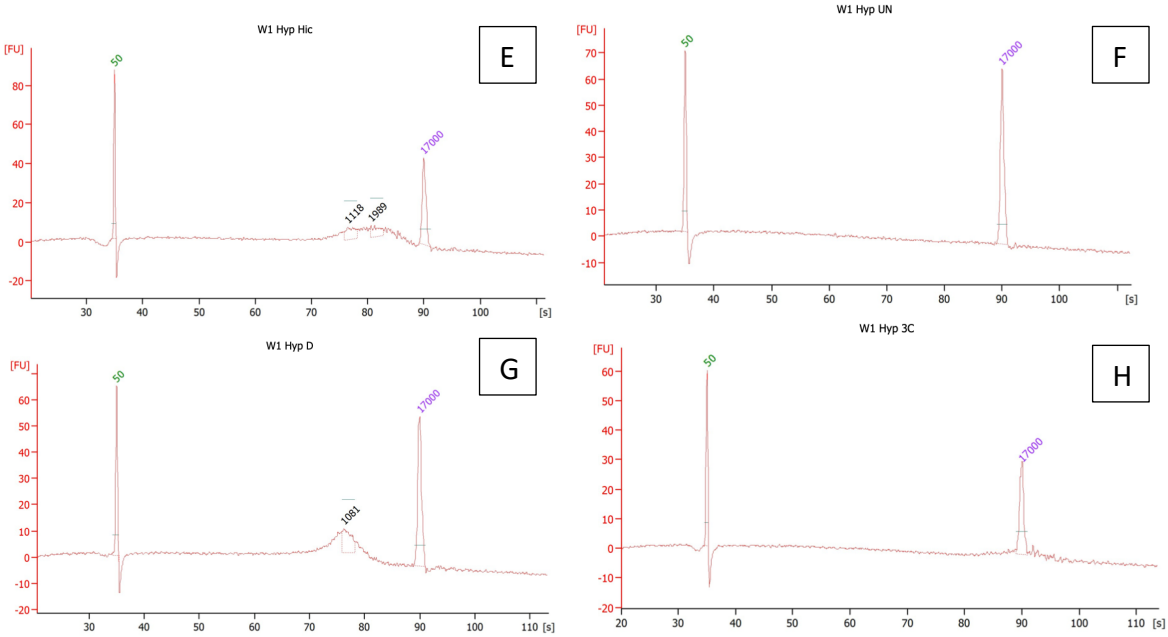


Figure 3.2: The four graphs are obtained from 2100 Bioanalyzer systems (Agilent) for the second Hi-C performed on sample Wolf 1 hypothalamus. Graph E depict the Hi-C sample and graph F, G and H are the three controls of undigested DNA, digested DNA and 3C control. The upper and lower markers indicate at 50 bp and 1700 bp. Due to erroneous ladder, x-axis could not be set to bp and is displayed in seconds.

3.1.1.3 Overloaded Hi-C

For one Hi-C performed on Wolf 2 prefrontal cortex, the Hi-C did not work, see Figure 3.3. In graph I, the Hi-C results show a peak around 2000-3000 bp and in graph K, the digested control shows a peak at 1947 bp. As there is almost no difference in fragment length between these two graphs, the ligation of the digested fragments seems to have failed.
As can be seen in Table 5.1 in appendix B, the starting material for this Hi-C was almost 6 times higher than for the other Hi-Cs which could explain these poor results. The Hi-C was probably overloaded meaning that there was not enough ligase or Klenow Fragment for this number of cells to allow for proper end-repair and ligation.

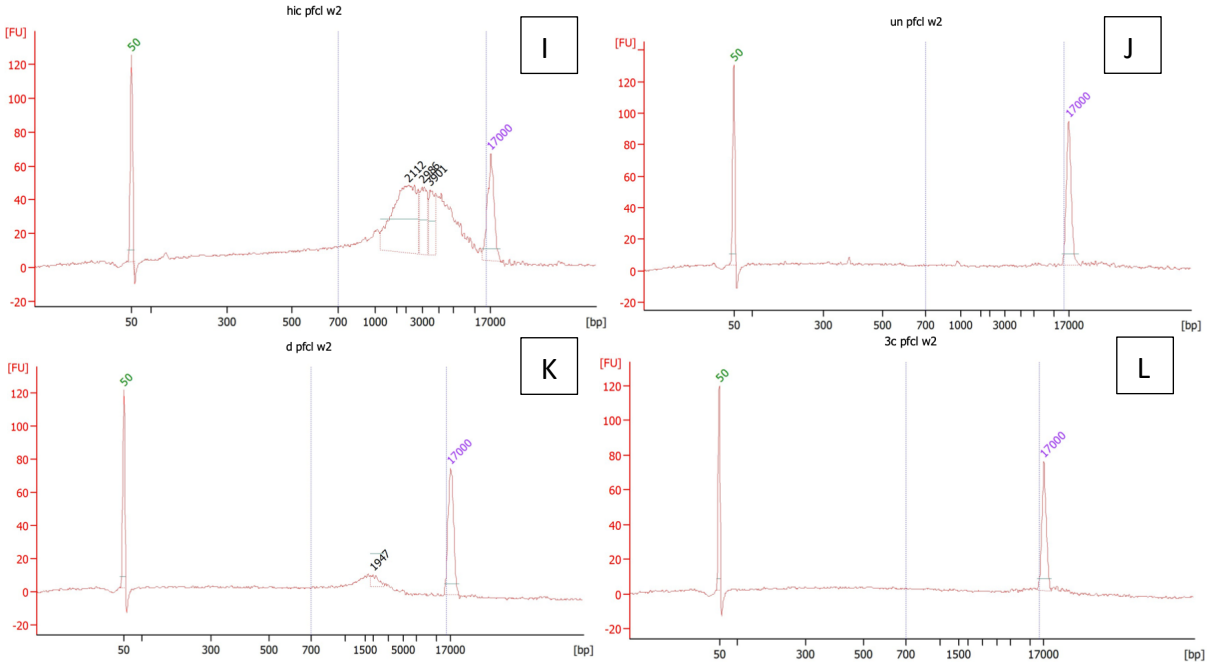


Figure 3.3: The four graphs are obtained from 2100 Bioanalyzer systems (Agilent) for Hi-C performed on sample Wolf 2 prefrontal cortex. Graph I depict the Hi-C sample and graph J, K and L are the three controls of undigested DNA, digested DNA and 3C control. The upper and lower markers indicate at 50 bp and 17000 bp.

3.1.2 Data from nuclei extraction and chromosome conformation capture

As can be seen in Table 3.1, the tissue samples from dog generated generally more DNA than wolf tissue samples. The average ng of DNA extracted per million cells for dog are 902.0 ng and 3100 ng for prefrontal cortex and hypothalamus. For wolf, the same values are 626.3 ng and 446.2 ng. For dogs tissue samples, hypothalamus generated higher amounts of DNA than prefrontal cortex, however, for wolfs the opposite applies. Dog G has the lowest DNA content of the dog tissues for both hypothalamus and prefrontal cortex. See Table 5.1 in Appendix B for the data from which these averages are calculated.

For samples with a DNA amount lower than 3 μg, an additional Hi-C was performed to reach a total concentration of 3 μg or higher. For the initial amounts of DNA used for library prep, see Table 5.1 in Appendix B.

Table 3.1: Average initial cell concentrations and average DNA amounts extracted per species and tissue. These values were calculated based on Table 5.1 in Appendix B. Values for samples with an initial cell concentration higher than 6.2 million have been excluded from these calculations as well as Wolf 21 hypothalamus where 1/3 of the sample was lost.

Species	**Tissue**	**Average initial cell concentration (10^6 cells)**	**Average DNA amount extracted (ng)**	**Average DNA amount extracted per million cells (ng)**	**Average cell viability (%)**
Wolf	Prefrontal cortex	5.02	3144	626.3	39.0
Wolf	Hypothalamus	5.60	2498	446.1	73.9
Dog	Prefrontal cortex	5.37	4844	902.0	59.5
Dog	Hypothalamus	5.50	17050	3100	29.3

3.2 Library preparation

3.2.1 Biotin removal of unligated ends

During biotin removal of unligated ends, 8-49% of the sample amount was removed, for DNA concentrations measured before and after biotin removal see Table 5.2 in Appendix C.

3.2.2 Sonication and AMPure XP beads purification

Figure 3.4 shows the bioanalyzer results after sonication and AMPure XP bead purification. All samples have fragment lengths ranging from 100-1000 bp, which is a clear decrease in size from the initial length peaking around 7000 bp, see Figure 3.1. This indicates that the shearing by sonication has been successful. Smaller fragments such as primer dimers has been removed through purification using AMPure XP beads. For DNA amounts measured after sonication and AMPure XP bead purification, see Table 5.3 in Appendix D.

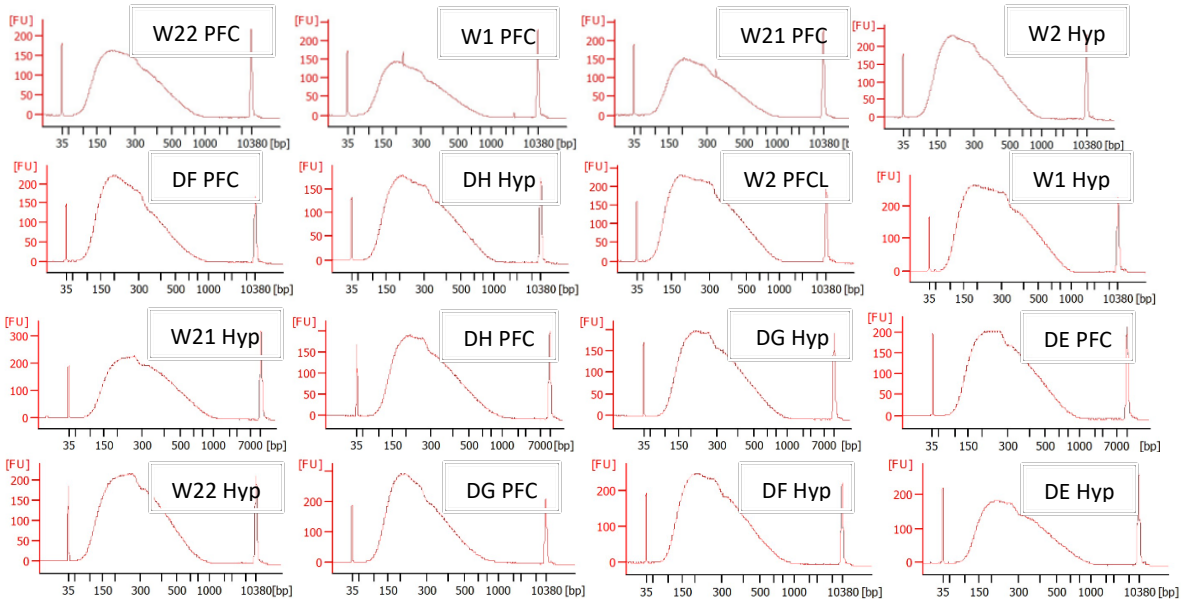


Figure 3.4: Bioanalyzer graphs performed on samples after sonication and AMPure XP bead purification. For DNA content for each samples, see Table 5.3 in Appendix D. All graphs are fitted to the same x-axis ranging from 35 bp to 10380bp. Due to limitations of the bionalyzer software, some of the graphs wrongfully seem to range from 35 bp to 7000 bp, however the 10380 marker peak clearly indicated the range in all samples. D stands for dog sample and W for wolf. PFC is an abbreviation of prefrontal cortex and Hyp is short for hypothalamus.

3.2.3 Final results of library preparation

As can be seen in Figure 3.5, seven of the libraries display fragment lengths of 200-1000 bp which are expected (samples referred to: DE PFC, W21 Hyp, DG Hyp, DH PFC, W1 Hyp, W2 PFC and W22 PFC). The remaining nine samples also display longer unexpected fragments with lengths up to 10 000bp.

As can be seen in Table 3.2, the DNA content provided by eight cycles generates, as expected, lower DNA amounts than nine cycles. Four of the measured DNA amounts are 3.36 μg or over. This corresponds to a concentration of 120 ng/ μl which is the highest concentration the Qubit fluorometric quantification (Invitrogen) can measure using the High Sensitivity dsDNA kit. These DNA amounts are probably not identical, but the upper measurement threshold resulted in them being estimated to the same concentration. However, the concentrations are probably close to the measured values.

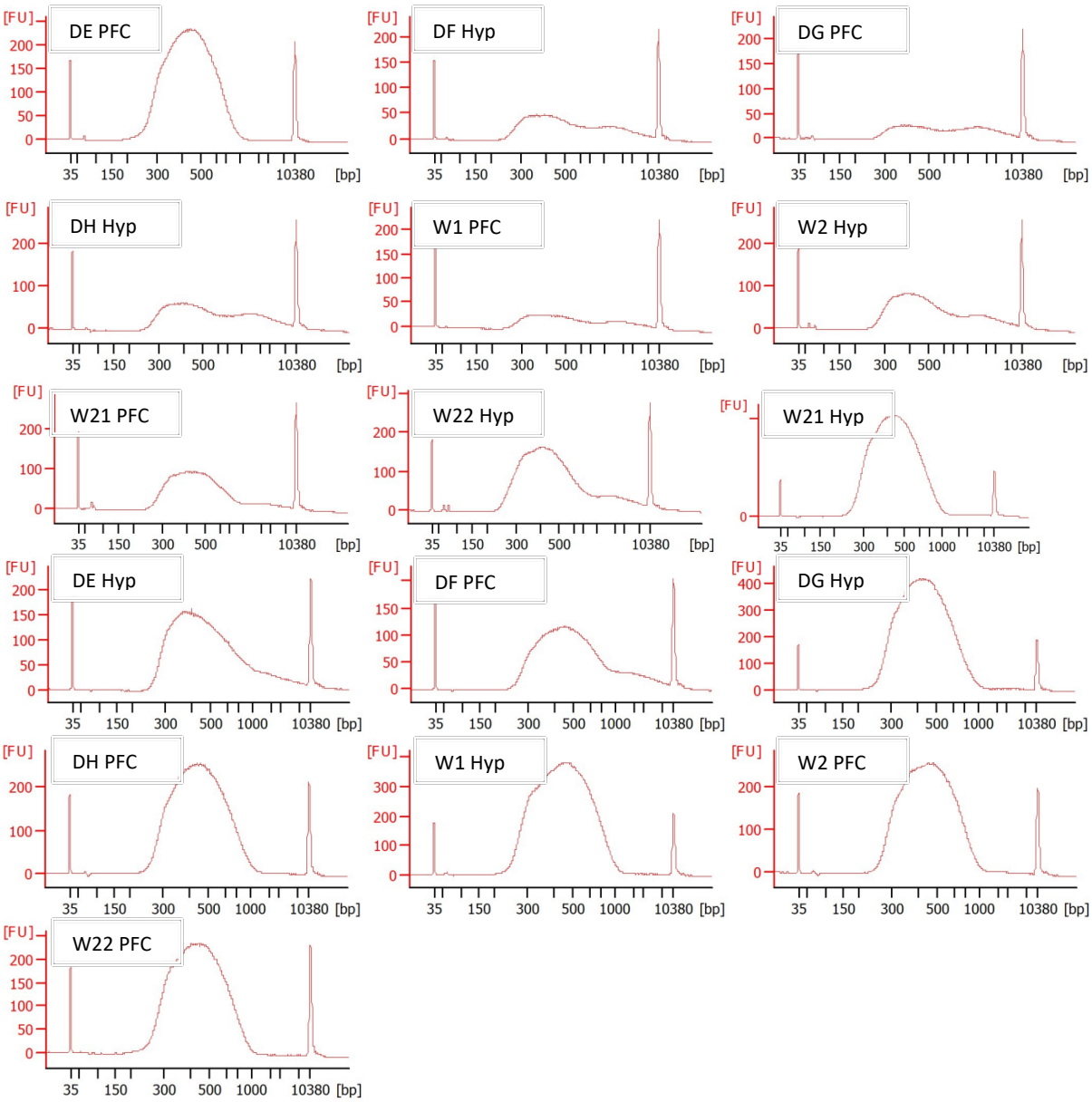


Figure 3.5: Bioanalyzer results from library preparation. Seven of the samples display expected fragment lengths between 200-1000 bp. Nine of the samples also display longer, unexpected fragments.

Table 3.2: DNA amounts of the final product of the library preparation measured using both a dsDNA kit and a ssDNA kit. The adaptor index assigned to each sample and the number of cycles used using amplification are also displayed.

Sample	Tissue	dsDNA amount after library preparation (μg)	ssDNA amount after library preparation (μg)	Cycles during amplification	Adaptor index assigned
Wolf 1	Prefrontal cortex	0.90	47.2	8	13
Wolf 1	Hypothalamus	3.36	23.7	9	5
Wolf 2	Prefrontal cortex	>3.36	16.4	9	6
Wolf 2	Hypothalamus	0.87	61.6	8	14
Wolf 21	Prefrontal cortex	0.82	61.6	8	15
Wolf 21	Hypothalamus	3.36	26.0	9	7
Wolf 22	Prefrontal cortex	1.18	9.9	9	8
Wolf 22	Hypothalamus	1.35	9.7	8	16
Dog E	Prefrontal cortex	2.16	18.0	8	9
Dog E	Hypothalamus	2.78	14.6	9	1
Dog F	Prefrontal cortex	2.80	9.8	9	2
Dog F	Hypothalamus	0.91	7.1	8	10
Dog G	Prefrontal cortex	0.60	4.9	8	11
Dog G	Hypothalamus	3.36	23.2	9	3
Dog H	Prefrontal cortex	3.08	12.5	9	4
Dog H	Hypothalamus	0.83	5.8	8	12

4. Discussion

4.1 Conclusions from HiCap and library preparation

4.1.1 Unexpected tails in bioanalyzer results from library preparation

Successful chromosome conformation capture was performed on all 16 samples as can be seen in the clear shift in the bioanalyzer results of the Hi-C samples in Figure 5.1 in Appendix A. The bioanalyzer plots after sonication of the samples shows fragment sizes of 100-1000 bp which indicates successful shearing, see Figure 3.4. The results after adaptor and primer ligation can be seen in Figure 3.5. As mentioned, seven of these graphs display the expected fragment lengths after adaptor and primer ligation, however, nine of them have an upper tail which was not expected. These tails could be single stranded fragments as single stranded fragments migrate more slowly on the gel and will therefore wrongfully appear as longer double stranded fragments (Stellwagen & Stellwagen, 2009). To check this hypothesis, a Qubit analysis was performed using a kit for ssDNA, these values can be seen in Figure 3.2. The ssDNA kit is not capable of differentiating dsDNA from ssDNA and will therefore measure the presence of both these molecules (Scientific, 2022). The ssDNA kit used was expired and the standards used for calibration belonged to another ssDNA kit. Therefore, these values are not very reliable, but they do provide some indication of the DNA amounts in the samples. As the DNA amounts are estimated much higher when using the ssDNA kit than for the dsDNA kit, it is possible to conclude that there are ssDNA fragments present in the product. Thus, strengthening the hypothesis that these tails in the graphs consists of ssDNA fragments. These ssDNA fragments are possibly products from failed hybridisation and reannealing during PCR amplification. One way to possible reduce the occurrence of ssDNA fragments would be to run an additional PCR cycle for the samples. This can also be seen by comparing the library preparation batch which had nine cycles to the batch using eight cycles. For nine cycles, two of eight samples display an upper tail in their graphs. For the batch using eight cycles, seven of eight samples had this tail. However, by increasing the number of cycles, the complexity of the samples would be decreased which could result in the loss of unique interactions in the chromosome. Another way to remove these ssDNA would be to perform an AMPure XP bead separation.

4.1.2 Varying DNA concentrations extracted from wolf and dog

As mentioned in *3.1.2 Data from nuclei extraction and chromosome conformation capture* the tissue samples from dog generated higher amounts of DNA than wolf tissue samples. This could be due to differences between the species, perhaps dog tissues generate more DNA or are more easily lysed. However, it is more likely that it is due to differences in handling of the tissues. To see if the freshness of the samples has had an impact on these values, their storage time and level of DNA were compared. The wolfs were euthanised in 2019 and the dogs in 2019-2021 (Dog H in 2019, Dog E in 2020 and Dog F and G in 2021). By comparing the dogs' tissues storage time and DNA content, no correlation could be seen as Dog H is the oldest sample but still generated more DNA than Dog G which is one of the newest tissues. The most likely explanation would be that this difference in the tissues is due to thawing and freezing. As the wolf tissues were delivered in bigger pieces, they have been subject to thawing (from -80°C to -20°C) when being cut in the cryostat. This additional cycle of

thawing and freezing could have decreased the quality of the tissues which could have impacted the efficiency of the nuclei extraction of the samples. This external factor would also explain why there is no pattern of which tissue type generates the highest amount of DNA.

By comparing the average cell viability with the DNA amounts, no correlation is seen, see Table 3.1. This would argue that the state of the tissue samples did not impact the DNA extracted from the tissues. However, during measurement of cell concentration, the cell suspension might not have been completely homogeneous which may have contributed to this variation in both DNA amount extracted per million of cells and estimate of cell viability, seen in Table 5.1 in Appendix B.

Another contributing factor could be that since only one small piece of each dog tissue was delivered for each sample, these tissues might have been chosen with more care than the tissue pieces randomly cut from the larger wolf tissue samples. If a more central piece of the prefrontal cortex and hypothalamus was chosen for the dog it could possibly explain why these pieces had a greater concentration of both cells and DNA.

4.2 Optimisations of protocol

4.2.1. Increased volumes for nuclei extraction

Initially, 10 ml of Cell lysis buffer and 1% formaldehyde 1xPBS buffer was prepared per sample during the nuclei extraction. According to the protocol, 1 ml per million cells should be added for both these solutions. However, since the tissue samples often generated higher cell concentrations than assumed during preparation of buffer solutions, there was often not enough solution to add 1 ml per million cells. These volumes were therefore increased by 50% to provide for a better nuclei extraction. To see which nuclei extractions were performed with increased buffer volumes, see Table 5.1 in Appendix B. By increasing these volumes, the cell lysis and chromosome fixation might be more efficient for tissue samples with higher cell concentrations. This optimisation was only applied on dog samples, as they were processed after the wolf samples, and might contribute to the increased DNA content provided from the cells. If the cells were lysed to a greater extent, more DNA might have been released from the nucleus, resulting in higher DNA concentrations. However, the nuclei extractions of Dog E Hyp and Dog G Hyp did not have increased buffer volumes but still provided higher DNA amounts than some other dog nuclei extracted with this optimisation. Thus, this increase in buffer solution might not have affected the DNA contents provided from the tissues. Perhaps the initial volume of cell lysis buffer was in such an excess that it could properly lyse higher cell concentrations.

4.2.2 Increased volume of Klenow Fragment

The end-repair performed by the Klenow Fragment is a crucial step in the Hi-C. Poor end-repair will result in fewer fragments being successfully ligated during blunt end ligation and important interactions being lost during reversal of crosslinking. To increase the efficiency of the end-repair, the addition of Klenow Fragment was increased from 1.2 µl to 2.5 µl. To see to which samples this optimisation was applied, see Table 5.1 in Appendix B. As mentioned in *3.1.1.3 Overloaded Hi-C*, the Hi-C performed on Wolf 2 PFC had a very poor shift, indicating a failed ligation event which was probably caused by a large initial number of

cells. Perhaps this Hi-C would not have been overloaded if this optimisation would have been applied to it. With an increase addition of Klenow Fragment, more fragments would have been end-repaired which could have resulted in a higher ligation efficiency.

4.2.3 Decreased number of cycles for amplification

During adaptor ligation, the samples are amplified to ensure adequate DNA amounts for future hybridisation and probe capture. The library preparation was performed in two rounds with eight samples in each round. For the first round, nine cycles were used for the amplification. As can be seen in Table 3.2, nine cycles generated 1.18-3.36 μg of DNA which is more than the 0.5-1 μg which is needed. As mentioned, overamplification reduces the complexity of the sample which can result in loss of sensitivity. Thus, the amount of cycles were reduced to eight which generated DNA amounts of 0.6-2.16 μg which is sufficient for the following steps of library preparation.

4.3 Future perspectives

As probes necessary for the last steps of the library preparation were not able to be designed and delivered during the time scope of the project, the laboratory work was cut short and no major results were obtained. The next step to be performed is the removal of ssDNA fragments from the libraries using bead separation before the continuation of the library preparation. The remaining steps of the library preparation to be performed are hybridising the DNA to the designed biotin labeled probes, capturing the libraries to streptavidin coated beads followed by amplification of the libraries and multiplex sequencing.

From this, the active promoters and enhancers in prefrontal cortex and hypothalamus could be mapped for both dog and wolf. By comparing these epigenetic maps, potential differences could be found which could help explain the behavioural and morphological differences the species exhibit. This could also provide new insightful information regarding the evolution of the dog and to some extent, humans. Perhaps this data could also provide a better understanding of gene regulations role in evolution.

To provide more general data, more biological replicates would need to be performed. It would also be interesting to use samples from individuals from different habitats and geographic locations. This could eliminate certain sample biases as all wolves used were taken from a zoo and all dogs were taken from Hungary. Also, by performing HiCap on additional tissue types, more information could be provided.

Bibliography

Galibert, F., Quignon, P., Christophe, H., & Catherine, A. (2011). Towards understanding dog evolutionary and domestication history. *Comptes Rendus Biologies, 334*(3), 190-196.

Perri, A. (2016). A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. *Journal of Archaeological Science, 68*, 1-4.

Vila, C., & Savolainen, P. M. (1997). Multiple and Ancient Origis of the Domesticated Dog . *Science, 276*(5319), 1687-1689.

Miklósi, Ádám. (2015). The story of domestication: archaeological and phylogenetic evidence. In *Dog Behaviour, Evolution, and Cognition* (p. Chapter 6). Oxford Scholarship Online.

Serpell, J. (1995). *The domestic dog* (1 ed.). Cambridge: Cambride University.

Paxton, D. W. (2000). A Case for a Naturalistic Perspective. *Taylor & Francis, 13*(1), 5-8.

Caniglia, R., Fabbri, E., Hulva, P., Bolfíková, B., Jindřichová, M., Stronen, A., . . . Galaverni, M. (2018). Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfdog. *BMC Genomics , 19*(1), 533.

Krzeminska, P., Nowacka-Woszuk, J., & Switonski, M. (2022). Copy number variation of the SRY gene showed an association with disorders of sex development in Yorkshire Terrier dogs. *Animal Genetics, 53*(2), 152-155.

Alexander, R., Fang, G. R., Snyder, M., & Gerstein, M. (2012). Annotating non-coding regions of the genome. *Nature Reviews Genetics, 11*, 559-571.

Anshul, K., Patrick, C. D., & Khatun, J. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74.

Palazzo, A., & Ryan, G. (2014). The case for junk DNA. *PLoS Genetics, 10*(5).

Strachan, T., Goodship, J., & Chinnery, P. (2015). *Genetics and Genome in Medicine* (1 ed.). New York, London: Garlans Science.

Sahlén, P., Yanhu, L., Xu, J., Kubinyi, E., Wang, G.-D., & Savolainen, P. (2021). Variants That Differentiate Wolf and Dog Populations Are Enriched in Regulatory Elements. *Genome Biology and Evolution, 13*(4).

Clifford, S., & Bradford, L. (2014). The hypothalamus. *Current biology , 24*(23), R1111-R1116.

Murray, M., & Nowicki, J. (2020). Prefrontal cortex. In J. Pizzorno, & M. Murray (Eds.), *Textbook of natural medicine (Fifth edition).* Churchill Livingstone.

Klein, J. C., Keith, A., Agarwal, V., Durham, T., & Shendure, J. (2018). Functional characterization of enhancer evolution in the primate lineage. *Genome Biology, 19*(1), 99.

Belton, J.-M., McCord, R., Gibcus, J., Naumova, N., Zhan, Y., & Dekked, J. (2012). Hi–C: A comprehensive technique to capture the conformation of genomes. *Elsevier, 58*(3), 268-276.

Mifsud, B., Tavares-Cadete, F., & Young, A. e. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics, 47*, 598-606.

de Laat, W., & Dekker, J. (2013). 3C-based technologies to study the shape of the genome. *Methods (San Diego California), 58*(3), 189-191.

Gavrilov, A., Eivazova, E., Pirozhkova, I., Lipinski, M., Razin, S., & Vassetzhy, Y. (2009). Chromsome Conformation Capture (from 3C to 5C) and Its ChIP-based Modification. *Collas P., 567*, 171-188.

Han, J., Zhiliang, Z., & Wang, K. (2018). 3C and 3c-based techniques: the powerful tools for spatial genome organization deciphering. *Molecular Cytogenetics, 11*(1), 21.

Belaghzal, H., Dekker, J., & Gibcus, J. H. (2017). Hi-C 2.0: an optimized Hi-C procedure for high-resolution genome-wide mapping of chromsome conformation. *Methods(San Diego, Calif.), 123*, 56-65.

Cambridge Dictionary. (2022). *Cambridge dictionary*. Retrieved 04 11, 2022, from https://dictionary.cambridge.org

Pilot, M., & Malewski, T. M. (2015). On the origin of mongrels: evolutionary history of free-breeding dogs in Eurasia. *The royal society publishing, 282*(1820).

Pescini-Marshall, S., Besserdich, I., Kratz, C., & Range, F. (2016). Exploring Differences in Dogs' and Wolves' Preference for Risk in a Foraging Task. *Frontiers in Pshycology, 7*(1241).

Pradhananga, S. (2020). *Towards understanding of complex disease etiology using sequence Capture Hi-C (HiCap) and associated high throughput functional assays.* Stockholm: KTH Royal Institute of Technology.

Saetre, P., Lindberg, J., Leonard, J., Olsson, K., Petterson, U., Ellegren, H., . . . Jazin, E. (2004). From wild wolf to domestic dog: gene expression changes in the brain. *Brain Research. Molecular Brain Research, 126*(2), 198-206.

Janssens, L., Perri, A., Crombe, P. V., & Lawler, D. (2019). An evaluation of classical morphologic and morphometric parameters reported to distinguish wolves and dogs. *Journal of Archaeological Science, 23*, 501-553.

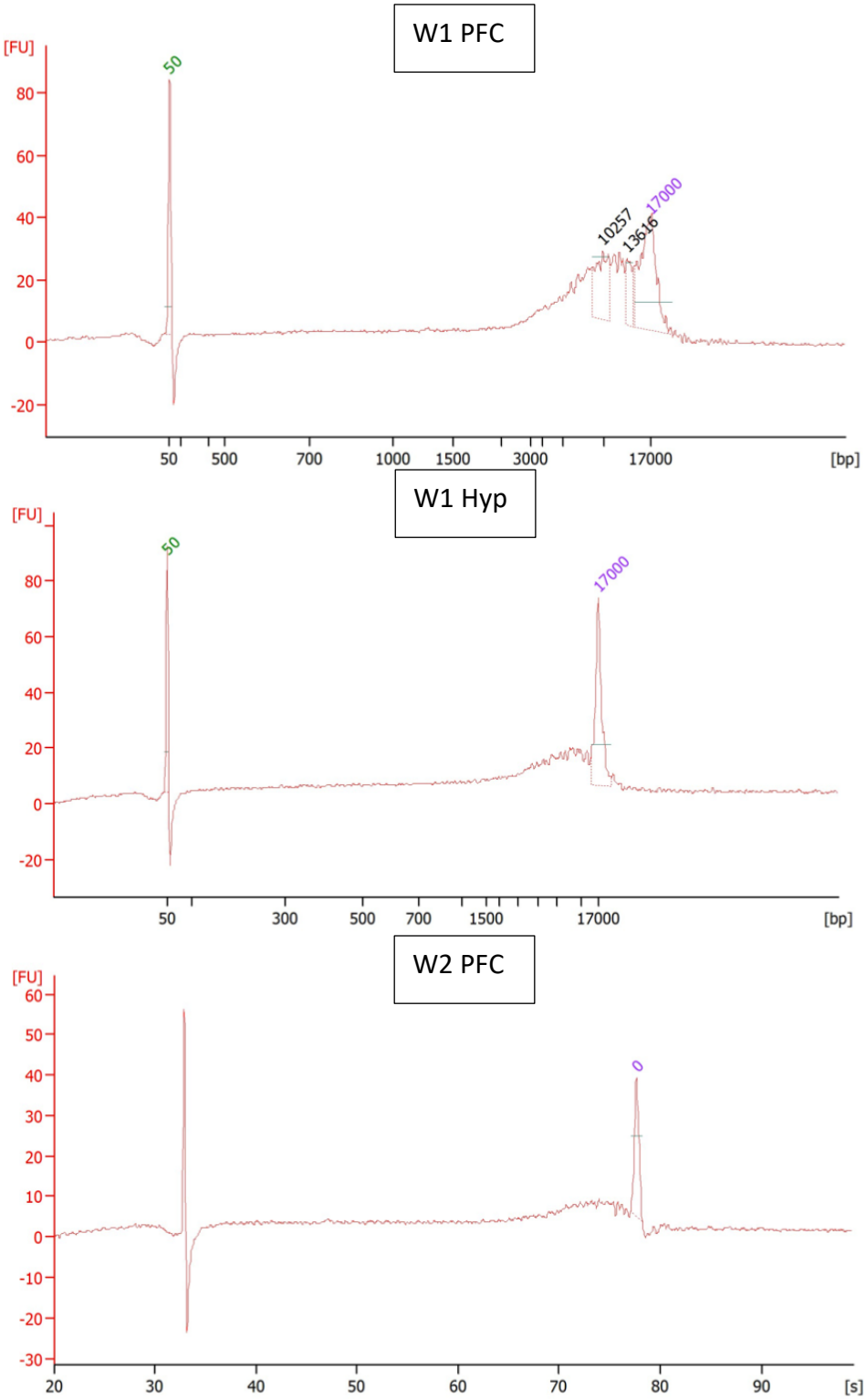
Matelot, M., & Noordermeer, D. (2016). Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq). *Polycomb Group Proteins, 1480*, 223-241.

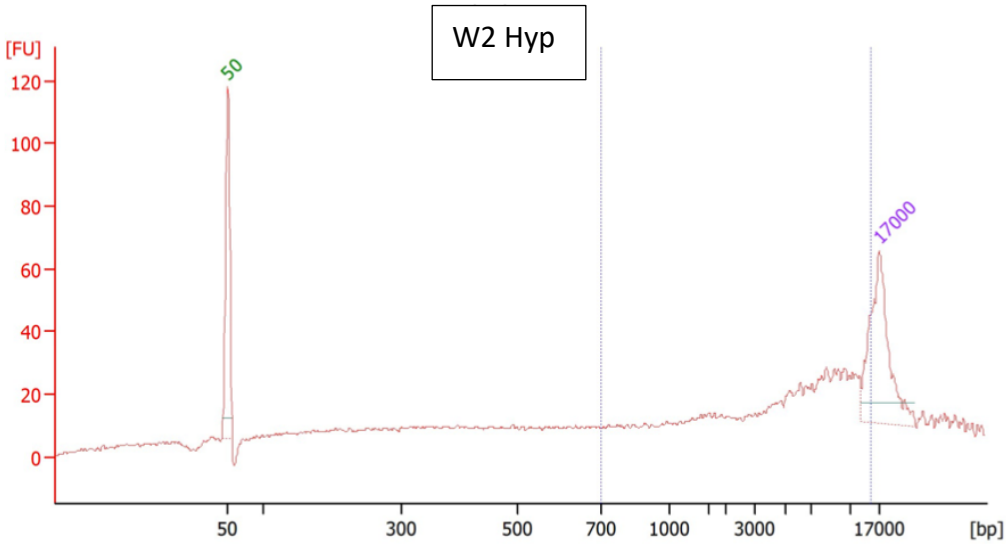
Scientific, T. (2022). *Qubit ssDNA Assay Kit*. Retrieved 05 24, 2022, from https://www.thermofisher.com/order/catalog/product/cn/en/Q10212

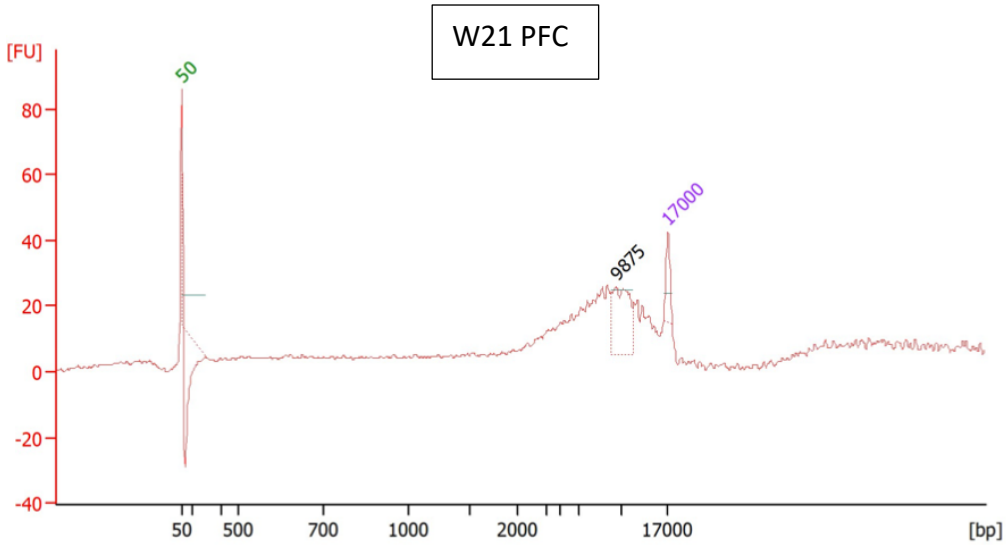
Stellwagen, N., & Stellwagen, E. (2009). Effect of the matrix on DNA electrophoretic mobility. *Journal of chromatography, 1216*(10), 1917-1929.

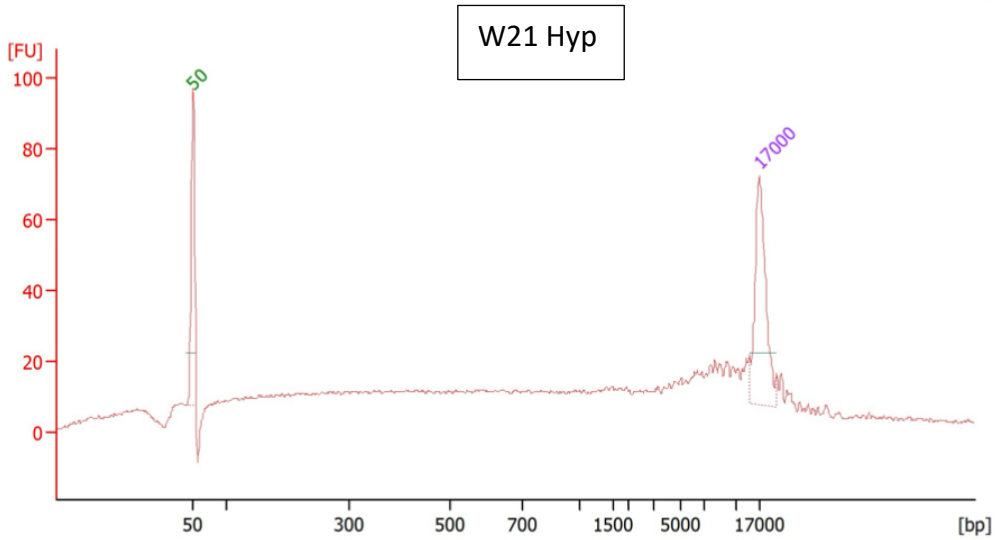
Appendices

Appendix A: Bioanalyzer results from Hi-C on all samples

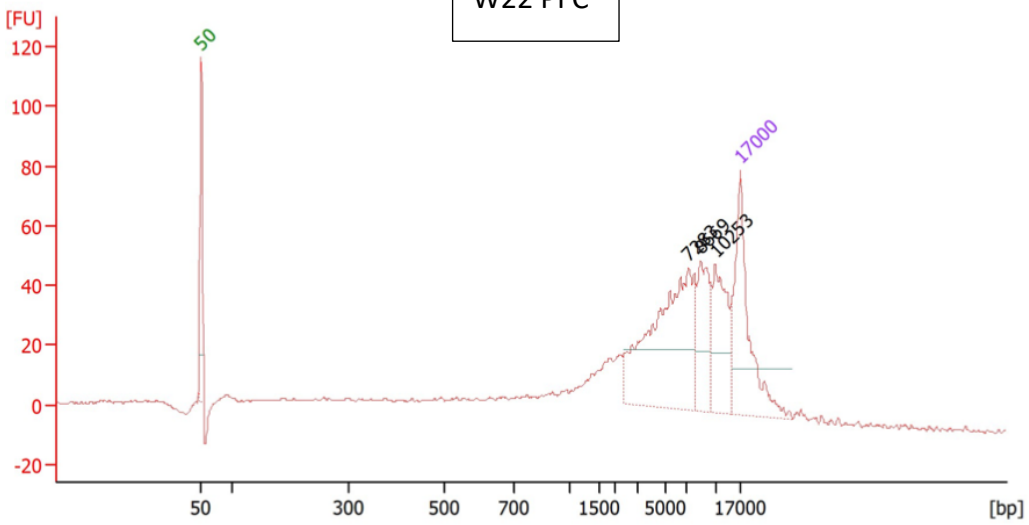



W2 Hyp
[FU]
120
100
80
60
40
20
0
50
17000
50
300
500
700
1000
3000
17000
[bp]

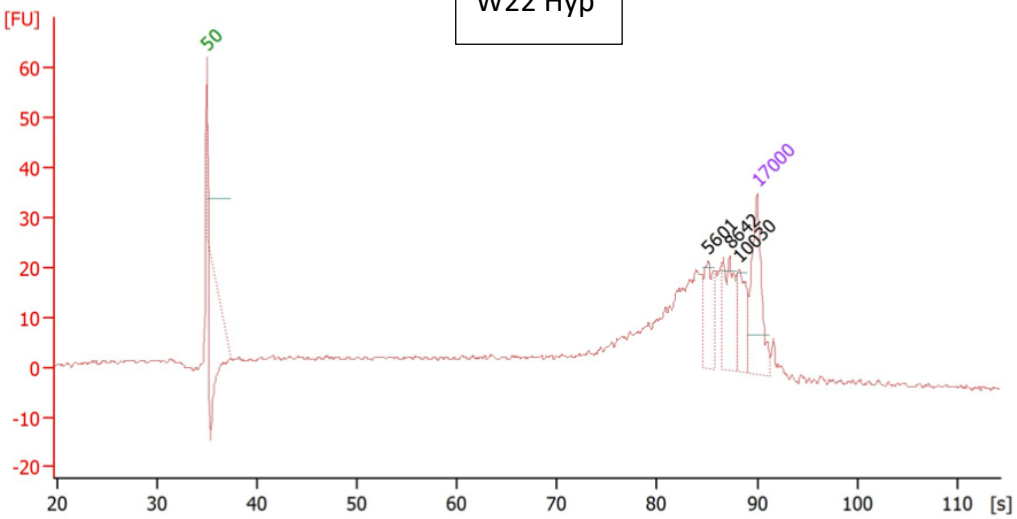

W21 PFC
[FU]
80
60
40
20
0
-20
-40
50
9875
17000
50
500
700
1000
2000
17000
[bp]


W21 Hyp
[FU]
100
80
60
40
20
0
50
17000
50
300
500
700
1500
5000
17000
[bp]

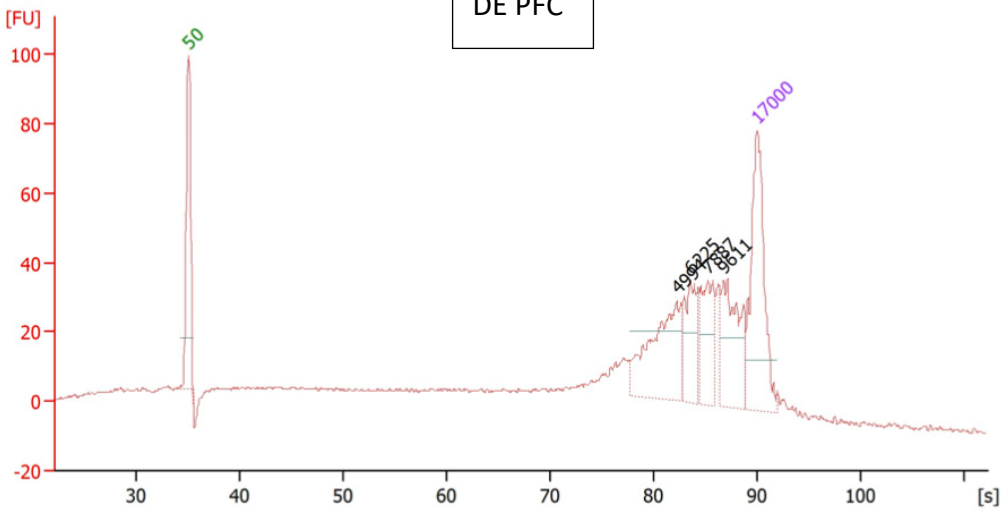
W22 PFC



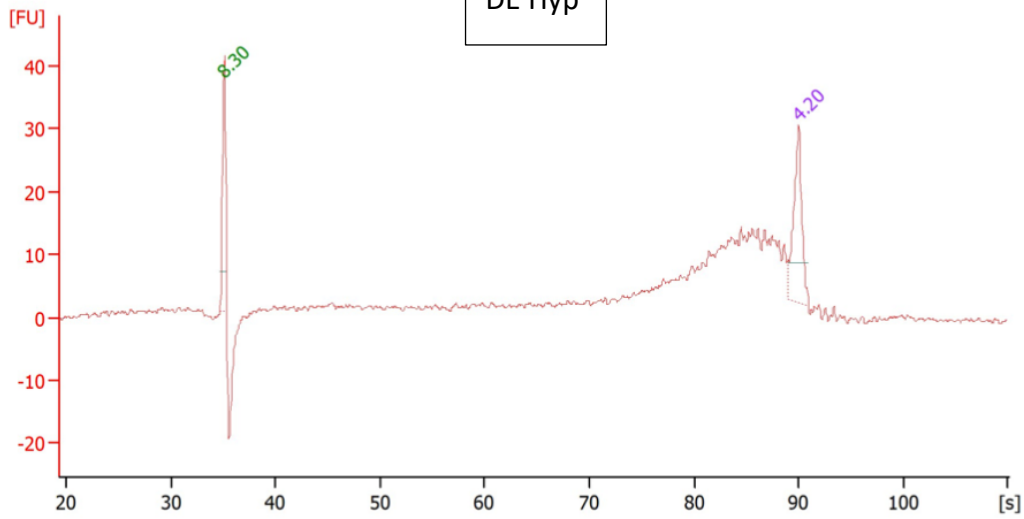
W22 Hyp



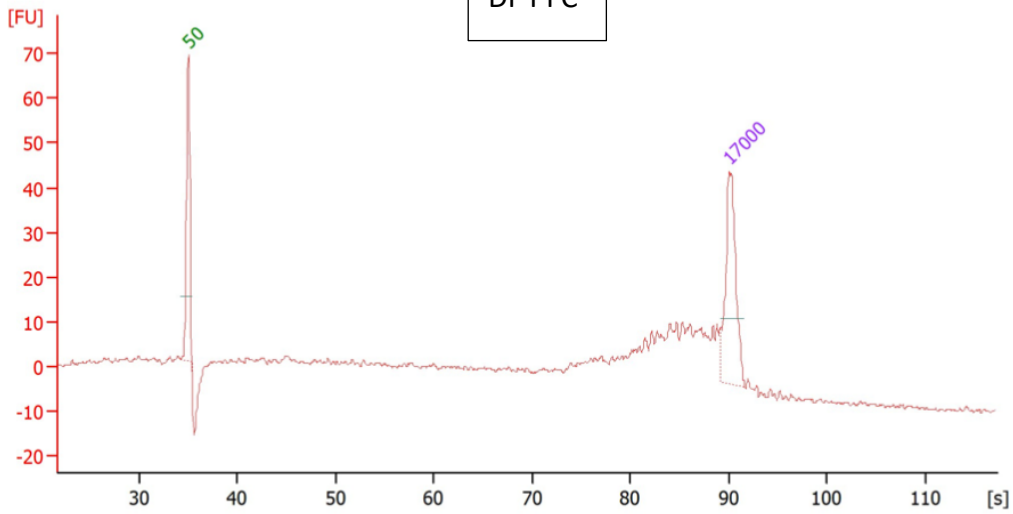
DE PFC



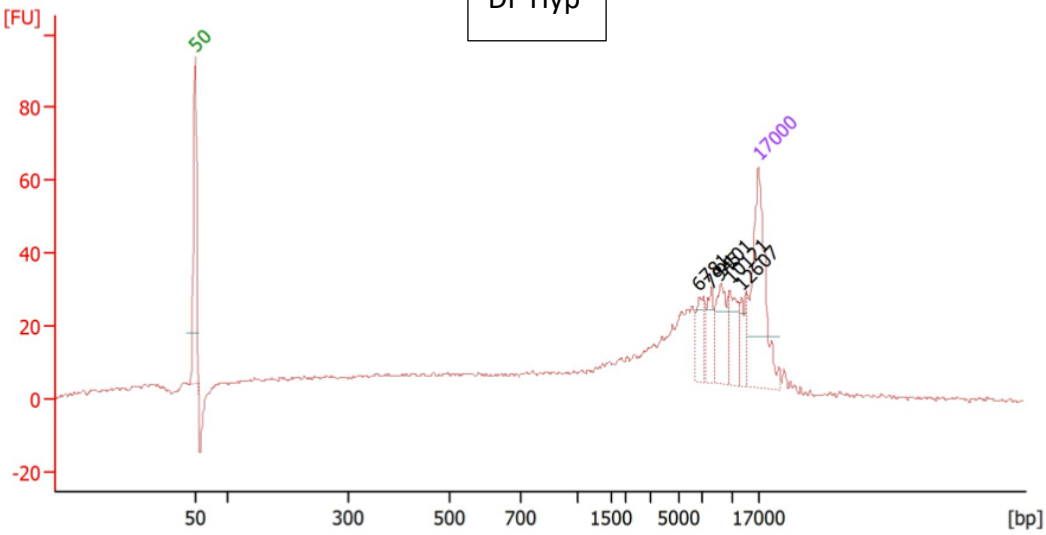
DE Hyp



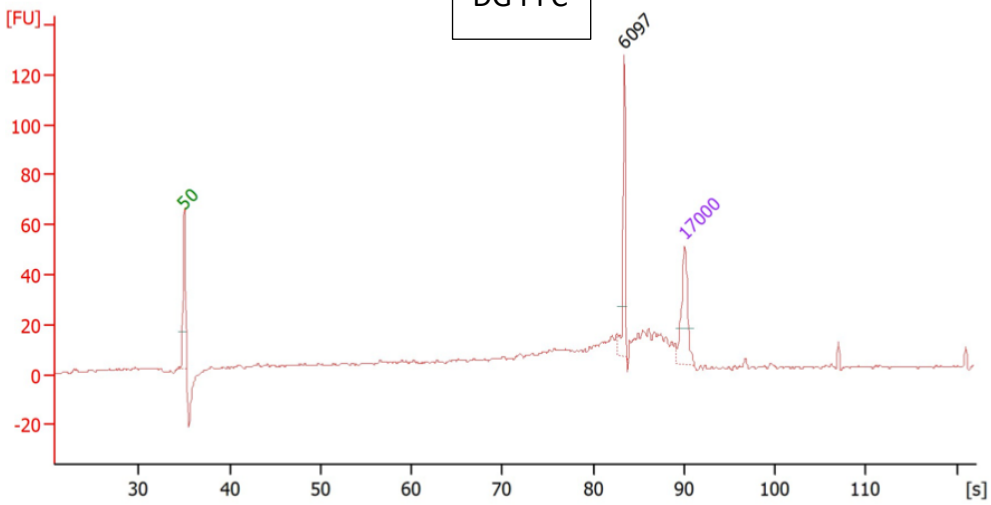
DF PFC



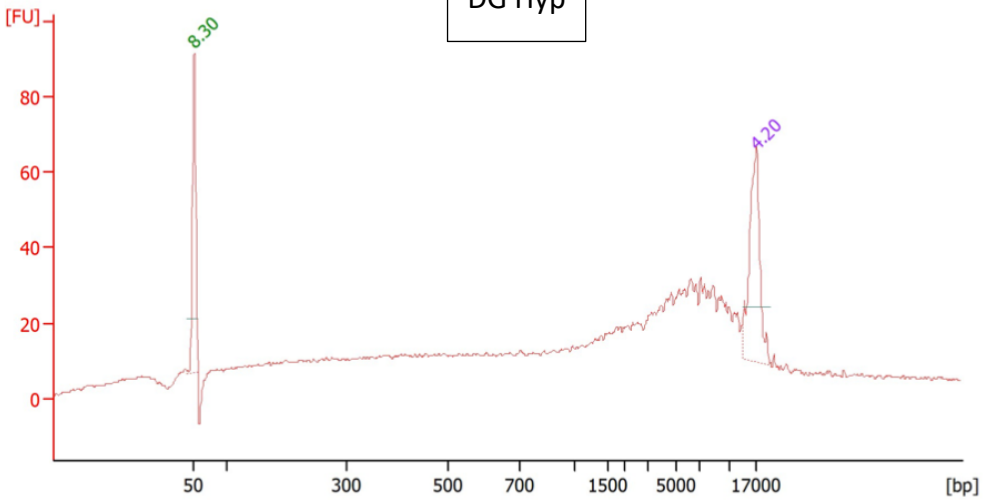
DF Hyp



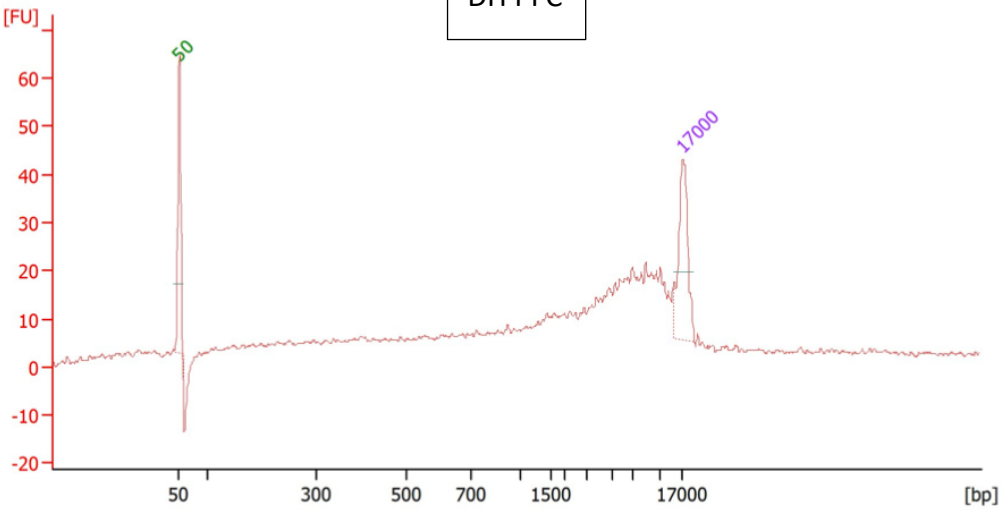
DG PFC



DG Hyp



DH PFC



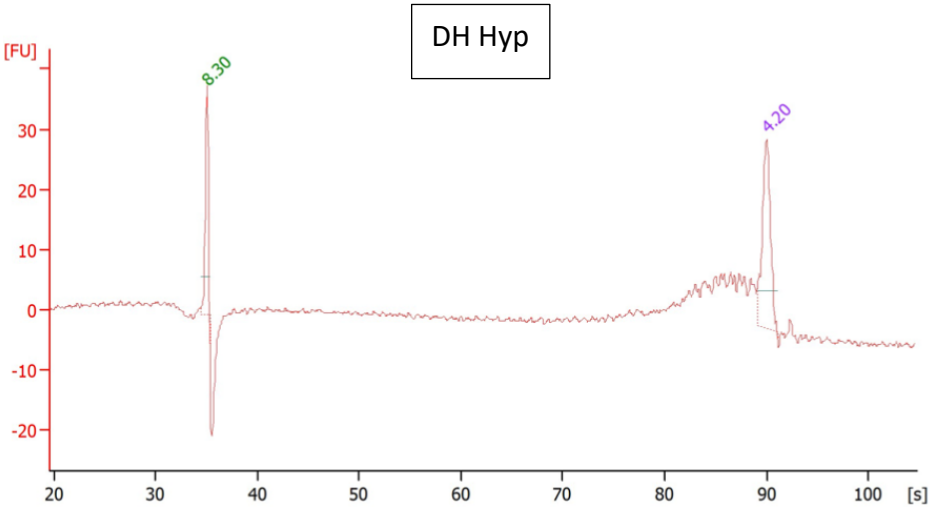


Figure 5.1: The figure displays Hi-C bioanalyzer results for all samples. For those samples which was run more than one time, only one bioanalyzer graph is presented. Due to erroneous ladder application, some of the graphs' x-axis is in seconds and not base pair. As the upper marked is known to be 17000 bp, it is still possible to see if the shift is great enough to consider the Hi-C successful.

Appendix B: Data from nuclei extraction and Hi-C

Table 5.1: The table presents cell concentrations and viabilities of the homogenate after douncing of tissue samples along with their resulting DNA concentrations after chromosome conformation capture. The table also presents if optimisations such as increasing volumes during nuclei extraction or increasing the addition of Klenow Fragment, has been implemented for those samples. See *4.2 Optimisation of protocol* for motivation behind these optimisations. The last column contains comments regarding the outcome of the run. If not stated otherwise, the Hi-C was successful.

Sample	**Tissue type**	**Initial cell concentration (10^6 cells)**	**DNA amount (ng)**	**Viability (%)**	**Increased volumes during extraction**	**Amount of Klenow Fragment (μl)**	**Comment**
Wolf 2	Prefrontal cortex	4.41	1640	37	No	1.2	
Wolf 2	Prefrontal cortex	29.19	9292	40	No	1.2	Overloaded,
Wolf 2	Prefrontal cortex	5.5	1620	40	No	2.5	
Wolf 1	Prefrontal cortex	6.18	3410	-	No	1.2	
Wolf 21	Prefrontal cortex	4.76	3330	-	No	1.2	
Wolf 22	Prefrontal cortex	4.23	5720	-	No	1.2	
Wolf 2	Hypothalamus	19.62	6260	48	No	1.2	
Wolf 21	Hypothalamus	6.00	1470	65	No	1.2	1/3 of sample lost
Wolf 21	Hypothalamus	5.5	1360	65	No	2.5	
Wolf 1	Hypothalamus	5.94	2010	77	No	2.5	
Wolf 1	Hypothalamus	5.94	1960	77	No	2.5	Smaller shift
Wolf 22	Hypothalamus	5.30	1720	75	No	2.5	
Wolf 22	Hypothalamus	5.30	5440	75	No	2.5	
Dog H	Hypothalamus	5.30	6780	55	No	2.5	
Dog E	Hypothalamus	5.19	17127	45	No	2.5	
Dog G	Hypothalamus	5.50	3250	75	Yes	2.5	
Dog F	Hypothalamus	5.50	7620	63	Yes	2.5	
Dog H	Prefrontal cortex	5.50	11200	42	Yes	2.5	
Dog E	Prefrontal cortex	5.50	29600	29	Yes	2.5	
Dog F	Prefrontal cortex	5.50	20750	20	Yes	2.5	
Dog G	Prefrontal cortex	5.49	6650	26	Yes	2.5	

Appendix C: DNA amounts before and after biotin removal

Table 5.2: shows the DNA amounts measured before and after biotin removal from unligated ends during library preparation.

Sample	**Tissue**	**Initial DNA amount (*µ*g)**	**Total amount of DNA (*µ*g) after biotin removal**	**Loss during biotin removal (%)**
Wolf 1	Prefrontal cortex	3.41	2.43	29
Wolf 1	Hypothalamus	3.97	2.89	27
Wolf 2	Prefrontal cortex	3.28	1.73	47
Wolf 2	Hypothalamus	5.0	3.09	38
Wolf 21	Prefrontal cortex	3.33	2.33	30
Wolf 21	Hypothalamus	2.83	1.84	35
Wolf 22	Prefrontal cortex	5.0	2.56	49
Wolf 22	Hypothalamus	5.0	3.59	28
Dog E	Prefrontal cortex	5.0	3.00	40
Dog E	Hypothalamus	5.0	4.58	8.4
Dog F	Prefrontal cortex	5.0	3.42	32
Dog F	Hypothalamus	5.0	2.82	44
Dog G	Prefrontal cortex	5.0	3.46	31
Dog G	Hypothalamus	3.25	1.86	43
Dog H	Prefrontal cortex	5.0	2.81	44
Dog H	Hypothalamus	5.0	2.86	43

Appendix D: DNA amounts measured after sonication

Table 5.3: DNA amounts measured after sonication and AMPure XP beads purification during library preparation.

Sample	**Tissue**	**DNA amount (μg) after sonication**
Wolf 1	Prefrontal cortex	1.66
Wolf 1	Hypothalamus	2.39
Wolf 2	Prefrontal cortex	1.21
Wolf 2	Hypothalamus	2.33
Wolf 21	Prefrontal cortex	1.54
Wolf 21	Hypothalamus	1.80
Wolf 22	Prefrontal cortex	1.58
Wolf 22	Hypothalamus	1.95
Dog E	Prefrontal cortex	2.29
Dog E	Hypothalamus	2.80
Dog F	Prefrontal cortex	1.93
Dog F	Hypothalamus	2.02
Dog G	Prefrontal cortex	1.75
Dog G	Hypothalamus	1.38
Dog H	Prefrontal cortex	1.97
Dog H	Hypothalamus	1.96