Genomic analysis of Chinese hamster ovary (CHO) cells (CHO 細胞における染色体配列に関する研究)

木村 修一

2018

Chapter 1

General introduction

1.1. Chinese hamster ovary (CHO) cells as host cells producing biopharmaceuticals

The demand for protein-based biopharmaceuticals such as monoclonal antibodies (mAbs) is growing. The market for biopharmaceutical products reached $140 billion in total cumulative sales in 2013 (Walsh 2014). These biopharmaceuticals are produced by the cultivation of recombinant cells. *Escherichia coli*, *Saccharomyces cerevisiae* (yeast), and mammalian cells are commonly used as host cells for industrial protein production. Among these host cells, Chinese hamster ovary (CHO) cell lines are particularly widely used for the production of protein-based biopharmaceuticals. As of 2014, biopharmaceuticals produced by CHO cells made up a 35% share of all approved biopharmaceuticals (Walsh 2014).

CHO cell lines have several advantages as host cells. First, they can achieve post-translation modification that is appropriate for biopharmaceuticals because they are derived from Chinese hamster, a higher eukaryote. Consequently, they can produce recombinant proteins with the correct glycoforms for compatibility with humans (Koo et al. 2009; Li and Stern 2005; Ono et al. 2003). Many protein-based biopharmaceuticals require specific modifications that can only be performed in higher eukaryotic cells (Maksimenko et al. 2015). Second, CHO cells have already been demonstrated to be relatively safe as host cells for clinical use, potentially making it

easier to obtain approval from regulatory agencies. Third, CHO cells can be easily adapted to serum-free suspension culture, which is suitable for large-scale cultivation for biopharmaceutical manufacturing (Omasa et al. 2010). Moreover, as approval for the production of many protein-based biopharmaceuticals using CHO cells has been granted, CHO cell production systems have steadily improved (Reinhart et al. 2015). Now, chemically defined (CD) medium is typically demanded for producing biopharmaceuticals using recombinant host cells. In the case of production using CHO cell lines as host cells, various CD media have been developed for the production of high levels of protein-based biopharmaceuticals.

1.2. Previous study on basic biology of CHO cells

Systems for the production of biopharmaceuticals using CHO cells are now well developed, but the basic biological background of CHO cells has not been investigated. Human and mouse genomic sequences were analyzed in the early 2000s (International Human Genome Sequencing Consortium 2004; Mouse Genome Sequencing Consortium, 2002), but the first draft CHO genomic sequence was only reported in 2011 (Xu et al. 2011). Moreover, this CHO genomic sequence does not refer to CHO chromosomes; in contrast, human and mouse genomic sequences were analyzed with their chromosomal information such as genomic sequence and relationship between genomic sequence and chromosomal location. The reason for this is that CHO cells exhibit aneuploidy in chromosome number and instability of chromosomal structure (Omasa et al. 2010; Cao et al. 2012a). In the manufacturing of recombinant proteins using host cells, the transfection and effective screening for highly productive cells are essential steps. In the case of using CHO cell lines as host cells, this screening is dependent on the knowhow and techniques of researchers. Genome targeting techniques such as the use of zinc finger nuclease, TALEN, and CRISPR/Cas9 (Lombardo et al. 2007; Kim et al. 2013; Qi et al. 2013) have been developed and refined; these techniques require a specific genome sequence as a targeted site. Consequently, it is

important to be able to refer to the particular chromosomal location when defining genomic sequences.

1.3. Our previous study involving construction of CHO DR-1000L-4N cell line and bacterial artificial chromosome (BAC) library

The workflow of our previous study is shown in Figure 1.3.1.

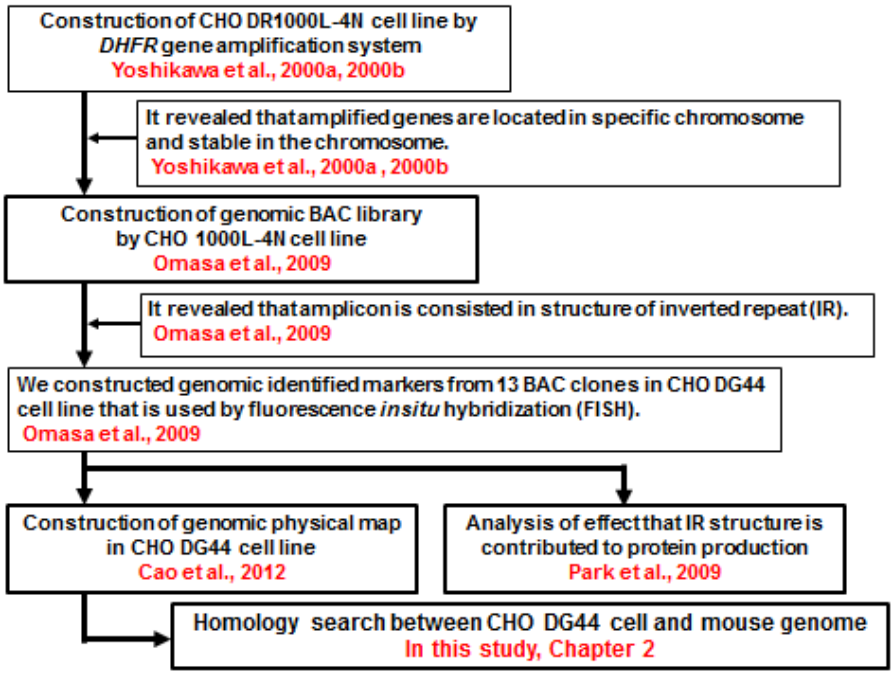

Figure 1.3.1 Workflow of our previous study

We investigated an effective procedure to construct cell lines that produce high levels of recombinant protein under a dihydrofolate reductase (DHFR) gene amplification system (Yoshikawa et al. 2000; Park et al. 2010). Our results revealed the construction of the CHO DR-1000L-4N cell line, which stably produced recombinant protein at a high level. We thus investigated the chromosomal location of the gene-amplified region and the relationship between the copy numbers of recombinant protein and DHFR gene on

the one hand and the specific production rate of recombinant protein on the other. However, the construction of a cell line that stably produces recombinant protein is extremely time-consuming and expensive. Consequently, in 2006, a bacterial artificial chromosome (BAC) library from CHO DR-1000L-4N cell line chromosomes was constructed to determine the genomic sequence at the gene-amplified location (Omasa et al. 2009). BAC libraries have been used for the analysis of human and mouse genomic sequences (International Human Genome Sequencing Consortium 2004; Mouse Genome Sequencing Consortium 2002), it was only procedure for determined genomic sequences in those days. Using this BAC library, we revealed that the amplified region consisted of an inverted repeat (Omasa et al. 2009). In human genome analysis, the development of a clone-based physical map was indispensable for the Human Genome Project because of the presence of numerous repeat sequences in the genome (McPherson et al. 2001). The DNA sequencing information of the CHO genome should be coupled with physical chromosomal locations because of the instability of the chromosomal structure of CHO cells. These locations should be obtained from BAC physical maps that can be derived from the sequence of BAC library clones and their comparison with the genomic sequences of Chinese hamster, various Chinese hamster-derived CHO cell lines, and other related species. Based on the

CHO genomic BAC library, we constructed a chromosomal physical map of the CHO DG44 cell line using 303 BAC clones and investigated the chromosomal rearrangements between the two most widely used CHO cell lines, CHO K1 and CHO DG44 (Cao et al. 2012a). It was revealed that the two longest chromosomes did not feature significant rearrangements among CHO cells.

1.4. Overview of the present study

In this study, my colleagues and I determined the BAC end sequences (BESs) of 303 clones, which were used as BAC probes for the construction of a chromosomal physical map in the CHO DG44 cell line. Chapter 2 describes a comparison between these BESs and mouse genomic sequences. It has been reported that Chinese hamster cDNAs have high homology to mouse cDNAs (Wlaschin et al. 2005; Melville et al. 2011). However, cDNA analyses of Chinese hamster have not revealed whether its genome has high homology to the mouse genome. Using BAC-FISH and BESs, I confirmed that the genomic sequences of CHO DG44 cells have regions that are highly homologous to the mouse genome.

Chapter 3 describes a comparison between these BESs and genomic sequences of the CHO K1 cell line (Xu et al. 2011). Many CHO cell lines are used for host cells. However, the karyotype and the genomic sequences differ among CHO cell lines. It is thus difficult to couple the DNA sequencing information of the CHO genome with physical chromosomal locations. Against this background, this thesis describes a comparison of these BESs and CHO K1 cell line genomic sequences. The genomic comparison between CHO DG44 and CHO K1 cell lines can clarify the differences of karyotypes and genomic structures in CHO cells. In the final chapter of this thesis,

Chapter 4, a discussion of the implications of the obtained results and future research aspects is presented.

Chapter 2

Genome sequence comparison between Chinese hamster ovary (CHO) DG44 cells and mouse using end sequences of CHO BAC clones based on BAC-FISH

2.1. Introduction

The basic biology of model mammalian organisms as represented by human and mouse has been well researched. The complete genomic sequences of human and mouse were also determined in the early 2000s. Moreover, these genomic sequences were almost all referenced to their chromosomes by constructed their physical maps. (Collins et al. 2004; Waterston et al. 2002). Chinese hamster ovary (CHO) cells play an important role as hosts for the commercial-scale production of protein-based pharmaceuticals (Greber and Fussenegger 2007; Hacker et al. 2009; Omasa et al. 2010). Two subclones of CHO cells, CHO K1 (Kao and Puck 1968) and CHO DG44 (Urlaub and Chasin 1980), are the most widely used for both scientific research and industrial applications (Wurm 2004; Griffin et al. 2007).

Recently, following the development of next-generation sequencing (NGS) techniques and their impact on genome research, whole-genome sequencing has become more economical and faster. Several groups have reported NGS analyses of the genomic sequences of Chinese hamster and/or CHO cells (Xu et al. 2011; Lewis et al. 2013; Kaas et al. 2015; Feichtinger et al. 2016; Vishwanathan et al. 2016). However, the genome of Chinese hamster includes various repeat sequences that are distributed on the same or different chromosomes (Ono and Sonta 2001). In human genome analysis, a

clone-based physical map was indispensable for the Human Genome Project because of the presence of numerous repeat sequences in the genome (McPherson et al. 2001). In view of this, the DNA sequencing information of the CHO genome should be coupled with physical chromosomal locations because of the instability of its chromosomal structure. These locations should be obtained from BAC physical maps that can be derived from BAC libraries and compared with the genomic sequences of Chinese hamster, various CHO cell lines, and other related species.

Previously, my colleagues and I constructed a genomic BAC library from the CHO DR1000L-4N cell genome, which provided 5-fold coverage of this genome (Omasa et al. 2009). This cell line can be used for the high and stable production of recombinant proteins using a dihydrofolate reductase (DHFR) gene amplification system (Yoshikawa et al. 2000a, 2000b). The above-mentioned BAC library consisted of 122,281 clones, based on which my colleagues and I constructed a chromosomal physical map of the CHO DG44 cell line using 303 BAC clones and investigated the chromosomal rearrangements between the two most widely used CHO cell lines, CHO K1 and CHO DG44 (Cao et al. 2012a). It was revealed that the two longest chromosomes did not feature significant rearrangements among CHO cells.

In the work described in this chapter, I determined the BAC end sequences (BESs) of

303 clones, which were used as BAC probes to construct a chromosomal physical map in the CHO DG44 cell line. Moreover, I compared these BESs and mouse genomic sequences.

2.2. Materials and methods

2.2.1. Cell lines and media

The CHO DG44 (*dhfr*$^-$) (Urlaub and Chasin 1980) cell line, provided by Dr. L.A. Chasin of Columbia University, was used in this study. It was maintained in Iscove's modified DMEM (IMDM) (Sigma-Aldrich, St. Louis, MO) with 10% dialyzed fetal bovine serum (FBS) (SAFC Biosciences, Lenexa, KS), 13.6 mg/L hypoxanthine (Yamasa, Choshi, Japan), and 2.42 mg/L thymidine (Yamasa). This cell line was cultivated at 37 °C in a humidified atmosphere containing 5% CO_2.

2.2.2. Extraction of bacterial artificial chromosome (BAC) clones

my colleagues and I previously constructed a genomic BAC library from the CHO DR1000L-4N cell genome, which provided 5-fold coverage of this genome (Omasa et al. 2009). BAC clones were cultured in 10 mL of Luria–Bertani (LB) medium containing chloramphenicol (12.5 μg/mL) at 37 °C for 14–16 h, and harvested by centrifugation (7500× g, 5 min at room temperature). BAC DNA was purified from *E. coli* using the JETstar 2.0 Plasmid Purification mini kit (Genomed GmbH, Löhne, Germany), in accordance with the manufacturer's protocol. The extracted BAC DNA

was purified by isopropanol precipitation and dissolved in 40 µL of sterilized ultrapure water. The DNA concentration was estimated by measuring the UV absorption at 260 nm. The final DNA concentration was adjusted in the range from 0.1 to 0.2 µg/µL.

2.2.3. Fluorescence *in situ* hybridization

Chromosome preparation and fluorescence *in situ* hybridization (FISH) were performed as described previously (Cao et al. 2012b; Omasa et al. 2009; Yoshikawa et al. 2000a; Yoshikawa et al. 2000b). The protocol of two-color BAC-FISH is summarized in Fig. 2.2.1.

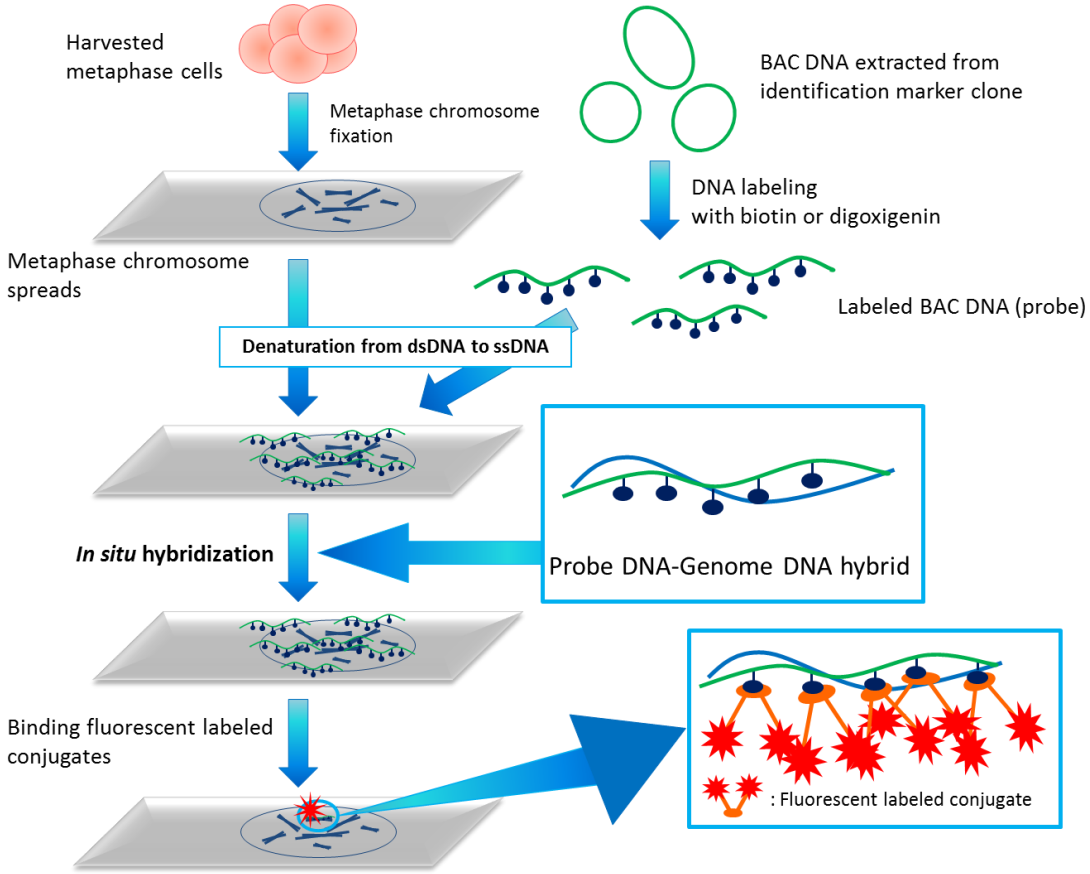

Fig. 2.2.1. Experimental procedure of fluorescence in situ hybridization using BAC DNA as a probe (BAC-FISH, figure from Omasa et al. 2009; Cao et al. 2012b).

Probes derived from BAC DNA were finally detected using fluorescein isothiocyanate (FITC)-labeled streptavidin and/or anti-digoxigenin (DIG)-rhodamine antibody. One microgram of BAC DNA was labeled by incubation using biotin or DIG-Nick translation mix (Roche Diagnostics, Basel, Switzerland) at 15 °C for 8 h following the manufacturer's protocol.

Metaphase chromosome spreads were prepared from exponential-phase cultures using standard techniques. CHO cells were treated with 10 μg/mL colcemid solution for 4.5 h.

After harvesting the metaphase cells, these cells were suspended in 40 mL of hypotonic solution (0.75 M potassium chloride solution). Approximately 20 µL drops of the cell suspension were placed on a slide and dried with a metaphase spreader (Hanabi, Adstec, Funabashi, Japan).

For hybridization, the probe DNA should be heat-denatured to single-stranded DNA. Ethanol-precipitated probe DNA was incubated for 10 min at 80 °C in a water bath. After incubation, the denatured probe DNA was incubated for more than 5 min on ice. Metaphase chromosome spreads on a slide glass were dried in an air incubator for 3 h at 37 °C. A denaturation solution (5 mL of 20× SSC, 10 mL of ultrapure water, and 35 mL of deionized formamide) was prepared and incubated in a tray at 70 °C in a water bath. The slide with spread metaphase chromosomes was incubated in the tray for 90 s and denatured. After incubation, the slide was subjected to dehydration. Denatured probe DNA on a 22 × 22 mm square microscope cover glass (Matsunami Glass Ind., Ltd.) was mounted on a slide glass with the spread metaphase chromosomes. Both slide and cover glasses were sealed with rubber cement. These glasses were incubated at 37 °C overnight in a moist chamber humidified with 2× SSC.

After incubation, the slide glass was washed with washing solution I (deionized formamide: 2× SSC: ultrapure water = 5:4:1) at 43 °C for 5 min three times. Then, the

slide was washed with washing solution II (0.8× SSC) at 60 °C for 5 min three times. After washing, 150 µL of blocking solution [3% (w/v) Block Ace™ (DS Pharma Biomedical, Suita, Japan), 0.1% (v/v) Tween 20 in 4× SSC] was mounted on the 24 × 50 mm cover glass. The cover glass was mounted on the slide glass, which was then incubated for 1 h at 37 °C in a moist chamber with 2× SSC. After incubation, 150 µL of detection solution [4× SSC solution containing 1% (w/v) Block Ace™, 0.1% (v/v) Tween 20, 0.375 µL of 1 mg/L fluorescein-labeled avidin (Fluorescein Avidin DN; Vector Laboratories, Inc., Burlingame, CA, USA), and 0.375 µL of 1 mg/L anti-DIG-rhodamine Fab fragment (Roche Diagnostics)] was mounted on the slide glass. The slide glass was again incubated for 40 min at 37 °C in a moist chamber with 2× SSC. Finally, the slide glass was washed with washing solution III [0.1% (v/v) Tween 20 with 4× SSC] at 43 °C for 5 min four times.

The washed slide glass was transferred into a tray with DAPI solution [25 µL of stock solution (DAPI 400 µg/mL, Sigma-Aldrich, stored at 4 °C in the dark) was diluted with 50 mL of 2× SSC] and incubated for 5 min at room temperature. The slide was then washed with 2-fold-diluted washing solution III for 5 min at room temperature, followed by air-drying in the dark. Finally, 40 µL of antifade solution, which prevents fluorescence degradation, was mounted on the slide glass, followed by its sealing with a

cover glass. The antifade solution was composed of 0.233 g of 1,4-diazabicyclo[2.2.2] octane (Sigma-Aldrich), 0.2 mL of 1 M Tris–HCl solution, 0.8 mL of ultrapure water, and 9 mL of glycerol.

The prepared FISH slide was observed under an Axioskop 2 fluorescence microscope (Carl Zeiss, Jena, Germany) with an oil immersion objective lens (Plan-Apochromat w0.8 63 × oil Ph3) and fluorescence filter sets for DAPI (excitation filter G365, dichroic mirror FT395, barrier filter BP445/50), FITC (excitation filter BP485/20, dichroic mirror FT510, barrier filter BP515-565), and rhodamine (excitation filter BP545/25, dichroic mirror FT570, barrier filter BP625/53). Photographs were taken with an AxioCam MRm CCD camera (Carl Zeiss, basic resolution: 1388 × 1040 = 1.4 megapixels). Image processing was performed using Adobe Photoshop CS3. ImageJ software (http://rsbweb.nih.gov/ij/) was used to analyze the chromosomal loci of the BAC clone probes and the positions of the centromeres on the chromosomes, which were expressed as FLpter values (relative distance from the short-arm telomere to the signal fractional length p-terminal; Lichter et al., 1990).

2.2.4. Determination of BAC end sequences (BESs)

The end sequencing of BAC DNA was performed with the forward sequencing primer 5′-CGCCAGGGTTTTCCCAGTCACGAC-3′ and the reverse sequencing primer 5′-CAGGAAACAGCTATGACC-3′ using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Inc., Foster City, CA). After the PCR reaction, the substrate DNA was purified by ethanol precipitation.

2.2.5. Searching for highly homologous sequences between BAC end sequences and mouse genomic sequences using BLAST

A sequence homology search with the mouse genome (*Mus musculus* strain C57BL/6J, GRCm38.p2, March 2013) was performed using the BLAST algorithm (http://blast.ncbi.nlm.nih.gov/), setting the threshold for a significant value (e-value) as $<10^{-2}$. The threshold is typically set to $<10^{-10}$ for homology searches with genomic sequences. However, I set the threshold to $<10^{-2}$ in this study because the determined BES lengths are short (less than 1,100 bp) and it was expected that regions with greater homology to mouse genomic sequences would be obtained by setting a low threshold. In the case of the threshold to $<10^{-4}$, the number of homologous BESs decreased from

465 BESs to 457 BESs. And, in the case of the threshold to $<10^{-10}$, the number of homologous BESs decreased from 465 BESs to 425 BESs.

2.3 Results and discussion

2.3.1. Determination of BAC end sequences (BESs) and their analysis

It has been reported that Chinese hamster cDNAs have high homology to mouse cDNAs (Melville et al. 2011; Wlaschin et al. 2005). However, cDNA analyses of Chinese hamster have not revealed whether its genome has high homology to the genome of mouse. Using BAC-FISH and BESs, my colleagues and I confirmed that the genomic sequences of CHO DG44 cells have regions that are highly homologous to the mouse genome. For Chinese hamster and the Chinese hamster-derived CHO cell line, less information on the genome sequence is available than for human and mouse. Thus, my colleagues and I constructed a chromosome physical map in the CHO DG44 cell line by BAC-FISH (Cao et al. 2012a). This physical map is based on the chromosomal locations of 303 BAC clones containing genomic fragments of CHO cells. I determined the BESs of BAC clones that were used for physical map construction in the CHO DG44 cell line. I attempted to determine all BESs of the 303 BAC clones (606 BESs) by end sequencing; I succeeded in this for 558 BESs. Consequently, both BESs of 266 BAC clones were determined (Table 2.3.1).

Table 2.3.1 Results of determined BAC end sequences

	Number of BAC clones
Determined both end sequences	266
Determined only forward end sequence	0
Determined only reverse end sequence	26
Determined neither end sequence	11
Total	**303**

The average length of the determined sequences was 532 bp and the minimum and maximum sequence lengths were 175 and 1,100 bp, respectively. However, both BESs of 11 BAC clones could not be determined. Among these 11 clones, 6 are located in chromosomes A and B, which are the pair of longest chromosomes of the CHO DG44 cell line (Omasa et al. 2009; Cao et al. 2012a). I also analyzed whether BESs have high homology with the mouse genomic sequence. The results showed that 465 BESs exhibit high homology with the mouse genome (Fig. 2.3.1). It appeared that the lengths of the determined sequences did not affect this high homology. A previous study on comparative genomic hybridization (CGH) between Chinese hamster and mouse showed that the entirety of each chromosome was not conserved between the two species (Yang et al. 2000).

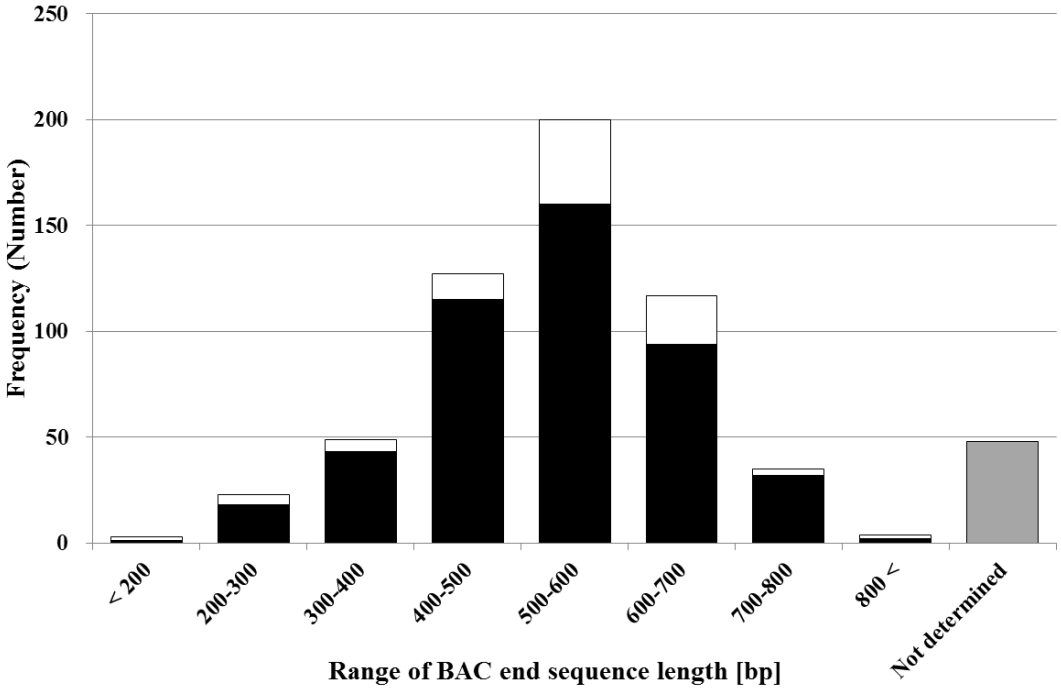

Fig. 2.3.1 Size distribution of determined BAC end sequences. Black: number of sequences homologous to mouse chromosome sequences. White: number of sequences not homologous to mouse chromosome sequences.

2.3.2. Determination of highly homologous sequences between CHO DG44 cells and mouse

In view of the findings described above, I compared the locations of BESs between the CHO DG44 cell line and mouse chromosomes using BLAST (Fig. 2.3.2). Figure 2.3.2 shows the proportion of regions highly homologous (e-value of less than 10^{-2}) to the mouse genome in each CHO DG44 chromosome (A to T). The results show that each chromosome in the CHO DG44 cell line exhibits high homology to various mouse chromosomes. Yang et al. (2000) also reported similar results from a comparative chromosome map between Chinese hamster and mouse.

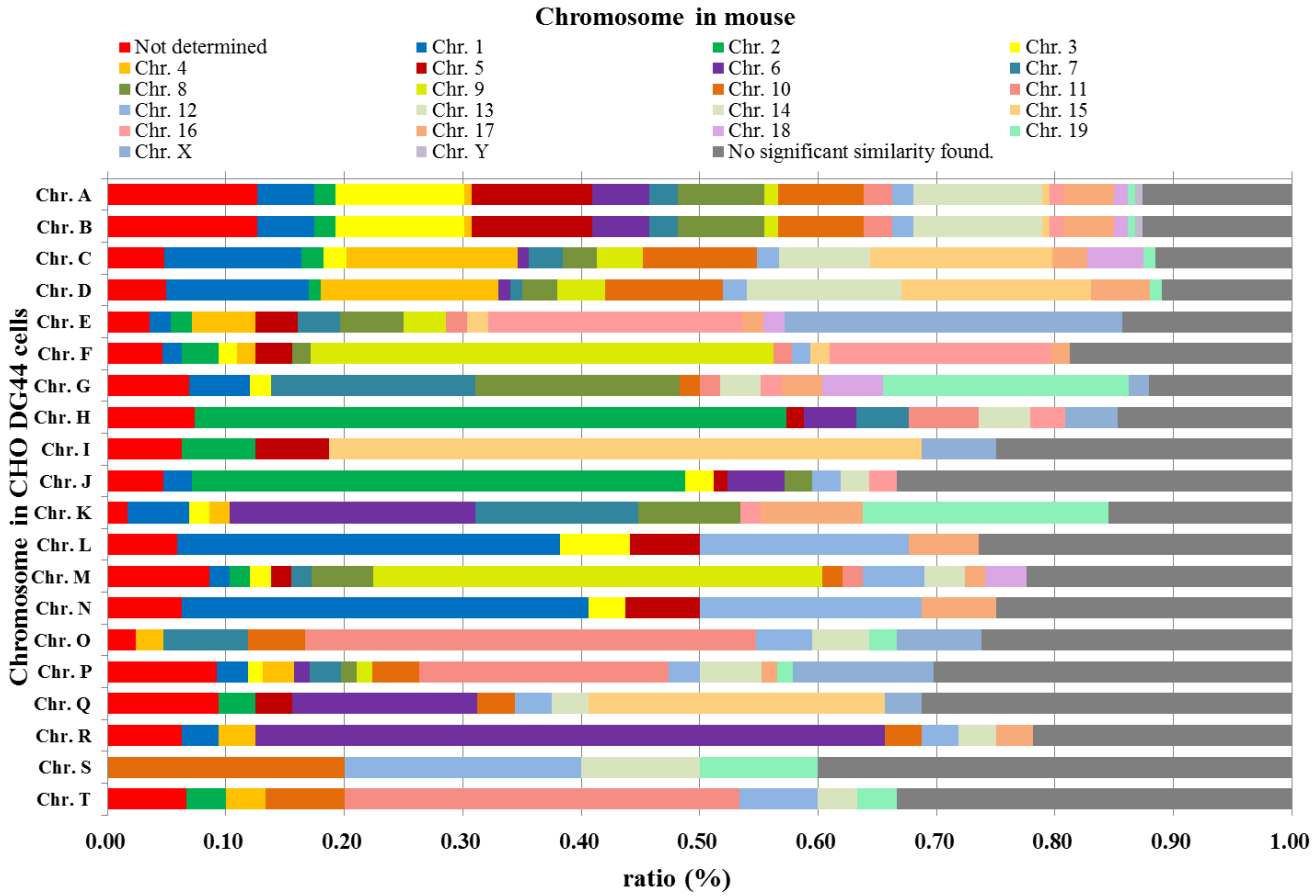

Fig. 2.3.2 Distribution of homologous chromosome sequences between CHO DG44 cell line and mouse chromosome. Legends indicate highly homologous loci in mouse chromosome. CHO DG44 chromosomes are arranged in order from A to T based on chromosome length, as reported previously (Omasa et al. 2009; Cao et al. 2012a)

2.3.3. Analysis of relationship between CHO DG44 cells and mouse genomic sequences

I have determined the loci of BESs in the chromosomes of the CHO DG44 cell line by BAC-FISH (Cao et al. 2012a). Although the chromosomal structures of Chinese hamster and mouse are not the same, I investigated the relationship of BESs between positions in the CHO DG44 chromosome and sequences in mouse chromosomes. The results showed that 23 specific regions in 13 chromosomes of the CHO DG44 cell line had similarities to specific mouse chromosomes ($r^2 \geq 0.850$).

Chromosome A and B

Eighty-three BAC clones are located in chromosome A and B, and about 83% of these BESs have homology to mouse genomic sequences (Fig. 2.3.2). In particular, three regions in chromosomes A and B were found to be correlated to mouse chromosomes 3 and 8 (Fig. 2.3.3a and 2.3.3b). Several BESs are not located in these specific regions; however, the results showed a tendency for the sequences of these two regions to be conserved in the mouse genome. Sixteen BESs showed high homology to mouse chromosome 14. Among these, 15 BESs are located in a specific narrow region of chromosomes A and B (FLpter value: 0.04–0.22, Fig. 2.3.3c). However, the correlation

coefficient between BESs and the mouse genome was quite low (r^2=0.208) here and no significant correlation was identified in this narrow region. Consequently, this region of chromosomes A and B was not found not to be conserved compared with the mouse genomic sequence. These results suggest that some genomic modifications such as translocation have occurred in the homologous regions of mouse chromosome 14 and/or CHO chromosomes A and B.

According to the report of Yang et al. (2000), Chinese hamster chromosome 1 shows high homology to mouse chromosome 3, 8, and 14. Besides, short sections of Chinese hamster chromosome 1 show homology to other mouse chromosomes. The results of Yang et al. correspond (2000) to this analysis. Consequently, chromosome A and B are considered to correspond to Chinese hamster chromosome 1.

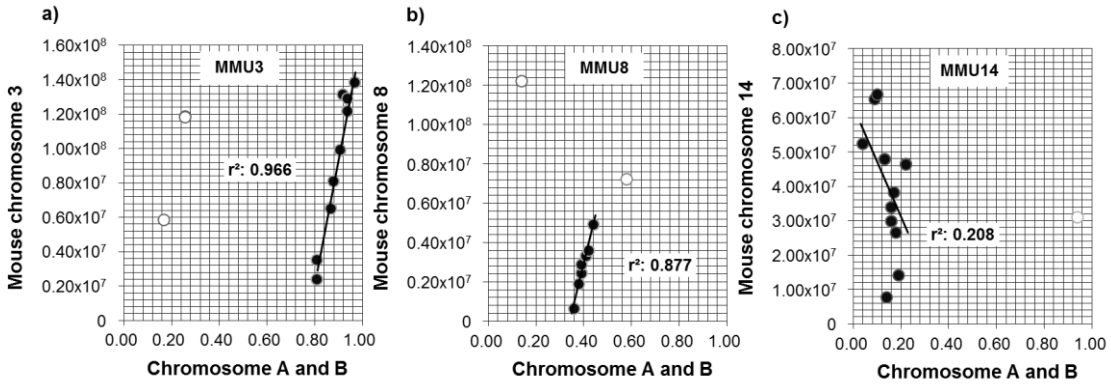

Fig. 2.3.3 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome A and B and mouse chromosome 3 (MMU3). b) Relationship between chromosome A and B and MMU8. c) Relationship between chromosome A and B and MMU14.

Chromosomes C and D

As shown in Fig. 2.3.4, homologous BESs were located symmetrically in chromosomes C and D. Even considering the error in determining BAC-FISH positions in chromosomes in the CHO DG44 cell line, specific regions in chromosomes C and D have high similarities to mouse chromosomes 1, 4, 10, and 15. My colleagues and I previously estimated that chromosome D is chromosome C but with a partial deletion, based on BAC-FISH results (Cao et al. 2012a). This BES analysis also confirmed that chromosome D is a mutated form of chromosome C, from which part of the long arm has been deleted.

According to the report of Yang et al. (2000), Chinese hamster chromosome 2 shows high homology to mouse chromosome 1, 4, 10, and 15. The results of Yang et al. (2000) correspond to the findings of the homology analysis here. Consequently, chromosome C and D are considered to correspond to Chinese hamster chromosome 2, and a large deletion might have occurred in chromosome D to create chromosome C.

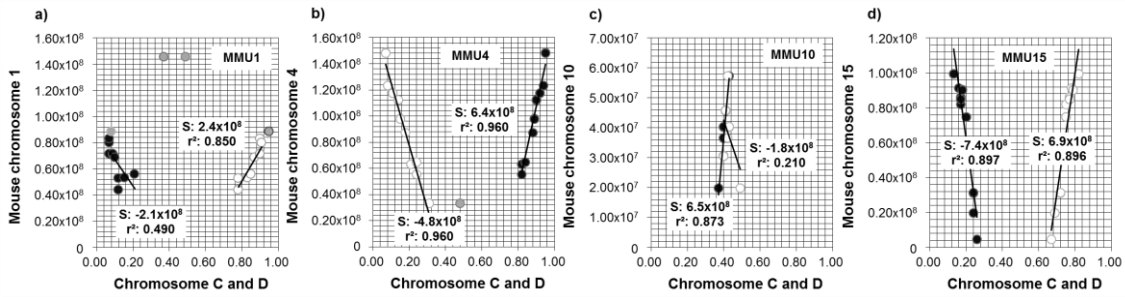

Fig. 2.3.4 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome C and D and MMU1. b) Relationship between chromosome C and D and MMU4. c) Relationship between chromosome C and D and MMU10. d) Relationship between chromosome C and D and MMU15. Black closed circles indicate the positions of chromosome C; white closed circles indicate the positions of chromosome D.

Chromosome F

There are two correlations between chromosome F and the mouse genomic sequence. The region of the long arm (FLpter value: 0.55–0.95, Fig. 2.3.5a) shows high homology to mouse chromosome 9 (MMU9), while the region of the short arm in chromosome F (FLpter value: 0.05–0.22, Fig. 2.3.5b) shows high homology to mouse chromosome 16. The correlation between chromosome F and mouse chromosome 9 is located in a large part of the long arm of chromosome F. This indicates that this area is conserved between Chinese hamster and mouse genomic sequences.

According to a the report of Yang et al. (2000), Chinese hamster chromosome 5 has high homology to mouse chromosome 9 and mouse chromosome 16. The results reported here are the same as those reported by Yang et al. (2000). This indicates that chromosome F originated from Chinese hamster chromosome 5.

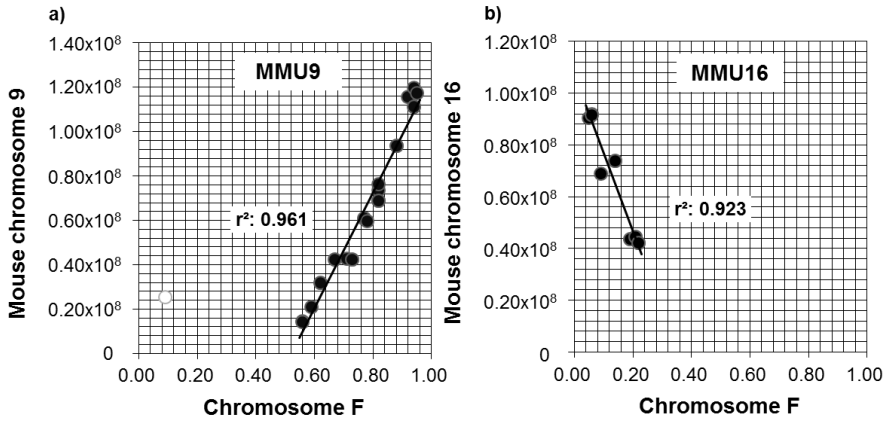

Fig. 2.3.5 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome F and MMU9. b) Relationship between chromosome F and MMU16.

Chromosome G

In chromosome G, there are three correlations of mouse genomic sequences. The region from 0.61 to 0.80 has high homology to mouse chromosome 7 (Fig. 2.3.6a), the region from 0.25 to 0.33 has high homology to mouse chromosome 8 (Fig. 2.3.6b), and the region from 0.80 to 0.93 has high homology to mouse chromosome 19 (Fig. 2.3.6c).

According to the report by Yang et al. (2000), Chinese hamster chromosome 3 has high homology to mouse chromosome 7, 8, and 19. This suggests that chromosome G corresponds to Chinese hamster chromosome 3.

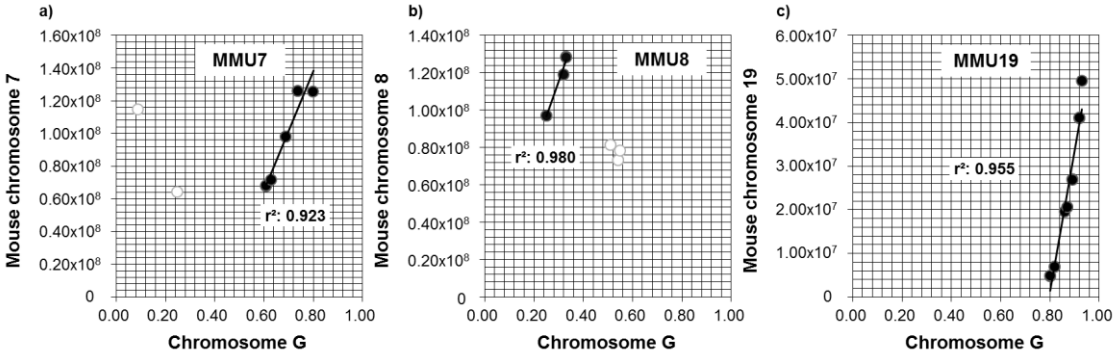

Fig. 2.3.6 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome G and MMU7. b) Relationship between chromosome G and MMU8. c) Relationship between chromosome G and MMU19.

Chromosome H and J

Chromosomes H and J showed homology to mouse chromosome 2 in large parts of each of their long arms (Fig. 2.3.7). This is unique to chromosome H and J of the CHO DG44 cell line because other regions homologous to the mouse genome are short. It was reported that large parts of Chinese hamster chromosome 6 and rat (*Rattus norvegicus*) chromosome 3 show high homology to mouse chromosome 2 (Yang et al. 2000). Consequently, these regions in chromosomes H and J might be conserved among rodent species. However, it is known that multiple chromosomal modifications are prone to occur in chromosomes of the CHO cell line. Moreover, large parts of chromosomes H and J in the CHO DG44 cell line in BAC-FISH and BES sequences show high homology to mouse chromosome 2. Thus, these regions might be conserved among several CHO cells, such as the CHO K1 cell line.

According to the report by Yang et al. (2000), a large part of CGR6 has high homology to mouse chromosome 2. These findings suggest that chromosome H and J correspond to Chinese hamster chromosome 6.

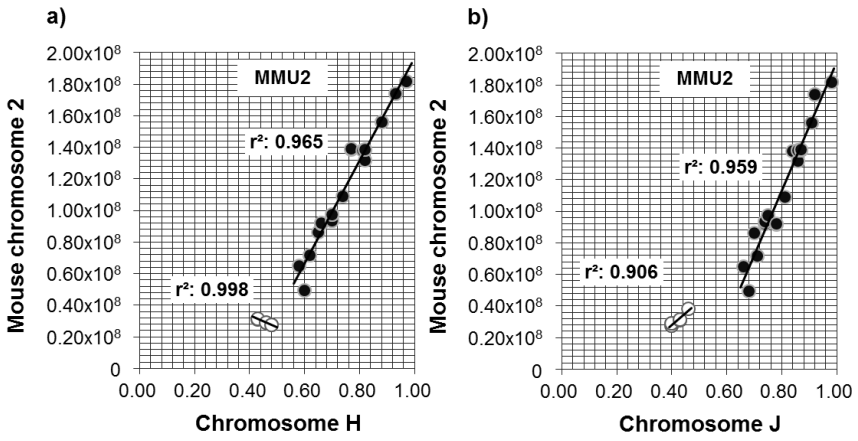

Fig. 2.3.7 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome H and MMU2. b) Relationship between chromosome J and MMU2. There are two correlated regions among chromosome H, J and MMU2: one of them is narrow (FLpter value: 0.43–0.48) and the other is wide (FLpter value: 0.58–0.97). White closed circles indicate the narrow region and black closed circles indicate the wide one.

Chromosome K

In chromosome K, there are two correlations of mouse genomic sequences. The region from 0.06 to 0.16 has high homology to mouse chromosome 6 (Fig. 2.3.8a) and the region from 0.50 to 0.74 has high homology to mouse chromosome 7 (Fig. 2.3.8b).

According to the report of Yang et al. (2000), the entirety of Chinese hamster chromosome 8 has high homology to mouse chromosome 6 and doesn't have homology to mouse chromosome 7. The result obtained here thus corresponds to the work of Yang et al. (2000), but chromosome K has correlation to mouse chromosome 7. These results indicate that chromosome K corresponds to Chinese hamster chromosome 6, but chromosomal modification such as insertion from Chinese hamster chromosome 8 might have occurred in chromosome K.

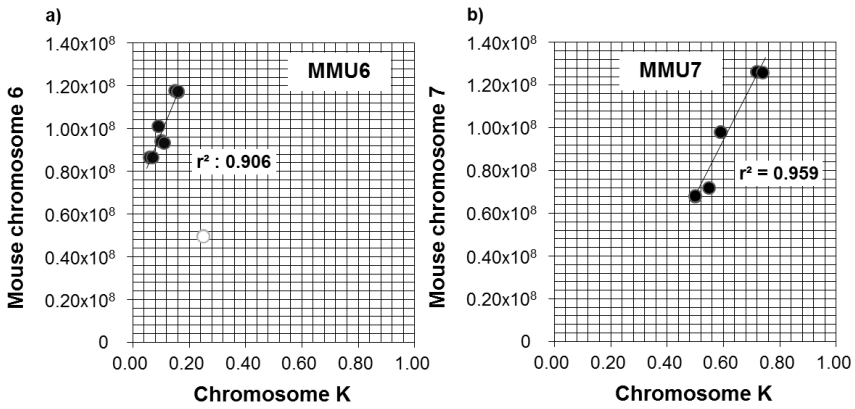

Fig. 2.3.8 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome K and MMU6. b) Relationship between chromosome K and MMU7.

Chromosome L and N

Chromosomes L and N showed homology to mouse chromosome 1 in large parts of each of their long arms (Fig. 2.3.9). This is unique to these two chromosomes of the CHO DG44 cell line because other regions homologous to the mouse genome are short. It was also reported that large parts of Chinese hamster chromosome 5 and rat (*Rattus norvegicus*) chromosome 13 show high homology to mouse chromosome 1 (Yang et al. 2000). Consequently, these regions in chromosomes L and N might be conserved among rodent species. CHO DG44 cell line chromosomes are not conserved 22 chromosomes, they are mainly 20 chromosomes (Omasa et al. 2009), and it is considered that many chromosomal modifications such as translocations have occurred in them. The above findings indicate that chromosome L and N may be paired chromosomes in Chinese hamster.

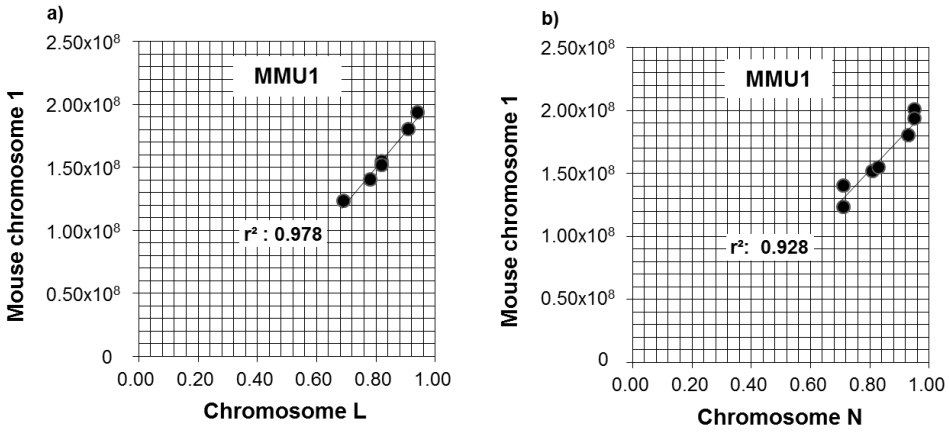

Fig. 2.3.9 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome L and MMU1. b) Relationship between chromosome N and MMU1.

Chromosome M

In a large part of chromosome M, there is high homology to mouse chromosome 9 (Fig. 2.3.10), which is the same result as obtained for chromosome F (Fig. 2.3.5a). However, chromosome M does not have high homology to mouse chromosome 16, which does not correspond to the result for chromosome F. Consequently, chromosome M might be based on chromosome F but underwent a large deletion and/or translocation.

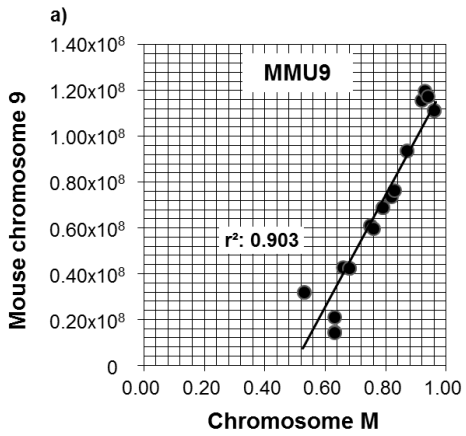

Fig. 2.3.10 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosome sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs. a) Relationship between chromosome M and MMU9.

Chromosome R

In chromosome R, there is a correlation to the mouse genomic sequence. The region from 0.11 to 0.77 has high homology to mouse chromosome 6 (Fig. 2.3.11), which is the same as the result for chromosome K (Fig. 2.3.8a), but chromosome R does not have homology to mouse chromosome 7.

According to the report of Yang et al. (2000), the entirety of Chinese hamster chromosome 8 has high homology to mouse chromosome 6, which is the same as the result obtained by Yang et al. (2000). These results indicate that chromosome K corresponds to chromosome F, but that a chromosomal modification such as an insertion from chromosome F might have occurred in chromosome R.

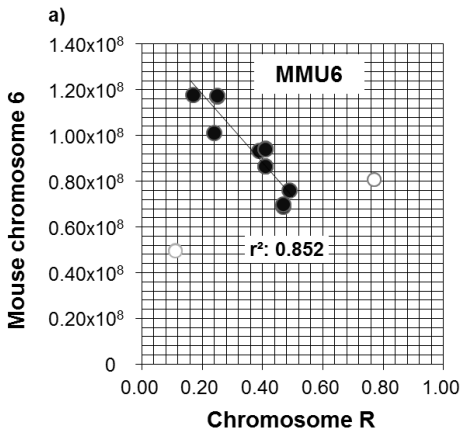

Fig. 2.3.11 Relationship between the chromosomes of the CHO DG44 cell line and mouse chromosomal sequences based on BESs. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient.. a) Relationship between chromosome R and MMU6.

Sex chromosome

It was reported that the X chromosome, a sex chromosome, is conserved to a large extent among Chinese hamster, mouse, and rat, as determined by CGH analysis (Yang et al. 2000). Most BESs that have homology to the X chromosome in mouse are located in chromosomes E and P (Fig. 2.3.12), but no structural similarity among them was identified by BES analysis. This indicates that the genetic composition as revealed by CGH analysis is similar among rodent X chromosomes, but the locations of genetic sequences in the chromosome revealed by BESs are quite different.

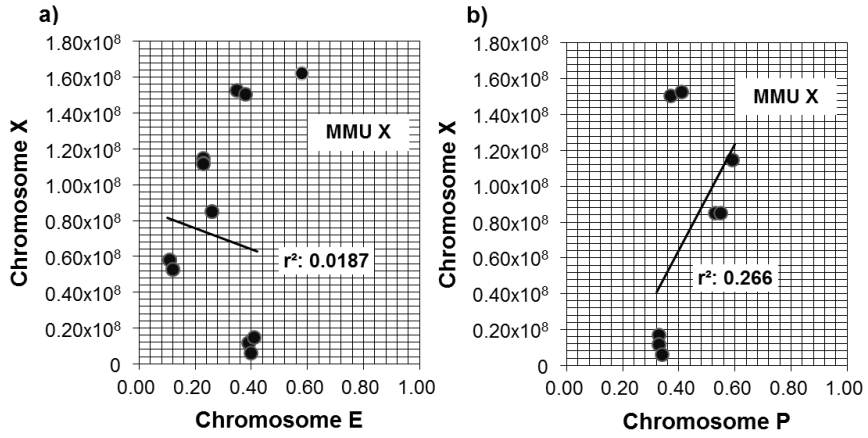

Fig. 2.3.12 Relationship between mouse X chromosome (MMUX) and CHO DG44 cell line chromosome a) E and b) P. The horizontal axis reflects the chromosomal position in each CHO DG44 cell chromosome (the mean of the FLpter value; Lichter et al. 1990), while the vertical axis represents the position on the mouse chromosome [bp] in all graphs. In these graphs, S is the slope of the fitted curve and r^2 is the correlation coefficient. Gray closed circles indicate the positions excluded from the fitted curve in all graphs.

2.4. Summary

The results presented here indicate that the CHO DG44 chromosomes exhibit high homology to mouse chromosomes. The results also enable estimation of which chromosomes in mouse correspond to CHO DG44 chromosomes. These findings were not revealed by hybridization analysis, such as CGH and SKY, because the exact relationship between sequences and chromosomal locations could not be revealed using hybridization analysis. To overcome this problem, BAC-FISH with BES analysis can be used, which involves the combination of BAC-based hybridization and sequence data; it is a useful technique to link sequences and chromosomal locations. Moreover, using BAC-FISH with BES analysis, it is possible to perform comparisons between chromosomal locations in several cell lines. Recently, the genomic sequences of Chinese hamster and its cell lines were revealed by next-generation sequencing (NGS; Xu et al. 2011; Vishwanathan et al. 2016), and the genomic scaffolds constructed using the NGS results were compared with mouse genomic sequences and chromosomal locations. However, their scaffolds could not be linked to chromosomal locations in Chinese hamster or Chinese hamster-derived cell lines. The genomic sequences determined by NGS should thus be combined with the chromosomal information revealed by BAC-FISH and BES data. Recently, Lewis et al. (2013) analyzed the CHO

genome sequence and compared the sequence scaffolds based on our previous BAC-FISH data (Lewis et al. 2013). Genomic sequences are essential information for understanding cells and useful for genome engineering technologies, such as genomic editing. It is expected that BAC-FISH and BES data can contribute to the development of scientific research on CHO cells and the construction of cell lines for producing protein-based pharmaceuticals.

Chapter 3

Determination of rearrangements

between CHO DG44 cells and CHO K1 cells

3.1. Introduction

Recently, following the development of next-generation sequencing (NGS) techniques and their impact on genomic research, whole-genome sequencing has become more economical and faster. Several groups have reported NGS analyses of the genomic sequences of Chinese hamster and/or CHO cells ((Xu et al. 2011; Lewis et al. 2013; Kaas et al. 2015; Feichtinger et al. 2016; Vishwanathan et al. 2016). According to the results of Xu et al. (2011), the estimated size of the CHO K1 genome is about 2.45 Gb and it was predicted to contain 24,383 genes.

However, the genome of Chinese hamster includes various repeat sequences that are distributed on the same or different chromosomes (Ono and Sonta 2001). In analysis of the human genome, a clone-based physical map was indispensable for the Human Genome Project because of the presence of numerous repeat sequences in the genome (McPherson et al. 2001). The DNA sequencing information of the CHO genome should thus be coupled with physical chromosomal locations because of the instability of the chromosomal structure. These locations should be obtained from BAC physical maps that can be derived from bacterial artificial chromosome (BAC) libraries and compared with the genomic sequences of Chinese hamster, various CHO cell lines, and other related species.

In the work described in this chapter, I determined the BAC end sequences (BESs) of 303 clones, which were used as BAC probes for a chromosomal physical map in the CHO DG44 cell line. Moreover, I compared these BESs and CHO K1 cell genomic sequences (Xu et al. 2011).

3.2. Materials and methods

The procedures regarding the extraction of BAC DNA and BAC end sequences are described in chapters 2.2.2 and 2.2.4.

3.2.1. Searching for highly homologous sequences between BAC end sequences and genomic sequences of CHO K1 cells using BLAST

A sequence homology search with the CHO K1 cell genomic sequences [*Cricetulus griseus* assembly CriGri_1.0 scaffolds (reference assembly in build 1.1), July 2011] was performed using the BLAST algorithm (http://blast.ncbi.nlm.nih.gov/), setting the threshold for significance (e-value) as $<10^{-10}$.

3.3. Results and discussion

3.3.1. Analysis of BAC end sequences

I determined the BESs of BAC clones that were used for physical map construction in the CHO DG44 cell line in the work described in chapter 2.3.1. Specifically, I attempted to determine all BESs of the 303 BAC clones (606 BESs) by end sequencing (Table 2.3.1). Then, I also analyzed whether BESs have high homology with the CHO K1 genomic sequence (Fig. 3.3.1). In general, all BESs show homology to the CHO K1 genome because these two cell lines have been derived from Chinese hamster. However, 13 BESs do not exhibit homology to CHO K1 genomic sequences (referred to here as "not homologous BESs"). Thus, I determined the differences between homologous and not homologous BESs based on their sequences. First, I evaluated the average sequence lengths: the average homologous sequence length was 533 bp, but the average not homologous sequence length was 479 bp. The significance of this difference was evaluated by *t*-test.

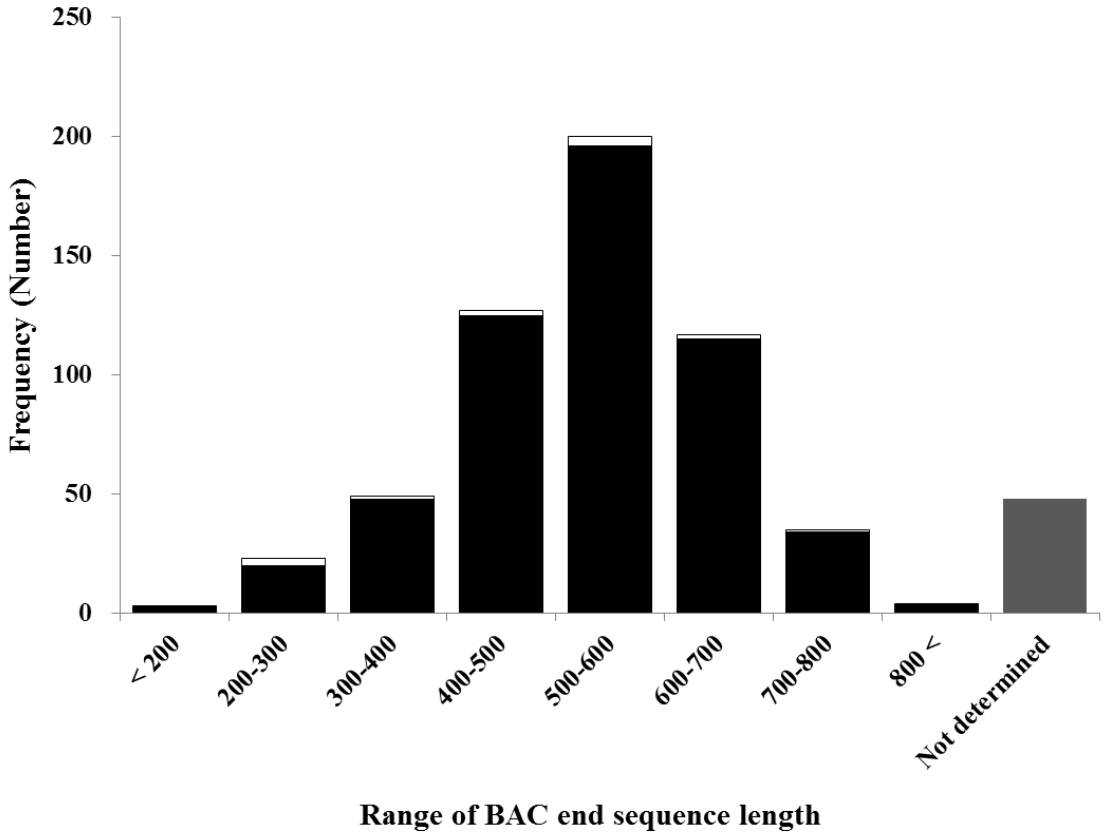

Fig. 3.3.1 Size distribution of determined BAC end sequences. Black: number of sequences homologous to CHO K1 genomic sequences. White: number of sequences not homologous to CHO K1 genomic sequences. Gray: number of sequences not determined to BESs.

But the difference was not significant. Thus, I analyzed some features of not homologous BES sequences. This revealed that nine BESs have low GC content (Table 3.3.1). Generally, genomic modifications such as single-nucleotide polymorphisms (SNPs) do not readily occur in sequences with high GC content. It is known that the A-T bond are weaker hydrogen bond than the G-C band. So, the region in low GC content is tend to occur chromosome modification. Consequently, it was estimated that some modifications occurred in these sequences. Two other BESs have some repeat sequences of at least 10 bp. In general, short repeat sequences are very difficult to characterize using a sequence analyzer. For example, the reverse sequence of BAC clone Cg0180E19 contains a $(TTAGGG)_n$ sequence (Cao et al. 2012a), which is a typical repeat sequence in human and Chinese hamster telomeres (Moyzis et al. 1988; Slijepcevic and Hande 1999; Bolzán and Bianchi 2006). It was thus considered that this sequence is potentially useful for establishing genomic information on Chinese hamster. The remaining two BESs, the reverse sequence of Cg0160D05 and the forward sequence of Cg0160L12, did not show significant features. It appears that these sequences are modified versions of sequences derived from the Chinese hamster genome.

Table 3.3.1 List of BESs not homologous to CHO K1 genomic sequences

BAC number	Direction of end sequence	DDBJ number	Insertion size [kb]	Length of end sequence [bp]	GC contents (%)
Cg0160D05	Reverse	DE989456	140	493	53.2
Cg0160F01	Forward	DE989336	135	515	36.5
Cg0160F07	Forward	DE989505	160	323	38.3
Cg0160L12	Forward	DE989585	110	517	47.4
Cg0180A19	Forward	DE990260	115	223	47.7
Cg0180B10	Reverse	DE989985	115	296	42.9
Cg0180D06	Forward	DE989864	140	654	38.4
Cg0180E19	Reverse	DE990269	35	257	49.5
Cg0180F03	Reverse	DE989776	115	651	39.1
Cg0180G09	Reverse	DE989963	90	714	41.6
Cg0180I19	Forward	DE990276	125	491	35.0
Cg0180K06	Forward	DE989878	125	580	36.1
Cg0180L20	Forward	DE990313	105	517	38.4

I determined the distribution of homologous BESs in each CHO DG44 chromosome (Fig 3.3.2). Each chromosome has BESs that do not show homology to the CHO K1 genome or has end sequences that were not characterized.

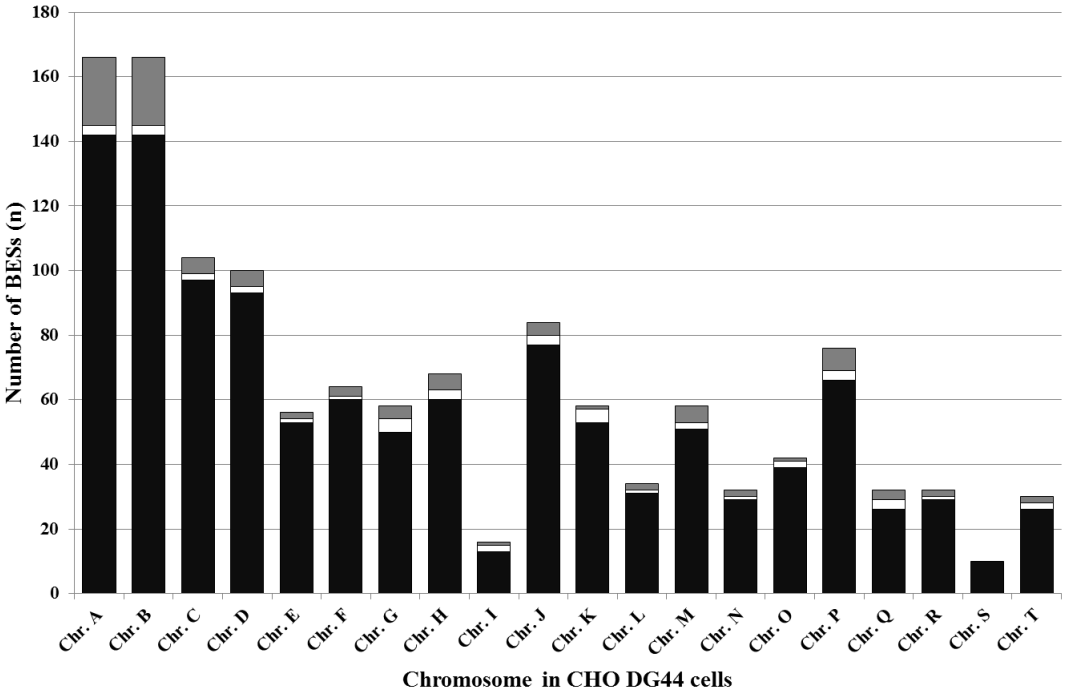

Fig. 3.3.2 Distribution of number of homologous sequences between CHO DG44 chromosome and CHO K1 genomic sequence. CHO DG44 chromosomes are arranged in order from A to T based on chromosome length, as reported previously (Omasa et al. 2009; Cao et al. 2012a). Black: number of BESs homologous to CHO K1 genomic sequences. White: number of BESs not homologous to CHO K1 genomic sequences. Gray: number of not determined BESs.

3.3.2. Analysis of BAC end sequences

The CHO genome is easily modified in previous studies (Omasa et al. 2009; Cao et al. 2012a). I determined the single-nucleotide polymorphisms and insertions/deletions (indels) between BESs and CHO K1 genomic sequences. Mismatches and gaps between BES and CHO K1 genomic sequences were identified (Table 3.3.2). SNPs constitute approximately 0.60% of the entire determined sequence length and indels approximately 0.17%. It can thus be estimated from the CHO genome size that the amounts of SNPs and indels are approximately 15 and 4.1 Mbp in the entire CHO CG44 genome, respectively.

Table 3.3.2 Total SNPs and indels between BESs and CHO K1 genomic sequences

SNPs (n)	**SNPs (%)**	**indels (n)**	**indels (%)**
1,525	0.598%	422	0.166%

3.4. Summary

The work described in this chapter does not clarify the relationship between the CHO DG44 and K1 genomes, but it does provide important information linking the genomic sequence to the chromosomes. The findings also show that the sequences homologous to CHO K1 genomic sequences contain many SNPs and indels. These results indicate that our BES and BAC-FISH data are useful for showing the correspondence between sequence data and chromosomal locations.

Chapter 4

Conclusions

Recently, following the development of next-generation sequencing (NGS) techniques and their impact on genome research, whole-genome sequencing has become more economical and faster. Several groups have reported NGS analyses of the genomic sequences of Chinese hamster and/or CHO cells (Xu et al. 2011; Lewis et al. 2013; Kaas et al. 2015; Feichtinger et al. 2016; Vishwanathan et al. 2016). According to the results of Xu et al. (2011), the estimated size of the CHO K1 genome is about 2.45 Gb and it was predicted to contain 24,383 genes. Nowadays, it is possible to obtain substantial data from NGS analysis, but it is difficult to use such data in Chinese hamster without basic biological information such as a physical map. Thus, a BAC library and a physical map for CHO cells were constructed to obtain basic information on the biology of these cells; several important findings were made using these BAC library data (Omasa et al. 2009; Park et al. 2010; Cao et al. 2012a; Cao et al. 2012b; Takagi et al. 2017). For example, it was discovered that the sequence of *DHFR* gene amplification site consists of an inverted repeat structure (Omasa et al. 2009), which is effective for the construction of "cell lines producing high levels of recombinant protein (Park et al. 2010). In addition, this sequence was shown to contain a motif that was estimated to be an insulator (Takagi et al. 2017). The physical map of CHO DG44 cells revealed that these cells mainly have 20 chromosomes, the modification of which has

frequently occurred (Omasa et al. 2009; Cao et al. 2012a). However, the longest pair of chromosomes (named chromosomes A and B) were shown to be conserved in various CHO cell lines (Cao et al. 2012a). In this study, I also determined the relationship between BESs derived from CHO DG44 cells and the mouse genome, and found that some regions in CHO DG44 chromosomes show high homology to mouse chromosomes.

Human and mouse genomic sequences have been linked to their positions on the chromosomes, and this information is widely available and used in many fields of research, including genome editing (Kim et al. 2013). It is possible that genome editing techniques can also contribute to improved knowledge of the basic biology in Chinese hamster. To enable this, the Chinese hamster or Chinese hamster-derived genomic sequences determined by NGS should be combined with chromosomal information revealed by BAC-FISH and BES data. Recently, Lewis et al. (2013) analyzed the CHO genome sequence and compared the sequence scaffolds based on our previously reported BAC-FISH data (Lewis et al. 2013). Moreover, very recently, third-generation PAC-BIO NGS analysis was reported (Rupp et al. 2018). Using these PAC-BIO data, it may be possible to precisely compare the Chinese hamster genome and the CHO DG44 genome or genome sequences of other cell lines. Genomic sequences are essential

information for understanding cells and useful for genome engineering technologies, such as genomic editing. It is expected that BAC-FISH and BES data can contribute to the progression of scientific research on CHO cells and the construction of cell lines for producing protein-based pharmaceuticals.

Publications

Shuichi Kimura and Takashi Omasa (2018) Genome sequence comparison betwe en Chinese hamster ovary (CHO) DG44 cells and mouse using end sequences of CHO BAC clones based on BAC-FISH results

Cytotechnology, (in press) doi: 10.1007/s10616-018-0233-5

Related publications

Takeshi Omasa, Yihua Cao, Joon Young Park, Yasuhiro Takagi, **Shuichi Kimura**, Hidenori Yano, Kohsuke Honda, Shuichi Asakawa, Nobuyoshi Shimizu, Hisao Ohtake (2009) Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells.

Biotechnol Bioeng 104(5):986-994. doi:10.1002/bit.22463

Yihua Cao, **Shuichi Kimura**, Takayuki Itoi, Kohsuke Honda, Hisao Ohtake, Takeshi Omasa (2012) Construction of BAC-based physical map and analysis of chromosome rearrangement in chinese hamster ovary cell lines.

Biotechnol Bioeng 109:1357–1367. doi: 10.1002/bit.24347

Yihua Cao*, **Shuichi Kimura***, Takayuki Itoi, Kohsuke Honda, Hisao Ohtake, Takeshi Omasa (2012) Fluorescence in situ hybridization using bacterial artificial chromosome (BAC) clones for the analysis of chromosome rearrangement in Chinese hamster ovary cells.

Methods 56:418–423. doi: 10.1016/j.ymeth.2011.11.002

* Equal contribution

References

Bolzán AD, Bianchi MS (2006) Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. Mutat Res - Rev Mutat Res 612:189–214. doi: 10.1016/j.mrrev.2005.12.003

Cao Y, Kimura S, Itoi T, Honda K, Ohtake H, Omasa T (2012a) Construction of BAC-based physical map and analysis of chromosome rearrangement in chinese hamster ovary cell lines. Biotechnol Bioeng 109:1357–1367. doi: 10.1002/bit.24347

Cao Y, Kimura S, Itoi T, Honda K, Ohtake H, Omasa T (2012b) Fluorescence in situ hybridization using bacterial artificial chromosome (BAC) clones for the analysis of chromosome rearrangement in Chinese hamster ovary cells. Methods 56:418–423. doi: 10.1016/j.ymeth.2011.11.002

Collins FS, Lander ES, Rogers J, Waterson RH (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945. doi: 10.1038/nature03001

Feichtinger J, Hernández I, Fischer C, Hanscho M, Auer N, Hackl M, Jadhav V, Baumann M, Krempl PM, Schmidl C, Farlik M, Schuster M, Merkel A, Sommer A, Heath S, Rico D, Bock C, Thallinger GG, Borth N (2016) Comprehensive genome

and epigenome characterization of CHO cells in response to evolutionary pressures and over time. Biotechnol Bioeng 113:2241–2253. doi: 10.1002/bit.25990

Greber D, Fussenegger M (2007) Mammalian synthetic biology: Engineering of sophisticated gene networks. J Biotechnol 130:329–345. doi: 10.1016/j.jbiotec.2007.05.014

Griffin TJ, Seth G, Xie H, Bandhakavi S, Hu WS (2007) Advancing mammalian cell culture engineering using genome-scale technologies. Trends Biotechnol 25:401–408. doi: 10.1016/j.tibtech.2007.07.004

Hacker DL, De Jesus M, Wurm FM (2009) 25??years of recombinant proteins from reactor-grown cells - Where do we go from here? Biotechnol Adv 27:1023–1027. doi: 10.1016/j.biotechadv.2009.05.008

Kaas CS, Kristensen C, Betenbaugh MJ, Andersen MR (2015) Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. 1–9. doi: 10.1186/s12864-015-1391-x

Kao FT, Puck TT (1968) Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. Proc Natl Acad Sci U S A 60:1275–1281. doi: 10.1073/pnas.60.4.1275

Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, Kim S, Lee C, Jeong E, Chung E, Kim D, Lee MS, Go EM, Song HJ, Kim H, Cho N, Bang D, Kim S, Kim J-S (2013) A library of TAL effector nucleases spanning the human genome. Nat Biotechnol 31:251–258. doi: 10.1038/nbt.2517

Koo TY, Park JH, Park HH, Park TH (2009) Beneficial effect of 30Kc6 gene expression on production of recombinant interferon-β in serum-free suspension culture of CHO cells. Process Biochem 44:146–153. doi: 10.1016/j.procbio.2008.09.018

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili I, Xu X, Wang J, Palsson BO (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat Biotechnol 31:759–765. doi: 10.1038/nbt.2624

Li J, Stern DF (2005) Regulation of CHK2 by DNA-dependent protein kinase. J Biol Chem 280:12041–12050. doi: 10.1074/jbc.M412445200

Lichter P, Tang C, Call K, Hermanson G, Evans G, Housman D, Ward D (1990) High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science (80-) 247:64–69. doi: 10.1126/science.2294592

Lombardo, A; Genovese, P; Beausejour, CM; Colleoni, S; Lee, YL; Kim, KA; Ando, D; Urnov, FD; Galli, C; Gregory, PD; Holmes MN, Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee Y-L, Kim K a, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25:1298–1306. doi: 10.1038/nbt1353

Maksimenko O, Gasanov NB, Georgiev P (2015) Regulatory elements in vectors for efficient generation of cell lines producing target proteins. Acta Naturae 7:15–26.

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen H-C, Jang W, Green ED, Idol JR, Maduro VVB, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati R, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, Jong PJ de, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng

J-F, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen X-N, Fujiyama A, Hattori M, Toyoda A, Yada T, Park H-S, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H, McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen H-C, Jang W, Schuler G, Chen H-C, Jang W, Green ED, Idol JR, Maduro VVB, Green ED, Idol JR, Maduro VVB, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati R, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati R,

Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, Jong PJ de, Catanese JJ, Nowak N, Osoegawa K, Jong PJ de, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng J-F, Hawkins T, Branscomb E, Olsen A, Doggett N, Cheng J-F, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen X-N, Korenberg J, Chen X-N, Fujiyama A, Hattori M, Toyoda A, Yada T, Park H-S, Sakaki Y, Fujiyama A, Hattori M, Toyoda A, Yada T, Park H-S, Sakaki Y, Shimizu N, Asakawa S,

Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H (2001) A physical map of the human genome. Nature 409:934–941. doi: 10.1038/35057157

Melville M, Doolan P, Mounts W, Barron N, Hann L, Leonard M, Clynes M, Charlebois T (2011) Development and characterization of a Chinese hamster ovary cell-specific oligonucleotide microarray. Biotechnol Lett 33:1773–1779. doi: 10.1007/s10529-011-0628-2

Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci 85:6622–6626. doi: 10.1073/pnas.85.18.6622

Omasa T, Cao Y, Park JY, Takagi Y, Kimura S, Yano H, Honda K, Asakawa S, Shimizu N, Ohtake H (2009) Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. Biotechnol Bioeng 104:986–994. doi: 10.1002/bit.22463

Omasa T, Onitsuka M, Kim W-D (2010) Cell engineering and cultivation of chinese hamster ovary (CHO) cells. Curr Pharm Biotechnol 11:233–40. doi: doi 10.2174/138920110791111960

ONO K-I, KAMIHIRA M, KUGA Y, MATSUMOTO H, HOTTA A, ITOH T, NISHIJIMA K-I, NAKAMURA N, MATSUDA H, IIJIMA S (2003) Production of Anti-Prion scFv-Fc Fusion Proteins by Recombinant Animal Cells. J Biosci Bioeng 95:231–238. doi: 10.1263/jbb.95.231

Ono T, Sonta S (2001) Chromosome map of cosmid clones constructed with Chinese hamster genomic DNA. Cytogenet Cell Genet 95:97–102. doi: 10.1159/000057025

Park JS, Kim H, Park J, Yu S, Kim D, Lee J, Oh H, Baek K, Yoon J (2010a) Overproduction of recombinant human hepatocyte growth factor in Chinese hamster ovary cells. Protein Expr Purif 70:231–235. doi: 10.1016/j.pep.2009.10.004

Park JY, Takagi Y, Yamatani M, Honda K, Asakawa S, Shimizu N, Omasa T, Ohtake H (2010b) Identification and analysis of specific chromosomal region adjacent to exogenous Dhfr-amplified region in Chinese hamster ovary cell genome. J Biosci Bioeng 109:504–511. doi: 10.1016/j.jbiosc.2009.10.019

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi: 10.1016/j.cell.2013.02.022

Rupp O, MacDonald ML, Li S, Dhiman H, Polson S, Griep S, Heffner K, Hernandez I, Brinkrolf K, Jadhav V, Samoudi M, Hao H, Kingham B, Goesmann A, Betenbaugh MJ, Lewis NE, Borth N, Lee KH (2018) A reference genome of the Chinese hamster based on a hybrid assembly strategy. Biotechnol Bioeng. doi: 10.1002/bit.26722

Slijepcevic P, Hande MP (1999) Chinese hamster telomeres are comparable in size to mouse telomeres. Cytogenet Genome Res 85:196–199. doi: 10.1159/000015292

Takagi Y, Yamazaki T, Masuda K, Nishii S, Kawakami B, Omasa T (2017) Identification of regulatory motifs in the CHO genome for stable monoclonal antibody production. Cytotechnology 69:451–460. doi: 10.1007/s10616-016-0017-8

Urlaub G, Chasin LA (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc Natl Acad Sci U S A 77:4216–20. doi: VL - 77

Vishwanathan N, Bandyopadhyay AA, Fu HY, Sharma M, Johnson KC, Mudge J, Ramaraj T, Onsongo G, Silverstein KAT, Jacob NM, Le H, Karypis G, Hu WS (2016) Augmenting Chinese hamster genome assembly by identifying regions of high confidence. Biotechnol J 11:1151–1157. doi: 10.1002/biot.201500455

Walsh G (2014) Biopharmaceutical benchmarks 2014. Nat Biotechnol 32:992–1000. doi: 10.1038/nbt0910-917

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki

Y, LaHillier DW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S,

Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. doi: 10.1038/nature01262

Wlaschin KF, Nissom PM, De Leon Gatti M, Ong PF, Arleen S, Tan KS, Rink A, Cham B, Wong K, Yap M, Hu WS (2005) EST sequencing for gene discovery in Chinese hamster ovary cells. Biotechnol Bioeng 91:592–606. doi: 10.1002/bit.20511

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22:1393–1398. doi: 10.1038/nbt1026

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Christina Fan H, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang J (2011a) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29:3–5. doi: 10.1038/nbt.1932

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Christina Fan H, Wang JJJ, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang JJJ

(2011b) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29:3–5. doi: 10.1038/nbt.1932

Yang F, O'Brien PCM, Ferguson-Smith MA (2000) Comparative chromosome map of the laboratory mouse and Chinese hamster defined by reciprocal chromosome painting. Chromosom Res 8:219–227. doi: 10.1023/A:1009200912436

Yoshikawa T, Nakanishi F, Itami S, Kameoka D, Omasa T, Katakura Y, Kishimoto M, Suga K (2000a) Evaluation of stable and highly productive gene amplified CHO cell line based on the location of amplified genes. Cytotechnology 33:37–46. doi: 10.1023/A:1008111328771

Yoshikawa T, Nakanishi F, Ogura Y, Oi D, Omasa T, Katakura Y, Kishimoto M, Suga KI (2000b) Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. Biotechnol Prog 16:710–715. doi: 10.1021/bp000114e

Acknowledgements

Undertaking this PhD has been a life-changing experience for me and it wouldn't have been possible to do without the support and guidance that I received from many people.

I would like to first express my sincere appreciation to my supervisors Professor Takeshi Omasa and Professor Yoshitoshi Nakamura, for the continuous support for my PhD study. Their suggestions, constructive criticism and incredible patience have guided me through my research and writing this thesis. I never forget their encouragement and support that I have received.

I would like to thank Professor Yoshihiro Uto and Professor Hitoshi Matsuki for reviewing this thesis as well as for their insightful comments and suggestions.

My deep appreciation goes to Akihiro Shirai, Masayoshi Onitsuka, Noriko Yamano and Jana Frank for their valuable cooperation in my experiments. I am also very grateful to my colleagues of Labs in Osaka university and the university of Tokushima.

Finally I would like to express my greatest appreciation and gratitude to my parents and sister for their endless love, understanding, support, encouragement and sacrifice throughout my study.

September, 2018

Shuichi Kimura